

**Erstellung von Metadatenbanken zu Umweltchemikalien und  
vergleichende Bewertung von Online Datenbanken und CD-ROMs**

Den Naturwissenschaftlichen Fakultäten  
der Friedrich-Alexander-Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades

vorgelegt von

Kristina Voigt

aus Berlin

Als Dissertation genehmigt von den Naturwissenschaftlichen Fakultäten der  
Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 03.06.97

Vorsitzender der Promotionskommission:

Professor Dr. D. Kölzow

Erstberichterstatter: Professor Dr. J. Gasteiger

Zweitberichterstatter: Professor Dr. M. Matthies

Meinem Doktorvater,

Herrn Prof. Dr. Johann Gasteiger,

danke ich sehr für die gewährte Unterstützung und die wertvollen Anregungen während der Durchführung der Arbeit.

Mein außerordentlicher Dank gilt darüber hinaus

Herrn Dr. Rainer Brüggemann

für die stets gewährte Unterstützung, für vielfältige Anregungen und Diskussionen insbesondere im Zusammenhang mit den theoretischen Aspekten der in der Arbeit angewandten Hasse-Diagramm-Technik.

Mein besonderer Dank gilt auch Herrn Prof. Dr. Michael Matthies für die hilfreiche Unterstützung der Arbeit.

Ebenso bedanke ich mich bei Herrn Dr. Joachim Benz für die Hilfe bei statistischen Fragestellungen.

Weiterhin danke ich in besonderem Maße Herrn Dr. Joachim Altschuh für zahlreiche Anregungen und wertvolle Diskussionen.

Mein besonderer Dank gilt auch Frau Annemarie Helmer für die Hilfe bei der Erstellung von Graphiken und Formatierungen sowie allen anderen Kollegen der Projektgruppe Umweltgefährdungspotentiale von Chemikalien der GSF - Forschungszentrum für Umwelt und Gesundheit.

## Liste der Abkürzungen

### Abkürzungen allgemein

$\Gamma$	=	Gesamtmaß $\Gamma$ für Objekte
$\vartheta$	=	Gesamtmaß $\vartheta$ für Kriterien
AN	=	Anzahl an Chemikalien
AK	=	Antikette
AR	=	Art/ Anwendung der Chemikalien
BI	=	bibliographische Datenbanken
BWK	=	Bewertungskriterium
BWZ	=	Bewertungsziffer
CA	=	CAS-Nummer
CD	=	CD-ROM
CE	=	Chemikalien Entwicklung
CN	=	Chemikaliennamen-Verzeichnisse
CT	=	Chemikalien-Testdatensatz
FBDB	=	faktenbasierte Datenbanken
FD	=	Forschungsdatenbanken
IB <sub>all+inh</sub>	=	Informationsbasis mit dem vollständige Satz an Bewertungskriterien (allgemeine und inhaltliche)
IB <sub>all</sub>	=	Informationsbasis mit dem Satz an allgemeinen Bewertungskriterien
IB <sub>inh</sub>	=	Informationsbasis mit dem Satz an inhaltlichen Bewertungskriterien
ID	=	Identifikationsmerkmale für Chemikalien
$I_i$	=	Shannon-Index
INDB	=	Integrierte Datenbanken
IP	=	Informationsparameter für Umweltchemikalien
Iso. E.	=	isolierte Elemente
K	=	Kette
KO	=	Kosten der Recherche / der CD-ROM
MAX	=	maximales Objekt, Maximal
MIN	=	minimales Objekt, Minimal
MW	=	Mittelwert
NEL	=	Anzahl der Elemente in der breitesten Ebene
NL	=	Anzahl der Ebenen
NU	=	Faktendatenbanken (numerische Datenbanken)
OM	=	Objektmenge
ON	=	Online Datenbank
PE	=	Parameter Entwicklung
QU	=	Validität/Güte der Datenbank
SA	=	Strukturformel nur Ausgabe
SF	=	Summenformel
SS	=	Strukturformel suchbar
STDEV	=	Standardabweichung
SY	=	Synonyma
TBDB	=	textbasierte Datenbanken
U(N)	=	Unvergleichbarkeiten
$U_i$	=	Umfangindex
UM	=	Umfang der Datenbank / CD-ROM
UP	=	Häufigkeit der Aktualisierung (Update)

V (N)	=	Vergleichbarkeiten
VE	=	Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien
VT	=	Volltextdatenbanken
VZ	=	Firmenverzeichnisse und sonstige Verzeichnisse

## Abkürzungen für Datenquellen

### Online Datenbanken

AGR	=	Agrochemicals Handbook (Agrochemicals Database)
ANA	=	ANABST (Analytical Abstracts)
API	=	APILIT (American Petroleum Institute Literature Database)
AQU	=	AQUIRE (Aquatic Information Retrieval Databank)
ASF	=	ASFA (Aquatic Science and Fisheries Abstracts)
BEI	=	BEILSTEIN (BEILSTEIN Database)
BIO	=	BIOSIS (BIOSIS Previews)
CAB	=	CAB (CAB Abstracts)
CAS	=	CA (Chemical Abstracts)
CBN	=	CBNB (Chemical Business News Database)
CEA	=	CEABA (Chemical Engineering and Biotechnology Abstracts)
CEN	=	CEN (Chemical Engineering News)
CES	=	CESARS (Chemical Evaluation Search and Retrieval System)
CHC	=	CHCD (Chapman and Hall Chemical Database)
CHE	=	CHEMSAFE (Chemical Safety Database)
CHR	=	CHRIS (Chemical Hazard Response Information System)
CIN	=	Chemical Industry Notes
CIT	=	Chem-Intell Trade (Chem-Intell Trade Database)
CJA	=	CJACS (Chemical Journals of the American Chemical Society)
CMS	=	CHEMSEARCH
COM	=	COMPENDEX
CON	=	CONFSCI (Conference Papers Index)
CSN	=	CSNB (Chemical Safety Newsbase)
DET	=	DET (DETERM)
DIP	=	DIPPR (Design Institute for Physical Property Data)
DIS	=	DISSABS (Dissertation Abstracts)
ECD	=	ECDIN (Environmental Chemicals Data and Information Network)
EMA	=	EMA (Engineered Materials Abstracts)
EMB	=	EMBASE (Excerpta Medica Database)
ENE	=	ENERGY (Energy Abstracts)
ENF	=	ENVIROFATE (Environmental Fate Databank)
ENV	=	ENVIROLINE (Environment Abstracts Online)
ERN	=	ERNS (Emergency Response Notification System Database)
FST	=	FSTA (Food Science and Technology Abstracts)
GEO	=	GEOREF (Geology References)
HAZ	=	HAZINF (Hazardous Chemicals Information and Disposal Guide)
HOD	=	HODOC (Handbook of Data on Organic Compounds)
HSD	=	HSDB (Datenbank Hazardous Substances Databank)
IPA	=	IPA (International Pharmaceutical Abstracts)
ISH	=	ISHOW (Information System for Hazardous Organics in Water)
JAN	=	JANAF (Joint Army-Navy-Air-Force)
JIC	=	JICST-E (JICST File on Science, Technology, and Medicine in Japan)
KKF	=	KKF (Kunststoffe Kautschuk Fasern)
MED	=	MEDLINE (Medical Literature Online)

MER	=	MERCK (The Merck Index Online)
MSD	=	MSDS-OHS (Material Safety Data Sheets from Occupational Health Service)
NAP	=	NAPRALERT (NAPRALERT)
NLD	=	NLDB (The Newsletter Database)
NTI	=	NTIS (National Technical Information Service)
OCE	=	OCEAN (Oceanic Abstracts)
PAP	=	PAP (PAPERCHEM)
PAS	=	PAS (PASCAL)
PHY	=	PHYTOTOX (Toxic Effects on Plants Database)
PIR	=	PIRA (Pira International Publishing and Information Services)
PLA	=	PLASPEC (PLASPEC Materials Selection Database)
POL	=	POLLUTION (Pollution Abstracts)
PRO	=	PRO (PROMT)
PST	=	PSTA (Packaging Science and Technology)
RAP	=	RAPPRA (RAPPRA Abstracts)
RTE	=	RTECS (Registry of Toxic Effects of Chemical Substances)
SCI	=	SCISEARCH (Science Citation Search System)
SPE	=	SPECINFO(SPECINFO)
TOX	=	TOXLINE (Toxicology Literature Online)
TRC	=	TRCTHERMO (Thermodynamic Research Center Database)
TSC	=	TSCAPP (Toxic Substances Control Act Plant and Production Search System)
UFO	=	UFOR (Umweltforschungsdatenbank)
ULI	=	ULIT (Umweltliteraturdatenbank)
USA	=	USAN (USA Adopted Names of Drugs)
VTB	=	VTB (Verfahrens-technische Berichte)

### CD-ROMs

AGI	=	AGRIS (International Information System for Agricultural Sciences and Technology)
AGR	=	AGREP (Agricultural Research Projects CD-ROM)
ARR	=	ARRIP (Australian Rural Research in Progress CD-ROM)
ARS	=	AGRSEARCH (AgriSearch on CD-ROM)
ASF	=	ASFA (Aquatic Sciences and Fisheries Abstracts on CD-ROM)
AST	=	ASTI (Applied Science and Technology Index)
BIA	=	Biolo. & Agri. Index (Biological and Agricultural Index on CD-ROM)
BIO	=	Biological Abstracts (BIOSIS on CD)
CAB	=	CABCD (CAB Abstracts on CD)
CAC	=	CASurveyor Cancer (CASurveyor Cancer Chemical Research)
CAE	=	CABPESTCD (CAB PEST CD-ROM)
CAF	=	CASurveyor Food (CASurveyor Food and Feed Chemistry)
CAH	=	CASurveyor Hazardous Materials
CAN	=	CANDATA (CCINFODisc Series B1, CANDATA)
CAV	=	CASurveyor Pollution (CASurveyor Pollution Control and the Environment)
CAW	=	CASurveyor Waste (CASurveyor Waste Management and Remediation)
CDR	=	CD-Römpp (Römpp Chemie Lexikon auf CD)
CER	=	CERCLA & RCRA (Environmental Factor System CD-ROM)
CHB	=	CHEM-BANK (CHEM-BANK on SilverPlatter)
CHS	=	CHEM Source (CCINFODisc Series A2, CHEM Source)
CR1	=	ChemReact 10 (ChemReact 10 with InfoChem CD-ROM)
CR3	=	ChemReact 32 (ChemReact 32 with InfoChem CD-ROM)
CRI	=	CRIS (Current Research Information System CD-ROM)
CUR	=	Current Facts (Beilstein Current Facts in Chemistry CD-ROM)
DID	=	Dictionary of Drugs (Dictionary of Drugs on CD-ROM)

DIN	=	Dictionary of Natural Products (Dictionary of Natural Products on CD-ROM)
DOC	=	Dictionary of Organic Compounds (Dictionary of Organic Compounds on CD-ROM)
ECD	=	E-CD (Environmental Quality from CAB International CD-ROM)
ECH	=	Environmental Chemistry (KR OnDisc Environmental Chemistry, Health and Safety)
ECO	=	ECODISC (Environmental Impact Statements)
ECP	=	Ei Compendex Plus (Computerized Engineering Abstracts)
EIC	=	Ei CHEM Disc (KR Information OnDisc Ei CHEN Disc)
EID	=	Ei Energy & Environment (KR Ondisc Ei Energy & Environment)
EIN	=	EINECS Plus (EINECS Plus CD-ROM)
EIS	=	Environmental Impact (Environmental Impact Statements CD-ROM)
EMB	=	EMBASE-CD: PolTox (Excerpta Medica Database on CD-ROM)
ENE	=	Environmental Manag. (Environmental Management CD-ROM)
ENM	=	Environmental Manag. (Environmental Management CD-ROM)
ESD	=	Earth Science Disc (Earth Science Disc)
ETI	=	Ei Tech Index (Engineering Information Tech Index)
GEB	=	GEOBASE (GEOBASE CD-ROM)
GEF	=	Gefahrgut CD-ROM (Dangerous Goods CD-ROM)
GIF	=	Giftliste (Giftliste, krebserregende, gesundheitsschädliche und reizende Stoffe)
GSA	=	General Science Abstracts CD-ROM
IAR	=	IARCancer Disk (International Agency for Research on Cancer CD-ROM)
IRI	=	IRIS (Integrated Risk Information System CD-ROM)
KIR	=	Kirk-Othmer (Kirk-Othmer Encyclopedia of Chemical Technology CD-ROM)
LAB	=	LABORDOC (Labordoc Database CD-ROM)
LSI	=	Life Science Collection (Life Science Collection CD-ROM)
MSD	=	MSDS (Materials Safety Data Sheets on SilverPlatter)
NIO	=	NIOSH/TIC/DIDS (CCINFODisc Series C1, NIOSH/TIC/DIDS)
OSD	=	OSH-CD (Occupational Health and Safety CD-ROM)
OEL	=	OCLC Environment (OCLC Environment Library CD-ROM)
OID	=	OSH INTERDATA (CCINFODisc Series B2, InterData)
OOO	=	OSH-OFFSHORE bundle ( OSH-CD / OSH Offshore CD-ROM bundle)
OSR	=	OSH-ROM (International Occupational Health and Safety CD-ROM)
PES	=	PEST-BANK (PEST-BANK on SilverPlatter)
POC	=	PolTox III: CAB (Pollution & Toxicology Database on CD-ROM)
PON	=	PolTox I: NLM, CSA, IFIS (Pollution and Toxicology Database on CD-ROM)
RDI	=	RDIR (Research and Development in Russia CD-ROM)
RTE	=	RTECS ((CCINFODisc Series C2, RTECS)
SAH	=	Sax, Hawley (Sax's Dangerous Properties of Industrial Materials / Hawley's Condensed Chemical Dictionary on CD-ROM)
SIG	=	SIGEDA (Siemens Gefahrstoff-Datenbank CD-ROM)
SOI	=	SOILCD (SOILCD on SilverPlatter)
SOR	=	SORBE (Sicherheitstechnische Kenndaten chemischer Stoffe CD-ROM)
TOP	=	TOXLINE Plus CD-ROM
TOX	=	TOXLINE on SilverPlatter (Toxicology Literature Online CD-ROM)
TRA	=	TRANSPORT (TRANSPORT on CD-ROM)
TRE	=	TREECD (TREECD on SilverPlatter)
TRI	=	TRIS (Toxic Release Inventory CD-ROM)
TRO	=	TROPAC & RURAL (Tropical Agriculture & Abstracts on Rural Development)
TSC	=	TSCA (TSCA Chemical Inventory on CD-ROM)
WAR	=	Water Resources Abstracts (Water Resources Abstracts on CD-ROM)
WAS	=	WasteInfo (Waste Information CD-ROM)
WTL	=	WATERLIT (Water Literature CD-ROM)

## Liste der Abkürzungen

## Inhaltsverzeichnis

<b>1. EINLEITUNG</b>	<b>14</b>
1.1 Proliferation an Chemikalien und korrespondierenden Publikationen	14
1.2 Notwendigkeit zur Erstellung von Metadatenbanken	15
1.3 Problemstellung: Notwendigkeit für die Auswahl und Bewertung von Datenquellen	15
1.4 Vorgehensweise bei der Bewertung von Datenquellen	15
<b>2. GRUNDLAGEN ÜBER DATENQUELLEN IN CHEMIE UND UMWELTSCHUTZ</b>	<b>17</b>
2.1 Anzahl der verfügbaren Datenquellen	17
2.2 Begriffserklärung Information, Datenbank, Informationssystem	18
2.2.1 Meta-Information, Metadatenbank, Meta-Informationssystem	18
2.2.2 Chemie-Information, Chemiedatenbank, Chemie-Informationssystem	19
2.2.3 Umwelt-Information, Umweltdatenbank, Umwelt-Informationssystem	20
2.2.4 Metadatenbanken, Meta-Informationssysteme in Chemie und Umweltschutz	20
2.3 Grundlagen der Online Datenbanken	21
2.4 Elektronische Informationsdienste	22
2.4.1 Datenbankproduzent (Datenbankhersteller)	23
2.4.2 Datenbankanbieter (Host)	23
2.5 Typisierung von Online Datenbanken gemäß Chemikalien- und Umweltschutzaspekten	24
2.5.1 Umweltchemikalien für die Recherche in verschiedenen Datenbanktypen	25
2.5.2 Faktendatenbanken (Modelldatenbanken der Naturwissenschaften)	26
2.6 CD-ROMs (Compact Disc Read Only Memory)	27
2.7 Internet Ressourcen	28
<b>3. HASSE-DIAGRAMM-TECHNIK ALS METHODE ZUR VERGLEICHENDEN BEWERTUNG</b>	<b>30</b>
3.1 Kurzer Überblick über alternative Bewertungsmethoden	30
3.2 Begründung der Wahl der Hasse-Diagramm-Technik als Bewertungsmethode	30
3.3 Kurzbeschreibung der Hasse-Diagramm-Technik	31
3.4 Schritte bei der vergleichenden Bewertung mittels der Hasse-Diagramm-Technik	32
3.4.1 Allgemeine Auswertung des Hasse-Diagramms	32
3.4.2 Analyse der Struktur des Hasse-Diagramms (D-Matrix: Schnittmengen-Matrix)	32
3.4.3 Bedeutung der Kriterien beim Rangfolgeverfahren (Sensitivitätsanalyse)	32

<b>4. ERSTELLUNG VON METADATENBANKEN FÜR UMWELTCHEMIKALIEN</b>	<b>33</b>
<b>4.1 Metadatenbank der Online Datenbanken - DADB</b>	<b>33</b>
4.1.1 Datenbankdefinition von DADB	33
4.1.2 Chemikalien-Testdatensatz	36
4.1.2.1 Auswahlverfahren	36
4.1.2.2 Klassifikation nach Strukturen bzw. funktionellen Gruppen	38
4.1.3 Parameter-Testdatensatz	43
4.1.3.1 Auswahlverfahren	43
4.1.3.2 Einteilung der Parameter in Anlehnung an das Chemikaliengesetz	44
4.1.4 Auswertung der Metadatenbank der Online Datenbanken (DADB) gemäß der Typisierung der Online Datenbanken	46
<b>4.2 Metadatenbank der CD-ROMs - DACD</b>	<b>47</b>
4.2.1 Datenbankdefinition von DACD	47
4.3.2 Auswertung der Metadatenbank der CD-ROMs (DACD) gemäß der Typisierung der CD-ROMs	48
<b>4.3 Gegenüberstellung der Datenbanken-Typen in DADB und DACD</b>	<b>48</b>
<b>5. AUSWAHL VON ONLINE DATENBANKEN UND CD-ROMS IN CHEMIE UND UMWELTSCHUTZ</b>	<b>49</b>
<b>5.1 Auswahl der Gruppen an Online Datenbanken</b>	<b>49</b>
5.1.1 Vorbemerkung	49
5.1.2 Chemie-Gruppe	50
5.1.2.1 Übersicht	50
5.1.2.2 Einengende Auswahl	50
5.1.3 Umwelt-Gruppe	52
5.1.3.1 Übersicht	52
5.1.3.2 Einengende Auswahl	53
5.1.4 Diskussion und Gegenüberstellung der Chemie- und Umwelt-Gruppe	54
5.1.5 Auswahl der Online Datenbankgruppen für das Bewertungszifferverfahren	55
5.1.5.1 Allgemeine Grundsätze zur Auswahl	55
5.1.5.2 Bildung der Datenbankgruppen 1-5	55
<b>5.2 Auswahl der Gruppen an CD-ROMs</b>	<b>56</b>
5.2.1 Vorbemerkung	56
5.2.2 Chemie- bzw. Chemikalien spezifische CD-ROMs	57
5.2.3 Umwelt- bzw. Ökologie-spezifische CD-ROMs	58
5.2.4 Diskussion und Gegenüberstellung der Chemie- und Umwelt-Spezialgebiete	60
5.2.5 Auswahl der Gruppen an CD-ROMs für das Bewertungszifferverfahren	61
5.2.5.1 Allgemeine Grundsätze zur Auswahl	61
5.2.5.2 Bildung der CD-ROM Gruppen 1-5	61
<b>6. BEWERTUNGSVERFAHREN FÜR DATENQUELLEN</b>	<b>62</b>
<b>6.1 Notwendigkeit und Bedeutung der Bewertung</b>	<b>62</b>
<b>6.2 Bewertungsziffern</b>	<b>62</b>
<b>6.3 Einsatz ordinaler und nominaler Kriterien bei der Bewertung von Datenquellen</b>	<b>63</b>
<b>6.4 Bewertungskriterien für Online Datenbanken und CD-ROMs</b>	<b>64</b>
6.4.1 Allgemeine Bewertung der Datenquellen	64
6.4.1.1 Umfang der Datenbank (UM)	64
6.4.1.2 Häufigkeit der Aktualisierung (UP)	65
6.4.1.2.1 Häufigkeit der Aktualisierung bei Online Datenbanken	65

	10
6.4.1.2.2 Häufigkeit der Aktualisierung bei CD-ROMs	65
6.4.1.3 Kosten (KO)	66
6.4.1.3.1 Kosten der Recherche	66
6.4.1.3.2 Kosten der CD-ROM	67
6.4.1.4 Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien (VE)	67
6.4.2 Bewertung nach inhaltlichen Kriterien der Chemie	69
6.4.2.1 Anzahl der Chemikalien (AN)	70
6.4.2.2 Identifikationsmerkmale (ID)	70
6.4.2.3 Chemikalien-Testdatensatz (CT)	73
6.4.2.4 Chemikalien Entwicklung (CE)	74
6.4.3 Bewertung nach inhaltlichen Kriterien des Umweltschutzes	75
6.4.3.1 Informationsparameter für Umweltchemikalien (IP)	75
6.4.3.2 Parameter Entwicklung (PE)	77
6.4.4 Umweltchemikalien-bezogene Bewertungskriterien	77
6.4.4.1 Art der Chemikalien (AR)	77
6.4.4.2 Validität/Güte der Datenquellen (QU)	78
6.4.4.2.1 Validität/Güte der Online Datenbanken	78
6.4.4.2.2 Validität/Güte der CD-ROMs	80
<b>6.5 Gesamtmaß <math>\Gamma</math> als grobes Einstufungsmaß für Objekte</b>	<b>82</b>
<b>6.6 Bedeutung der Bewertungskriterien und deren Ausprägungsmengen</b>	<b>82</b>
6.6.1 Umfangsindex $U_i$	83
6.6.2 Shannon-Index $I_i$	83
6.6.3 Interpretation der Indizes	83
<b>7. HASSE-DIAGRAMME FÜR DIE AUSGEWÄHLTEN GRUPPEN AN DATENQUELLEN</b>	<b>1</b>
<b>7.1 Besonderheiten der Interpretationen und Übersicht</b>	<b>84</b>
<b>7.2 Hasse-Diagramme für Online Datenbanken</b>	<b>85</b>
7.2.1 Bibliographische Online Datenbanken der Chemie-Gruppe (Gruppe 1)	85
7.2.1.1 Allgemeine und inhaltliche Bewertungskriterien	86
7.2.1.1.1 Auswertung des Hasse-Diagramms	87
7.2.1.1.2 Analyse anhand der D-Matrix	88
7.2.1.1.3 Bedeutung der Bewertungskriterien	89
7.2.1.2 Allgemeine Bewertungskriterien	92
7.2.1.2.1 Auswertung des Hasse-Diagramms	94
7.2.1.2.2 Analyse anhand der D-Matrix	94
7.2.1.2.3 Analyse anhand der W-Matrix	96
7.2.1.3 Inhaltliche Bewertungskriterien	98
7.2.1.3.1 Auswertung des Hasse-Diagramms	99
7.2.1.3.2 Analyse anhand der D-Matrix	100
7.2.1.3.3 Bedeutung der Bewertungskriterien	101
7.2.1.4 Zusammenfassende Analyse der Gruppe 1: Online Datenbanken	105
7.2.1.4.1 Vergleichende Interpretation der Hasse-Diagramme	105
7.2.1.4.2 Gegenüberstellung einzelner Datenbanken im Hasse-Diagramm mit dem Gesamtmaß $\Gamma$	107
7.2.1.4.3 Bewertungskriterien und deren Ausprägungsmengen	108
7.2.2 Faktendatenbanken der Chemie-Gruppe (Gruppe 2)	110
7.2.3 Gruppe 1 und Gruppe 2 der Online Datenbanken im Vergleich	113
7.2.4 Textbasierte Datenbanken der Umwelt-Gruppe (Gruppe 3)	114
7.2.5 Umweltrelevante Faktendatenbanken (Gruppe 4)	117
7.2.6 Gruppe 3 und Gruppe 4 der Online Datenbanken im Vergleich	120
7.2.7 Umweltchemikalien-relevante Datenbanken der Chemie/Umwelt-Gruppe (Gruppe 5)	121
<b>7.3 Hasse-Diagramme für CD-ROMs</b>	<b>123</b>
7.3.1 Textbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 1)	123

7.3.2	Faktenbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 2)	126
7.3.2.1	Allgemeine und inhaltliche Bewertungskriterien	127
7.3.2.2	Allgemeine Bewertungskriterien	129
7.3.2.3	Inhaltliche Bewertungskriterien	130
7.3.2.4	Gegenüberstellung einzelner CD-ROMs im Hasse-Diagramm mit dem Gesamtmaß $\Gamma$	131
7.3.2.5	Bewertungskriterien und deren Ausprägungsmengen: Gruppe 2 (CD-ROMs)	132
7.3.3	Gruppe 1 und Gruppe 2 der CD-ROMs im Vergleich	133
7.3.4	Textbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 3)	133
7.3.5	Faktenbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 4)	136
7.3.6	Gruppe 3 und Gruppe 4 der CD-ROMs im Vergleich	139
7.3.7	Umweltchemikalien-relevante CD-ROMs des Chemie/Umwelt-Spezialgebietes (Gruppe 5)	139
<b>8.</b>	<b>DISKUSSION DER ERGEBNISSE DER BEWERTUNG</b>	<b>143</b>
8.1	Analyse der Ergebnisse der Bewertung der Online Datenbanken	143
8.2	Analyse der Ergebnisse der Bewertung der CD-ROMs	144
8.3	Auswertung und Gegenüberstellung der Medien Online/CD-ROM	145
8.3.1	Bedeutung der einzelnen Bewertungskriterien in zehn Gruppen	145
8.3.2	Kriterienmaß $\vartheta$ als grobes Einstufungsmaß für Kriterien	148
8.3.3	Hasse-Diagramm für zehn Gruppen an Datenquellen	148
8.3.4	Vergleich des Hasse-Diagramms mit dem Gesamtmaß $\Gamma$	151
8.3.5	Auswertung der Ergebnisse des Kriterienmaßes $\vartheta$ für acht Bewertungskriterien	152
8.4	Reflexion der Bewertungskriterien	152
8.4.1	Nicht-berücksichtigte Bewertungskriterien	152
8.4.2	Qualität der Bewertungskriterien	154
8.4.3	Einteilung und Vergabe der Bewertungsziffern	154
<b>9.</b>	<b>ZUSAMMENFASSUNG</b>	<b>155</b>
9.1	Gruppierung von Datenquellen	155
9.2	Auswahl von Bewertungskriterien	155
9.3	Bewertungsziffern	157
9.4	Bedeutende Datenquellen	157
9.5	Bedeutung der Bewertungskriterien bei der Bewertung	158
9.6	Vergleichende Bewertung der Medien Online/CD-ROM	158
<b>10.</b>	<b>AUSBLICK</b>	<b>159</b>
	<b>LITERATURVERZEICHNIS</b>	<b>160</b>
	<b>ANHANG A. RECHERCHEBEISPIELE ZU DATENQUELLEN IN CHEMIE UND UMWELTSCHUTZ</b>	<b>169</b>
A.1	Faktenbasierte Datenbanken	169
A.1.1	Faktendatenbanken ( Modelldatenbanken der Naturwissenschaften)	169
A.1.2	Chemikaliennamen-Verzeichnisse	169
A.1.3	Firmenverzeichnisse und andere Verzeichnisse	170

	12
A.1.4 Forschungsdatenbanken	171
A.1.5 Strukturdatenbanken (Formalismendatenbanken der Naturwissenschaften)	172
A.1.6 Metadatenbanken	174
<b>A.2 Textbasierte Datenbanken</b>	<b>175</b>
A.2.1 Bibliographische Datenbanken	175
A.2.2 Volltextdatenbanken	176
<b>A.3 Integrierte Datenbanken</b>	<b>177</b>
A.3.1 Reaktionsdatenbanken	178
A.3.2 Patentdatenbanken	180
<b>A.4 Zusammenfassung</b>	<b>182</b>
<b>ANHANG B. HASSE-DIAGRAMME ALS WERKZEUG ZUR BEWERTUNG</b>	<b>183</b>
<b>B.1 Grundlagen der Hasse-Diagramm-Technik</b>	<b>183</b>
B.1.1 Permutationsdiagramm versus Hasse-Diagramm	184
B.1.2 Definition von Grundbegriffen	185
B.1.3 Menge der Attribute	186
<b>B.2 Graphen und Hasse-Diagramme</b>	<b>186</b>
B.2.1 Lesen von Hasse-Diagrammen	186
B.2.2 Verbandstheoretische Bezeichnungen	187
B.2.3 Kenngrößen (Maßzahlen) zur Unterstützung der Interpretation von Hasse-Diagrammen	187
<b>B.3 Schlüsselemente und Nachfolgermengen</b>	<b>187</b>
<b>B.4 D-Matrix - Schnittmengen-Matrix</b>	<b>188</b>
<b>B.5 Bedeutung der Attribute beim Rangfolgeverfahren</b>	<b>188</b>
<b>ANHANG C. DETAILLIERTERE BEWERTUNGSERGEBNISSE ZU ZEHN GRUPPEN AN DATENQUELLEN</b>	<b>190</b>
<b>C.1 Bibliographische Online Datenbanken der Chemie-Gruppe (Gruppe 1)</b>	<b>190</b>
<b>C.2 Faktendatenbanken der Chemie-Gruppe (Gruppe 2)</b>	<b>190</b>
C.2.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 2 (Online Datenbanken)	191
C.2.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 2 (Online Datenbanken)	192
<b>C.3 Textbasierte Datenbanken der Umwelt-Gruppe (Gruppe 3)</b>	<b>193</b>
C.3.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 3 (Online Datenbanken)	194
C.3.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 3 (Online Datenbanken)	195
C.3.3 Besonderheiten in der Struktur des Hasse-Diagramms der Gruppe 3: allgemeine Bewertungskriterien	196
<b>C.4 Umweltrelevante Faktendatenbanken (Gruppe 4)</b>	<b>198</b>
C.4.1 Besonderheiten in der Struktur des Hasse-Diagramms der Gruppe 4: Online Datenbanken	198
C.4.2 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 4 (Online Datenbanken)	200
C.4.3 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 4	201
<b>C.5 Umweltchemikalien-relevante Datenbanken (Gruppe 5)</b>	<b>202</b>
C.5.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 5 (Online Datenbanken)	203
C.5.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 5	204
<b>C.6 Textbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 1)</b>	<b>205</b>
C.6.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 1 (CD-ROMs)	206

C.6.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 1 CD-ROMs	207
<b>C.7 Faktenbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 2)</b>	<b>208</b>
<b>C.8 Textbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 3)</b>	<b>208</b>
C.8.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 3 (CD-ROMs)	209
C.8.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 3 CD-ROMs	210
<b>C.9 Faktenbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 4)</b>	<b>211</b>
C.9.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 4 (CD-ROMs)	212
C.9.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 4 CD-ROMs	213
<b>C.10 Umweltchemikalien-relevante CD-ROMs (Gruppe 5)</b>	<b>213</b>
C.10.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 5 (CD-ROMs)	214
C.10.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 5 CD-ROMs	215

## Lebenslauf

## 1. Einleitung

### 1.1 Proliferation an Chemikalien und korrespondierenden Publikationen

Im Zeitraum zwischen 1830 und 1960 wurden ungefähr 1.5 Millionen chemische Verbindungen beschrieben, das bedeutet durchschnittlich 11.500 Verbindungen pro Jahr. Zwischen 1960 und 1980 wurden weitere 3,5 Millionen Verbindungen publiziert, das wiederum korrespondiert mit einer durchschnittlichen Anzahl von 175.000 Verbindungen pro Jahr [Lankenau 1992]. Neben der Proliferation der beschriebenen chemischen Verbindungen ist auch eine starke Zunahme der Produktion der Menge an Chemikalien zu verzeichnen. Während der vergangenen 50 Jahre hat sich die Produktionsmenge an synthetischen organischen Chemikalien in den Vereinigten Staaten um das 20fache erhöht, d.h. von einem jährlichen Produktionsvolumen von ca. 4,5 Millionen Tonnen 1943 auf ein Produktionsvolumen im Jahre 1989 von ca. 61 Millionen Tonnen [Metcalf 1993]. Die aktuellen Zahlen für das Produktionsvolumen der 50 höchstproduzierten Chemikalien in den USA beträgt für das Jahr 1995 sogar 340 Millionen Tonnen [Kirschner 1996].

Die Industrieländer versuchen seit einigen Jahren bzw. Jahrzehnten durch die Etablierung von Gesetzen, die Probleme zu steuern, die sich insbesondere in bezug auf den Umwelt-, Gesundheits- und Arbeitsschutz in Zusammenhang mit der enormen Anzahl und Menge der Chemikalien ergeben. In der Bundesrepublik ist beispielsweise das Chemikaliengesetz (ChemG) seit 1981 in Kraft. Zweck dieses Gesetzes ist es, den Menschen und die Umwelt vor schädlichen Einwirkungen gefährlicher Stoffe und Zubereitungen zu schützen, insbesondere sie erkennbar zu machen, sie abzuwenden und ihrem Entstehen vorzubeugen [ChemG 1994]. Im Chemikaliengesetz werden alte und neue Stoffe unterschieden. Unter alten Stoffen versteht man Stoffe, die im Altstoffverzeichnis der Europäischen Gemeinschaften EINECS (European Inventory of Existing Commercial Chemical Substances) in der jeweils neuesten veröffentlichten Fassung verzeichnet sind. Es sind dort ca. 101.000 Chemikalien erfaßt. Neue Stoffe, sind solche Stoffe, die nicht in dem EINECS Verzeichnis aufgeführt sind. Für die sog. neuen Stoffe besteht eine Anmeldepflicht, die neben Angaben zur Identität der Chemikalie, deren Herstellung, Entsorgung etc. auch noch umfangreiche Prüfnachweise für eine Anzahl von Kriterien vorsieht. Diese Kriterien werden im wesentlichen auch die Grundlage für die Auswahl der in dieser Arbeit berücksichtigten Parameter bilden.

Die Zunahme an Chemikalien und die damit im Zusammenhang stehenden Umweltgesetze bedingen selbstverständlich auch eine Erhöhung der Publikationen auf diesen Fachgebieten. Jeden Tag erscheinen ca. 20.000 wissenschaftlich/technische Publikationen weltweit, das ergibt ca. 5 Millionen Publikationen pro Jahr. Auf dem Gebiet der Chemie und des Umweltschutzes liegt die Schätzung bei ca. einer Million wissenschaftlichen Artikeln pro Jahr [Lankenau 1992]. Betrachtet man alleine die Chemie, so sollen ca. 500.000 Veröffentlichungen pro Jahr erscheinen [Luckenbach 1996].

Die Proliferation an Information auf dem Gebiet der Chemie und des Umweltschutzes führt somit konsequenterweise zu Bestrebungen, die Daten geeignet zu speichern, zu verwalten, zu pflegen und aufzubereiten. Wie die Autoren Zupan und Gasteiger formulieren, wird "Information" in Zukunft ein extrem bedeutsamer Wertfaktor in vielen Aktivitäten des täglichen Lebens bis hin zu den Naturwissenschaften werden [Zupan 1993, S. 3]. Es ist jedoch vielfach festzustellen, daß wir von einer Datenflut überschwemmt werden, und daß es immer schwieriger wird, die richtige Information für eine spezielle Fragestellung zu erhalten. Aus diesem Grunde wurden und werden Datenbanken aufgebaut und angeboten.

## 1.2 Notwendigkeit zur Erstellung von Metadatenbanken

Wie im Laufe der Arbeit ausgeführt wird, gibt es eine Vielzahl an Datenbanken, in denen Informationen zu Chemikalien und Umweltschutzaspekten vorgehalten werden. Diese Datenbanken unterscheiden sich z.B. in ihrer Datenbankstruktur, ihrem Umfang, ihren Themenschwerpunkten, ihren Inhalten, ihrer Aktualität, ihrem Zugriffsmedium, ihrem Preis etc. Zur Lösung von Fragestellungen auf dem Gebiet der Chemie und des Umweltschutzes ist die Kenntnis von geeigneten Datenquellen von entscheidender Bedeutung. Es muß daher zunächst eine Unterstützung zur Findung der geeigneten Datenquelle(n) geschaffen werden. Dafür eignen sich sog. **Metadatenbanken** (Datenbanken der Datenquellen). In diesen Metadatenbanken werden Informationen über Datenquellen, in dieser Arbeit auf den Gebieten der Chemie und des Umweltschutzes, strukturiert aufbereitet und vorgehalten. Die Metadatenbanken werden nach den Zugriffsmedien (online und CD-ROM) getrennt. Es gibt dementsprechend eine Metadatenbank der Online Datenbanken und eine Metadatenbank der CD-ROMs. Die Frage "In welcher Datenquelle bzw. in welchen Datenquellen findet man Daten über Bioabbau oder Fischtoxizität zu Pestiziden" kann beispielsweise durch Nutzung von Metadatenbanken in Chemie und Umweltschutz zufriedenstellend beantwortet werden.

## 1.3 Problemstellung: Notwendigkeit für die Auswahl und Bewertung von Datenquellen

Da die Anzahl der in den Metadatenbanken für Chemie und Umweltschutz enthaltenden Datenbanken sehr umfangreich ist und verschiedene Datenbanktypen vorliegen, ist eine **Auswahl** und **Gruppierung** der Datenquellen vordringlich. Hier muß zunächst eine themenspezifische und Datenbanktyp-spezifische Auswahl getroffen werden.

Von übergeordneter Bedeutung ist die **Bewertung** der Auswahl der Gruppen an Datenquellen. Die Beantwortung der Frage "Wie gut ist die gefundene Datenquelle bzw. sind die gefundenen Datenquellen im Vergleich zu anderen?" ist das **wesentliche Ziel** dieser Arbeit. Für den Informationssuchenden sind gerade die Antworten auf die Fragen zur Qualität der Datenquellen von größtem Interesse. Die Erarbeitung von geeigneten Kriterien (**Bewertungskriterien**) und deren Bedeutung ist für die vergleichende Bewertung ein entscheidender Schritt. Die Güte hängt nicht alleine von den Inhalten der Datenquellen sondern auch von allgemeinen Kriterien z.B. der Aktualität und der Verfügbarkeit der Datenquelle ab.

Neben der Bewertung von einzelnen Gruppen an Datenquellen ist auch von großem Interesse, welches Medium - Online Datenbanken oder CD-ROMs - geeigneter erscheint.

## 1.4 Vorgehensweise bei der Bewertung von Datenquellen

Das hier aufgezeigte Bewertungsverfahren für Datenquellen in Chemie und Umweltschutz umfaßt die Arbeitsschritte:

- Auswahl bzw. Gruppierung von Datenquellen,
- Auswahl geeigneter Bewertungskriterien für die Datenquellen,
- Zuordnung von Bewertungsziffern für Datenquellen entsprechend der Bewertungskriterien und
- die eigentliche vergleichende Bewertung mittels des mathematischen Verfahrens der Hasse-Diagramm-Technik.

Bei dieser Vorgehensweise stellen sich folgende Fragen:

1. Wie lassen sich die Datenquellen (Online Datenbanken und CD-ROMs) sinnvoll gruppieren?

2. Welche Bewertungskriterien sind für die Datenquellen in Chemie und Umweltschutz besonders bedeutsam?
3. Wie können die qualitativen und quantitativen Bewertungskriterien in einer in sich konsistenten Weise klassifiziert werden?
4. Gibt es innerhalb der erstellten Gruppen von Datenquellen unter den gewählten Kriterien besonders empfehlenswerte bzw. besonders unbedeutende Datenquellen, die durch einen Bewertungsansatz herausgearbeitet werden können?
5. Welche Bedeutung haben die Bewertungskriterien?
6. Welche Vergleiche lassen sich zwischen den Gruppierungen finden?

Die Beantwortung dieser Fragen ist der wesentliche Bestandteil dieser Arbeit und eine Zusammenfassung hierüber wird am Ende in Kapitel 9 gegeben.

Die Auswertungen der Datenquellen sind von Mitte 1995 bis Anfang 1996 erbracht worden. Neuere Entwicklungen auf dem Datenbanksektor können selbstverständlich zu anderen Ergebnissen im Bewertungsansatz führen.

## 2. Grundlagen über Datenquellen in Chemie und Umweltschutz

### 2.1 Anzahl der verfügbaren Datenquellen

Die intensiven Forschungsarbeiten auf dem Gebiet der organische Chemikalien in der Umwelt haben zur Einführung der Disziplin "Environmental Organic Chemistry" (Organische Chemie im Umweltschutz) geführt. Dieses Gebiet ist historisch aus der Analytischen und Organischen Chemie hervorgegangen, insbesondere der Möglichkeit in der Analytischen Chemie kleinste Konzentrationen von Schadstoffen in komplexen Matrizes wie z.B. Böden, atmosphärischen Partikeln, Tiergeweben und Pflanzen zu analysieren. In diesem Zusammenhang ist auch die Anzahl der Forschungsarbeiten auf dem Gebiet der organischen Reaktionen unter Umweltbedingungen gestiegen. Viele Arbeiten zeigen, daß die Abbauprodukte der Chemikalien viel gefährlicher sind als die Ausgangsstoffe. Eine kürzlich publizierte Übersichtsdarstellung der Arbeiten auf dem Gebiet der Organischen Chemie in bezug auf Umweltschutz zeigt verschieden Reaktionstypen, die unter Umweltbedingungen auftreten [Larson 1994]. Hier werden beispielsweise Hydrolyse und nukleophile Reaktionen, Reduktionen, Oxidationen, Reaktionen mit Desinfektionsmitteln, photochemische Reaktionen (Photochemie) und Diels-Alder-Reaktionen betrachtet. In einer weiteren Publikation aus dem Bereich der Umweltchemie bzw. Umweltchemikalien wird speziell auf das Umweltmedium Atmosphäre (Luft) eingegangen. Die Chemikalien werden nach ihren chemischen Strukturen geordnet, deren Quellen und Vorkommen dargestellt [Graedel 1986]. Ebenso werden differenzierte Einteilungen in ausgewählte Substanzgruppen von Umweltchemikalien in bezug auf deren toxikologische Eigenschaften vorgenommen [Eisenbrand 1994].

Die ständig anwachsende Informationsflut erfordert auf der Nachfragerseite ein geeignetes Instrument zu ihrer Bewältigung, ein gezieltes Informationsmanagement, u.a. für die Ermittlung, Beschaffung und Verteilung von Informationen. Wesentlich ist hier die genaue Kenntnis aller Informationsquellen (Informationsprodukte), die für spezielle Fragestellungen von Interesse sein können [Hoppenstedt Verlag 1992, S. 18]. Dazu gehören heute zunehmend Online Datenbanken (online beschreibt die Art des Zugangs zur externen Datenbank) und CD-ROMs (Compact Disc Read Only Memory). Mit der Verbreitung des Internets ist selbstverständlich auch die Anzahl der Datenquellen im Internet in den Fachgebieten Chemie und Umweltschutz in letzter Zeit stark gestiegen und wird auch weiterhin zunehmen.

Man unterscheidet also folgende Informationsprodukte:

- gedruckte Dokumente
- Online Datenbanken
- CD-ROMs und Datensammlungen auf Diskette
- Internet Ressourcen

Die Anzahl der Online Datenbanken und CD-ROMs ist im Zeitraum von 1975 bis 1995 von 301 bis 9.207 d.h. um den Faktor 31 gestiegen [Williams 1996]. In der größten Zusammenstellung von Datenbanken weltweit, dem Gale Directory of Databases [Williams 1996, S.1], das sowohl Online Datenbanken als auch CD-ROMs umfaßt, unterteilt man die 9.207 Datenbanken nach folgenden Sachgebieten: Naturwissenschaft, Technologie, Ingenieurwesen, Wirtschaft, Gesundheitswesen, Sozialwissenschaften, Geisteswissenschaften, Gesetze und Nachrichten allgemein. Die Sachgebiete Naturwissenschaft, Technologie, Ingenieurwesen sowie Gesundheitswesen sind für die Thematik der Chemie und des Umweltschutzes von Interesse. Einige Datenbanken aus anderen Bereichen, z.B. Gesetzestexte oder Nachrichten allgemein könnten in diesem Zusammenhang ebenso wichtig sein.

Die Auswertung des Gale Directories of Databases ergibt für die Fachdisziplin Wissenschaft, Technologie, Ingenieurwesen einen Anteil von 19 % an der Gesamtheit der Datenbanken und für die Thematik Gesundheitsschutz einen Prozentsatz von 9. Die in Kapitel 4 noch näher be-

schriebenen Metadatenbanken für Chemie und Umweltschutz, die Metadatenbank der Online Datenbanken - DADB und die Metadatenbank der CD-ROMs - DACD [Voigt 1992] haben einen Dokumentenstand (Februar 1997) DADB: 453 und DACD: 347 Dokumente, d.h. 800 Online Datenbanken/CD-ROMs enthalten Informationen über das Thema Chemikalien und Umweltschutz.

## 2.2 Begriffserklärung Information, Datenbank, Informationssystem

In der Fachliteratur wird heute von Informationen, Datenbanken und Informationssystemen gesprochen. Bevor auf die speziellen Begriffe der Meta-, Chemie- und Umwelt- Informationen, Datenbanken und Informationssysteme eingegangen wird, müssen zunächst die Ausgangsbegriffe erläutert werden. Die in Tabelle 1 aufgeführte Klassifikation von Informationsbegriffen wird im Kontext, d.h. Bewertung von Chemie- und Umweltdatenbanken mittels der Auswertung der Metadatenbanken für Chemikalien und Umweltschutz, behandelt.

**Informationen** im Sinne von Wissen ist für Zwecke der Informationsarbeit im Hinblick auf seine Wirkung beim Informationsaufnehmer beschrieben worden: Information ist die Verringerung von Ungewißheit aufgrund von fachlichen Kommunikationsprozessen [Seeger 1991]. Unter dem Begriff **Datenbank** versteht man die Zusammenfassung von computerlesbaren Daten in organisierter Form aus einem bestimmten Themengebiet [Barth 1992, S.402].

**Tabelle 1: Klassifikation von Informationsbegriffen**

Information	Datenbank	Informationssystem
Meta-Information	Metadatenbank	Meta-Informationssystem
Chemie-Information	Chemiedatenbank	Chemie-Informationssystem
Umwelt-Information	Umweltdatenbank	Umwelt-Informationssystem

Ein konventionelles **Informationssystem** ist ein System zur Speicherung, Organisation, Integration und Wiedergabe von Daten unterschiedlichster Herkunft und Struktur: Strukturierte Daten z.B. Strukturformeln und Graphiken, Meßwerte aus Überwachungssystemen, formatierte Faktendaten, unformatierte Dokumente [Page 1994].

### 2.2.1 Meta-Information, Metadatenbank, Meta-Informationssystem

Eine verbindliche Begriffserklärung des Begriffes Metadatenbank bzw. Meta-Informationssystem oder eine anerkannte begriffliche Abgrenzung von Sachinformation, Kontextinformation und Meta-Information existieren bislang nicht. Das gilt in gleichem Maße für die Schreibweisen der Begriffe. Mit den Begriffen **Meta-Information** und **Meta-Informationssystem** wird eines der wichtigsten Konzepte zur Identifizierung und Nutzbarmachung von Information in umfangreichen Datenbeständen gekennzeichnet [Greve 1995].

Die Meta-Informationen können wie folgt eingeteilt werden:

- semantische Meta-Information, d.h. die Beschreibung des Inhaltes der Informationsquelle
- syntaktische Meta-Information, d.h. die formale Struktur der Daten (Datentypen oder Zugriffsmethoden)

- pragmatische Meta-Information, d.h. Ort der Quellen, technische und organisatorische Anforderungen für den Zugriff auf die Informationen [Greve 1994].

Ein **Meta-Informationssystem** ist ein Informationssystem, das Informationen über das Vorhandensein und den Zugang zu anderen Informationen bzw. Daten speichert [Denzer 1994].

In diesem Kontext wird unter dem Begriff **Metadatenbanken** Datenbanken verstanden, welche Inhalte über andere Datenquellen umfassen und vorhalten. Eine Einteilung in verschiedene Metadatenbank-Klassen wurde anlässlich einer eigens zu diesem Thema durchgeführten Veranstaltung auf der Internationalen Online Konferenz 1993 in London vorgenommen. Es werden hier Metadatenbanken in bezug auf das Zugriffsmedium (z.B. Online Datenbanken, CD-ROMs) nach Regionen, nach Sachgebieten etc. unterschieden [Voigt 1993].

Die umfangreichste Metadatenbank ist die weltweit größte Zusammenstellung von Datenbanken (Online Datenbanken und CD-ROMs), das Gale Directory of Databases 1996 [Young Marcaccio, 1996]. In dieser Metadatenbank sind die Standardwerke Computer-Readable Databases, Cuadra/Gale's Directory of Online Databases und Directory of Portable Databases vereinigt.

### 2.2.2 Chemie-Information, Chemiedatenbank, Chemie-Informationssystem

Die Chemie ist ein Wissensgebiet, das besonders geeignet ist, mit Hilfe des Computers neue Möglichkeiten der Problemlösung erschlossen zu bekommen. Chemisches Wissen gründet sich weniger auf explizit angebbare, lösbare mathematische Gleichungen als vielmehr auf Gesetzmäßigkeiten und Regeln, die aus vielen Einzelbeobachtungen, aus einer Fülle von Informationen abgeleitet werden.

Die GDCh - Fachgruppe "Chemie-Information" erkannte die Zeichen der Zeit und rief im November 1985 eine Arbeitsgruppe ins Leben, die sich die Bezeichnung "CIC - Computer in der Chemie" gab. Diese Fachgruppe ist für alle traditionellen Einzeldisziplinen offen. Ein Leitgedanke dieser Arbeitsgruppe ist "**Information ist die Grundlage jeder chemischen Wissenschaft**". Zur ersten wissenschaftlichen Bestandsaufnahme und um fachübergreifende Kontakte zu knüpfen, wurde im November 1986 in Hochfilzen/Tirol der Workshop "Software-Entwicklung in der Chemie" veranstaltet [Gasteiger 1987]. Seitdem finden jährlich Workshops zu diesem Thema an unterschiedlichsten Orten statt.

Die Aufgabe der Chemie-Information im engeren Sinn ist es, publizierte Information (wieder) zu finden; im weiteren Sinn, Information in der Chemie ganz allgemein verfügbar zu machen [Zass 1994].

Wie schon erwähnt, erschwert das rasante Anwachsen der chemisch-technischen Forschung in den letzten Jahren dem Chemiker das Wiederauffinden veröffentlichter Forschungsergebnisse. Es ist jedoch nicht allein die Informationsflut, die auf dem Gebiet der Chemie sowohl Informations- als auch Kommunikationsprobleme aufwirft.

Vielmehr sind es

- spezielle Informationskategorien,
- spezielle Arten von Fragestellungen und
- die Art und Weise, wie sich Chemie-fachliche Informationen in der Literatur darstellen, die zu einer, dieses Fachgebiet kennzeichnenden, zusätzlichen Problematik führen.

Insgesamt hat die **Chemie-Information** vier Kategorien von Informationen zu berücksichtigen:

1. chemische Verbindungen und Verbindungstypen (= strukturelle Informationen)
2. chemische Reaktionen und Reaktionstypen (= Informationen über das chemische Verhalten)
3. nichtstrukturbezogene Informationen (= Sachverhalte) und
4. die ein Dokument formal kennzeichnenden bibliographischen Daten.

Die beiden zuletzt genannten Informationskategorien finden sich auch in allen Informationssystemen anderer Fachgebiete, während die strukturellen Informationen und die Informationen zu Reaktionen ausschließlich in der Chemie-Information eine Rolle spielen [Berger 1991].

Eine *Chemiedatenbank* ist demnach eine Datenbank, die mindestens eine der oben genannten vier Kategorien von Chemie-Informationen beinhaltet, abspeichert und vorhält.

Unter einem integrierten *Chemie-Informationssystem* versteht man ein System zur Speicherung und Wiederfindung von chemierelevanten Informationen, das u. a. folgende Komponenten umfaßt: direkte Verbindung zwischen verschiedenen Datenbanken, Integration von Anwendungssoftware, Verbindung mit anderen, heterogenen Informationssystemen und Unterstützung durch dezentrale Benutzersoftware [Barth 1992, S. 401].

### **2.2.3 Umwelt-Information, Umweltdatenbank, Umwelt-Informationssystem**

*Umwelt-Informationen* besitzen einen fachlichen, räumlichen und zeitlichen Bezug [Pillmann 1994]. Der Begriff *Umweltdatenbank* ist schwierig einzugrenzen und zu definieren. Generell gesprochen ist eine Umweltdatenbank ein spezieller Typ von Datenbanken, die hauptsächlich Umweltdaten speichert [Page 1994, S. 40]. Im "Handbuch der Informatik: Umweltinformatik" [Page 1994] ist dieser Begriff näher erläutert.

Als Umweltdatenbank wird eine Datenbank bezeichnet, die die folgenden Bedingungen erfüllt:

1. Es sind überwiegend Umweltdaten gespeichert.
2. Es wird ein Datenbanksystem zur Speicherung der Daten verwendet.
3. Die Datenbank wurde als Grundlage für umweltbezogene Anwendungen oder Auskünfte angelegt [Baumewerd-Ahlmann 1994].

Werden Chemikalien im Umweltbereich betrachtet, muß die Definition entsprechend erweitert werden:

4. Die Mehrheit der Daten sind Informationen und Daten über Chemikalien.
5. Die Datenbank wurde als Grundlage für Chemikalien-bezogene Anwendungen oder Auskünfte angelegt.

*Umwelt-Informationssysteme* können aufgrund des häufigen Raumbezuges der gespeicherten Daten, d. h. ihrer geometrischen bzw. geographischen Beschreibung, als erweiterte Geographische Informationssysteme (GIS) angesehen werden, die zusätzlich (Sach-)Daten mit thematischem Bezug (Umweltfakten wie Meßwerte, Stoffdaten oder Umweltdokumente wie Forschungsvorhaben, Gesetzestexte oder Literaturstellen) und zeitlichem Bezug (z.B. Nutzungsänderungen von Sanierungsflächen, saisonale Schwankungen der Schadstoffbelastungen) enthalten können. Darüber hinaus besitzen Umweltdatenobjekte (z.B. chemische Stoffe) vielfach eine komplexe Struktur und vielfältige Verknüpfungen. Will man solche Datenobjekte auf die Datenstrukturen herkömmlicher (Standard-)Datenbanksysteme abbilden, so stößt man auf prinzipielle Schwierigkeiten, da unterschiedliche Aspekte der Umwelt-Informationsverarbeitung bisher nur durch unterschiedliche Typen von Datenbanksystemen adäquat unterstützt werden können. Objekt-orientierte Datenbanken bieten die Möglichkeiten, die heterogenen Umwelt-Informationen besser zu verarbeiten [Page 1996].

### **2.2.4 Metadatenbanken, Meta-Informationssysteme in Chemie und Umweltschutz**

Wegen der Interdisziplinarität der Fragestellungen und Problematik in der Chemie und im Umweltschutz nehmen Metadatenbanken, Meta-Informationen und Meta-Informationssysteme in diesen Fachgebieten an Bedeutung immer mehr zu. Einen Abriss über den Stand der Forschung und Entwicklung auf dem Gebiet der Meta-Informationssysteme und Umwelt-Informationen in

Deutschland gibt Greve 1995 [Greve 1995]. Ein Meta-Informationssystem für Umweltdaten ist ein Informationssystem, das Informationen über das Vorhandensein und den Zugang zu anderen Informationen bzw. Daten speichert. Ein solches Meta-Informationssystem wird häufig auch als Umweltdatenkatalog (UDK) bezeichnet [Denzer 1994].

Das im Rahmen eines vom Bayerischen Staatsministerium für Landesentwicklung und Umweltfragen geförderten Forschungsvorhabens entwickelte Informationssystem Umweltchemikalien besteht neben den noch näher zu beschreibenden Metadatenbanken aus mehreren Komponenten. Das Network-Pass-Through Facility (NPF) Programm, die Kommunikationssoftware und das Suchhilfesystem CASS dienen zur Automatisierung und Unterstützung der Online Recherchen [Benz 1989]. Die folgenden Metadatenbanken bilden das Kernstück des Informationssystems Umweltchemikalien:

- DALI: Metadatenbank der Literatur
- DADB: Metadatenbank der Online Datenbanken
- DACD: Metadatenbank der CD-ROMs
- DAIN: Metadatenbank der Internet Ressourcen

Die Inhalte der Metadatenbanken der Online Datenbanken und der CD-ROMs bilden die Grundlage für das hier angewandte Bewertungsverfahren (siehe Kapitel 3).

### 2.3 Grundlagen der Online Datenbanken

Der Begriff "**Online Datenbanken**" hat sich durchgesetzt für Datenbanken, die der Allgemeinheit von darauf spezialisierten Einrichtungen über die nationalen und internationalen Datenfernübertragungsnetze zur Abfrage vom eigenen Rechner aus angeboten werden. Online Datenbanken waren bis vor etwa 12 Jahren **bibliographische Datenbanken** (textbasierte Datenbanken), die das Zitat mit oder ohne Kurzreferat umfassen, das die gewünschte Information beinhaltet. Nun gewannen aber in den letzten Jahren "nichttextliche" Online Datenbanken immer mehr an Bedeutung. Diese **Faktendatenbanken** (Datenbanken, in denen die gesuchten Fakten enthalten sind, z.B. der Schmelzpunkt einer Chemikalie oder Konzentrationsangaben in Umweltmedien) beruhen auf Modellierungstechniken, die sich deutlich von denen der textbasierten Datenbanken unterscheiden. Diese werden im folgenden herausgearbeitet.

Die klassischen Datei- und Datenbanktechniken beruhen immer noch auf einem Dateikonzept, das sich am Beispiel einer sequentiellen Datei folgendermaßen beschreiben läßt:

- Eine **Datei** besteht aus einer größeren Anzahl strukturell identischer **Datensätze**.
- Jeder **Datensatz** besteht normalerweise aus mehreren **Feldern**.
- Die **Felder** eines **Datensatzes** beschreiben mit bestimmten Feld- bzw. **Datentypen** jeweils ein **Objekt**, z.B. eine Chemikalie, ein Forschungsprojekt, eine Online Datenbank etc.
- Alle **Datensätze** der **Datei** beschreiben dann die Gesamtheit dieser **Objekte**, die **Objektklasse**.

Es gibt selbstverständlich daneben zahlreiche andere Dateikonzepte, z.B. Dateien von Textsystemen, von Statistikpaketen, Dateien von geographischen Informationen usw. Es ist daher durchaus sinnvoll, Datenbanken mit Dateien der oben beschriebenen Struktur als **Datensatzorientierte Datenbanken** (record-oriented databases) zu bezeichnen, wie das in der neueren Datenbank-theoretischen Literatur geschieht und sie damit von den übrigen (nicht-Datensatzorientierten Datenbanken) abzugrenzen [Staud, SS. 13-26, 1991]. Unter einer **Datei** versteht man eine Sammlung zugehöriger Datensätze. Liegt nicht mehr nur eine Datei vor, sondern mehrere, die zusammengehörig und inhaltlich (oder auf andere Weise) verknüpft sind (z.B. entsprechend dem relationalen Datenmodell) dann liegt eine **Datenbank** vor. Entsprechendes gilt für die Systemtypen. Systeme, die nur die Einrichtung und Verwaltung von Dateien erlau-

ben, sind *Dateisysteme*, solche die entsprechende weiterführende Techniken anbieten, sind *Datenbanksysteme*.

Es spielen zwei Typen von Datenbanksystemen eine Rolle: das *klassische Datenbanksystem* und das *Information Retrieval-System*. Bei Online Datenbanken spielen die Information Retrieval-Systeme eine führende Rolle (näherere Erklärungen siehe [Staud SS. 13-26, 1991]). Als Beispiel für einen Datensatz (Dokument) und dessen Felder dient das Rechercheergebnis in Chemical Abstracts, der größten und wichtigsten bibliographischen Datenbank auf dem Gebiet der Chemie und deren angrenzenden Fachdisziplinen inklusive der Umweltwissenschaften, in welcher eine Autoren-Recherche durchgeführt wurde. Die Fragestellung lautet: Welche Publikationen gibt es von Gasteiger J. zu "databases". Ein Dokument wird als Beispiel gegeben.

### **Recherchebeispiel 1: STN File: CA Item: 123:32205**

**L3 ANSWER 1 OF 32 CA COPYRIGHT 1995 ACS**

**AN** 123:32205 CA

**TI** Chemistry information in the year 2000

**AU** \*\*\*Gasteiger, J.\*\*\*

**CS** Org.-Chem. Inst., Tech. Univ. Muenchen, Munich, D-85748, Germany

**SO** Mitteilungsbl. - Ges. Dtsch. Chem., Fachgruppe Chem.-Inf.-Comput.

(1993), 27, 4-25

CODEN: MGCCEQ; ISSN: 0178-4927

**DT** Journal; General Review

**LA** German

**CC** 20-0 (History, Education, and Documentation)

**AB** A review without refs. concerning present and future developments in chem. \*\*\*databases\*\*\*. Included are technol., sociol. and economical aspects. For more acceptance of information \*\*\*databases\*\*\* by scientists, improved graphical systems for information presentation, intelligent retrieval systems, the development of neuronal networks and more complex systems are necessary.

**ST** review chem reaction literature information \*\*\*database\*\*\*

**IT** Chemistry, (chem. information in the year 2000)

**IT** Information science and technology, (system, computerized, chem. information in the year 2000)

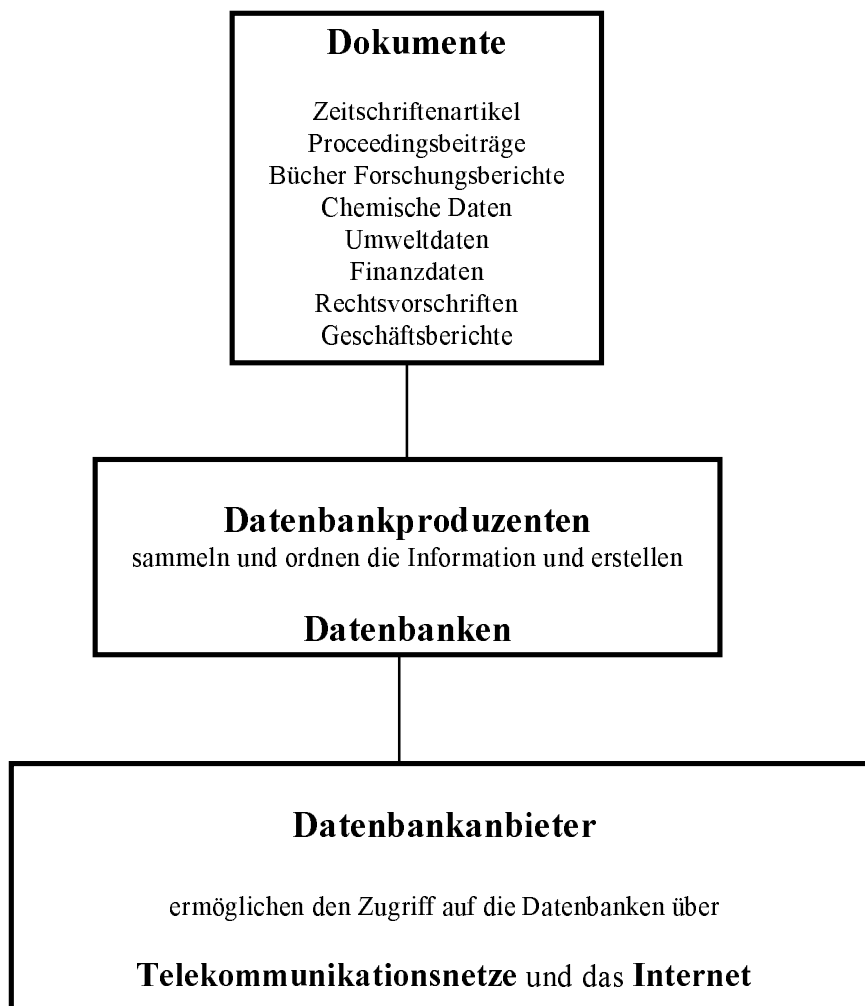
**Legende:** AN = Akzessionsnummer ist der Schlüssel für den Zugang zu den Dokumenten, TI = Titel der Publikation, AU = Autor, CS = Corporate Source (Adresse der Institution, in der der Autor arbeitet), SO = Source (Quelle der Publikation), DT = Document Type (Art der Quelle z.B. Journal), LA = Language (Sprache, in der der Artikel abgefaßt wurde), CC = Classification Code (vom Hersteller der Datenbank vergeben), AB = Abstract, (Kurzfassung der Arbeit), ST = Supplementary Terms (vom Hersteller vergeben), IT = Index Terms (vom Hersteller vergeben).

## **2.4 Elektronische Informationsdienste**

*Elektronische Informationsdienste* sind "Periodika", die an elektronische, magnetische oder optische Informationsträger gebunden sind. Die Information ist auf dem Informationsträger in Form einer Datenbank angelegt. Zum Ablegen und Wiederauffinden der auf dem Informationsträger gespeicherten Information bedient man sich eines Datenbanksystems oder eines Information Retrieval-Systems. Die Herstellung eines elektronischen Informationsdienstes ist ein arbeitsteiliger Prozeß, in der Datenbankproduzent (Datenbankhersteller) und Hosts (Datenbankanbieter) involviert sind. Die Grundstruktur dieses arbeitsteiligen Prozesses läßt sich in der Abbildung 1 verfolgen.

### 2.4.1 Datenbankproduzent (Datenbankhersteller)

Informationen, die auf elektronischem Wege angeboten werden sollen, müssen von einem *Datenbankproduzenten* in einer Datenbank erfaßt und gespeichert werden. Wie schon erwähnt, ist die Anzahl der öffentlich zugänglichen Datenbanken von 301 (1975) auf 9207 (1996) gestiegen.



**Abbildung 1: Grundstruktur elektronischer Informationsdienste**

Ähnlich erhöht hat sich in diesem Zeitraum auch die Anzahl der Datenbankproduzenten (von 200 auf 2860) [Williams 1996]. Datenbankproduzenten können private oder öffentliche Einrichtungen oder Verlage sein. Der Datenbankproduzent ist für den Inhalt der Datenbank verantwortlich, der Datenbankanbieter (Host) ist für die informationstechnische Aufbereitung und Verbreitung zuständig.

### 2.4.2 Datenbankanbieter (Host)

Hauptaufgabe des *Hosts* ist die Bereitstellung eines elektronischen Zugriffssystems auf die von Datenbankproduzenten erstellten Datenbanken. Hierzu bedient sich der Host eines entspre-

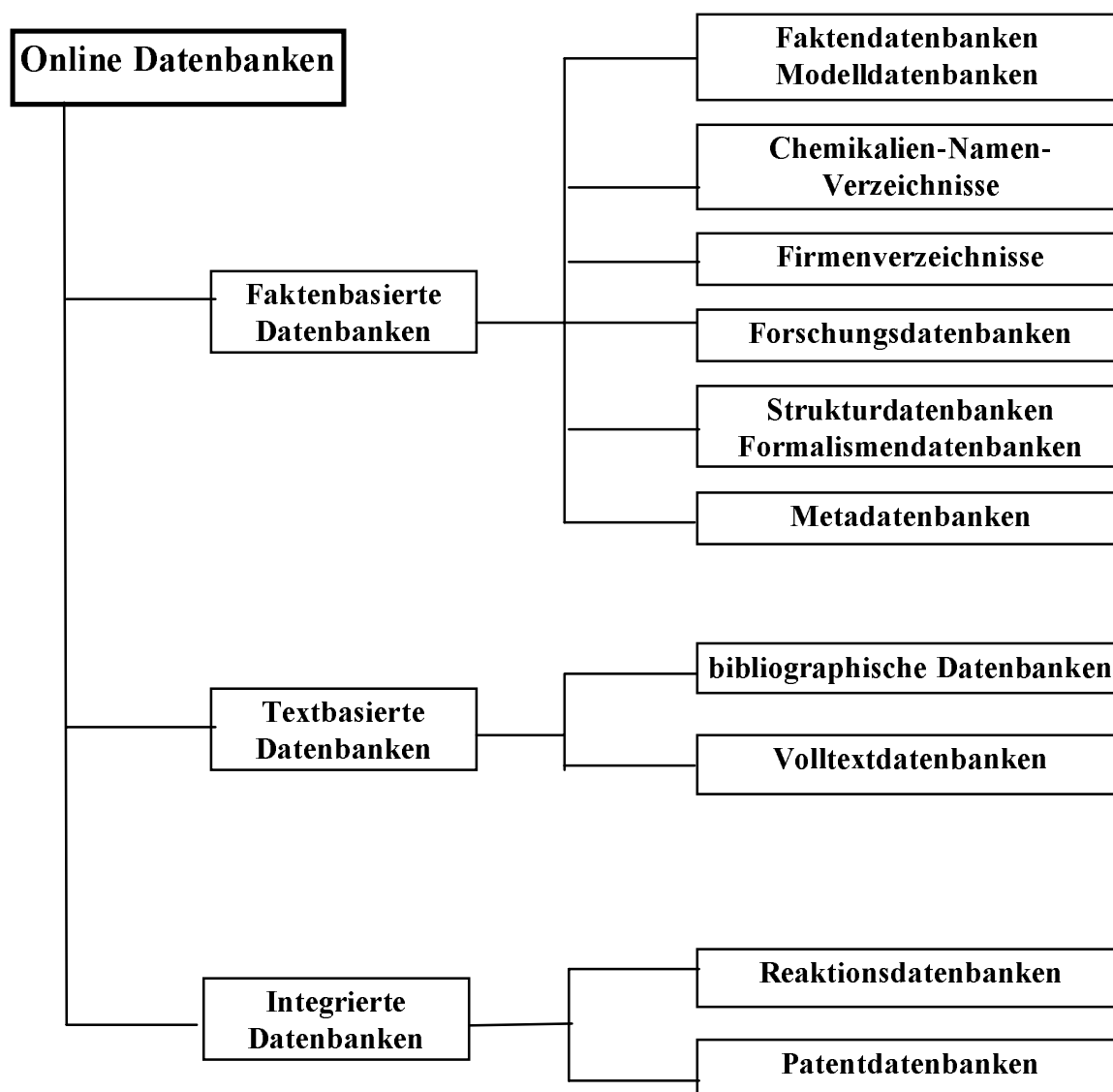
chenden Computersystems. Die Hostcomputer sind entweder an einem Ort konzentriert oder über mehrere Orte verteilt. So befinden sich beispielsweise die Computer des Hosts DIALOG (Knight Ridder) alle in Palo Alto, Kalifornien, USA, während die Rechner des Hosts STN International in der Bundesrepublik, in den Vereinigten Staaten und in Japan installiert sind [Keitz 1993, S. 4].

Ein Host stellt meist eine Vielzahl von Datenbanken zur Verfügung. Oft spezialisiert sich ein Datenbankanbieter auf ein Fachgebiet. So bietet beispielsweise der deutsche Host DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) Datenbanken aus dem Bereich Medizin und Toxikologie und deren angrenzenden Gebieten schwerpunktmäßig an. Der amerikanische Datenbankanbieter CIS (The Chemical Information System) ist spezialisiert auf Datenbanken über Chemikalien und Umweltschutz. Auch STN International hat in seinem Datenbankangebot überwiegend solche Datenbanken mit den Schwerpunkten Naturwissenschaft, Technik inklusive Umweltschutz, Gesundheitsschutz, Arbeitsschutz.

## 2.5 Typisierung von Online Datenbanken gemäß Chemikalien- und Umweltschutzaspekten

Die Online Datenbanken sind in sich nicht homogen. Daher erscheint es sinnvoll, die Datenbanken zu unterteilen. Es besteht die Möglichkeit einer Gliederung gemäß Informationstypen und Struktur. Zunächst werden faktenbasierte und textbasierte Datenbanken unterschieden. Texte liegen oft nicht als Quellmaterial (selbst) vor, sondern nur in verdichteter Form. Textverdichtung bedeutet bis heute meist immer noch die intellektuelle Erstellung eines Kurzreferates, das neben beschreibenden Fakten in der Datenbank abgespeichert wird. Die dabei entstehenden **bibliographischen Datenbanken** geben dann, aus der Sicht des Informationssuchenden, nur indirekte Hinweise auf die gesuchte Information ("Literatur zum Thema"), während im Gegensatz dazu in **faktenbasierten Datenbanken** "die Information selbst" vermutet wird. Diese Unterscheidung erfaßt allerdings nicht den Unterschied "Text/Fakten", sondern den von verweisender und direkter Information, wie spätestens **Volltextdatenbanken** deutlich machen, die, ohne Faktendatenbanken zu sein, die "Informationen selbst" enthalten. Die Obergruppe **textbasierte Datenbanken** enthält damit Datenbanken, in denen der Informationstyp "Text" auf verschiedene Weise eine zentrale Rolle spielt. Zum einen wird hier oft Text modelliert (oder repräsentiert), zum anderen geschieht dies meist auch wieder durch Text (Kurzreferat).

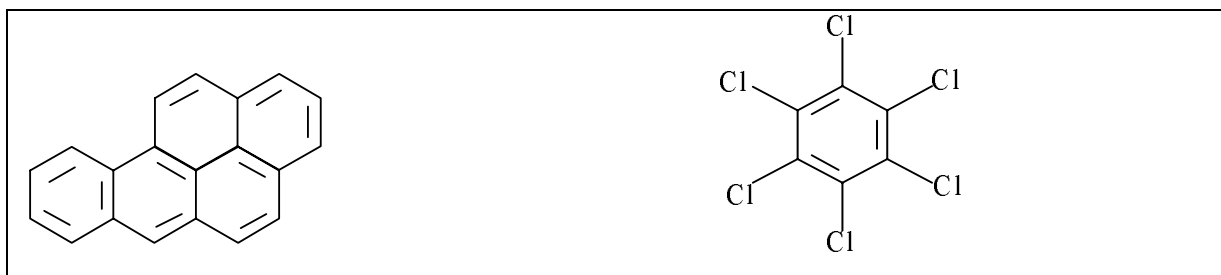
Die Konzentration auf Text und Fakten ist nur pragmatisch, von der gegenwärtigen Situation bei den Online Datenbanken her, begründet. Der Stand der Technik erlaubt heute bereits die effiziente Berücksichtigung zahlreicher weiterer Informationstypen, von denen einige in der letzten Zeit zunehmend auch ihren Weg in Online Datenbanken finden. Dies ist speziell für die hier betrachteten Online Datenbanken auf dem Gebiet der Naturwissenschaft und Technik der Fall. Als dritte Obergruppe in dieser Typisierung werden noch die **integrierten Datenbanken** unterschieden. Diese umfassen in der Regel verschiedene Dateien unterschiedlicher Informationstypen, z.B. Merkmalsinformationen, Zeitreihen, Texte und Tabellen, aber auch Graphiken. In Abbildung 2 wird die Typisierung der Online Datenbanken weiter vertieft. Hier werden, die in Chemie und Umweltschutz unterschiedenen Datenbanktypen dargestellt, die nun anhand eines Beispiels aus dem Themenbereich erläutert werden. Dabei wird nur exemplarisch auf faktenbasierte Datenbanken eingegangen. Eine vollständige Zusammenstellung aller in Abbildung 2 aufgeführten Datenbanktypen nebst Recherchebeispielen und Erläuterungen befindet sich im Anhang A.



**Abbildung 2:** Typen von Datenbanken in Chemie und Umweltschutz

### **2.5.1 Umweltchemikalien für die Recherche in verschiedenen Datenbanktypen**

Bei dem Recherchebeispiel sowie bei den im Anhang aufgelisteten Beispielen werden Informationen über Chemikalien und Umweltschutzaspekte zu den beiden in Abbildung 3 aufgeführten Umweltchemikalien gesucht. Es werden diese beiden Chemikalien zur Erläuterung der Typen der Datenbanken herangezogen, da sie wegen ihrer ubiquitären Verbreitung in der Umwelt eine große Relevanz sowohl für die Ökologische Chemie [Korte 1987] als auch für die Toxikologie [Eisenbrand 1994] haben. Aufgrund ihrer Bedeutung ist eine Fülle von Datenmaterial in sämtlichen Typen von Datenbanken zu vermuten.

**Benzo[a]pyren**

Benzo[a]pyrene

CAS-Nummer: 50-32-8

Summenformel: C<sub>20</sub>H<sub>12</sub>

Molekulargewicht: 252,3

**Hexachlorbenzol**

Hexachlorobenzene

CAS-Nummer: 118-74-1

Summenformel: C<sub>6</sub>Cl<sub>6</sub>

Molekulargewicht: 284,8

**Abbildung 3: Benzo[a]pyren und Hexachlorbenzol als Testchemikalien für verschiedene Datenbanktypen**

Benzo[a]pyren gilt als einer der bestuntersuchten polycyclischen aromatischen Kohlenwasserstoffe (PAK oder PAH nach dem englischen polycyclic aromatic hydrocarbon). Benzo[a]pyren und zahlreiche andere polycyclische aromatische Kohlenwasserstoffe sind kanzerogen im Tierversuch [Eisenbrand 1994, S. 175].

Hexachlorbenzol ist ein Vertreter der Halogenkohlenwasserstoffe [Hulpke 1993, S. 166] und seit mehreren Jahren als ubiquitär vorkommende Umweltchemikalie erkannt. Die Chemikalie wurde aufgrund ihrer fungiziden Eigenschaften als Saatgutbeizmittel, besonders für Weizen, angewandt [Korte 1987, S. 33]. Darüber hinaus findet Hexachlorbenzol Anwendung als Weichmacher für PVC [Bundesminister für Umwelt, Naturschutz und Reaktorsicherheit 1991]. Seit einigen Jahren ist Hexachlorbenzol als Pflanzenschutzmittel in Deutschland nicht mehr zugelassen [Falbe 1990, S. 1791]. Die Chemikalie hat eine hohe Persistenz, d.h. sie wird nur sehr langsam abgebaut. Somit kann sie sich im tierischen und menschlichen Organismus anreichern und zu Stoffwechselstörungen und Organschäden führen.

Die Recherche kann mit den oben in Abbildung 3 aufgeführten Identifikationsmerkmalen, Strukturformel, CAS-Nummer, Summenformel, Molekulargewicht oder mit der englischen Bezeichnung des Chemikaliennamens durchgeführt werden. Da ca. 95 % der Online Datenbanken in englischer Sprache aufgebaut sind, ist die Recherche mit der deutschen Bezeichnung einer Chemikalie als wenig aussichtsreich anzusehen [Voigt 1993].

**2.5.2 Faktendatenbanken (Modelldatenbanken der Naturwissenschaften)**

Naturwissenschaftliche Faktendatenbanken enthalten typischerweise drei Teile, einen bibliographischen Teil, einen Informationsteil und einen Datenteil.

Der **bibliographische Teil** enthält Hinweise auf die Publikationen, aus der die Fakteninformationen entnommen wurden. Der **Informationsteil** enthält die Informationen über die eigentlich im Mittelpunkt stehende Information. Hier wären beispielsweise folgende Identifikationsmerkmale zu nennen: Chemikaliename, Synonyma, CAS-Nummer, sonstige Registriernummern (z.B. Beilstein-Nummer, RTECS-Nummer, EG-Nummer, EINECS-Nummer, NIOSH-Nummer etc.), Molekulargewicht und Summenformel. Der **Datenteil** enthält die eigentlichen Informationen, so im vorliegenden Recherchebeispiel der Datenbank CHCD Chapman & Hall

Chemical Database bei dem Host DIALOG (Knight-Ridder): Schmelzpunkt, Siedepunkt, Stoffbeschaffenheit, Verwendungsart und Gesundheitsschutzaspekte.

**Recherchebeispiel 2: DIALOG File: CHCD Chapman & Hall Chemical Database (303)  
Item: 0071484**

<p>DIALOG(R)File 303:Chapman &amp; Hall Chemical Database (c) 1989 Chapman &amp; Hall. All rts. reserv. 0071484 CHCD Acc No: B-20048 Subfile: Dictionary of Organic Compounds, 5th Edition, Second Supplement, 1984 <b>CHCD Name:</b> Benzo(a)pyrene; 9CI <b>Synonyms:</b> 3,4-Benzopyrene; 3,4-Benzopyrene; Benzo(a)pyrene,1,2-Benzopyrene; <b>CAS Registry No:</b> 50-32-8 <b>Formula:</b> C20H12 <b>Weight:</b> 252.315 <b>Structure:</b> Available for display or see printed work. <b>RTECS:</b> DJ3675000 <b>Hazard:</b> Carcinogenic <b>Source:</b> A principal carcinogenic constit. of coal tars, air pollution, etc. <b>Use/Importance:</b> Extensively used in cancer research. <b>State:</b> Pale-yellow needles (C6H6/MeOH) <b>Mp:</b> 177 deg C <b>Bp:</b> 310-312 deg C at 10 mm <b>References:</b> Cook, J.W. et al, J. Chem. Soc., 1933, 395, 398; isol,synth, Mayneord, et al, Proc. R. Soc. London, Ser. A, 1935, 152, 299; uv, Buchta, E. et al, Justus Liebigs Ann. Chem., 1968, 716, 102; synth, Bhatia, K., Anal. Chem., 1971, 43, 609; glc, Popp, F.A., Z. Naturforsch., C, 1973, 28, 165; uv,tox, Buchanan, G.W. et al, Can. J. Chem., 1975, 53, 1829; Phillips, D.H., Nature (London), 1983, 303, 468; rev, Unkefer, C.J. et al, J. Am. Chem. Soc., 1983, 105, 733; pmr,cmr <b>Reference Tags:</b> Carbon (13C) NMR Spectrum; Gas-Liquid Chromatography; General; Hazard; Proton (1H) NMR Spectrum; Review; Synthesis; Toxicity; Ultraviolet Spectrum, Data Tags: Boiling Point; CAS Registry Number; Hazard/Toxicity Data; Melting Point; Physical State; RTECS Reference; Source; Structure; Use/Importance</p>
---

Am Beispiel dieser Faktendatenbank CHCD Chapman & Hall Chemical Database ist die oben beschriebene Aufteilung in Informationsteil, Datenteil und bibliographischen Teil deutlich zu erkennen. Zum Informationsteil gehören die Datenfelder CHCD Name, Synonym, CAS Registry No., Formula, Weight, RTECS, zum Datenteil die Felder Hazard, Source, Use/importance, State, Mp (Melting point), Bp (Boiling point). Der bibliographische Teil ist unter dem Feld References subsumiert und wird nicht in einzelne Datenfelder aufgliedert.

## 2.6 CD-ROMs (Compact Disc Read Only Memory)

Seit länger als zehn Jahre gibt es CD-ROMs ("Compact Disc Read Only Memory" Medium "Nur-Lese-Medium") als Speichermedium. Eine einzige CD-ROM hat eine Kapazität von ca. 650 MB. Man kann man z.B. auf einer einzigen CD-ROM Daten von 200.000 Blatt Papier oder eine Million Katalogkarten von Bibliotheken oder ca. 500 3 1/2 Zoll Disketten speichern [Jacso 1990]. So ist es nicht erstaunlich, daß der Markt an CD-ROM Titeln ebenfalls exponentiell angestiegen ist. 1994 waren ca. 14.000 CD-ROM Titel auf dem Markt. Diese Zahl wird nach Schätzungen 1995 verdoppelt bis verdreifacht [Herther 1995].

Als Hilfestellung bei der Auffindung der geeigneten CD-ROMs bieten sich sog. Verzeichnisse, Directories an. Es handelt sich bei diesen Standardwerken um Metadatenbanken, die sowohl in

gedruckter Form als auch als CD-ROM angeboten werden. Die Standardverzeichnisse von CD-ROMs sind "CD-ROMs in Print" [Meckler 1995], "The CD-ROM Directory" [TFPL Publishing 1996], und "Handbuch lieferbarer CD-ROMs" [Scientific Consulting 1995a]. CD-ROMs in Print umfaßt in seiner neuesten Ausgabe über 8.000 CD-ROM Produkte. Im CD-ROM Directory werden ca. 13.000 Datensätze zur Verfügung gestellt. Das Handbuch lieferbarer CD-ROMs verfügt über Informationen zu ca. 7.500 CD-ROM Produkten, ca. 10 % davon stammen von Informationsanbietern aus Deutschland.

Analog der Vorgehensweise bei den Online Datenbanken können auch die CD-ROMs eingeteilt werden. Wie bereits im Kapitel 2.5 näher ausgeführt, unterscheidet man die in Abbildung 2 aufgeführten Datenbanktypen. Die einzige Ausnahme bilden die Chemikaliennamen-Verzeichnisse, die bei CD-ROMs nicht vorkommen.

Entgegen der gängigen Praxis der Recherche bei den Online Datenbanken mit Hilfe einer Retrievalsprache verfügen die meisten CD-ROMs über die Möglichkeit der Abfrage mit einem menügestützten Retrieval. In der Regel wird bei CD-ROM Retrievalsystemen auf einen besonders benutzerfreundlichen menügesteuerten Dialog Wert gelegt.

Menüsysteme desselben CD-ROM Produzenten (z.B. DIALOG) haben dasselbe menügesteuerte Retrieval. Andere CD-ROM Produzenten bieten jedoch eine andere Menüsteuerung an. Der Nutzer mehrerer CD-ROM Datenbanken von unterschiedlichen Produzenten muß sich mit unterschiedlichen Menüsystemen vertraut machen. Eine Abstimmung der Online Datenbanken und CD-ROMs desselben Anbieters bzw. CD-ROM Produzenten besteht hingegen. Als Beispiel sei hier der Host DIALOG angeführt, der in seinem Datenbankangebot, sowohl Online Datenbanken als auch CD-ROMs der gleichen Datenbanken vorhält. In DIALOG CD-ROMs kann man sowohl mit der Menütechnik als auch mit der DIALOG Retrievalsprache recherchieren. Eine Übertragung der Recherche von der CD-ROM auf die DIALOG Online Datenbanken ist mit Hilfe der Kommunikationssoftware DIALOGLINK gegeben [Keitz SS. 199-221, 1993]. DIALOG sieht also seine CD-ROMs als Ergänzung des Online Datenbankangebotes und umgekehrt [Vogt 1992].

In den vergangenen Jahren wurden in verstärktem Maße Chemiedatenbanken auf CD-ROM herausgegeben. In einer Studie über das Informationsverhalten und den Informationsbedarf in der chemischen Forschung wurde festgestellt, daß der Anteil an CD-ROM Datenbanken auf dem Gebiet der Chemie mit 38 % der verwendeten Systeme (d.h. inhouse Datenbanken, Online Datenbanken und CD-ROMs) deutlich höher lag als in den USA und Canada. Die Ausstattung der Informationsvermittlungstellen mit CD-ROM Laufwerken ist in Europa besser als in Nordamerika [Bucher 1993].

Es ist also ein leicht positiver Trend der Nutzung dieses Mediums in der Chemie-Information zu verzeichnen. Die CD-ROM wird in Zukunft das bevorzugte Nutzungsmedium in vielen Bereichen der Chemie-Information sein [Heller 1993]. Einige Beispiele von Chemiedatenbanken auf CD-ROM werden in einer neueren Publikation von Zass gegeben. Es wird die Frage diskutiert, ob und inwieweit das Medium CD-ROM vorteilhafter ist als Papierversionen oder Online Datenbanken. Der Autor kommt zu dem Schluß, daß die Stärke der CD-ROMs bisher vor allem bei günstigen Einzelplatzversionen liegt [Zass 1995].

## 2.7 Internet Ressourcen

Das Internet als größtes Computer-Netzwerk der Welt birgt ein ungeheuer großes Potential an Daten und Informationen, die heute noch teilweise kostenlos zur Verfügung gestellt werden. Mitte 1995 wurden über 40 Millionen Nutzer in 160 Ländern gezählt [Williams 1995]. Experten gehen davon aus, daß es 1998 ca. 100 Millionen Internet Nutzer geben wird. Auch für die Thematik Chemie und Umweltschutz ist das Internet eine wertvolle Quelle an Informationen [Voigt 1996a]. Man unterscheidet im Internet sog. freie, d.h. kostenlose und kostenpflichtige

Ressourcen. Auf die frei verfügbaren Ressourcen sollen sich die weiteren Ausführungen beziehen. Bezüglich der Informationen in diesen freien Ressourcen müssen jedoch zwei Überlegungen berücksichtigt werden: 1. Was ist der Hintergrund dieser Ressource? und 2. Wie lange wird diese Ressource vorhanden sein? Darüber hinaus muß sich der Nutzer der Internet Ressourcen darüber bewußt sein, daß die meisten Internet Quellen mehr aus Hobbygründen als für einen konkreten Forschungsauftrag erstellt werden, so daß sie in vielen Fällen nur wenige und z.T. sehr spezielle Daten und Informationen beinhalten [Heller 1996]. Die Anzahl der Internet Ressourcen, die sich mit der Thematik Chemie und Umweltschutz auseinandersetzen ist nicht genau bekannt. Die Schätzungen in Bereich Chemie variieren von mehreren hundert Seiten [Krieger 1995] [Heller 1996] bis ca. 50.000 chemierelevante Seiten (Web pages) [Glock 1996]. Das erste Struktur-orientierte Indexierungssystem für chemische Strukturen im World Wide Web wurde kürzlich im Computer-Chemie-Centrum der Universität Erlangen-Nürnberg entwickelt und implementiert [Ihlenfeldt 1996]. Auch für die Quantität der Umweltinformationen im Internet erhöht sich kontinuierlich. Eine gute Übersicht über Umweltressourcen im Internet bieten Reichardt und Voigt [Reichardt 1996] [Voigt 1996b].

Wegen der Größe und Unübersichtlichkeit des Internets ist die Unterstützung des Auffindens von geeigneten Ressourcen mittels Verzeichnissen hier von besonderer Bedeutung. Diese Verzeichnisse im Internet umfassen Informationen über andere Ressourcen, auf die sie in der Regel auch verweisen. Dieser Problematik tragen verschiedene Ansätze von unterschiedlichen Produzenten Rechnung. Eine umfassendere Unterstützung für die Auffindung von Umweltchemikalien-relevanten Ressourcen bietet DAIN - Metadatenbank der Internet Ressourcen [Voigt 1996b]. Es handelt sich bei dieser Ressource (URL: <http://dino.wiz.uni-kassel.de/dain.html>) um eine Metadatenbank, in der in mehreren Datenfeldern mittels Verknüpfung mit Hilfe von logischen Operatoren AND/OR recherchiert werden kann. Das dieser Datenbank zugrunde liegende Indexierungssystem wird von Benz näher erläutert [Benz 1995]. Jedoch verfügt diese Metadatenbank erst über eine geringe Anzahl von Datensätzen. Somit können die Informationen aus dieser Metadatenbank noch nicht einem Bewertungsverfahren, wie es in dieser Arbeit vorgestellt wird, unterzogen werden.

### 3. Hasse-Diagramm-Technik als Methode zur vergleichenden Bewertung

#### 3.1 Kurzer Überblick über alternative Bewertungsmethoden

Die in Kapitel 2 erörterte Vielzahl und Vielfalt an Datenquellen in Chemie und Umweltschutz erschwert den Umgang mit den Informationsquellen erheblich und somit auch den Zugang zu den gewünschten Informationen. Eine vergleichende Bewertung von Datenquellen ist somit ein wichtiger Schritt hin zu einem effektiven Informationsmanagement auf diesem Gebiet.

Versteht man unter Bewertungsverfahren, Verfahren, die möglichst eine Einordnung von Objekten in "gut und böse" hinsichtlich vorgegebener Bewertungsziele erlauben, so sind zwei typische Schritte erforderlich:

A. Die Bereitstellung von Einzelinformationen, die für das Bewertungsziel als wichtig erachtet werden. Diese Bereitstellung ist kontextspezifisch und die Bewertung von Objekten ist nur so gut, wie die Einzelinformationen die Objekte hinsichtlich des Bewertungsziels charakterisieren. In vielen Fällen, wie z.B. zur Chemikalienbewertung hinsichtlich ihrer Exposition kann beträchtlicher Aufwand betrieben werden, um Einzelinformationen durch kausale Modellierung zu erhalten. Als Beispiel hierfür sei das Programmpaket E4CHEM genannt [Matthies 1989].

B. Sind die Einzelinformationen in einer quantitativen Weise verfügbar, so entsteht das Problem "Vergleich von Objekten" auf der Basis von vielen Informationen durchzuführen. Die einfachsten Methoden aggregieren die Einzelinformationen operational zu einer Gesamtgröße (Gütefunktion, Qualitätsfunktion), um die im allgemeinen nicht begründbare Bildung eines Gesamtindex zu vermeiden [Opperhuizer 1982]. (Siehe hierzu auch die Ausführungen zum Gesamtmaß  $\Gamma$  in Kapitel 6.5 und Kriterienmaß  $\vartheta$  in Kapitel 8.3.2). Ein wesentlicher Schritt für viele Verfahren ist die möglichst datengetreue Projektion auf niedrig dimensionale Merkmalsräume, um erst anschließend eine Bewertung, etwa auf Basis eines Expertensystems durchzuführen. Als ein Beispiel unter mehreren möglichen sei die Clusteranalyse genannt. Neuere Ansätze mittels der Fuzzy-Clusterung, in welchen Chemikalien gemäß ihrer ökotoxikologischen Eigenschaften bewertet werden, sind in einer Publikation von Friedrichs beschrieben [Friedrichs 1996].

Andere Verfahren (z.B. neuronale Netze) geben für gut bekannte Objekte eine Bewertung vor und trainieren den Rest der Objekte, um zu Regeln für eine Bewertung zu gelangen.

#### 3.2 Begründung der Wahl der Hasse-Diagramm-Technik als Bewertungsmethode

Die Hasse-Diagramm-Technik stellt ein Verfahren dar, welches die "Parallelität von Einzelinformationen" beibehält (siehe auch Punkt A des Kapitels 3.1). Dies erfolgt anhand einer verallgemeinerten Ordnung, nämlich - wie hier - mit Hilfe der Produktordnung (komponentenweiser Vergleich), in der zwei Objekte nur noch unter bestimmten Bedingungen als "vergleichbar" angesehen werden. Hierdurch vermeidet die Hasse-Diagramm-Technik die meist nicht begründete Verknüpfung unterschiedlicher Kriterien und erlaubt es, Besonderheiten in den Ausprägungen der Einzelinformationen zu erkennen. Die Hasse-Diagramm-Technik ist aber selbstverständlich darauf angewiesen, daß die Einzelinformationen selbst, einen sinnvollen Beitrag zum Bewertungsziel liefern. Eine Mittelstellung zwischen den Ansätzen, die auf dem Konzept von Gesamtindizes beruhen und solchen, die jede Einzelinformation separat verarbeiten (Produktordnung) nehmen Konzepte einer verallgemeinerten Ordnung ein, wie sie beispielsweise von Yu beschrieben werden [Yu 1973]. Die Hasse-Diagramm-Technik ist deshalb auf die Anwendung von Fragestellungen, in welchen viele Objekte - in diesem Fall Datenquellen - mit mehreren Kriterien gleichzeitig bewertet werden sollen prädestiniert.

Auf dem Gebiet der Datenbanken wurden bereits einige Hasse-Diagramm Ansätze publiziert. Beispielsweise wurden 15 chemische und umweltrelevante Online Datenbanken eingestuft und bewertet [Voigt 1994]. Ebenso wurde eine Auswertung mit Hilfe der Hasse-Diagramm-Technik von toxikologischen Datenbanken vorgenommen [Voigt 1995a]. Auch wurden bereits einige theoretische Aspekte der Hasse-Diagramme in bezug auf Datenbanken veröffentlicht [Brüggemann 1994, 1995b].

### 3.3 Kurzbeschreibung der Hasse-Diagramm-Technik

Die Technik der Hasse-Diagramme stellt ein mathematisches Verfahren dar, welches mehrere Kriterien gleichzeitig berücksichtigen kann. Die Grundlage hierfür ist die Theorie *partiell geordneter Mengen* (im englischen auch *posets* von *partially ordered sets*). Sie wird oft im Rahmen der Verbandstheorie mitabgehandelt, bekommt aber zunehmendes Gewicht als eigenständige Disziplin.

Im Anhang B wird die Theorie der Hasse-Diagramme im Hinblick auf Datenbanken näher erläutert. Weitergehende Erklärungen zur Hasse-Diagramm-Technik sind von Brüggemann ausgeführt worden [Brüggemann, 1995a]. In dieser Publikation geht Brüggemann auch ausführlich auf das in der GSF-Forschungszentrum für Umwelt und Gesundheit erarbeitete Programm "Hasse" ein.

Im folgenden wird nur auf einige wenige grundsätzliche Begriffe, die bei der Auswertung der Hasse-Diagramme der Online Datenbanken und CD-ROMs zum Tragen kommen, eingegangen. *Kriterien* oder auch *Bewertungskriterien* genannt, umfassen sowohl quantitative als auch qualitative Eigenschaften. *Objekte*, in dieser Arbeit *Datenquellen* in Chemie und Umweltschutz, sind Gegenstand des Interesses und werden von den *Bewertungskriterien* charakterisiert.

*Hasse-Diagramme* werden von oben nach unten gelesen. Die Objekte werden in Form von Kreisen dargestellt, wobei das Akronym des Objektens im Kreisinnern geschrieben steht. Benachbart sind Objekte, die genau durch eine Linie miteinander verbunden sind.

Unter dem Begriff der *äquivalenten Objekte* in einem Hasse-Diagramm versteht man unterschiedliche Objekte, die dieselben Ausprägungen in bezug auf einen bestimmten Satz an Kriterien aufweisen.

*Maximale Objekte* oder auch *Maximale* genannt, sind Elemente der Objektmenge  $E$ , die keine nach oben benachbarten Elemente aufweisen. *Minimale Objekte* oder *Minimale* genannt, sind Elemente der Objektmenge  $E$ , die keine nach unten benachbarten Elemente aufweisen.

Man spricht vom *kleinsten Objekt*, wenn nur ein Minimal und vom *größten Objekt*, wenn nur ein Maximal vorhanden ist.

Falls nur ein Objekt (oder eine Äquivalenzklasse) einen Teil des Hasse-Diagramms darstellt, nennt man dieses Objekt ein *isoliertes Element*.

Objekte, die alle untereinander *vergleichbar* sind, werden auch *Ketten* in der Verbandstheorie genannt. Eine Menge an paarweise *unvergleichbaren* Objekten wird *Antikette* genannt.

Als ein *Artikulationspunkt* wird ein Objekt bezeichnet, welches bei dessen Auslassung zu einem Zerfall des Hasse-Diagramms in mehrere Hierarchien führt.

Die *Anzahl der Äquivalenzklassen* ( $NECA = \text{number of equivalence classes}$ ) bedeutet die Anzahl der Äquivalenzklassen mit mehr als einem Objekt und wird auch als Anzahl der *nicht-trivialen Äquivalenzklassen* bezeichnet.

Die *Anzahl der Ebenen* ( $NL = \text{number of levels}$ ) bedeutet die Höhe des Hasse-Diagramms und wird definiert als  $H(E) = L(E) + 1$ .

Die *Anzahl der Objekte in der breitesten Ebene* ( $NEL = \text{number of elements in level}$ ) ist ein Maß für die Breite des Hasse-Diagramms.

Die Anzahl der *Vergleichbarkeiten*  $V(N)$  und *Unvergleichbarkeiten*  $U(N)$  sind nützliche Hilfsmittel zur Interpretation von Hasse-Diagrammen.

### 3.4 Schritte bei der vergleichenden Bewertung mittels der Hasse-Diagramm-Technik

Es werden folgende Schritte bei der Auswertung von Hasse-Diagrammen vorgenommen.

#### 3.4.1 Allgemeine Auswertung des Hasse-Diagramms

Diese Auswertung umfaßt die oben beschriebenen maximalen, minimalen, isolierten Objekte, Ketten, Vergleichbarkeiten, Unvergleichbarkeiten, Anzahl der Ebenen und strukturelle Besonderheiten wie z.B. Artikulationspunkte, isolierte Hierarchien.

#### 3.4.2 Analyse der Struktur des Hasse-Diagramms (D-Matrix: Schnittmengen-Matrix)

Es interessiert die Zahl der Objekte, die zu einem herausgegriffenen Objekt vergleichbar sind, oder es ist die Frage zu beantworten, welche Objekte gleichzeitig vergleichbar sind zu zwei herausgegriffenen Objekten. Vorzugsweise wird ein maximales Objekt als Ausgangspunkt für die Analyse gewählt. Das ausgewählte Element wird **Schlüsselement** (key element) genannt. Es können auch mehrere Schlüsselemente gleichzeitig gewählt werden. Die Analyse eines Schlüsselementes impliziert eine Suche nach allen Objekten, die im Hasse-Diagramm unterhalb des Schlüsselementes lokalisiert sind, d.h. alle Objekte, die von einem Schlüsselement nach unten erreicht werden können. Diese Objekte zusammen mit den Objekten, die äquivalent aber nicht identisch mit dem Schlüsselement sind, werden **Nachfolger** (successors) genannt. Die Nachfolgermengen (successor sets) und ihre Kardinalitäten sind das Kernstück der Hasse-Analyse.

Die Schnittmengen- oder **D-Matrix** dient der Analyse der Struktur des Hasse-Diagramms. Es sollen Relationen zwischen Schlüsselementen aufgezeigt werden. Weitere Ausführungen siehe Anhang B.

#### 3.4.3 Bedeutung der Kriterien beim Rangfolgeverfahren (Sensitivitätsanalyse)

Die **W-Matrix** beschreibt den Einfluß der Kriterien auf das Rangfolgeverfahren.

Um die Bedeutung eines Kriteriums zu bestimmen, werden die Resultate des Rangfolgeverfahrens unter Nutzung verschiedener Untermengen an Kriterien verglichen. Dies impliziert einen Vergleich von Hasse-Diagrammen. Dieser wiederum läuft auf den Vergleich von Nachfolgermengen hinaus. Dazu werden Distanzen zwischen Nachfolgermengen in einer sog. **W-Matrix** gespeichert. Details zur Distanzberechnung sind in [Brüggemann 1996a] zu finden.

## 4. Erstellung von Metadatenbanken für Umweltchemikalien

### 4.1 Metadatenbank der Online Datenbanken - DADB

Wie schon oben erwähnt, wurden Metadatenbanken zu den Fragestellungen Chemikalien und Umweltschutz entwickelt, gepflegt und ausgebaut [Voigt 1992]. Eine dieser Metadatenbanken ist die Metadatenbank der Online Datenbanken DADB. Der Dokumentenstand beträgt **453 Einträge** (Stand 31.01.97) von nationalen und internationalen Datenbanken auf dem Gebiet der Informationen über Chemie und Umweltschutz.

#### 4.1.1 Datenbankdefinition von DADB

Diese Metadatenbank wurde mit der kommerziellen Software LARS erstellt. Die Abkürzung LARS steht für "Leistungsstarkes Archivierungs- und Recherchesystem" [WEKA, 1992]. Es werden folgende Feldwerte unterschieden: Alphanumerisch, Text, ganze Zahl, Volltext, Kommazahl, Formel (zum Rechnen), Datum, Zeit und Betrag.

In der Tabelle 2 ist die Datenbankdefinition von DADB festgehalten.

Als Feldwerte sind für diese Datenbank "alphanumerisch", "ganze Zahl", "Datum", "Betrag" und "Volltext" gewählt worden.

Neben den *administrativen* Datenfeldern (1-4) unterscheidet man die *bibliographischen* Felder (5-26) und die *inhaltserschließenden* Datenfelder (fettgedruckt).

Unter den bibliographischen Datenfeldern findet man die *Kurzbezeichnung* und die *Bezeichnung* der Online Datenbank. Danach schließen sich die Felder *Host* und *sonstige Hosts* an. Auf die Hosts (Datenbankanbieter) wurde bereits im Kapitel 2.4.2 eingegangen.

**Tabelle 2: Datenbankdefinition DADB** (inhaltserschließende Datenfelder fett gedruckt)

Position	Feldname	Kurzbez.	Feldwerte
01	Dokumentennummer	DO	ganze Zahl
02	Datum	DA	Datum
03	Standort	ST	alphanumerisch
04	Bearbeiter	BE	alphanumerisch
05	Kurzbezeichnung	KU	alphanumerisch
06	Bezeichnung	BZ	alphanumerisch
07	Host (Datenbankanbieter)	HO	alphanumerisch
08	sonstige Hosts	SH	alphanumerisch
09	Produzent (Datenbankhersteller)	PR	Volltext
10	Typ der Datenbank	TY	alphanumerisch
11	Typ 2	TP	alphanumerisch
12	Sachgebiet	SA	alphanumerisch
13	Quellen	QU	alphanumerisch
14	Sprache	SP	alphanumerisch
15	Retrievalsprache	RE	alphanumerisch
16	ab wann (sind Einträge vorhanden)	AB	ganze Zahl

17	bis wann (sind Einträge vorhanden)	BI	ganze Zahl
18	Aktualisierung (Update)	UP	alphanumerisch
19	Umfang	UM	ganze Zahl
20	Zuwachs/Jahr	ZU	ganze Zahl
21	Kosten/h	KO	Betrag
22	Online Kosten	ON	Betrag
23	Offline Kosten	OF	Betrag
24	SDI Kosten	SD	Betrag
25	sonstige Kosten	SK	Betrag
26	Arbeitsunterlagen	AR	alphanumerisch
<b>27</b>	<b>Informationstyp</b>	<b>IT</b>	<b>alphanumerisch</b>
<b>28</b>	<b>Anwendung der Chemikalien</b>	<b>AW</b>	<b>alphanumerisch</b>
<b>29</b>	<b>Anzahl der Stoffe</b>	<b>AN</b>	<b>ganze Zahl</b>
<b>30</b>	<b>Deskriptoren (Parameter-Testdatensatz)</b>	<b>DE</b>	<b>alphanumerisch</b>
<b>31</b>	<b>Chemikalien-Testdatensatz</b>	<b>TE</b>	<b>ganze Zahl</b>
32	Sonstiges	SO	alphanumerisch
33	Pflege der DADB	PF	alphanumerisch

In DADB werden Datenbanken der folgenden Hosts ausgewertet: CDP, CIS (Chemical Information Service), DATASTAR, DIALOG, DIMDI (Deutsches Institut für Medizinische Dokumentation und Information), ECHO (European Communities Host Organization), ESA (European Space Agency), FIZ Technik, GBI (German Business Information), ICSTI (International Centre for Scientific and Technical Information), ORBIT, Questel, STN (Scientific and Technical Information Network), VINITI (The All Russian Institute of Scientific and Technical Information) und einige kleinere Hosts, die nur ein oder zwei Datenbanken anbieten. Es zeichnet sich auf dem Datenbankanbieter-Markt eine Konzentrierung ab. Beispielsweise sind DIALOG und DataStar in der Firma Knight-Ridder Information zusammengefaßt. Ebenso fand ein Zusammenschluß von ORBIT und QUESTEL statt. BRS wird seit kurzem unter dem Namen CDP Technologies geführt. Da jedoch die früheren Strukturen, u.a. auch unterschiedliche Retrievalsprachen bestehen blieben, werden die Hosts weiter getrennt aufgeführt.

Der **Produzent** oder auch Datenbankhersteller wird in DADB mit der vollständigen Adresse angegeben. Zu diesem Feld wurden bereits in Kapitel 2.4.1 Ausführungen gemacht. Ebenso wurde im Kapitel 2.5 die Typisierung der Datenbanken ausführlich besprochen. Zwei Datenfelder, **Typ** der Datenbank und **Typ 2** wurden eingerichtet. Oftmals ist eine eindeutige Zuordnung zu einem einzigen Typ nicht möglich. Beispielsweise ist die Datenbank BEILSTEIN nicht nur als eine Faktendatenbank, sondern auch als eine Strukturdatenbank und als Reaktionsdatenbank eingestuft.

Im Feld **Sachgebiet** werden die Hauptthemen der Datenbank im Freitext wiedergegeben. Im Feld **Quellen** findet man die Art der erschlossenen Publikationen. Beispiele hierfür sind: Fachzeitschriften, Fachbücher, Forschungsberichte, Proceedingsbände, Dissertationen, Patente, Firmenberichte, graue Literatur etc. Es folgen die Datenfelder Sprache und Retrievalsprache, die deutlich voneinander zu unterscheiden sind. Unter dem Feld **Sprache** ist die Sprache zu verstehen, in der die Datenbank aufgebaut wurde. Über 90 % der in DADB inkorporierten Datenbanken sind auf englisch. Unter dem Begriff **Retrievalsprache** ist die jeweilige Abfragesprache für die Datenbanken zu verstehen. Für die Suche in Online Datenbanken muß der Host Retrievalprogramme bereitstellen. Jeder Host verfügt über seine eigene Retrievalsprache. Der Rechercheur benutzt für seine Suche einen Satz von Kommandos, die als Retrievalsprache bezeichnet werden. Trotz Bemühungen einiger europäischer Hosts gibt es bisher keine allge-

mein verbindliche Retrievalsprache, so daß der Rechercheur mehrere Retrievalsprachen beherrschen muß [Kind 1991]. Einige Hosts bieten die Möglichkeit der menügesteuerten Abfrage in einigen ihrer angebotenen Datenbanken. In diesen Fällen entfällt das Erlernen der Retrievalsprache.

In den Datenfeldern *ab wann* und *bis wann* wird die Zeitspanne vermerkt, seit wann die Datenbank aufgebaut wird und bis wann sie weiter gepflegt wird. In der Regel ist keine Eintragung im Feld bis wann zu finden. Es gibt jedoch einige Datenbanken, die nicht weiter gepflegt werden, aber dennoch weiter online verfügbar sind. In diesen Fällen ist im Datenfeld bis wann ein Eintrag des Termins der Schließung der Datenbank vermerkt. Es schließt sich das Datenfeld *Aktualisierung* (Update) an. Hier wird die Frequenz der Pflege der Datenbank vermerkt. Online Datenbanken werden unterschiedlich häufig gepflegt. Die Aktualisierungsfrequenz reicht von täglich bis jährlich. Einige Datenbanken werden auch unregelmäßig gepflegt. Das Feld Nr. 33 *Pflege der DADB* bezieht sich hingegen auf die Pflege der Metadatenbank. *Umfang* und *Zuwachs pro Jahr* beziehen sich auf die Anzahl der Dokumente, die in der betreffenden Online Datenbank zu finden sind und auf die Anzahl der jährlich hinzukommenden Einträge. Es schließen sich fünf Datenfelder, die sich mit den Kosten der Recherche befassen, an. Zunächst einmal handelt es sich um die *Kosten/h*, d.h. die Anschaltkosten, die man für die Verbindung mit dem Host in der Datenbank zahlt. Die Bezeichnungen *Online Kosten* und *Offline Kosten* beziehen sich auf die Preise für die Ausgabe der einzelnen Dokumente online, d. h. am eigenen Rechner inklusive. des Herunterladens der Dokumente bzw. offline, d.h. bei der Übersendung der Dokumente durch den Host. Unter dem Begriff *SDI* (Selective Dissemination of Information) bezieht sich auf die selektive Informationsverbreitung. Es handelt sich um einen besonderen Vorgang zur regelmäßigen Ermittlung von Informationen aus Online Datenbanken. Der Nutzer formuliert ein Suchprofil in einer Datenbank als SDI-Suche. Diese Suche wird von einem Informationsretrieval-System bei jeder Aktualisierung der betreffenden Datenbank durchgeführt. Als Resultat erhält der Nutzer nach jeder Aktualisierung die gewünschten Antworten zu seinem Suchprofil [Barth 1992, S. 412]. Die Preise für die SDI Suche werden unter *SDI Kosten* aufgeführt. Für die Recherche in Online Datenbanken werden verschiedene Arbeitsunterlagen hergestellt. Dieses sind zunächst einmal die kurzen Datenbankbeschreibungen, die in Form von sog. Datenbankblättern, auch "fact sheets", "cream sheets" etc. genannt, herausgegeben werden. Auch gibt es für viele Datenbanken umfassende Manuals. Darüber hinaus ist für die inhaltliche Recherche in Datenbanken die Konsultation von Thesauri und Deskriptorenlisten zu empfehlen. In dem Datenfeld *Arbeitsunterlagen* werden diese vermerkt.

Es schließen sich die sog. inhaltserschließenden Datenfelder (27-31) an. Letztere sind in bezug auf die Thematik "Informationen über Chemie und Umweltschutz" oder "Umweltchemikalien" von übergeordneter Bedeutung. Die Eintragungen in diesen Feldern werden für den im weiteren beschriebenen Bewertungsansatz im besonderen herangezogen.

Unter dem Begriff *Informationstyp* oder *Informationsparameter* versteht man beispielsweise physikalisch-chemische Eigenschaften oder Strukturmerkmale. Es gibt Online Datenbanken, deren Schwerpunkte auf einzelnen Informationstypen liegen. Ebenso sind einige Datenbanken für spezielle *Anwendungen von Chemikalien* (z.B. Pestizide) ausgelegt. Darüber hinaus dient das Datenfeld *Anzahl der Chemikalien* als grobe Einstufung der Bedeutung der entsprechenden Online Datenbanken für das Thema. Es gibt Datenbanken, die zu über 100.000 Chemikalien Informationen vorhalten; bei anderen Quellen liegt die Anzahl der Stoffe bei einigen Hundert. Für das wichtigste inhaltserschließende Datenfeld, das Datenfeld *Deskriptoren (Parameter-Testdatensatz)*, wurde ein Thesaurus entwickelt, der ca. 100 Schlagwörter oder Suchbegriffe aus den Fachgebieten des Umwelt-, Gesundheits- und Arbeitsschutzes enthält.

Das Datenfeld *Testdatensatz Chemikalien* dient zur Darstellung, welcher Stoff aus der Testchemikalien-Liste mit 68 umweltrelevanten Chemikalien in der entsprechenden Online Datenbank enthalten ist.

Beide Testdatensätze werden in den Kapiteln 4.1.2 und 4.1.3 näher erläutert.

#### **4.1.2 Chemikalien-Testdatensatz**

Für die pragmatische Vorgehensweise im Zusammenhang mit der Problematik der Umweltchemikalien wird versucht, die Anzahl der zu betrachtenden Stoffe einzugrenzen. National und international existieren hierzu einige Auswahlverfahren von prioritären Stoffen [GDCh 1986, Walker 1993, Hushon 1984]. Solche Auswahlssysteme für Chemikalien sind nützlich bei frühen Stufen in der Bewertung von neuen Stoffen und bei der Überprüfung von großen Chemikalienlisten an alten Stoffen z.B. des European Inventory of Existing Commercial Chemical Substances (EINECS) [Klöppfer 1994]. Es wurde in Deutschland in einem mehrstufigen Selektierungsprozeß eine Stoffliste von 512 Stoffen erarbeitet [GDCh 1992].

##### ***4.1.2.1 Auswahlverfahren***

Die Liste der 512 Stoffe wurde weiter stufenweise eingegrenzt. Ziel war es, durch selektives Anlegen spezifischer Stoffkriterien, potentielle Lebensmittelkontaminanten zu erhalten. Unter Berücksichtigung des Vorkommens in der Umwelt, des Bioakkumulationspotentials und der Abbaubarkeit (Kriterien, die insbesondere auch potentielle Lebensmittelkontaminanten qualifizieren) wurden schließlich 68 Stoffe ausgewählt [Mücke 1987, Mücke 1989].

Diese Stoffliste, die als Grundlage für die weiteren Arbeiten dient, ist in Tabelle 3, alphabetisch geordnet nach dem systematischen Chemikaliennamen aufgelistet. Darüber hinaus werden die laufende Nummer (Testdatensatz Nummer) und die Chemical Abstracts Service Registry Number (CAS-Nummer) aufgeführt. In den letzten beiden Spalten werden die Anzahl der Treffer in DADB, d.h. in wievielen Online Datenbanken die einzelnen Chemikalien abgedeckt sind und deren prozentualer Anteil an der Gesamtheit der in DADB enthaltenen Datenbanken (Dokumentenstand: 453 Januar 1997) aufgeführt. Hierauf wird in Kapitel 6.4.2.3 noch näher eingegangen.

**Tabelle 3: Chemikalien-Testdatensatz (68 Stoffe Testdatensatz)**

Nr.	Name der Substanz	CAS-Nummer	Anzahl der Treffer in DADB	Anteil der Treffer in %
01	Acenaphthylene	208-96-8	171	38
02	Anthracene	120-12-7	280	62
03	Benz[a]anthracene	56-55-3	190	42
04	Benzo[k]fluoranthene	207-08-9	140	31
05	Benzene, 1,3-Bis(1-methylethyl)-	99-62-7	91	20
06	Benzene, 1-Chloro-2,4-dinitro-	97-00-7	134	30
07	Benzene, 1-Chloro-2-methyl-	95-49-8	120	27
08	Benzene, 1-Chloro-3-methyl-	108-41-8	97	21
09	Benzene, 1-Chloro-4-methyl-	106-43-4	11	24
10	Benzene, 1-Chloro-2-methyl-3-nitro-	83-42-1	56	12
11	Benzene, 4-Chloro-1-methyl-2-nitro-	89-59-8	47	10
12	Benzene, 1,4-Dibromo-	106-37-6	125	28
13	Benzene, 1,2-Dichloro-	95-50-1	209	46
14	Benzene, 1,3-Dichloro-	541-73-1	180	40
15	Benzene, 1,4-Dichloro-	106-46-7	222	49
16	Benzene, 1,2-Dichloro-3-nitro-	3209-22-1	76	16

17	Benzene, Ethenyl-	100-42-5	307	67
18	Benzene, Ethyl-	100-41-4	256	57
19	Benzene, Hexachloro-	118-74-1	266	59
20	Benzene, Methylbis(phenylmethyl)-	26898-17-9	34	8
21	Benzene, (1-Methylethyl)-	98-82-8	210	46
22	Benzene, (1-Methylethenyl)-	98-83-9	142	31
23	Benzene, 1,1'-Oxybis(methyl)-	28299-41-4	45	10
24	Benzene, Propyl-	103-65-1	160	35
25	Benzene, 1,2,4-Trichloro-	120-82-1	200	44
26	Benzene, 1,2,4-Trimethyl-	95-63-6	174	38
27	Benzeneamine, 2-Chloro,4-nitro-	121-87-9	81	18
28	Benzeneamine, N,N-Diethyl-	91-66-7	144	32
29	Benzeneamine, N-Phenyl-	122-39-4	216	48
30	Benzo[a]pyrene	50-32-8	221	49
31	1,1'-Biphenyl,3,3'-Dimethyl-	612-75-9	60	13
32	1,1'-Biphenyl,4,4'-Dimethyl-	631-33-2	54	12
33	1,1'-Biphenyl, Ethyl-	40529-66-6	47	10
34	1,1'-Biphenyl, Methyl-	28652-72-4	85	19
35	1,1'-Biphenyl, 3-Methyl-	643-93-6	93	21
36	9H-Carbazole	86-74-8	172	38
37	Cyclohexane	110-82-7	279	62
38	Cyclohexane, 1,1-Dimethyl-	590-66-9	76	17
39	Cyclohexane, 1,2-Dimethyl-	583-57-3	69	15
40	Cyclohexane, 1,3-Dimethyl-	591-21-9	67	15
41	Cyclohexane, 1,4-Dimethyl-	589-90-2	73	16
42	Cyclohexane, Methyl-	108-87-2	188	42
43	Dibenz[a,h]anthracene	53-70-3	157	35
44	Ethane, 1,1,2-Trichloro-1,2,2,-trifluoro-	76-13-1	151	33
45	Fluoranthene	206-44-0	207	46
46	9H-Fluorene	86-73-7	169	37
47	Hexadecane	544-76-3	196	43
48	Methanone, Diphenyl-	119-61-9	222	49
49	Naphthalene	91-20-3	267	59
50	Naphthalene, 1-Chloro-	90-13-1	143	32
51	Naphthalene, 2-Chloro-	91-58-7	128	28
52	Naphthalene, Dimethyl-	28804-88-8	92	20
53	Naphthalene, 1,3-Dimethyl-	575-41-7	76	17
54	Naphthalene, 2,6-Dimethyl-	581-42-0	104	23
55	Naphthalene, 2-Ethyl-	939-27-5	78	17
56	Naphthalene, Heptachloro-	32241-08-0	40	9
57	Naphthalene, 1-Methyl-	90-12-0	168	37
58	Naphthalene, 2-Methyl-	91-57-6	168	37
59	Naphthalene, Trichloro-	1321-65-9	71	16
60	Phenanthrene	85-01-8	209	46
61	Phenol, 4-(1,1-dimethylethyl)-	98-54-4	143	32
62	Phenol, 4-Dodecyl-	104-43-8	50	11
63	Phenol, 4,4'-(1-Methylethylidene)bis-	80-05-7	228	50
64	Phenol, 4-Nonyl-	104-40-5	142	31
65	Phenol, 4-(1,1,3,3-Tetramethylbutyl)-	140-66-9	107	24
66	Phenol, 2,4,5-Trichloro-	95-95-4	190	42
67	Pyrene	129-00-0	260	58

68	1,1',3',1''-Terphenyl,5'-Phenyl-	612-71-5	80	18
----	----------------------------------	----------	----	----

Erläuterungen siehe Text

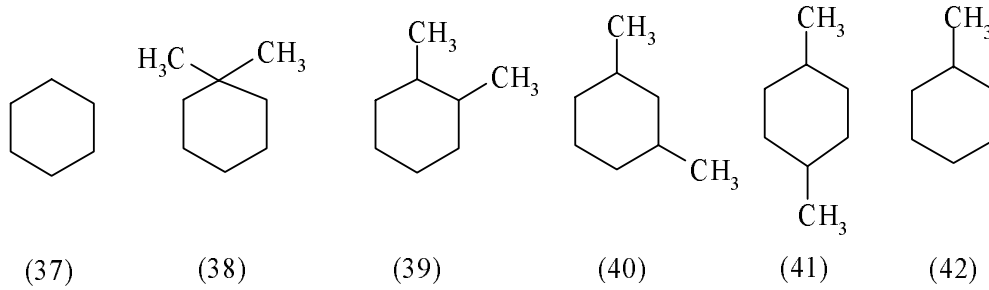
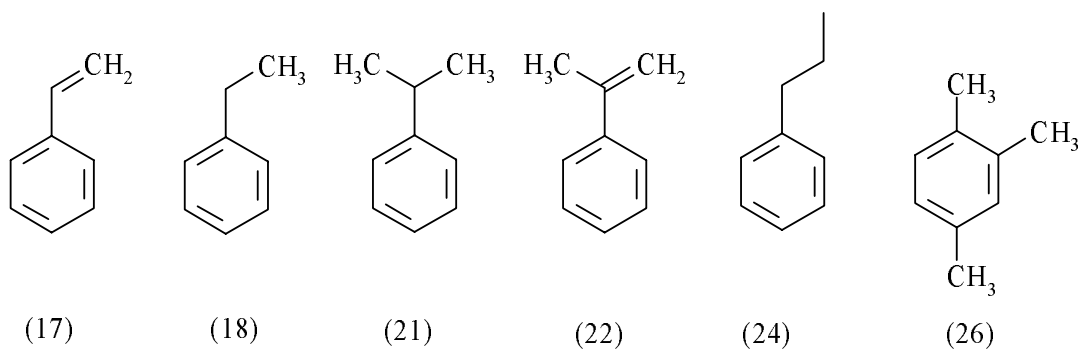
Die Chemikalien sind unter ihrer englischen Bezeichnung aufgeführt, da der überwiegende Anteil der zu betrachtenden Datenbanken in englischer Sprache aufgebaut sind. Es wird deutlich, daß die Liste eine ganze Reihe von unterschiedlichen z.T. sehr bekannten Umweltchemikalien enthält z.B. Polycyclische Aromaten, halogenierte Aromaten, Biphenyle etc. enthält.

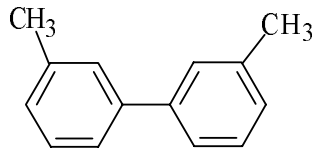
#### 4.1.2.2 Klassifikation nach Strukturen bzw. funktionellen Gruppen

**Tabelle 4: Chemische Klassifikation der Testdatensatz Stoffe** (Erläuterungen siehe Text)

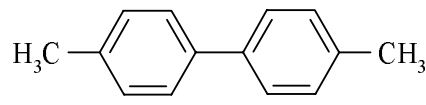
chemische Klassifikation			Anzahl der Testdatensatz Stoffe
<b>Kohlenwasserstoffe</b>			<b>38</b>
	Alkane		1
	Monocyclische Kohlenwasserstoffe		6
	Monocyclische Aromatische Kohlenwasserstoffe		14
	Polycyclische Aromaten		17
		Bicyclische Aromaten	7
		Tricyclische Aromaten	6
		Höhere Polycyclische Aromaten	4
<b>Ether</b>			<b>1</b>
<b>Alkohole</b>			<b>5</b>
<b>Ketone</b>			<b>1</b>
<b>Stickstoff enthaltende Organische Verbindungen</b>			<b>3</b>
<b>Halogen enthaltende Organische Verbindungen</b>			<b>20</b>

Die einzelnen Stoffe des Testdatensatzes von 68 umweltrelevanten organischen Chemikalien werden in der Tabelle 4 nach einer von Graedel vorgeschlagenen Einteilung der Chemikalien [Graedel 1986] zugeordnet. Den größten Anteil haben die Kohlenwasserstoffe mit einer Anzahl von 38 Stoffen. Hier überwiegen die Aromaten und die Polycyclischen Aromaten. Eine weitere starke Vertretung haben die Testdatensatz Stoffe mit einer Anzahl von 20 in der Gruppe der Halogen enthaltenden organischen Verbindungen. Diese Verbindungsklassen sollen nochmals beispielhaft in folgender Abbildung skizziert werden:

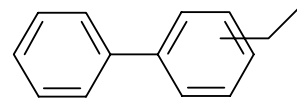
**Monocyclische Kohlenwasserstoffe: Substituierte Cyclohexane****Monocyclische Aromatische Kohlenwasserstoffe**

**Substituierte Biphenyle**

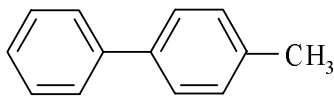
(31)



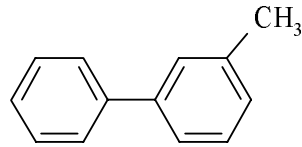
(32)



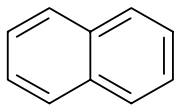
(33)



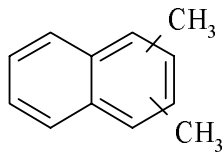
(34)



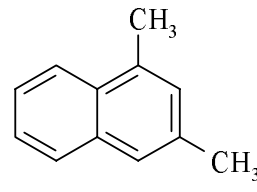
(35)

**Polycyclische Aromaten: Bicyclen**

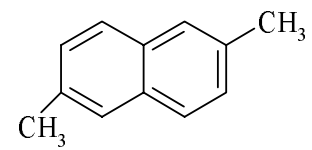
(49)



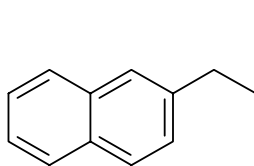
(52)



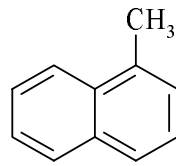
(53)



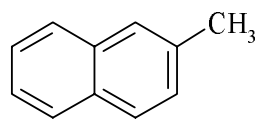
(54)



(55)



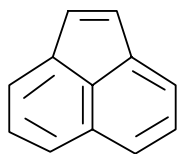
(57)



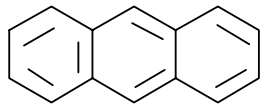
(58)

(52) Gemisch nicht-definierter Dimethyl-Naphthalene

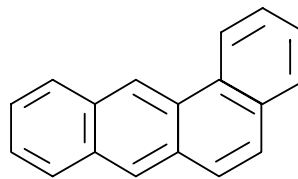
(53) und (54) definierte Dimethyl-Naphthalene

**Polycyclische Aromaten: Tricyclen**

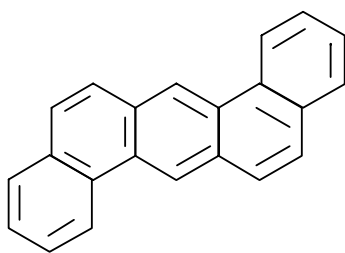
(1)



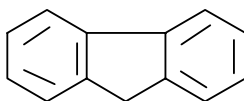
(2)



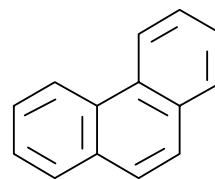
(3)



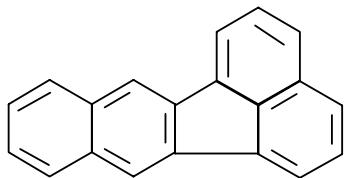
(43)



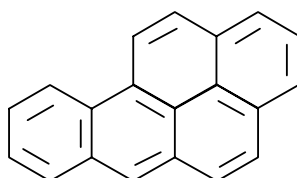
(46)



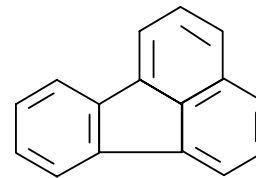
(60)

**Höhere Polycyclische Aromaten**

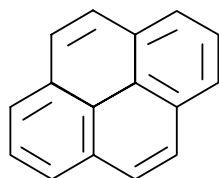
(4)



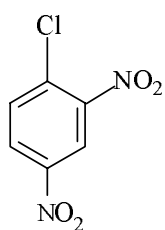
(30)



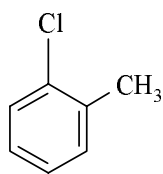
(45)



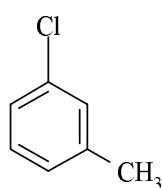
(67)

**Halogenierte Benzolderivate**

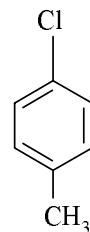
(6)



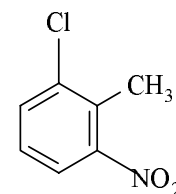
(7)



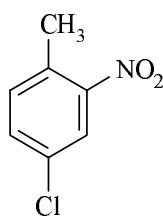
(8)



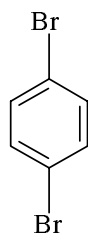
(9)



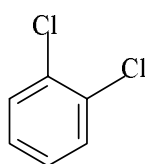
(10)



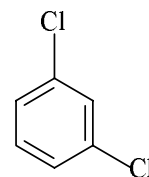
(11)



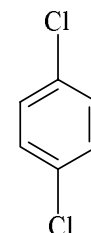
(12)



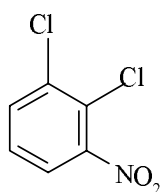
(13)



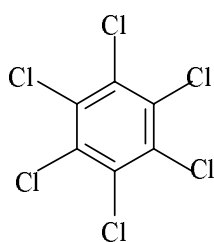
(14)



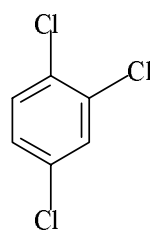
(15)



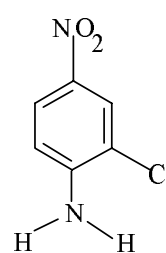
(16)



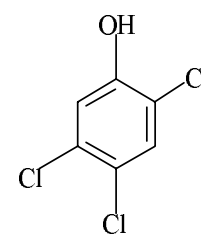
(19)



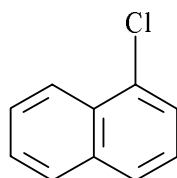
(25)



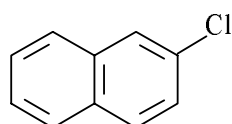
(27)



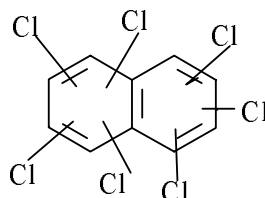
(66)

**Halogenierte Naphthalin-Derivate**

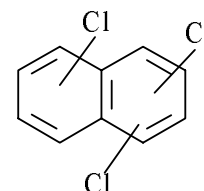
(50)



(51)



(56)



(59)

**Abbildung 4: Strukturen einiger ausgewählter Chemikalien des Testdatensatzes**

(Testdatensatz Nummern von Tabelle 3 sind in Klammern gesetzt)

Die oben aufgeführten und klassifizierten Umweltchemikalien umfassen nicht nur industriell hergestellte Stoffe (Industriechemikalien) sondern auch Stoffe, die nicht industriell produziert werden. Beispielsweise entstehen die Polycyclischen Aromatischen Kohlenwasserstoffe (Polycyclische Aromaten) bei der unvollständigen Verbrennung von praktisch allen organi-

schen Stoffen, z.B. durch Waldbrand, Großfeuerungsanlagen, Grillen, Räuchern und Rauchen [Larson 1994, S. 30] [Falbe 1992, S. 3521].

Bei den Halogenkohlenwasserstoffen handelt es sich um Kohlenwasserstoffe, bei denen H-Atome durch Halogene ersetzt sind [Falbe 1993, S. 1717]. Sie sind nützliche Zwischenprodukte für die organische Synthese, z.B. für Grignard-Reaktionen, für nucleophile Substitutionen, für Eliminierungen und für elektrophile Substitutionen. Außerdem finden sie Verwendung als Lösemittel, Anästhetika, Feuerlöschmittel, Kältemittel und Treibmittel. Besonders die Chlorfluorkohlenwasserstoffe (FCKWs) sind wegen ihrer wahrscheinlichen Umweltbelastung insbesondere wegen der schädlichen Einflüsse auf die Ozonschicht der Atmosphäre in Mißkredit gekommen [Falbe 1992, S. 1717], [Hulpke 1994, S. 326]. Das Auftreten von Halogenkohlenwasserstoffen ist aufgrund der zahlreichen Verwendungen und der Persistenz dieser Chemikalienklasse praktisch ubiquitär. Hinsichtlich des Umweltverhaltens ist die Persistenz der hochhalogenierten Vertreter (z.B. Hexachlorbenzol Stoff Nr. 19) bedeutungsvoll, die zur Anreicherung dieser Stoffe in der unbelebten Natur und auch zur Bioakkumulation führt. Die Dichlorbenzole (Stoffe Nr. 13,14,15) werden ebenso als Zwischenprodukt für die chemische Industrie, sowie als Lösemittel eingesetzt [Hulpke 1994, S. 326]. Fünf der im Testdatensatz aufgeführten Chemikalien gehören zu den höchstproduzierten Chemikalien in den USA [Kirschner 1996]. Es handelt sich dabei um Bisphenol A (Nr. 63), Cumol (Nr. 21), Cyclohexan (Nr. 37), Ethylbenzol (Nr. 18) und Styrol (Nr. 17).

Eine eindeutige Zuordnung von Chemikalien zu einem Anwendungsgebiet ist in der Regel demnach nicht möglich, da Chemikalien entweder nicht durch die chemische Industrie produziert werden oder falls sie hergestellt werden, meist nicht nur eine sondern mehrere Anwendungen haben. Man spricht in diesem Zusammenhang dann von einem Anwendungsmuster einer Chemikalie. Solche Anwendungsmuster ändern sich ständig und werden aus diesem Grund von verantwortlicher Seite aus selten publiziert. Ein nennenswerter Ansatz zur Einteilung von Chemikalien in Anwendungsbereiche wurde jedoch vor dem Hintergrund der Abschätzung der Eintragungen von Chemikalien in die Umwelt vorgenommen. Es wurden hier 25 Anwendungsklassen von Chemikalien herausgearbeitet [Brüggemann 1989, 1992].

### **4.1.3 Parameter-Testdatensatz**

Zur inhaltlichen Auswertung von Chemie- und Umweltschutz-Datenbanken wird neben einem Chemikalien-Testdatensatz auch noch ein Parameter-Testdatensatz benötigt. Dieser Testdatensatz soll Aufschluß darüber geben, welche Umwelt- und Chemikalien-relevanten Parameter in den einzelnen Datenbanken zu finden sind.

#### ***4.1.3.1 Auswahlverfahren***

Die Auswahl der Parameter erfolgt in Anlehnung an die für die Anmeldung von Chemikalien geforderten Angaben nach dem Chemikaliengesetz [ChemG 1994].

Im Gesetz werden folgende Punkte aufgeführt:

- Identitätsmerkmale
- Nachweis-, Bestimmungs- und Analysemethoden
- Angaben zur Herstellung, Verwendung, Exposition und Verbleib
- schädliche Wirkung bei der Verwendung
- Hinweise zur Toxikokinetik
- Einstufung, Verpackung und Kennzeichnung
- Empfehlungen über die Vorsichtsmaßnahmen beim Verwenden und über Sofortmaßnahmen bei Unfällen
- die Menge des Stoffes, die jährlich produziert wird

- Verfahren zur Entsorgung und zur möglichen Wiederverwendung
- *Prüfnachweise* in Abhängigkeit von der Produktionsmenge

Die Prüfnachweise erstrecken sich auf:

- physikalisch, chemische und physikalisch-chemische Eigenschaften
- akute Toxizität, subakute Toxizität, subchronische Toxizität, chronische Toxizität
- krebserzeugende Eigenschaften, erbgutverändernde Eigenschaften, fortpflanzungsgefährdende Eigenschaften
- peri- und postnatale Wirkungen
- Organ- und Systemtoxizität
- reizende und ätzende Eigenschaften, sensibilisierende Eigenschaften
- abiotische Abbaubarkeit, biologische Abbaubarkeit, Bioakkumulation
- Adsorption, Desorption
- Toxizität gegenüber Wasserorganismen, Hemmung des Algenwachstums, Bakterieninhibition
- Toxizität gegenüber Bodenorganismen, Toxizität gegenüber Pflanzen
- Toxizität gegenüber Fischen, Toxizität gegenüber Vögeln

Neben der Aufarbeitung dieser im Chemikaliengesetz genannten Parameter wurde der Umweltthesaurus des Umweltbundesamtes zur Aufstellung des Parameter-Testdatensatzes mit herangezogen [Umweltbundesamt 1991]. Dieser Thesaurus dient als Grundlage zur inhaltlichen Auswertung der Dokumente, die in der Umweltliteraturdatenbank (ULIDAT), der umfangreichsten deutschsprachigen Umweltdatenbank enthalten sind.

Es handelt sich hierbei um einen hierarchischen Thesaurus, der in den Begriff, dessen Oberbegriff (OB) und Unterbegriffe (UB) unterteilt ist.

Für den Begriff "Schadstoffbewertung" findet man hier folgende Einteilung:

*Schadstoffbewertung*

OB: Schadstoff

UB: Chemikalienprüfung, Hautverträglichkeit, Ökotoxikologische Bewertung, Toxikologische Bewertung, Pflanzenverträglichkeit

Für den Begriff "Chemikalienprüfung" ist vermerkt:

*Chemikalienprüfung*

OB: Schadstoffbewertung, Umweltchemikalien

UB: In-vitro, In-vivo, Karzinogenitätsprüfung, Mutagenitätsprüfung, Teratogenitätsprüfung, Life-cycle-Toxizitätstest, Miti-Test, Pflanzenschutzmittelprüfung.

Die oben aufgeführten Begriffe des Umweltthesaurus sind im Vergleich zum Gesetzestext lückenhaft. Hier wird nur auf die ökotoxikologische Bewertung und auf die toxikologische Bewertung, nicht aber auf Identitätsmerkmale, Produktionsmengen, Verwendungsarten, physikalisch-chemische Eigenschaften, Arbeitsschutzaspekte etc. eingegangen. Im gesamten Umweltthesaurus des Umweltbundesamtes sind diese Begriffe jedoch verzeichnet.

#### **4.1.3.2 Einteilung der Parameter in Anlehnung an das Chemikaliengesetz**

Der Parameter-Testdatensatz umfaßt über 100 Begriffe, die im Zusammenhang mit dem Umweltschutz, Arbeitsschutz und Gesundheitsschutz von Chemikalien zu sehen sind. Neben den oben aufgeführten Begriffen kommen als wichtige Schwerpunkte noch die Parameter zum Vorkommen von Chemikalien in Umweltmedien hinzu.

Die zum Teil vom Gesetzestext abweichende Bezeichnungsweise ist dadurch zu erklären, daß in der Literatur meist die nachstehend aufgeführten Bezeichnungen zu finden sind.

Die gewählten Parameter des Testdatensatzes lassen sich somit in folgende in Tabelle 5 aufgelisteten Kategorien A bis J einteilen:

**Tabelle 5: Einteilung der Parameter in Kategorien**

Kategorie	Parameter
A	Identifikationsmerkmale
B	Angaben zur Verwendung
C	Ökonomische Daten
D	Vorkommen in der Umwelt
E	Physikalisch-chemische Eigenschaften
F	Abbau / Akkumulation
G	Ökotoxizität
H	Toxizität
I	Arbeitsschutz
J	Sonstiges z.B. gesetzliche Regelungen, experimentelle Methoden

Im folgenden werden die ca. 100 Parameter diesen Kategorien zugeordnet.

**Tabelle 6: Zuordnung der Testdatensatz Parameter zu Kategorien**

<b>A. Identifikationsmerkmale</b>	Identifikationsnummern Beilstein-Nummer, CAS-Nummer, EG-Nummer, RTECS-Nummer, UN-Nummer, WLN Nummer, Chemikaliename, Synonyma, Handelsname Molekulargewicht, Summenformel, Strukturformel Spektraldaten, Stoffbeschaffenheit
<b>B. Angaben zur Verwendung, Entsorgung</b>	Abfallaufkommen, Entsorgung Anwendung / Verwendung Anwendungsmuster
<b>C. Ökonomische Daten</b>	Einfuhr / Ausfuhr, Produktionsvolumen, Vermarktungsmenge
<b>D. Vorkommen in der Umwelt</b>	Atmosphäre, Biosphäre, Hydrosphäre (Meer, Süßwasser), Grundwasser, Klärschlamm, Pedosphäre, Sediment Lebensmittel, Trinkwasser
<b>E. Physikalisch-chemische Eigenschaften</b>	Brechungsindex, Dampfdruck, Dissoziationskonstante, Entzündlichkeit, Explosionspunkt, Flammpunkt, Henry-Konstante, Löslichkeit (Fett, organische Lösungsmittel), Wasserlöslichkeit, Oberflächenspannung, Relative Dichte, Schmelzpunkt, Siedepunkt, Thermodynamische Daten, Verteilungskoeffizienten ( $K_{ow}$ , $K_{oc}$ ), Viskosität, Zersetzungspunkt, Zündtemperatur
<b>F. Abbau / Akkumulation</b>	Abiotische Abbaubarkeit, Bioabbau, Bioakkumulation, Geoakkumulation, Photoabbau

<b>G. Ökotoxizität</b>	Alge, Biene, Bakterien, Daphnia, Fisch, Insekten, Krustaceen Mollusken, Pflanzen, Photosynthesehemmung, Protozoen, Würmer
<b>H. Toxizität</b>	Akute Toxizität orale, percutane dermale, inhalative, intravenöse, intraperitoneale Augenreizung, Hautreizung Carcinogenität, Mutagenität, Teratogenität, Reproduktionstoxizität Subakute, Subchronische Toxizität, Chronische Toxizität
<b>I. Arbeitsschutz</b>	Brand- und Explosionsschutz, Erste Hilfe, Gefahrensymbole, Gefahrenbezeichnung Grenzwerte: BAT-, MAK-, TLV-, TRK-, TWA-Werte Löschmittel, Maßnahmen bei Unfällen und Bränden, Maßnahmen nach Verschütten, Auslaufen, Schutzmaßnahmen, Lagerung, Handhabung, R,S-Sätze
<b>J. Sonstiges</b>	Experimentelle Methoden, Gesetzliche Regelungen, QSAR, etc.

#### **4.1.4 Auswertung der Metadatenbank der Online Datenbanken (DADB) gemäß der Typisierung der Online Datenbanken**

In Kapitel 2.5 wurde auf die Typisierung der Online Datenbanken gemäß Chemikalien- und Umweltschutzaspekten eingegangen. An dieser Stelle folgt nun die Auswertung der DADB - Metadatenbank der Online Datenbanken in bezug auf die Datenbanktypen. Von den in der Metadatenbank der Online Datenbanken und der Metadatenbank der CD-ROMs im Januar 1997 bearbeiteten und gespeicherten 453 Online Datenbanken und 347 CD-ROMs ist die Verteilung auf die einzelnen bereits näher erläuterten Typen in der Tabelle 7 angegeben. Neben der Anzahl der Datenquellen ist der prozentuale Anteil in bezug auf die Gesamtzahl der Dokumente in der korrespondierenden Metadatenbank aufgeführt.

Es zeigt sich allerdings, daß die textbasierten Datenbanken mit einer Anzahl von 269 Datenbanken (bei DADB) deutlich den 179 faktenbasierten Datenbanken und den integrierten Datenbanken mit einer Anzahl von 5 überlegen sind. Prozentual ausgedrückt gibt es in DADB 59 % textbasierte Datenbanken, 40 % faktenbasierte Datenbanken und nur 1 % integrierte Datenbanken. Bei den textbasierten Datenbanken liegt das Schwergewicht deutlich auf den bibliographische Datenbanken und bei den faktenbasierten Datenbanken sind die Faktendatenbanken (numerische Datenbanken) am Stärksten vertreten.

**Tabelle 7: Typen von Datenbanken in DADB und DACD**

Oberbegriff	Akronym	Datenbank Typ	Dokumente in DADB / %	Dokumente in DACD / %
<i>faktenbasierte Datenbanken</i>	NU	Faktendatenbanken (numerische Datenbanken)	106 / 29	88 / 25
	CN	Chemikaliennamen-Verzeichnisse	13 / 3	-
	VZ	Firmenverzeichnisse und sonstige Verzeichnisse	16 / 4	12 / 3
	FD	Forschungsdatenbanken	20 / 5	7 / 2
	ST	Strukturdatenbanken	6 / 1	2 / 1
	MD	Metadatenbanken	18 / 4	7 / 2
<i>textbasierte Datenbanken</i>	BI	Bibliographische Datenbanken	230 / 51	177 / 51
	VT	Volltextdatenbanken	39 / 8	47 / 13
<i>integrierte Datenbanken</i>	RE	Reaktionsdatenbanken	5 / 1	10 / 3
	MM	Multi-Media Discs	0 / 0	1 / 0

Dokumentenstand (Februar 1997): DADB: 453 Dokumente, DACD: 347 Dokumente  
Erläuterungen siehe Text

## 4.2 Metadatenbank der CD-ROMs - DACD

Die Metadatenbank der CD-ROMs wurde ebenfalls mit der kommerziellen Software LARS aufgebaut. Der Dokumentenstand beträgt **347 Einträge** (Stand: 31.01.97). Auch stimmen viele Datenfelder mit denen in DADB überein. Die Unterschiede zu dem Aufbau der Metadatenbank der Online Datenbanken DADB (vgl. hierzu 4.1) werden im folgenden herausgearbeitet.

### 4.2.1 Datenbankdefinition von DACD

Die Datenbankdefinition von DACD ist analog strukturiert wie die bereits beschriebene Metadatenbank der Online Datenbanken - DADB (vgl. Tabelle 2).

Unter dem Feld **Sprache** ist die Sprache zu verstehen, in der die Datenbank aufgebaut wurde. Über 80 % der in DACD inkorporierten Datenbanken sind auf englisch. Da einige deutsche Standardwerke als CD-ROM publiziert werden, z.B. das Römpp Chemie Lexikon [CD-Römpp 1995], die Sicherheitstechnischen Kenndaten chemischer Stoffe [Gefahrstoff CD 1995] das Handbuch der Gefährlichen Güter Hommel [Gefahrgut CD-ROM 1995] ist der Anteil an deutschen Datenquellen auf CD-ROM deutlich höher als bei den Online Datenbanken. Unter dem Begriff **Retrievalsprache** ist die jeweilige Abfragesprache für die CD-ROMs zu verstehen. In der Regel handelt es sich hierbei um menügestützte Retrievalsoftware, wie bereits in Kapitel 2.6 beschrieben wurde. Die Frequenz der Pflege der Datenbank wird im Datenfeld **Update** vermerkt. CD-ROMs werden in der Regel alle drei bis sechs Monate gepflegt. Die Aktualisierungsfrequenz unterscheidet die CD-ROMs daher deutlich von den Online Datenbanken. Erstere werden in der Regel häufiger gewartet und sind damit aktueller. Im Gegensatz zu den Recherchen in Online Datenbanken, ist in CD-ROMs die Dauer und Häufigkeit der Recherche, die Anzahl der ausgegebenen Dokumente, der Ausdruck der Dokumente unabhängig vom Preis. Der Preis einer CD-ROM umfaßt den einmaligen Anschaffungspreis und den Preis für die Aktualisierungen.

### **4.3.2 Auswertung der Metadatenbank der CD-ROMs (DACD) gemäß der Typisierung der CD-ROMs**

Von den in der Metadatenbank der CD-ROMs im Januar 1997 bearbeiteten und gespeicherten 347 CD-ROMs ist die Typisierung in Tabelle 7 (rechte Spalte) angegeben.

Auch bei den CD-ROMs überwiegen die textbasierten Datenbanken mit einer Anzahl von 224 gegenüber den faktenbasierten Datenbanken mit einer Anzahl von 116 und den integrierten Datenbanken mit nur 11 Einträgen. Prozentual ausgedrückt, gibt es in DACD 64 % textbasierte Datenbanken gegenüber 33 % faktenbasierten Datenbanken und 3 % integrierten Datenbanken.

Nur sehr wenige Chemiedatenbanken verfügen über Strukturformeln und über die Möglichkeiten der Struktursuche und der Substruktursuche. In einer weiteren Publikation zum Thema Chemiedatenbanken werden das Chapman & Hall CHCD Dictionary of Natural Products on CD-ROM und die Beilstein Current Facts in Chemistry on CD-ROM vorgestellt und erläutert. Da es sich hierbei um zwei CD-ROMs handelt, die auch über chemische Strukturen verfügen, wird besonders auf die Recherche nach diesen Merkmalen eingegangen. Es handelt sich bei den CD-ROMs nicht nur um Strukturdatenbanken sondern auch um Faktendatenbanken, die darüber hinaus auch noch bibliographische Informationen enthalten. Auch diese Autorin kommt zu der Schlußfolgerung, daß es eine zunehmend größere Nachfrage nach CD-ROMs mit Strukturinformation gibt [Warr 1994a]. In einem weiteren Artikel von W. Warr geht die Autorin auf Strukturdatenbanken und Reaktionsdatenbanken ein. Sie beschreibt die Reaktionsdatenbank CD-Select, eine der wenigen Reaktionsdatenbanken auf CD-ROM [Warr 1994b].

Die in den oben betrachteten Publikationen dargelegten Tatsache nach dem zunehmenden Interesse nach Struktur- und Reaktioneninformationen auf CD-ROM läßt sich ebenfalls in der Aufstellung der Tabelle 7 bestätigen. Es ist eine deutliche Verbesserung in dem Bereich der Reaktionsdatenbanken zu vermerken. Auch ist eine Zunahme der CD-ROMs, die Struktursuchen ermöglichen, zu verzeichnen.

### **4.3 Gegenüberstellung der Datenbanken-Typen in DADB und DACD**

Von zusätzlichem Interesse ist der Vergleich der Online Datenbanken mit den Datenbanken auf CD-ROM in bezug auf die Aufteilung der einzelnen Quellen auf die diskutierten Datenbank Typen. Aus Tabelle 7 läßt sich erkennen, daß sowohl bei Online Datenbanken als auch bei CD-ROMs die textbasierten Datenbanken deutlich gegenüber den faktenbasierten Datenbanken und integrierten Datenbanken überwiegen. Über die Hälfte der betrachteten Quellen sind bibliographischer Art. Bei den CD-ROMs überwiegen die Volltextdatenbanken mit einem Prozentsatz von 13 gegenüber 8 % bei den Online Datenbanken. Faktenbasierte Datenbanken kommen im Medium Online mehr vor als im Medium CD-ROM. Dieses bezieht sich vor allem auf Chemikaliennamen-Verzeichnisse (nicht bei CD-ROMs vertreten), Forschungsdatenbanken und Metadatenbanken. Hingegen dominieren die Faktendatenbanken (numerische Datenbanken) bei den CD-ROMs.

Bei den integrierten Datenbanken ist die Situation bei den CD-ROMs mit einem 3 %igen Anteil an der Gesamtheit gegenüber dem 1 %igen Anteil bei den Online Datenbanken deutlich positiver.

## 5. Auswahl von Online Datenbanken und CD-ROMs in Chemie und Umweltschutz

### 5.1 Auswahl der Gruppen an Online Datenbanken

#### 5.1.1 Vorbemerkung

Wie bei allen Bewertungsverfahren muß zunächst eine repräsentative Auswahl geeigneter Objekte getroffen werden, d.h. es muß die Objektmenge definiert werden. Anders ausgedrückt: Welche Online Datenbanken sind für die Beantwortung von Fragestellungen in der Chemie und im Umweltschutz geeignet? In diesem Zusammenhang spielt auch der Typ der Online Datenbanken eine Rolle.

Es gibt national und international einige Ansätze, aus der Gesamtheit der Online Datenbanken, diejenigen, die sich mit den Fragestellungen in Chemie und Umweltschutz beschäftigen, herauszuarbeiten.

Einige Datenbankführer spezialisieren sich auf Datenbanken in Naturwissenschaft und Technik, z.B. das jährlich als Buch erscheinende und vierteljährlich als CD-ROM publizierte "Handbuch der Datenbanken für Naturwissenschaft und Technik" [Scientific Consulting 1995b]. Ebenso seien an dieser Stelle das "Verzeichnis von Datenbanken in der Technik" [Pilkington 1992] und das von der British Library herausgegebene Standardwerk "Online Searching in Science and Technology" [Orton 1995] genannt. Der englischsprachige Umweltdatenbankführer Environment Databases [Cox 1991] beschreibt 58 bibliographische Umweltdatenbanken. Das deutschsprachige "Umwelt-Datenbank-Verzeichnis" [Ottahal 1989] liefert Informationen zu über 200 umweltrelevanten Datenbanken jeden Typs. Die Publikation "Datenquellen für Umweltchemikalien" befaßt sich mit der Untersuchung von 104 Faktendatenbanken [Voigt 1990]. Weiterhin wurde vor einigen Jahren das Thema Umweltdatenbanken in einem Sonderheft der Fachzeitschrift DATABASE abgehandelt [Online 1992]. Unter dem Titel "Environment Online - The Greening of the Databases" wurden die Datenbanken in folgende drei Gruppen inhaltlich aufgeteilt:

- Allgemeine Datenbanken [Stoss 1991]
- Wissenschaftlich und technische Datenbanken [Gayle Alston 1991]
- Wirtschafts- und Rechtsdatenbanken [Gayle Alston 1992].

Mit der Zusammenstellung von Chemiedatenbanken befaßt sich Orton im Kapitel Database Reviews: Chemistry [Orton S. 61-66, 1995].

Auf die Auswahl geeigneter Datenbanken für spezielle Themenbereiche gehen die Datenbankanbieter mit der Vorhaltung von hostspezifischen Datenbankgruppen ein. Diese Datenbankgruppen nennt man je nach Host Cluster (STN), Superbase (DIMDI), CROS-Groups (DATASTAR) oder OneSearch Categories (DIALOG). Zur Vereinfachung wird im folgenden der Begriff "**Gruppe**" benutzt. Bei den meisten Hosts gibt es sowohl Chemie- als auch Umwelt-Gruppen. Die Ausnahme bildet DIMDI, ein auf toxikologische und medizinische Datenbanken spezialisierter Host, der keine Umwelt-Gruppe vorhält. Gemäß einer Auswertung im Endbericht des Forschungsvorhabens Informationssystem Umweltchemikalien [Voigt, S. 31, 1995c] sind die Hosts DATASTAR, DIALOG, DIMDI und STN diejenigen, deren Datenbankangebot am meisten in der Metadatenbank der Online Datenbanken (Kapitel 4.1) vertreten sind. Aus diesem Grunde werden deren Gruppen analysiert und zur Grundlage des sich anschließenden Rangfolgeverfahrens herangezogen. Alle im folgenden genannten Datenbanken sind in der Metadatenbank der Online Datenbanken ausgewertet.

## **5.1.2 Chemie-Gruppe**

### ***5.1.2.1 Übersicht***

Für das umfangreiche Gebiet der Chemiedatenbanken werden von den Hosts entweder keine oder gleich mehrere Gruppen angeboten.

Bei **DATASTAR** unterscheidet man die Gruppen **Chemistry** (CROS Group 52, 74) mit 14 Datenbanken und **Chemical Industry** (CROS Group 2123) mit 10 Datenbanken [DATA-STAR 1995].

Beim Datenbankanbieter **DIALOG** unterscheidet man die folgenden sechs Gruppen: **CAS Registry Numbers - Chemical Files** (RNCHEM) mit 16 Datenbanken, **Chemical Business News** (CHEMBUS) mit 23 Datenbanken, **Chemical Engineering** (CHEMENG) mit 15 Datenbanken, **Chemical Literature** (CHEMLIT) mit 22 Datenbanken, **Chemical Properties** (CHEMPROP) mit fünf Datenbanken und **Chemical Substance Directories** (CHEMSUBS) mit 6 Datenbanken [Knight-Ridder Information 1995a].

**STN** stellt die Gruppen **Chem Properties and Data Cluster - Numeric Data** (CHEMDATA) mit 19 Datenbanken, **Chemical Engineering Cluster** (CHEMENG) mit 19 Datenbanken, den **Chemical Literature Cluster** (CHEMISTRY) mit 39 Datenbanken und den **Fulltext Chemical Journals Online Cluster** (CJO) mit sechs Datenbanken zur Verfügung [FIZ Karlsruhe 1995].

Die Hosts unterscheiden z.T. bei ihren Gruppen ebenfalls faktenbasierte und textbasierte Datenbanken (vgl. hierzu die Ausführungen in Kapitel 2.5). Darüber hinaus werden wirtschaftliche Informationen über Chemikalien in getrennten Gruppen vorgehalten. Jedoch kommt es bei diesen Aufteilungen zu zahlreichen Doppelnennungen.

### ***5.1.2.2 Einengende Auswahl***

Für weitere Untersuchungen zum Thema Chemikalien und Umweltschutz wurden einige Datenbanken nicht berücksichtigt. Es handelt sich dabei um Patentdatenbanken (vgl. hierzu die Ausführungen im Anhang A), Datenbanken, die sich mit einer Chemikalie befassen z.B. Kupfer, im Fall der Datenbank COPPERDATA oder Aluminium im Beispiel von AAASD. Ebenso werden Datenbanken in diesem Ansatz ausgelassen, die sich ausschließlich mit anorganischen Chemikalien befassen. Diese sind zwar von großer Bedeutung im Zusammenhang mit Fragestellungen zum Thema Chemikalien und Umweltschutz, jedoch sind in dem gewählten Chemikalien-Testdatensatz (siehe Kapitel 4.1.2) nur organische Chemikalien enthalten. Aus diesem Grunde würde es zu keiner objektiven Bewertung von umfangreichen anorganischen Datenbanken wie z.B. den GMELIN, ICSC und METADEX kommen. Auch werden Datenbanken nicht berücksichtigt, die nur für eingeschränkte Nutzerkreise online angeboten werden.

Somit wurden 65 Online Datenbanken herausgearbeitet, die in Tabelle 8 alphabetisch nach ihrem Namen bzw. deren gängiger Kurzbezeichnung aufgeführt werden. Neben der Nennung der Zugehörigkeit zu einem der drei Hosts DATASTAR, DIALOG und STN ist in der letzten Spalte der Typ der Datenbank genannt. Die fett gedruckten Datenbanken kommen nicht nur in der Chemie-Gruppe, sondern auch in der Umwelt-Gruppe (Kapitel 5.1.3) vor. Auffällig ist darüber hinaus, daß es nur sechs Datenbanken gibt, die von allen drei Hosts gleichzeitig als chemierelevante Datenbank angesehen werden.

**Tabelle 8: Chemie-Gruppen bei den Hosts DataStar, DIALOG und STN** (fettgedruckte Datenbanken sind auch in den Umwelt-Gruppen vorhanden)

Nr.	Datenbank	DATASTAR	DIALOG	STN	Typ
01	Agrochemicals Handbook		X		NU
<b>02</b>	<b>ANABST</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>BI</b>
<b>03</b>	<b>APILIT</b>		<b>X</b>	<b>X</b>	<b>BI</b>
04	BEILSTEIN		X	X	NU
05	BIOTECHABS			X	BI
<b>06</b>	<b>CA</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>BI</b>
<b>07</b>	<b>CAB</b>			<b>X</b>	<b>BI</b>
<b>08</b>	<b>CAPLUS</b>			<b>X</b>	<b>BI</b>
<b>09</b>	<b>CAPREVIEWS</b>			<b>X</b>	<b>BI</b>
10	CEABA		X	X	BI
<b>11</b>	<b>CBNB</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>BI</b>
<b>12</b>	<b>CEN</b>			<b>X</b>	<b>BI</b>
13	CHCD		X		NU
14	CHEM-INTELL Chemical Plant	X			VZ
15	CHEM-INTELL Trade & Prod.	X			NU
16	Chemical Registry Nomenclature	X			CN
17	CHEMSAFE			X	NU
18	CHEMSEARCH			X	NU
<b>19</b>	<b>CIN</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>BI</b>
<b>20</b>	<b>CJACS</b>			<b>X</b>	<b>VT</b>
21	CJAOAC			X	VT
22	CJELSEVIER			X	VT
23	CJRSC			X	VT
24	CJVCH			X	VT
25	CJWILEY			X	VT
<b>26</b>	<b>COMPENDEX</b>		<b>X</b>	<b>X</b>	<b>BI</b>
27	CONFSCI		X	X	BI
<b>28</b>	<b>CSNB</b>	<b>X</b>	<b>X</b>		<b>BI</b>
29	Derwent Crop Protection File	X			NU
30	DETERM			X	NU
31	DIPPR			X	NU
32	DISSABS			X	BI
33	EMA		X		BI
34	European Chemical News	X			BI
35	Fine Chemicals Database		X		NU
36	FSTA	X			BI
37	HODOC			X	NU
<b>38</b>	<b>HSDB</b>	<b>X</b>		<b>X</b>	<b>NU</b>
39	INSPEC		X	X	BI
40	INVESTEXT			X	VT
41	IPA			X	BI
42	JANAF			X	NU
<b>43</b>	<b>JICST-E</b>			<b>X</b>	<b>BI</b>
44	JPNEWS			X	BI
45	KIRK		X		VT
46	KKF			X	BI
47	KOSMET	X			BI
48	MERCK		X	X	NU
49	NAPRALERT			X	BI
<b>50</b>	<b>NTIS</b>		<b>X</b>	<b>X</b>	<b>BI</b>
<b>51</b>	<b>PAPERCHEM</b>		<b>X</b>	<b>X</b>	<b>BI</b>
52	Pascal		X		BI
<b>53</b>	<b>PIRA</b>		<b>X</b>		<b>BI</b>
54	PLASPEC		X		NU
55	Polymer Online		X		VT
<b>56</b>	<b>PROMT</b>			<b>X</b>	<b>BI</b>
57	RAPPRA		X	X	BI
58	RTECS	X	X	X	NU
<b>59</b>	<b>SCISEARCH</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>BI</b>
60	SPECINFO			X	NU

61	TOXLINE	X			BI
62	TRCTHERMO			X	NU
63	TSCA		X		NU
64	USAN			X	NU
65	VTB			X	BI

Abkürzungen für den Datenbanktyp siehe Abkürzungsverzeichnis

Es handelt sich hierbei um ANABST (Analytical Abstracts), CA (Chemical Abstracts), CBNB (Chemical Business Newsbase), CIN (Chemical Industry Notes), RTECS (Registry of Toxic Effects of Chemical Substances) und SCISEARCH (Science Citation Search System). Eine Übersicht über die Zugehörigkeit von Datenbanken zu einer oder mehrerer Gruppen ist aus Tabelle 7 schwer zu erhalten. Auch hier kann mit Erfolg die Visualisierung anhand der Hasse-Diagramm-Technik herangezogen werden, wobei hier nicht die komponentenweise  $\leq$ -Relation sondern die Inklusionsrelation die Realisierung der Ordnungsrelation darstellt. Für eine nähere Erörterung sei auf die Arbeit von Brüggemann et al. verwiesen [Brüggemann 1995b].

Es sei an dieser Stelle darauf hingewiesen, daß nur der Host STN die Datenbank CA auch mit Kurzreferaten (Abstracts) anbietet. Es kommt durchaus vor, daß chemierelevante Datenbanken zwar im Angebot eines Hosts sind, jedoch nicht der Chemie-Gruppe zugeordnet werden. Beispielsweise trifft dies für die Datenbank CAB (CAB Abstracts) zu: CAB ist zwar in der Gruppe von STN und wird auch von DATASTAR und DIALOG angeboten, ist aber nicht in deren Chemie-Gruppe. Ähnlich ist die Situation bei der Datenbank CEABA (Chemical Engineering and Biotechnology Abstracts), die zwar im Angebot von DATASTAR aber nicht in dessen Chemie-Gruppe ist. COMPENDEX wird zwar von DATASTAR vorgehalten, ist jedoch in keinem seiner Chemie-Gruppen zu finden. DISSABS (Dissertation Abstracts) gibt es im Angebot von DIALOG, jedoch nicht in dessen Chemie-Gruppe. FSTA (Food Science and Technology Abstracts) werden von DIALOG und STN angeboten, jedoch wiederum nicht in deren Chemie-Gruppen.

Die Datenbank IPA (International Pharmaceutical Abstracts) liegt zwar bei DIALOG auf, jedoch nicht in dessen Chemie-Gruppe. NTIS (National Technical Information Service) wird vom Host DATASTAR nicht zur Chemie-Gruppe gezählt. Für die Datenbanken PIRA (Pira International Publishing and Information Services) und PLASPEC (Plastics Technology Material Selection Database) gilt, daß beide zwar im Angebot bei STN sind, jedoch nicht in dessen Chemie-Gruppe. Schließlich gilt für TOXLINE, daß weder DIALOG noch STN diese große Datenbank für chemierelevant halten.

Diese Ausführungen lassen den Schluß zu, daß es keine Übereinkünfte über inhaltliche Einordnungen über Datenbanken gibt, und daß der Nutzer demgemäß nicht alle chemierelevanten Informationen erhält, wenn er nur in der Gruppe eines Anbieters recherchiert.

Es sei an dieser Stelle außerdem darauf hingewiesen, daß die Datenbankanbieter ihre Systeme und damit auch ihre Gruppen laufend überarbeiten, so daß sich die Situation ständig verändern kann.

Zusammenfassend sei hier festgestellt, daß nicht ausgeschlossen werden kann, daß es Datenbanken gibt, die chemierelevant sind, aber nicht in der Aufstellung in Tabelle 8 auftreten. Trotz dieser bewußt in Kauf genommenen Einschränkung, enthält die Vereinigungsmenge, gebildet aus den Datenbanken der Gruppen von DATASTAR, DIALOG und STN einen repräsentativen Querschnitt an chemierelevanten Online Datenbanken.

### 5.1.3 Umwelt-Gruppe

#### *5.1.3.1 Übersicht*

Für das Gebiet der Umweltdatenbanken werden von allen ausgewählten Hosts Gruppen angeboten. Bei *DATASTAR* gibt es die Gruppe *Energy & Environment* (CROS Group 751, 752) mit acht Datenbanken [DATASTAR 1995]. Der Datenbankanbieter *DIALOG* stellt die Grup-

pen *Environment* (ENVIRON) mit 27 Datenbanken und *POLLUTION* (POLLUT) mit 22 Datenbanken zur Verfügung [Knight-Ridder Information 1995a]. Beide Gruppen weisen viele gemeinsame Datenbanken auf. Der deutsche Host *DIMDI* nennt seine 22 Datenbanken umfassenden Gruppe Superbase *Umweltschutz*, *Umweltmedizin* (XENVIR) [DIMDI 1995]. *STN* stellt mit seiner Gruppe Environment (ENVIRONMENT) 45 Datenbanken zur Verfügung [FIZ Karlsruhe 1995]. Eine neuere Abhandlung über das Angebot an Umweltdatenbanken findet man in Koch et al. [Koch 1995].

### 5.1.3.2 Einengende Auswahl

Analog zur Vorgehensweise bei den Chemiedatenbanken werden auch bei den Umweltdatenbanken einige nicht in den engere Auswahl genommen, nämlich beispielsweise Patentdatenbanken. Das Resultat der umweltrelevanten Datenbanken ist in Tabelle 9 dargelegt. Hier wird neben den Host DATASTAR, DIALOG und STN auch der Datenbankanbieter DIMDI berücksichtigt.

**Tabelle 9: Umwelt-Gruppen bei den Hosts DataStar, DIALOG, DIMDI und STN**  
(fettgedruckte Datenbanken sind auch in den Chemie-Gruppen vorhanden)

Nr.	Name	DATA-STAR	DIALOG	DIMDI	STN	Typ
01	ABI-INFORM				X	VT
02	Academic Index		X			BI
03	AGRICOLA			X		BI
04	AGRIS			X		BI
<b>05</b>	<b>ANABST</b>		X			<b>BI</b>
<b>06</b>	<b>APILIT</b>				X	<b>BI</b>
07	ASFA		X	X	X	BI
08	ASI		X			BI
09	BALTIC			X		BI
10	BIOBUSINESS				X	BI
11	BIOLIS			X		BI
12	BIOSIS		X	X	X	BI
<b>13</b>	<b>CA</b>		X		X	<b>BI</b>
<b>14</b>	<b>CAB</b>		X	X	X	<b>BI</b>
15	CANCERLIT				X	BI
<b>16</b>	<b>CAPLUS</b>				X	<b>BI</b>
<b>17</b>	<b>CAPREVIEWS</b>				X	<b>BI</b>
<b>18</b>	<b>CBNB</b>				X	<b>BI</b>
<b>19</b>	<b>CEN</b>				X	<b>BI</b>
<b>20</b>	<b>CIN</b>				X	<b>BI</b>
<b>21</b>	<b>CJACS</b>				X	<b>VT</b>
<b>22</b>	<b>COMPENDEX</b>		X		X	<b>BI</b>
23	CRIS/USDA		X			FD
<b>24</b>	<b>CSNB</b>				X	<b>BI</b>
25	ELFIS			X		BI
26	EMBASE		X	X	X	BI
27	ENERGIE				X	BI
28	ENERGY		X		X	BI
29	ENVIROLINE	X	X	X		BI
30	Environmental Bibliography		X			BI
31	Federal Register		X			BI
32	Federal News Service		X			VT
33	FT Reports E. & Environ.	X				VT
34	GeoArchive		X			BI
35	GEOBASE		X			BI
36	GEOREF		X		X	BI
37	HEALSAFE				X	BI
<b>38</b>	<b>HSDB</b>				X	<b>NU</b>
<b>39</b>	<b>IPA</b>			X		<b>BI</b>

40	ISTP			X		BI
41	JIST-E				X	BI
42	LIFESCI				X	BI
43	MEDLINE			X	X	BI
44	NEI				X	BI
45	NLDB		X		X	VT
46	NTIS		X		X	BI
47	OCEAN		X		X	BI
48	OS&H		X			BI
49	PAIS		X			BI
50	PAPERCHEM				X	BI
51	PIRA				X	BI
52	POLLUTION	X	X		X	BI
53	PROMT				X	BI
54	PSTA			X	X	BI
55	SAE Automotive Standards	X				BI
56	SAE Global Mobility	X				BI
57	SCISEARCH		X	X	X	BI
58	SOMED			X		BI
59	TOXBIO			X		BI
60	TOXCAS			X		BI
61	TOXLINE		X	X	X	BI
62	TOXLIT				X	BI
63	UFOR	X			X	FD
64	ULIT	X			X	BI
65	UR-DB	X				BI
66	WATERNET		X			BI
67	WRA		X			BI

Abkürzungen für den Datenbanktyp siehe Abkürzungsverzeichnis

Besonders hervorzuheben ist, daß es keine einzige Datenbank gibt, die von allen vier Hosts gleichzeitig als umweltrelevant angesehen wird. Zu drei von vier Hosts zugehörig sind die folgenden sieben Online Datenbanken: ASFA (Aquatic Science and Fisheries Abstracts), BIOSIS (BIOSIS Previews), EMBASE (Excerpta Medica Database), ENVIROLINE (Environment Abstracts Online), POLLUTION (Pollution Abstracts), SCISEARCH und TOXLINE (Toxicology Literature Online). Es ist auffällig, daß der Host DATASTAR einige - von anderen Hosts als umweltrelevant angesehene - Datenbanken in seinem Angebot hat, wie z.B. BIOSIS (BIOSIS Previews), CA (Chemical Abstracts), CANCERLIT, EMBASE (Excerpta Medica Database), ENERGY (Energy Science and Technology), MEDLINE (Medical Literature Analysis and Retrieval System Online), SCISEARCH und TOXLINE (Toxicology Literature Online), diese aber nicht in der Umwelt-Gruppe anbietet. Das gleiche gilt für DIALOG. Dort sind z.B. AGRICOLA, AGRIS, APILIT (American Petroleum Institute Literature Database), BIOBUSINESS, CANCERLIT, IPA (International Pharmaceutical Abstracts), LIFESCI (Life Sciences Collection), MEDLINE, PSTA (Packaging Science and Technology) zwar aufgelegt, jedoch nicht in der Gruppe Umwelt. DIMDI bietet CANCERLIT an, jedoch auch nicht in der Umwelt-Gruppe. Bei STN gilt das für IPA. ANABST (Analytical Abstracts) ist nur bei DIALOG in der POLLUTION Gruppe, ansonsten bei keinem anderen Host, obwohl die Datenbank dort überall aufliegt.

#### **5.1.4 Diskussion und Gegenüberstellung der Chemie- und Umwelt-Gruppe**

Bei Betrachtung der Chemie- und Umwelt-Gruppe in den Tabellen 8 und 9 wird deutlich, daß die textbasierten Datenbanken deutlich die faktenbasierten Datenbanken überwiegen (siehe hierzu auch Kapitel 4.3). Die genaue Auswertung der Gruppen in bezug auf die Typen der

Datenbanken wird in Tabelle 10 verdeutlicht. Die Chemie-Gruppe verfügt über 65 Datenbanken, die Umwelt-Gruppe über 67.

**Tabelle 10: Datenbanktypen in der Chemie- und Umwelt-Gruppe**

Typ	Chemie-Gruppe	65 Datenbanken	Umwelt-Gruppe	67 Datenbanken
	Anzahl der DB	%	Anzahl der DB	%
BI	34	52	59	87
NU	20	30	1	2
VT	9	14	5	7
CN	1	2	-	-
VZ	1	2	-	-
FD	-	-	2	4

Abkürzungen für den Datenbanktyp siehe Abkürzungsverzeichnis

Bei der Chemie-Gruppe ist ersichtlich, daß weder Strukturdatenbanken (ST) noch Reaktionsdatenbanken (RE) von den Hosts zur dieser Gruppe gerechnet werden. Diese Tatsache kann mit der unterschiedlichen Struktur dieser chemierelevanten Datenbanken erklärt werden (vgl. hierzu Kapitel 2. 5 und Anhang A). Obwohl bei den ausgewählten Chemie-Gruppen auch die textbasierten Datenbanken, hier vor allem die bibliographische Datenbanken (BI) deutlich überwiegen, gibt es doch ca. ein Drittel der ausgewerteten Datenbanken, die der Gruppe der faktenbasierten Datenbanken, hier vor allem Faktendatenbanken (numerische Datenbanken) (NU) zuzurechnen sind. Weder die Chemie-Gruppe noch die Umwelt-Gruppe verfügen über integrierte Datenbanken (siehe auch Kapitel 4.1.4).

## **5.1.5 Auswahl der Online Datenbankgruppen für das Bewertungszifferverfahren**

### ***5.1.5.1 Allgemeine Grundsätze zur Auswahl***

Die Hasse-Diagramm-Technik beruht in erster Linie auf der Visualisierung der vergleichenden Beurteilung von Objekten, hier Datenbanken. Aus diesem Grund sind unübersichtliche Diagramme zu vermeiden, die erfahrungsgemäß dann entstehen, wenn eine zu große Anzahl an Datenbanken vorliegt. Techniken zur Verbesserung der Übersichtlichkeit von Hasse-Diagrammen werden entwickelt [Brüggemann 1996b], sind jedoch noch nicht abgeschlossen. Aus dem Gesagten folgt, daß eine weitere Auswahl aus den in den oben aufgeführten Datenbanken der Chemie- und Umwelt-Gruppe gefunden werden muß. Hierfür werden Auswahlprinzipien zur Reduktion des Umfangs angewandt, welche weiter unten erläutert werden. Anstelle der Betrachtung der "Gesamtheit der Gruppendatenbanken" in einer Objektmenge werden im Folgenden fünf thematisch geordnete Datenbankgruppen gebildet.

### ***5.1.5.2 Bildung der Datenbankgruppen 1-5***

Es werden nun im ersten Auswahlverfahren aus den Chemie- bzw. Umwelt-Gruppen der einzelnen Hosts folgende fünf Gruppen an Datenbanken herausgearbeitet:

- 34 bibliographische Datenbanken der Chemie-Gruppe (*Datenbankgruppe 1*)
- 18 Faktendatenbanken (numerische Datenbanken) der Chemie-Gruppe (*Datenbankgruppe 2*)
- 19 textbasierte Datenbanken der Umwelt-Gruppe (*Datenbankgruppe 3*)
- 14 umweltrelevante Faktendatenbanken (*Datenbankgruppe 4*)
- 21 Umweltchemikalien-relevante Datenbanken der Chemie-/Umwelt-Gruppe (*Datenbankgruppe 5*)

Die einzelnen Gruppen, in der Hasse-Diagramm-Technik auch *Objektmengen (OM)* genannt, sind in Kapitel 7. 2 aufgelistet zusammen mit den Bewertungsziffern (BWZ) für die einzelnen Bewertungskriterien (BWK). Die Gruppe 4 kann nicht aus den Umwelt-Gruppen bestimmt werden, da einerseits die Hosts in ihren Umwelt-Gruppen mit Ausnahme der Datenbank Hazardous Substances Databank HSDB keine Faktendatenbanken (NU) berücksichtigen, andererseits jedoch diese Art von Datenbanken im Umweltbereich häufig erforderlich ist, wird für die Auswahl der Faktendatenbanken eine von den Gruppen unabhängige Vorgehensweise eingeschlagen. Es werden die in den bereits erwähnten Publikationen [Cox 1991], [Orton 1995] und [Gayle Alston 1991] aufgeführten numerischen Umweltdatenbanken erfaßt. Hierbei werden nicht nur die von den oben besprochenen Hosts sondern auch die Datenbanken, die von dem amerikanischen Host CIS angeboten werden, berücksichtigt. Die geringfügige Überschneidung der Datenbankgruppen 2 und 4 (gemeinsam HSDB) wird in Kauf genommen.

Aus Gründen der Übersichtlichkeit werden die Datenbankgruppen noch einmal tabellarisch dargestellt.

**Tabelle 11: Zusammenstellung der fünf Online Datenbankgruppen**

Nr. der Gruppe	Auswahl aus:	Typ	einengendes Kriterium	Bemerkung
1ON	Chemie-Gruppen, Tabelle 8	BI	Vorgabe des Typs	-
2ON	Chemie-Gruppen, Tabelle 8	NU	Vorgabe des Typs	nur verfügbare Datenbanken
3ON	Umwelt-Gruppen, Tabelle 9	BI, VT	Vorgabe des Typs	bei mindestens zwei Hosts eingestuft
4ON	Umweltpublikationen	NU	Vorgabe des Typs	-
5ON	Chemie-/Umwelt-Gruppen Tabelle 8/9	BI, NU	Vorkommen in beiden Gruppen	-

1ON = 1. Online Datenbankgruppe etc.,

Abkürzungen für den Datenbanktyp siehe Abkürzungsverzeichnis

## 5.2 Auswahl der Gruppen an CD-ROMs

### 5.2.1 Vorbemerkung

Wie bei allen Bewertungsverfahren muß zunächst eine repräsentative Auswahl geeigneter Objekte getroffen werden, d.h. es müssen Gruppen definiert werden. Anders ausgedrückt: Welche CD-ROMs sind für die Beantwortung von Fragestellungen in der Chemie und im Umweltschutz geeignet? In diesem Zusammenhang spielt auch der Typ der CD-ROM eine Rolle.

Wie bereits in Kapitel 2.6 erwähnt, ist die Anzahl an CD-ROMs in den vergangenen Jahren exponentiell angestiegen. Beispielsweise umfaßt das Standardwerk "CD-ROM Directory" in

seiner 1996er Ausgabe 13.000 CD-ROM Titel [TFPL 1996]. Im Jahr 1992 umfaßte dieses Verzeichnis 2.200, 1993 3.600, 1994 6.000 und 1995 8.000 Titel.

Auf die Auswahl geeigneter Datenbanken für spezielle Themenbereiche gehen einige Hersteller und Vertreiber von CD-ROMs ein. Jedoch ist der Markt an Vertreibern von CD-ROMs wesentlich komplexer und unübersichtlicher als der der Online Hosts. In dem oben erwähnten CD-ROM Directory sind 4.000 Hersteller von CD-ROMs aufgelistet, die 13.000 Titel herstellen [TFPL 1996]. Es werden daher nur einige auf dem Gebiet der Naturwissenschaft und Technik bedeutende CD-ROM Anbieter und Hersteller analysiert. Hier seien die Firmen SilverPlatter [1996], KRInformation OnDisc [1996], U.S. Department of Commerce [1995], Springer Verlag [1995], The Royal Society of Chemistry [1996], Bowker Saur [1995], Elsevier Science [1995], HMSO [1995], CAS [1996], Chapman & Hall [1995], Canadian Centre for Occupational Health and Safety [1995] und ecomed Verlagsgesellschaft [1996] genannt. Einige dieser CD-ROM Anbieter halten Unterteilungen in verschiedene Sachgebiete vor, andere listen ihr CD-ROM Angebot nur alphabetisch auf. SilverPlatter, einer der größten Hersteller von CD-ROMs, hat z.Zt. 220 CD-ROMs im Angebot. Diese werden thematisch eingeteilt. Auf Seiten der Chemie werden die CD-ROMs in "Chemikalien" und "Chemie" eingeteilt, auf Seiten des Umweltschutzes unterscheidet SilverPlatter "Ökologie", "Pollution" und "Umwelt" [1996].

### **5.2.2 Chemie- bzw. Chemikalien spezifische CD-ROMs**

SilverPlatter (SP) teilt 26 CD-ROM Titel in das Spezialgebiet Chemikalien ein und sechs CD-ROMs in das Spezialgebiet Chemie [1996]. Knight-Ridder (KR) unterteilt lediglich in Wissenschaft und Technik [1996]. Der Springer-Verlag (SV) hat verschiedene spezielle chemiespezifische CD-ROMs im Angebot [1995]. Die Royal Society of Chemistry (RSC) liefert die Daten für vier CD-ROMs, die bei SilverPlatter oder Knight-Ridder angeboten werden [1996]. Chapman & Hall (CH) bietet spezielle Chemical Dictionaries an, die ebenfalls ausgewertet werden. Es werden jedoch nur diejenigen CD-ROMs in Betracht gezogen, die organische Chemikalien beinhalten [1995]. Aus diesem Grunde werden das Dictionary of Inorganic Compounds on CD-ROM und das Dictionary of Organometallic Compounds on CD-ROM nicht analysiert. Ferner werden die Auszüge der Datenbank Chemical Abstracts (CA) hier berücksichtigt [1996]. Auch bietet das Canadian Centre for Occupational Health and Safety (CCO) eine Reihe von Chemikalien-relevanten CD-ROMs an [1995]. Darüber hinaus wird das CD-ROM Angebot der ecomed Verlagsgesellschaft (ECO) ausgewertet [1996]. Die Firma Beilstein Informationssysteme GmbH bietet die Current Facts in Chemistry (BEI), das aktuelle Jahr der Beilstein Datenbank, an [1996]. Das Römpp Chemie Lexikon wird vom Georg Thieme Verlag (TH) auf CD-ROM herausgegeben [1995].

Es werden zehn Hersteller bzw. Anbieter von CD-ROMs auf dem Chemie- bzw. Chemikalien-Sektor ausgewertet und in Tabelle 12 zusammengestellt. Neben der Nennung der Zugehörigkeit zu einem der Hersteller ist in der letzten Spalte der Typ der Datenbank aufgeführt. Im Gegensatz zu der Betrachtung der Gruppen bei den Online Datenbanken treten hier nur wenige Datenquellen auf, die von mehreren Anbietern vertrieben werden. Der Grund hierfür liegt in der Tatsache, daß viele Fachverlage gleichzeitig Hersteller und Anbieter ihrer CD-ROM Produkte sind. Von den 45 auszuwertenden CD-ROMs sind 21 bibliographische Datenbanken, fünf Volltextdatenbanken, 17 Faktendatenbanken (numerische Datenbanken) und zwei Reaktionsdatenbanken.

**Tabelle 12: Spezialgebiet Chemie/Chemikalien bei verschiedenen CD-ROM Anbietern**

Nr.	CD-ROM	SP	KR	SV	RSC	C&H	CA	CCO	ECO	BEI	TH	Typ
01	Analytical Abstracts	X			X							BI
02	ASTI	X										BI
03	Biological Abstracts	X										BI
04	CANDATA							X				NU
05	CASurveyor Cancer						X					BI
06	CASurveyor Food						X					BI
07	CASurveyor Haz. Mat.						X					BI
08	CBNB		X		X							BI
09	CEABA		X		X							BI
10	CHEM-BANK	X										NU
11	ChemReact 10			X								RE
12	ChemReact 32			X								RE
13	CHEM Source							X				NU
14	Current Facts in Chem.											NU
15	CD-ROMPP										X	VT
16	Dict. of Drugs					X						NU
17	Dict. of Natural Pro.					X						NU
18	Dict. of Org. Comp.					X						NU
19	Dissertation Abstracts	X										BI
20	Ei CHEM Disc		X									BI
21	Ei Compendex Plus	X	X									BI
22	Ei Tech Index	X										BI
23	EINECS-Plus	X										NU
24	EMBASE-CD:PollTox	X										BI
25	Environ. Chemistry		X		X							BI
26	Gefährgut CD-ROM			X								NU
27	General Science Abstr.	X										BI
28	Giftliste								X			NU
29	IARCancer Disk	X										BI
30	Kirk-Othmer		X									VT
31	MSDS	X						X				NU
32	NIOSH/TIC/DIDS							X				NU
33	NTIS		X									BI
34	OSH-CD	X										VT
35	OSH/OFFSHORE	X										VT
36	OSH-ROM	X										VT
37	OHS INTERDATA							X				BI
38	PEST-BANK	X										NU
39	PolToxI:NLM,CSA,IFI	X										BI
40	RTECS							X				NU
41	Sax, Hawley					X						NU
42	SIGEDA								X			NU
43	SORBE								X			NU
44	TOXLINE on SP	X										BI
45	TOXLINE Plus	X										BI

Die Abkürzungen für den Datenbanktyp sind im Abkürzungsverzeichnis, die Abkürzungen für die Hersteller im Text verzeichnet.

### 5.2.3 Umwelt- bzw. Ökologie-spezifische CD-ROMs

Für das Gebiet der Umweltdatenbanken werden folgende Hersteller bzw. Vertreiber von CD-ROMs ausgewertet: SilverPlatter (SP) teilt seine CD-ROM Titel in die Spezialgebiete Umwelt, Pollution und Ökologie ein [1996]. Knight-Ridder (KR) unterteilt lediglich in Wissenschaft und Technik [1996]. Es werden die umweltrelevanten CD-ROMs gemäß der inhaltlichen Beschreibung ausgewählt. Der Springer-Verlag (SV) hat die Gefährgut - Dangerous Goods CD-ROM im Angebot [1995]. Die Royal Society of Chemistry (RSC) liefert die Daten für die Environ-

mental Chemistry, Health and Safety CD-ROM, die von Knight-Ridder (KR) angeboten wird [1996]. Ferner werden die Auszüge der Datenbank Chemical Abstracts (CA) hier berücksichtigt [1996]. Das Canadian Centre for Occupational Health and Safety (CCO) stuft seine CD-ROMs sowohl als Chemikalien- als auch als umweltrelevant ein [1995]. Darüber hinaus wird das CD-ROM Angebot der ecomed Verlagsgesellschaft (ECO) ausgewertet [1996]. Ebenso werden die umweltrelevanten Datenbanken der U.S. Environmental Protection Agency (EPA) aufgenommen [Department of Commerce 1995].

**Tabelle 13: Spezialgebiet Umwelt/Pollution/Ökologie bei verschiedenen CD-ROM Anbietern**

Nr.	CD-ROM	SP	KR	SV	RSC	CA	CCO	ECO	EPA	Typ
01	AGREP	X								FD
02	AGRICOLA on SP	X								BI
03	AGRIS	X								BI
04	AGRISearch	X								BI
05	ARRIP	X								FD
06	ASFA	X								BI
07	ASTI	X								BI
08	Biolo. & Agri.Index	X								BI
09	Biological Abstracts	X								BI
10	CABCD	X								BI
11	CABPESTCD	X								BI
12	CANDATA						X			NU
13	CASurveyor Haz. Mat.					X				BI
14	CASurveyor Pollution					X				BI
15	CASurveyor Waste					X				BI
16	CERCLA & RCRA								X	NU
17	CHEM-BANK	X								NU
18	CHEM Source						X			NU
19	CRIS	X								FD
20	E-CD	X								BI
21	Earth Science Disc	X								BI
22	ECODISC	X								BI
23	Ei Compendex Plus	X	X							BI
24	Ei Energy & Environ.		X							BI
25	EINECS plus-CD	X								NU
26	EMBASE CD: PolTox	X								BI
27	Environ. Chemistry		X		X					BI
28	Environmental Impact	X								BI
29	Environmental Manag.		X							BI
30	ETDE Energy Db	X								BI
31	Gefahrgut- CD-ROM			X						NU
32	General Science Ab.	X								BI
33	GEOBASE	X								BI
34	GeoRef	X								BI
35	Giftliste							X		NU
36	IRIS								X	NU
37	LABORDOC	X								BI
38	LifeScience Collection	X								BI
39	MSDS	X						X		NU
40	NIOSH TIC/DIDS						X			NU
41	OCLC Environment	X								BI
42	OSH/OFFSHORE	X								VT
43	OSH-ROM	X								VT
44	OHS INTERDATA						X			BI
45	PEST-BANK	X								NU
46	PolTox I: NLM, CSA, IFI	X								BI
47	PolTox III: CAB	X								BI
48	RDIR	X								FD
49	RTECS						X			NU

50	SIGEDA							X		NU
51	SOILCD	X								BI
52	SORBE							X		NU
53	TOXLINE on SP	X								BI
54	TOXLINE Plus	X								BI
55	TRANSPORT	X								BI
56	TREECD	X								BI
57	TRI								X	NU
58	TROPAC & RURAL	X								BI
59	TSCA								X	NU
60	WasteInfo	X								BI
61	Water Resources Abst.	X								BI
62	WATERLIT	X								BI

Abkürzungen für den Datenbanktyp siehe Abkürzungsverzeichnis, Abkürzungen Hersteller siehe Text

Es werden acht Hersteller bzw. Anbieter von CD-ROMs auf dem Umwelt- bzw. Ökologie-Sektor ausgewertet. Neben der Nennung der Zugehörigkeit zu einem der Hersteller ist in der letzten Spalte der Typ der Datenbank aufgeführt. Auch hier treten aus den oben genannten Gründen nur wenige CD-ROMs auf, die von mehreren Firmen angeboten werden. Von den 62 auszuwertenden CD-ROMs sind 40 bibliographische Datenbanken, 2 Volltextdatenbanken, 16 Faktendatenbanken (numerische Datenbanken) und 4 Forschungsdatenbanken.

#### **5.2.4 Diskussion und Gegenüberstellung der Chemie- und Umwelt-Spezialgebiete**

Bei Betrachtung der Chemie- und Umwelt-Spezialgebiete in den Tabellen 12 und 13 wird deutlich, daß die textbasierten Datenbanken deutlich die faktenbasierten Datenbanken überwiegen (siehe hierzu auch Kapitel 4.3). Die genaue Auswertung der beiden Spezialgebiete in bezug auf die Typen der Datenbanken wird in Tabelle 14 verdeutlicht.

**Tabelle 14: Typen von CD-ROMs im Chemie- und Umwelt-Spezialgebiet**

Typ	Chemie-Spezialgebiet	45 Datenbanken	Umwelt-Spezialgebiet	62 Datenbanken
	Anzahl der DB	%	Anzahl der DBs	%
BI	21	47	40	65
NU	17	38	16	26
VT	5	11	2	3
RE	2	4	-	-
FD	-	-	4	6

Abkürzungen für den Datenbanktyp siehe Abkürzungsverzeichnis, Erläuterungen siehe Text

Im Chemie-Spezialgebiet wurden 45 und im Umwelt-Spezialgebiet 65 CD-ROMs berücksichtigt.

Beim Chemie-Spezialgebiet wird deutlich, daß hier - im Gegensatz zu der Auswahl an Online Datenbanken - Reaktionsdatenbanken (RE) zum Spezialgebiet Chemie gerechnet werden. Reaktionsdatenbanken auf CD-ROM verfügen über eine spezielle für Strukturen geeignete und eigens vom jeweiligen Hersteller entwickelte Retrievalsoftware. Sowohl bei den ausgewählten Chemie- bzw. Chemikalien-relevanten CD-ROMs als auch bei den auf die Thematik Umweltschutz spezialisierten CD-ROMs überwiegen die textbasierten Datenbanken, hier vor allem die bibliographischen Datenbanken (BI) deutlich. Besonders auffällig wird dieser Umstand bei den umweltrelevanten Datenbanken. Hier sind 65 % der ausgewählten CD-ROMs textbasiert. Bei

den chemierelevanten sind es "nur" 47 %. Bei den chemierelevanten Datenbanken sind über ein Drittel der ausgewerteten Datenbanken faktenbasierte Datenbanken, bei den Umweltdatenbanken sind es hingegen nur 30 %. Nur bei den chemiespezifischen CD-ROMs gibt es - wie schon erwähnt - Reaktionsdatenbanken, die als integrierte Datenbanken gelten (siehe Kapitel 2.5).

## **5.2.5 Auswahl der Gruppen an CD-ROMs für das Bewertungszifferverfahren**

### ***5.2.5.1 Allgemeine Grundsätze zur Auswahl***

Hier gelten wieder die unter 5.1.5.1 dargelegten Grundsätze, nach denen es zweckmäßig ist, eine Auswahl aus den in den oben aufgeführten CD-ROMs der beiden Spezialgebiete Chemie und Umwelt zu treffen. Hierfür werden Auswahlprinzipien zur Reduktion des Umfangs angewandt, welche weiter unten erläutert werden. Anstelle der Betrachtung der "Gesamtheit der CD-ROM Spezialgebiet Datenbanken" in einer Gruppe werden im folgenden fünf thematisch geordnete Gruppen (auch Objektmengen genannt) gebildet.

### ***5.2.5.2 Bildung der CD-ROM Gruppen 1-5***

Analog der in Kapitel 5.1.5.2 erläuterten Vorgehensweise für Online Datenbanken sind hier nun im ersten Auswahlverfahren folgende fünf Gruppen an CD-ROMs herausgearbeitet worden:

- 26 textbasierte CD-ROMs des Chemie-Spezialgebietes (**Gruppe 1**)
- 19 faktenbasierte und integrierte CD-ROMs des Chemie-Spezialgebietes (**Gruppe 2**)
- 42 textbasierte CD-ROMs des Umwelt-Spezialgebietes (**Gruppe 3**)
- 20 faktenbasierte CD-ROMs des Umwelt-Spezialgebietes (**Gruppe 4**)
- 25 Umweltchemikalien-relevante CD-ROMs des Chemie-/Umwelt-Spezialgebietes (**Gruppe 5**)

Die einzelnen Gruppen werden mit ihren dazugehörigen Bewertungsziffern im Kapitel 7.3 aufgeführt. Aus Gründen der Übersichtlichkeit werden die Gruppen noch einmal tabellarisch dargestellt.

**Tabelle 15: Zusammenstellung der fünf CD-ROM Gruppen**

Nr. der CD-ROM Gruppe	Auswahl aus:	Typ	einengendes Kriterium	Anzahl der CD-ROMs
1CD	Chemie-Gruppe Tabelle 12	BI	Vorgabe des Typs	26
2CD	Chemie-Gruppe Tabelle 12	NU, IN	Vorgabe des Typs	19
3CD	Umwelt-Gruppe Tabelle 13	BI, VT	Vorgabe des Typs	42
4CD	Umwelt-Gruppe Tabelle 13	NU, FD	Vorgabe des Typs	20
5CD	Chemie-Gruppe Umwelt-Gruppe Tabellen 12/13	BI, NU	Vorkommen in beiden Gruppen	25

1CD = 1. CD-ROM Gruppe, Abkürzungen für den Datenbanktyp siehe Abkürzungsverzeichnis

## 6. Bewertungsverfahren für Datenquellen

### 6.1 Notwendigkeit und Bedeutung der Bewertung

Die Notwendigkeit der Bewertung von Datenbanken ist nicht nur durch die in Kapitel 2 ausgeführte Vielzahl der Datenbanken, Datenbankhersteller, der Unterschiede im Typ der Datenbanken, sondern auch durch die Art der Zugriffsmöglichkeit gegeben. Die Notwendigkeit der Bewertung wird in einem Buch über die Nutzer-orientierte Evaluation von Informationssystemen beleuchtet [Bawden 1990], worin der Begriff "Nutzer-orientierte Bewertung" (UOE = User-oriented Evaluation) eingeführt wird. Bawden stellt die Nutzerfreundlichkeit von Informationssystemen vor. Er leitet diese aus qualitativen allgemeinen Datenbank Bewertungskriterien aufbauend auf Fallstudien ab.

Neuere Ansätze beschäftigen sich mit der Bewertung von Datenbanken nach allgemeinen Kriterien der Informationswissenschaft. An dieser Stelle sei auf einen Ansatz für die Bewertung von Datenbanken auf CD-ROMs nach den Kriterien der Konsistenz, der effizienten Ausnutzung der Recherchezeit, der Einfachheit der CD-ROM, der Objektivität der Beschreibung der CD-ROM und der Flexibilität der Abfrage verwiesen [Harry 1993a,b].

Um eine Bewertung vornehmen zu können, ist die Klassifikation der Objekte (hier: Datenquellen) ein wichtiger vorbereitender Schritt. Die Theorie und Anwendung der nichtmetrischen Klassifikation von Objekten wird von Belke et al. ausgeführt [Belke 1979]. Belke beschreibt die nichtmetrische Klassifikation als eine spezifische Form der Objektbeschreibung und Modellierung. "Objekte werden über ihre Eigenschaften bzw. Merkmale erfaßt, bewertet und beschrieben". Anstelle des Begriffes Merkmal oder Eigenschaft wird im folgenden die Bezeichnung Kriterium bzw. Bewertungskriterium gewählt. Die Ausnahme bilden die Zitate von Behlke.

Es sei hier weiter Belke zitiert: "Durch die Klassifikation werden Objekte aufgrund ihrer Merkmale von anderen Objekten abgegrenzt und Klassen von Objekten zugeordnet. Die im allgemeinen notwendigerweise zu berücksichtigende Vielzahl von Merkmalen und die unterschiedliche Werteart (Modus) dieser Merkmale - ausgedrückt durch die Zulassung nominaler, ordinaler und kardinaler Wertgrößen der Merkmale - führen zu einer insgesamt betrachtet nichtmetrischen Objektbeschreibung, also zur "**nichtmetrischen Klassifikation**" [Belke 1979, Vorwort]. Nach Belke kennzeichnet ein **ordinales Merkmal** Objekte nach dem Ausmaß des Besitzes eines Merkmalwertes gegenüber einem anderen Objekt. Es wird auch als Rangordnungsmerkmal bezeichnet. Über einzelne Objekte ist damit weniger als über das Verhältnis zwischen den Objekten ausgesagt. "Ein **nominales Merkmal** kennzeichnet Objekte nach dem Besitz von Merkmalen/Merkmalwerten, die entweder vorhanden oder nicht vorhanden sind". "Ein **kardinales Merkmal** kennzeichnet Objekte hinsichtlich der Ausprägung eines Merkmals absolut, nicht nur im Unterschied zu anderen Objekten" [Belke 1979, S. 37-38].

Eine Einteilung nach qualitativen (mit oder ohne Ordnungsstrukturen, zwei- oder mehrstufig) und nach quantitativen Merkmalen nimmt H. Bock [Bock 1974] vor.

### 6.2 Bewertungsziffern

Um die Objekte, in diesem Fall Datenquellen zu bewerten, muß die Definition von charakterisierenden Zahlen so gewählt werden, daß eine gemeinsame Qualität festgestellt werden kann. In diesem Fall bedeutet eine hohe Ziffer in bezug auf ein Kriterium eine sehr gute und eine kleine Ziffer eine schlechte Datenquelle (siehe Tabelle 16).

Das Bewertungsverfahren (Bewertungszifferverfahren) wird nach einem sechs Zahlen umfassenden Bewertungsziffersystem durchgeführt [Voigt 1994, 1995b], d.h. eine Ausprägungsmenge für die Bewertungsziffern mit {0,1,2,3,4,5} angenommen.

**Tabelle 16: Bewertungsziffern und deren Bedeutung**

<b>5</b>	=	<b>sehr gut</b>
<b>4</b>	=	<b>gut</b>
<b>3</b>	=	<b>befriedigend</b>
<b>2</b>	=	<b>ausreichend</b>
<b>1</b>	=	<b>mangelhaft</b>
<b>0</b>	=	<b>ungenügend</b>

### 6.3 Einsatz ordinaler und nominaler Kriterien bei der Bewertung von Datenquellen

Für die Zuordnung und Vergabe der Bewertungsziffern zu den Bewertungskriterien werden die Eintragungen aus der Metadatenbank der Online Datenbanken und der Metadatenbank der CD-ROMs herangezogen. Die Einteilung der Klassen von Bewertungsziffern, die Klassengrenzen und die damit im Zusammenhang stehende Vergabe der Bewertungsziffern werden aus den Eintragungen in den entsprechenden Datenfeldern der Metadatenbanken vorgenommen. Hierbei werden in der Regel die Eigenschaften, die am seltensten auftreten (z.B. Vorhaltung von Strukturformeln) hoch bewertet, diejenigen die häufiger vorkommen mit niedrigeren Bewertungsziffern (z.B. Chemikaliennamen) versehen. Wie bei anderen Bewertungsverfahren spielt auch hier die Erfahrung im Umgang mit den Datenbanken, d.h. die Expertenmeinung eine wichtige Rolle [GDCh 1986].

Die folgenden Kapitel werden erkennen lassen, daß eine Vielfalt verschiedener Kriterien auftritt, die - nach Belke [Belke 1979] - eine nichtmetrische Klassifizierung induzieren. Neben dem Auftreten verschiedener Modi ist zu berücksichtigen, daß die Zahl der Kriterien für die vorgesehene Multikriterienanalyse zu groß wird. Daher werden Kriterien nach bestimmten Inhalten aggregiert und zwar so, daß das System der resultierenden Kriterien durchwegs optimal interpretierbar wird. Grundsätzlich erlaubt aber die Hasse-Diagramm-Technik auch die Bewertung nach einer nichtmetrischen Klassifikation, die aber in dieser Arbeit nicht weiter verfolgt wird.

Um die Aggregation von solchen nominalen Kriterien durchzuführen, die inhaltlich zusammengehören, wird folgende pragmatische Vorgehensweise gewählt:

- |   |
|---|
| <ul style="list-style-type: none"> <li>a) Auflistung aller Kombinationen, die sich durch Zutreffen oder Nicht-Zutreffen bestimmter Teilkriterien unterscheiden</li> <li>b) Aussonderung solcher Kombinationen, die sich in der Praxis der Datenbank Beurteilung als unrealistisch erweisen (falls notwendig)</li> <li>c) Vergabe einer Reihenfolge der Qualität der Teilkriterien</li> <li>d) Vergabe von Bewertungsziffern für die Teilkriterien gemäß der in c) erarbeiteten Reihenfolge</li> <li>e) Ermittlung eines Kombinationswertes für jede Kombination</li> <li>f) Zuordnung zur Ausprägungsmenge</li> </ul> |
|---|

## 6.4 Bewertungskriterien für Online Datenbanken und CD-ROMs

Die Bewertungskriterien für Datenquellen lassen sich in allgemeine und inhaltliche Bewertungskriterien einteilen. Die meisten Kriterien sind sowohl auf Online Datenbanken als auch auf CD-ROMs anwendbar. Diejenigen, die unterschiedliche Einteilungen erfordern, werden gesondert betrachtet. Wichtig ist in diesem Zusammenhang, daß die CD-ROMs nicht so umfassend wie Online Datenbanken getestet wurden, da nur die wenigsten CD-ROMs gekauft aus finanziellen Gründen gekauft werden konnten. Die Beurteilung erfolgte daher nur aufgrund von Testversionen und Informationsmaterial. Die Bewertungskriterien, die den Chemikalien Testdatensatz umfassen, werden bei den CD-ROMs nicht angewandt.

### 6.4.1 Allgemeine Bewertung der Datenquellen

Sowohl bei der Betrachtung der Online Datenbanken als auch CD-ROMs spielen der Umfang der Datenquelle (UM), die Häufigkeit der Aktualisierung (UP), die Kosten der Recherche in der Online Datenbank bzw. die Kosten der CD-ROM (KO) und die Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien (VE) eine Rolle. Die einzelnen Kriterien werden im folgenden näher erläutert.

#### *6.4.1.1 Umfang der Datenbank (UM)*

Dieses Kriterium ist ein ordinales Merkmal (Rangordnungsmerkmal).

Bei der Betrachtung des Umfanges von Datenbanken wird zwischen bibliographischen Datenbanken auf der einen Seite und Volltextdatenbanken und Faktendatenbanken (numerischen Datenbanken) auf der anderen Seite unterschieden, da letztere durch ihre Struktur generell über weniger Dokumente verfügen als bibliographische Datenbanken. In der Metadatenbank der Online Datenbanken gibt es eine große Spannweite in bezug auf die Anzahl der Dokumente. Sie reichen von 90 bis 17 Millionen Dokumenten. Der größte Anteil liegt hingegen bei bibliographischen Datenbanken im Bereich von 100.000 bis 1.000.000 Dokumenten bzw. bei Faktendatenbanken oder Volltextdatenbanken zwischen 2.500 bis 10.000 Dokumenten. Die Einstufung der Klassengrenzen ist dementsprechend vor dem Hintergrund der Eintragungen in die Metadatenbanken entstanden.

Viele Datenbanken liegen heute schon vollständig auf CD-ROM vor und sind daher in bezug auf den Umfang der Dokumente den Online Datenbanken gleichzustellen.

Auch hier spielt das Expertenwissen wieder hinein. Der hohe Grad an Subjektivität könnte aber teilweise anhand von Sensitivitätsstudien abgemildert werden, indem die Klassengrenzen zur Vergabe der Bewertungsziffern variiert werden und ein erneutes Bewertungszifferverfahren durchgeführt wird.

**Tabelle 17: Bewertungsziffern für das Kriterium Umfang der Datenbank (UM)**

Bibliographisch	Numerisch/Volltext	Bewertungsziffer
>5.000.000	>100.000	5
>1.000.000	> 25.000	4
> 500.000	> 10.000	3
> 100.000	> 2.500	2
> 10.000	> 500	1
< 10.000	< 500	0

### 6.4.1.2 Häufigkeit der Aktualisierung (UP)

Auch bei diesem Kriterium handelt es sich um ein ordinales Merkmal. Die wenigsten wissenschaftlich-technischen Datenbanken werden täglich bzw. wöchentlich, die meisten monatlich bis vierteljährig aktualisiert. Es ist jedoch in den vergangenen zwei Jahren ein deutlicher Trend zu einer schnelleren Aktualisierung zu verzeichnen [Voigt 1995b]. Die Frequenz der Aktualisierungen ist bei Online Datenbanken generell häufiger als bei anderen Datenquellen z.B. bei CD-ROMs. Deshalb ist bei diesem Kriterium eine unterschiedliche Einteilung der Bewertungsziffern für diese beiden Medien vorzunehmen.

#### 6.4.1.2.1 Häufigkeit der Aktualisierung bei Online Datenbanken

Es lassen sich die Online Datenbanken in die in der folgenden Tabelle aufgeführten zeitliche Aktualisierungsabschnitte einteilen. In der zweiten Spalte der Tabelle 18 ist die Anzahl der Datenbanken, die zu den einzelnen Einteilungsstufen gehören, vermerkt. Anzumerken sei in diesem Zusammenhang, daß in der Metadatenbank nicht für alle dort eingetragenen Datenbanken ein Eintrag hierfür vorliegt. Der Grund für des Fehlen der Einträge ist, daß von den Hosts keine Angaben hierzu gemacht wurden. Zu den in den Gruppen vorkommenden Datenbanken liegen jedoch Daten vor.

**Tabelle 18: Bewertungsziffern für das Kriterium Häufigkeit der Aktualisierung (UP)**

Häufigkeit der Aktualisierung	Anzahl der Datenbanken in DADB	Bewertungsziffer
täglich, wöchentlich	49	5
zweimal pro Monat, monatlich	159	4
alle drei Monate	50	3
alle sechs Monate	33	2
jedes Jahr	29	1
unregelmäßig	42	0

Je öfter die Datenbank aktualisiert wird, desto besser wird sie eingestuft. Datenbanken, die einmal im Jahr neue Einträge bekommen, sind für den Nutzer von geringer Bedeutung, da die Daten nicht mehr aktuell sind. Eine wöchentliche bzw. tägliche Aktualisierung ist bei Online Datenbanken als sehr gut zu bewerten, eine monatliche bzw. zweiwöchige als gut. Bei unregelmäßigen Aktualisierungen wird vielfach gar nicht mehr überarbeitet. Aus diesem Grunde ist die Datenbank in bezug auf dieses Bewertungskriterium als ungenügend eingestuft.

#### 6.4.1.2.2 Häufigkeit der Aktualisierung bei CD-ROMs

Die Aktualisierung der CD-ROMs erfolgt in der Regel in größeren Abständen als bei den Online Datenbanken. Eine tägliche Aktualisierung kommt bei CD-ROMs nicht vor. Ein wöchentliche bzw. vierzehntägige Aktualisierung ist die Ausnahme und demnach als sehr gut einzustufen. Die meisten CD-ROMs werden heutzutage alle drei Monate aktualisiert. In der zweiten Spalte der Tabelle ist die Anzahl der CD-ROMs, die zu den einzelnen Einteilungsstufen gehören, vermerkt.

Es lassen sich die Datenbanken in die in der folgenden Tabelle aufgeführten zeitlichen Rahmen einteilen.

**Tabelle 19: Bewertungsziffern für das Kriterium Häufigkeit der Aktualisierung von CD-ROMs (UP)**

Häufigkeit der Aktualisierung	Anzahl der Datenbanken in DACD	Bewertungsziffer
wöchentlich, alle 14 Tage	16	5
monatlich	49	4
alle drei Monate	142	3
alle sechs Monate	42	2
jedes Jahr	50	1
unregelmäßig	3	0

### 6.4.1.3 Kosten (KO)

Die Kosten sind für die Medien Online Datenbanken und CD-ROM unterschiedlich. Das Bewertungskriterium ist in beiden Fällen jedoch ein Rangordnungsmerkmal (ordinales Merkmal).

#### 6.4.1.3.1 Kosten der Recherche

Die Kosten für eine Recherche setzen sich aus verschiedenen Einzelkosten zusammen. Zunächst wird in der Regel eine Gebühr für die Anschaltzeit pro Stunde bzw. pro Minute berechnet. Darüber hinaus muß der Nutzer die Dokumente, die angezeigt werden, je nach Umfang der Ausgabe bezahlen. Bei einigen Datenbanken schlägt auch bereits der einzelne eingegebene Suchbegriff zu Buche. Hier wird aus pragmatischen Gründen nur der Preis pro Stunde betrachtet, der bei fast allen Hosts als wichtiger Preisfaktor einer Recherche angesetzt wird. Seit langem wird jedoch in der Fachwelt die nicht mehr zeitgemäße Berechnungsgrundlage nach Anschaltzeit beklagt. Vorteilhafter und auch von wenigen Hosts unterstützt, wäre die Berechnung der gewonnenen Informationen. In der zweiten Spalte der Tabelle 20 ist die Anzahl der Datenbanken, die zu den einzelnen Einteilungsstufen gehören, aufgeführt. Anzumerken sei wieder, daß in der Metadatenbank nicht für alle ausgewerteten Datenbanken ein Eintrag hierfür vorliegt. Für die in Kapitel 4.1 herausgearbeiteten Gruppen an Online Datenbanken liegen jedoch Einträge vor. Je höher die Kosten, desto schlechter wird die Datenbank eingestuft in bezug auf dieses Bewertungskriterium. Obwohl die meisten der betrachteten Datenbanken in den drei oberen Preiskategorien liegen, d.h. zwischen 150 und über 200 DM/h Anschaltzeit, werden diese als schlecht eingestuft.

**Tabelle 20: Bewertungsziffern für das Kriterium Kosten der Online Recherche (KO)**

Preis pro Recherchestunde in DM	Anzahl der Datenbanken in DADB	Bewertungsziffer
≤ 50	44	5
≤ 80	24	4
≤ 100	35	3
≤ 150	131	2
≤ 200	92	1
> 201	95	0

#### 6.4.1.3.2 Kosten der CD-ROM

Die Kosten für eine CD-ROM setzen sich aus dem Anschaffungspreis und dem Preis für die Aktualisierungsversionen zusammen. Die Kosten einer CD-ROM sind also unabhängig von der Dauer der Nutzung und der Anzahl der Recherchen. In Analogie zur Betrachtung bei den Online Datenbanken wird hier auch nur ein Preis, d.h. der Anschaffungspreis in Betracht gezogen. In der zweiten Spalte der Tabelle ist die Anzahl der CD-ROMs aus der Metadatenbank der CD-ROMs, die zu den einzelnen Einteilungstufen gehören, aufgeführt. Je höher die Kosten, desto schlechter wird die CD-ROM eingestuft in bezug auf dieses Bewertungskriterium. Die meisten der betrachteten CD-ROM Produkte liegen in den Preiskategorien zwischen 2.000 und 5.000 DM.

**Tabelle 21: Bewertungsziffern für das Kriterium Kosten der CD-ROM (KO)**

Anschaffungspreis in DM	Anzahl der Datenbanken in DACD	Bewertungsziffer
≤ 500	25	5
≤ 1000	38	4
≤ 2000	60	3
≤ 5000	112	2
≤ 10000	51	1
> 10001	41	0

#### 6.4.1.4 Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien (VE)

Während die bisher abgehandelten Kriterien Rangordnungsmerkmale (ordinale Merkmale) darstellen, ist das Kriterium Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien deutlich davon zu unterscheiden (vgl. hierzu Kapitel 6.3).

Zunächst sind vier nominale Merkmale, d.h. Internet (ja/nein), CD-ROM (ja/nein), andere Online Datenbanken (ja/nein) und Buchform (ja/nein) vorhanden. Danach werden die Nominalkriterien ordinal interpretiert in der Rangordnung

Internet > CD-ROM > Hosts > Buchform.

Die Verbreitung des Internets und der damit im Zusammenhang stehende exponentielle Anstieg an Ressourcen auf diesem Netzwerk der Netzwerke liefert die Begründung für diese hohe Einordnung des Mediums Internet. In diesem Zusammenhang sind jedoch nur die kostenlos zur Verfügung gestellten Datenbanken im Internet, d.h. die kostenfrei verfügbaren Internet Ressourcen angesprochen, nicht die aktuellen Entwicklungen im sog. Commercial Internet, d.h. die Möglichkeit des Zugriffes auf Online Datenbanken via Internet. Die technischen Möglichkeiten der Hostanbindungen im Internet sind in mehreren neueren Publikationen dargelegt [Gaehtgens 1995] [Green 1995]. Bei der Betrachtung von CD-ROMs ist der Zugriff auf die Daten einer CD-ROM über das Internet gemeint. Es handelt sich hierbei um den kostenpflichtigen Zugang zu den Daten. Kostenfreie Zugriffe auf diese Daten sind zwar prinzipiell denkbar jedoch nicht realistisch. So wird beispielsweise der Internet Zugriff auf einige SilverPlatter Datenquellen mit Hilfe der Software WebSpirs realisiert [SilverPlatter 1996].

Die CD-ROM, ebenfalls als ein gegenüber Online Datenbanken und Büchern an Bedeutung gewinnendes Medium [Zass 1995], folgt dem Internet. Danach schließen sich die weitere Verfügbarkeit über andere Hosts und die konventionelle, für Recherchen jedoch wenig geeignete Buchform an.

Je weiter verbreitet eine Datenbank ist, d.h. je mehr Kombinationen der einzelnen Nominalkriterien, desto besser ist die Bewertung. Ist die zu betrachtende Datenbank beispielsweise sowohl in Internet als auch als CD-ROM, als auch bei anderen Hosts, als auch in Buchform verfügbar, erhält sie die höchste Bewertungsziffer.

Auf die allgemeine Vorgehensweise im Zusammenhang mit der Aggregation nominaler Merkmale wurde im Kapitel 6.3 eingegangen. Hier werden die Teilschritte jeweils in Klammern gesetzt. Folgende Betrachtungen sollen zum Verständnis der Findung der Bewertungsziffern dienen: Es können theoretisch die in der Tabelle 22 aufgelisteten Kombinationen auftreten (Teilschritt a). Gemäß der oben beschriebenen Rangordnung erhält nun das Internet die Ziffer 4, CD-ROM 3, Host 2 und Buchform 1 (Teilschritte c und d).

Nach dieser Maßgabe erhält die 4er Kombination Internet, CD-ROM, Host, Buchform ( $4+3+2+1 = 10$ ), die 3er Kombination Internet, CD-ROM, Buchform ( $4+3+1 = 8$ ) etc. (Teilschritt e). Diese Bewertungsziffern sind in Tabelle 23 aufgeführt. Um zur eingangs besprochenen Ausprägungsmenge zu gelangen, werden die Kombinationswerte (die ordinal interpretiert werden sollen) durch 2 dividiert und die Werte jeweils auf- bzw. abgerundet (Teilschritt f).

**Tabelle 22: Kombinationen der Medien Internet, CD-ROM, Host, Buchform**

Kombinationen	Internet	CD-ROM	Host	Buchform	Kombinationswerte
4er Kombination	ja	ja	ja	ja	10
3er Kombination	nein	ja	ja	ja	6
	ja	nein	ja	ja	7
	ja	ja	nein	ja	8
	ja	ja	ja	nein	9
2er Kombination	nein	nein	ja	ja	3
	nein	ja	nein	ja	4
	ja	nein	nein	ja	5
	nein	ja	ja	nein	5
	ja	nein	ja	nein	6
	ja	ja	nein	nein	7
1er Kombination	ja	nein	nein	nein	4

	nein	ja	nein	nein	3
	nein	nein	ja	nein	2
	nein	nein	nein	ja	1

Es kommt demgemäß nicht in erster Linie auf die Anzahl der verfügbaren Medien an, sondern vielmehr auf die oben dargestellte Bedeutung der einzelnen Medien. Die vorgenommene Aggregation führt dazu, daß - entgegen der Hierarchie - bestimmte Kombinationen die gleiche Bewertungsziffer erhalten. Das Rangfolgeverfahren nach der Hasse-Diagramm-Technik verlangt diese Aggregation nicht, sondern kann auch nichtmetrische Klassifikationen (wie hier) verarbeiten. In der vorliegenden Arbeit wird jedoch der einfachere pragmatische Weg eingeschlagen. Aus diesem Grunde sind bei der 1er Kombination Internet und CD-ROM auf einer Stufe, sowie Host und Buch. Es ist deshalb beispielsweise eine 2er Kombination (Internet, CD-ROM 4+3) genauso hoch eingestuft worden wie eine 3er Kombination (Internet, Host, Buchform 4+2+1).

**Tabelle 23: Bewertungsziffern für das Kriterium Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien (VE)**

Weitere Verfügbarkeit	Bewertungsziffer
Internet, CD-ROM, Host, Buchform	5
Internet, CD-ROM, Host	5
Internet, CD-ROM, Buchform	4
Internet, Host, Buchform	4
Internet, CD-ROM	4
CD-ROM, Host, Buchform	3
Internet, Host	3
Internet, Buchform	3
CD-ROM, Host	3
CD-ROM, Buchform	2
Host, Buchform	2
Internet	2
CD-ROM	2
Host	1
Buchform	1
keine	0

Bedeutend für die Erschließung der Inhalte von Datenbanken sind die inhaltlichen Kriterien. In diesem Zusammenhang werden wiederum Kriterien unterschieden, die sich auf die Chemie- bzw. auf Umwelt-Parameter beziehen.

#### **6.4.2 Bewertung nach inhaltlichen Kriterien der Chemie**

Im folgenden werden die Bewertungskriterien mit Chemie- bzw. Chemikalien-Bezug vorgestellt.

### 6.4.2.1 Anzahl der Chemikalien (AN)

Das Bewertungskriterium Anzahl der Chemikalien ist ein ordinales Merkmal.

Dieses Kriterium gibt einen Hinweis über die Wahrscheinlichkeit, ob die gesuchte Chemikalie in der betrachteten Online Datenbank zu finden ist. Die Bandbreite reicht von einigen wenigen Chemikalien bis über 100.000 Chemikalien.

Das Kriterium ist dementsprechend ein quantitativer Wichtungsfaktor für Datenquellen.

**Tabelle 24: Bewertungsziffern für das Kriterium Anzahl der Chemikalien (AN)**

Anzahl der Chemikalien	Bewertungsziffer
> 100.000	5
> 25.000	4
> 10.000	3
> 5.000	2
> 2.000	1
< 2.000	0

### 6.4.2.2 Identifikationsmerkmale (ID)

Bei diesem Kriterium handelt es sich um ein nominales Merkmal, das in der in Kapitel 6.3 beschriebenen Vorgehensweise behandelt wird. Die einzelnen Teilschritte werden wieder vermerkt.

Zu den Identifikationsmerkmalen von Chemikalien gehören Name, Synonyma, Strukturformel, Summenformel, Molekulargewicht, diverse Registriernummern, insbesondere die CAS-Nummer (Chemical Abstract Service Registry Number). Weitere wichtige und häufig gebrauchte Registriernummern sind Beilstein-Nummer, EG-Nummer, EINECS-Nummer, RTECS-Nummer, NIOSH-Nummer etc. Die CAS-Nummer hat sich als Standard Registrier-Nummer durchgesetzt. Es ist jedoch an dieser Stelle anzumerken, daß immer noch über 50 % der Online Datenbanken in Naturwissenschaft und Technik diese Identifikationsnummer nicht enthalten [Voigt 1995b]. Diese Aussage wird durch die Arbeit von Ludl et al. [Ludl 1995] bestätigt, in der in 18 biomedizinischen Datenbanken Recherchen zu vier Umweltchemikalien mit Chemikaliennamen und ggf. mit CAS-Nummern durchgeführt und gegenübergestellt worden sind.

Die Strukturformel ist nur sehr selten in Online Datenbanken zu finden. In einigen Datenbanken ist sie zwar vorhanden, d.h. man erhält sie bei der Ausgabe eines Datensatzes, man kann jedoch nicht nach der Strukturformel suchen.

Zunächst treten hier sieben nominale Merkmale auf, d.h. Strukturformel suchbar (SS), Strukturformel nur Ausgabe (SA), Summenformel (SF), Molekulargewicht (MG), CAS Nummer (CA), Chemikaliename (CN) und Synonyma bzw. Handelsname (SY). Chemiedatenbanken, die Molekulargewichte vorhalten, beinhalten in der Regel auch Summenformeln. Aus diesem Grunde werden diese beiden Merkmale zusammengefaßt, so daß nur noch sechs nominale Merkmale ordinal zu interpretieren sind.

Dazu wird folgende Rangordnung definiert (Teilschritt c):

Strukturformel suchbar (SS) > Strukturformel nur Ausgabe (SA) > Summenformel (SU) > CAS-Nummer (CA), Chemikaliename (CN) > Synonyma bzw. Handelsname (SY).

Als wichtigstes Identifikationsmerkmal ist dementsprechend die Strukturformel einzustufen. In vielen "Umweltchemikalien-relevanten" Online Datenbanken kann man jedoch nur mit dem

Handelsnamen bzw. den Synonyma recherchieren. Diese Datenbanken sind als ungenügend in bezug auf die Abdeckung der Identifikationsmerkmale zu klassifizieren.

Folgende Betrachtungsweise soll die Vergabe der Bewertungsziffern verdeutlichen. Es können theoretisch nach der Binominal-Verteilung gemäß der folgenden Formel (Pascalsches Dreieck)

$$\frac{N!}{k!(N-k)!} \quad N = \text{Anzahl der Merkmale, } k = \text{Anzahl der Kombinationen}$$

eine 6er-, sechs 5er-, 15 4er-, 20 3er-, 15 2er und sechs 1er Kombinationen (Teilschritt a) auftreten. Es sind jedoch nicht alle Kombinationen sinnvoll.

Folgende Kombinationen sind nicht realistisch (Teilschritt b):

Strukturformel suchbar (SS) ja,	Strukturformel nur Ausgabe (SA) nein
Strukturformel suchbar (SS) ja,	Chemikaliennamen (CN) nein
Strukturformel suchbar (SS) ja,	Synonyma (SY) nein
Strukturformel nur Ausgabe (SA) ja,	Chemikaliename (CN) nein
Strukturformel nur Ausgabe (SA) ja ,	Synonyma (SY) nein
Summenformel (SF) / Molekulargewicht (MG) ja,	Chemikaliennamen (CN) nein
Summenformel (SF) / Molekulargewicht (MG) ja,	Synonyma (SY) nein
CAS Nummer (CA) ja,	Chemikaliename (CN) nein
CAS Nummer (CA) ja,	Synonyma (SY) nein

Beispielsweise bietet eine Datenbank, in der man nach Strukturformeln suchen kann, auch die Möglichkeit der Ausgabe der Strukturformel. Eine Kombination Strukturformel suchbar (ja), Strukturformel Ausgabe (nein) ist daher in der Praxis nicht vorhanden. Ebenso gibt es keine Datenbank, die beispielsweise über Strukturformeln, aber über keinen Chemikaliennamen verfügt. Weitere unrealistische Kombinationen sind oben aufgeführt.

Es ist daher festzustellen, daß viele der theoretisch möglichen 5-, 4-,3-,2-Kombinationen unrealistisch sind. Hersteller von Online Datenbanken, die die Suche nach Chemikalien in den Vordergrund stellen, bereiten die Dokumente nicht nur nach chemischen Strukturen, sondern darüber hinaus auch nach anderen chemischen Gesichtspunkten d.h. anderen Identifikationsmerkmalen wie z.B. Summenformel, Molekulargewicht, CAS-Nummer und Synonyma auf. Die mögliche Kombination Strukturformel (suchbar) ja und CAS-Nummer nein tritt erfahrungsgemäß jedoch bei einigen wenigen Datenbanken, vor allen Dingen CD-ROMs auf. Nur die 12 in Tabelle 25 dargestellten Kombinationen sind realistisch.

Gemäß der oben beschriebenen Rangordnung wird Strukturformel suchbar (SS) der Wert 5, Strukturformel nur Ausgabe (SA) der Wert 4, Summenformel (SF) / Molekulargewicht (MG) der Wert 3, CAS-Nummer (CA) der Wert 2, Chemikaliename (CN) der Wert 1 und Synonyma (SY) der Wert 0 zugeordnet (Teilschritt d). Die Kombinationswerte sind in der letzten Spalte der Tabelle 25 aufgeführt (Teilschritt e). Es ergeben sich die Werte 1-15.

**Tabelle 25: Realistische Kombinationen für das Kriterium Identifikationsmerkmale (ID)**

Kombinationen	SS	SA	SF	CA	CN	SY	Kombinationswerte
6er Kombination	ja	ja	ja	ja	ja	ja	15
5er Kombination	nein	ja	ja	ja	ja	ja	10
	ja	ja	nein	ja	ja	ja	12
	ja	ja	ja	nein	ja	ja	13
4er Kombination	nein	nein	ja	ja	ja	ja	6

	nein	ja	nein	ja	ja	ja	7
	nein	ja	ja	nein	ja	ja	8
	ja	ja	nein	nein	ja	ja	10
3er Kombination	nein	nein	nein	ja	ja	ja	3
	nein	nein	ja	nein	ja	ja	4
	nein	ja	nein	nein	ja	ja	5
2er Kombination	nein	nein	nein	nein	ja	ja	1

Es kommt also nicht in erster Linie auf die Tatsache Vorhandensein der einzelnen Identifikationsmerkmale an, sondern vielmehr auf die Bedeutung der Identifikationsmerkmale. Die vorgenommene Aggregation führt dazu, daß

die 5er Kombination: SA, SF, CA, CN, SY

und die 4er Kombination: SS, SA, CN, SY

die gleiche Kombinationsziffer 10 erhalten.

Die Strukturformel suchbar wird als das wichtigste Identifikationsmerkmal genauso hoch bewertet wie die Summenformel und die CAS-Nummer gemeinsam. So ist die Vergabe der selben Bewertungsziffer durchaus vertretbar.

Die einzig sinnvolle 1er Kombination, d.h. die bloße Möglichkeit der Suche nach Informationen über Chemikalien über einen Handelsnamen oder Synonym ist als unzulänglich einzustufen. Das Synonym hat die Bewertungsziffer 0. Für die Vergabe der Bewertungsziffern werden diese Kombinationswerte durch 3 geteilt, was durch Ab- bzw. Aufrundung zu der üblichen Einteilung in sechs Ziffern führt, die in Tabelle 26 aufgezeigt wird (Teilschritt f).

Eine Datenbank, die über alle aufgeführten Identifikationsmerkmale verfügt, bekommt die höchste Bewertungsziffer, eine Datenbank, in der eine Recherche nur mit dem Chemikaliennamen und dem Synonym möglich ist, erhält die Bewertungsziffer 0. Datenbanken, in denen die Strukturformel als Identifikationsparameter vorhanden ist, werden in der Vergabe der Bewertungsziffern bevorzugt.

**Tabelle 26: Bewertungsziffern für das Kriterium Identifikationsmerkmale (ID)**

Identifikationsmerkmale	Bewertungsziffer
Strukturformel (suchbar), Strukturformel (Ausgabe), Summenformel, CAS-Nummer, Chemikalienname, Synonyma	5
Strukturformel (suchbar), Strukturformel (Ausgabe), Summenformel, Chemikalienname, Synonyma	4
Strukturformel (suchbar), Strukturformel (Ausgabe), CAS-Nummer, Chemikalienname, Synonyma	4
Strukturformel (Ausgabe), Summenformel, CAS-Nummer, Chemikalienname, Synonyma	3
Strukturformel (suchbar), Strukturformel (Ausgabe), Chemikalienname, Synonyma	3
Strukturformel (Ausgabe), Summenformel, Chemikalienname, Synonyma	3
Summenformel, CAS-Nummer, Chemikalienname, Synonyma	2
Strukturformel (Ausgabe), CAS-Nummer, Chemikalienname, Synonyma	2
Strukturformel (Ausgabe), Chemikalienname, Synonyma	2
Summenformel, Chemikalienname, Synonyma	1

CAS-Nummer, Chemikalienname, Synonyma	1
Chemikalienname, Synonyma	0

#### 6.4.2.3 Chemikalien-Testdatensatz (CT)

Es handelt sich hier ebenfalls um ein ordinales Merkmal.

Dieses Bewertungskriterium unterstreicht die Notwendigkeit, Informationen über weniger gut untersuchte Substanzen des Testdatensatzes zu erhalten. Während die Kriterien Anzahl an Chemikalien und Identifikationsmerkmale die allgemeine Situation der Unterstützung der Recherche in bezug auf Chemie-Informationen charakterisieren, wird mit dem Parameter Chemikalien-Testdatensatz besonders auf den Testdatensatz an umweltrelevanten Chemikalien, der im Kapitel 4.1.2 erläutert wurde, eingegangen. Wie bereits an dieser Stelle ausgeführt, gehört eine Vielzahl der ausgewählten 68 Stoffe zu der Gruppe der Kohlenwasserstoffe (und hier zu den monocyclischen und polycyclischen Aromaten) und zu den halogenierten organischen Verbindungen (hier zu den halogenierten cyclischen und aromatischen Verbindungen). Diese Verbindungen kommen häufig in der Umwelt vor und sind aus diesem Grund auch in Online Datenbanken umfassend abgedeckt. Bedeutend ist es aber, gerade für diejenigen Chemikalien Informationen zu erhalten, über die weniger geforscht und somit publiziert wurde. Der Parameter CT quantifiziert somit die Chancen, auch "exotische" Chemikalien, d.h. solche, über die nur wenig Datenmaterial vorliegt, zu finden.

Die Auswertung der DADB - Metadatenbank der Online Datenbanken in bezug auf den Chemikalien-Testdatensatz ist in den beiden letzten Spalten der Tabelle 3 in Kapitel 4.1.2.1 zu finden. Teilt man nun die prozentualen Anteile der Treffer in sechs Gruppen ein (siehe Tabelle 27), so läßt sich ablesen, daß sieben Chemikalien in mehr als 50 % der Online Datenbanken vorkommen, 14 in mehr als 40 %, 16 in mehr als 30 %, 9 in mehr als 20 %, 17 in mehr als 10 % und fünf in weniger als 10 % der Datenbanken. Die Chemikalien werden mit ihren korrespondierenden Nummern in Tabelle 27 aufgelistet. Von großem Interesse ist es nun, diejenigen Datenbanken zu identifizieren, die Informationen über die selten vorkommenden Chemikalien beispielsweise über Methylbis(phenylmethyl)benzol (Chemikalie Nr. 20) oder Heptachlor-naphthalin (Chemikalie Nr. 56) vorhalten. Es handelt sich bei diesen Chemikalien um Stoffe, die schlecht untersucht sind und die aus diesem Grunde nur selten in Datenbanken erscheinen. Datenbanken, die solche Chemikalien vorhalten, sind höher einzustufen, als diejenigen, die nur die gängigen Chemikalien des Testdatensatzes enthalten z.B. Anthracen (Nr. 2), Styrol (Nr. 17), Ethylbenzol (Nr. 18), Hexachlorbenzol (Nr. 19) und Cyclohexan (Nr. 37). Die letztgenannten Chemikalien kommen in mehr als der Hälfte der angesprochenen Datenbanken vor.

Enthält eine zu bewertende Online Datenbank nur Chemikalien, die in mehr als 50 % der untersuchten Datenbanken vorkommen, d.h. gut untersuchte Stoffe, so ist dieser Umstand im Sinne der Auffindung von Datenbanken für wenig untersuchte Umweltchemikalien geringer einzustufen als eine Datenbank, die einen Stoff enthält, der in weniger als 10 % der DADB Datenbanken vorkommt. In der Praxis werden jedoch desöfteren Informationen gerade über diejenigen Umweltchemikalien gesucht, die nicht so häufig und eingehend untersucht worden sind und zu denen dementsprechend wenig Daten vorliegen. Nicht unwichtig ist in diesem Zusammenhang die berechnete Annahme, daß neben der selten gefundenen Chemikalie auch die anderen Chemikalien wahrscheinlich in der entsprechenden Datenbank abgedeckt sind.

Demgemäß wird die folgende Bewertungsziffereinteilung aufgrund der Auswertung der DADB in bezug auf die 68 Testdatensatz Chemikalien vorgenommen.

Dieses Kriterium steht in gewissem Zusammenhang mit der Anzahl der Chemikalien. Die Wahrscheinlichkeit, eine selten vorkommende Chemikalie zu recherchieren, ist selbstverständlich in Datenbanken, die eine große Anzahl von Chemikalien abdecken, größer als in einer

Datenbank, in der nur Daten über wenige Chemikalien vorliegen. Es ist jedoch auch möglich, daß einige Datenbanken, obwohl sie wenige Chemikalien umfassen, gerade Informationen zu den wenig in Datenbanken verbreiteten Stoffen haben.

**Tabelle 27: Bewertungsziffern für das Kriterium Chemikalien-Testdatensatz (CT)**

<b>Anzahl und Nummern der Testdatensatz Chemikalien (vgl. Tabelle 3)</b>	<b>Vorkommen der Testdatensatz Chemikalien in DADB in %</b>	<b>Bewertungsziffer</b>
<u>5</u> : 11,20,23,33,56	≤ 10	5
<u>17</u> : 5,10,16,27,31,32,34,38,39,40,41,52,53,55,59,62,68	11-20	4
<u>9</u> : 6,7,8,9,12,35,51,54,65	21-30	3
<u>16</u> : 1,4,14,22,24,26,28,36,43,44,46,50,57,58,61,64	31-40	2
<u>14</u> : 3,13,15,21,25,29,30,42,45,47,48,60,63,66	41-50	1
<u>7</u> : 2,17,18,19,37,49,67	> 50	0

#### 6.4.2.4 Chemikalien Entwicklung (CE)

Von großem Interesse ist auch die Fortentwicklung, die die Datenbanken im Laufe der vergangenen 8-10 Jahre durchgemacht haben. Jeder Datenbankhersteller gibt mit dem Umfang der Datenbank auch die Häufigkeit der Aktualisierungen an. Bedeutend ist jedoch in diesem Zusammenhang, wie sich die Situation in bezug auf den Chemikalien-Testdatensatz verändert hat. Um diese Veränderungen zu analysieren, wurden bei Datenbanken, die schon seit vielen Jahren auf dem Markt sind und in denen zunächst Ende der Achtziger Jahre recherchiert wurde, die Recherchen mit dem Chemikalien-Testdatensatz 1994 bzw. 1995 erneut durchgeführt. Die Vergleiche der Rechercheergebnisse führen zu folgender qualitativen Einstufung: Es wird bei dieser Betrachtungsweise darauf Wert gelegt, um wieviele Chemikalien die Datenbank zugenommen hat. Der Quotient wird aus dem Wert der vorher vorhandenen Anzahl der Testdatensatz-Chemikalien zu der neuen Anzahl der Testdatensatz-Chemikalien gebildet. Datenbanken, die seit Jahren eine umfangreiche Anzahl an den Testdatensatz-Chemikalien abdecken, werden dementsprechend wenig zunehmen. Auch können nur Datenbanken in diesem Sinne betrachtet werden, die es schon seit mehreren Jahren auf dem Markt gibt. Dieses Kriterium stellt wieder ein ordinales Merkmal dar.

Es ist jedoch anzumerken, daß Datenbanken, die bereits seit langem viele Chemikalien beschreiben, keine starke Zunahme mehr verzeichnen und somit in diesem Kriterium niedrig eingestuft werden. Diese "Ungerechtigkeit" wird anhand von Beispielen später diskutiert.

**Tabelle 28: Bewertungsziffern für das Kriterium Chemikalien Entwicklung (CE)**

Veränderungen	Quotient	Bewertungsziffer
umfangreiche Zunahme	> 1,7	5
größere Zunahme	1,5-1,6	4
deutliche Zunahme	1,3-1,4	3
kleine Zunahme	1,2	2
unbedeutende Zunahme	1,1	1
keine Zunahme		0

Eine statistische Einteilung der Bewertungsziffern etwa anhand empirischer Verteilungsfunktionen würde deren Vergabe abhängig machen von den fünf Gruppen. Diese Abhängigkeit soll jedoch zugunsten einer allgemeineren Verwendbarkeit der Bewertungskriterien vermieden werden.

### **6.4.3 Bewertung nach inhaltlichen Kriterien des Umweltschutzes**

Im folgenden werden die Bewertungskriterien mit Umwelt- bzw. Umweltparameter-Bezug vorgestellt.

#### ***6.4.3.1 Informationsparameter für Umweltchemikalien (IP)***

Für die Thematik der Umweltchemikalien sind einige Informationsparameter von besonderer Bedeutung. Hierzu gehören das Vorkommen in den Umweltmedien, Wasser, Boden, Sediment, Luft etc., Abbau- und Akkumulationsparameter, ökotoxikologische Parameter z.B. Fisch-, Algen-, Daphnientoxizität und einige physikalisch-chemische Eigenschaften, die insbesondere zur Abschätzung und Modellierung im Umweltbereich bedeutend sind [Matthies 1992]. Hier seien beispielsweise Verteilungskoeffizienten genannt, die in Expositionsmodellen eine wichtige Rolle spielen [Matthies 1991], [Behrendt 1993]. In der oben beschriebenen Metadatenbank der Online Datenbanken - DADB werden im Deskriptorfeld ca. 100 unterschiedliche Informationsparameter für Umweltchemikalien vorgehalten (vgl. hierzu Kapitel 4.1.3). Es ist einleuchtend, daß eine Online Datenbank höher einzustufen ist, die eine große Anzahl der relevanten Parameter enthält als eine, die nur über wenige bzw. über keine verfügt. Einige spezielle Parameter haben für die Beschreibung des Verhaltens von Chemikalien in der Umwelt eine extreme Bedeutung, so z.B. die Pflanzentoxizität, die Photosynthesehemmung, die Verteilungskoeffizienten  $\log K_{ow}$ ,  $\log K_{oc}$ , die Henry-Konstante etc. Einige dieser Parameter sind jedoch erfahrungsgemäß äußerst selten verfügbar. Aus diesem Grunde werden sie an dieser Stelle "Spezial-Umwelt-Parameter" genannt. Die Einteilungen werden deshalb unter Hinzunahme der "Spezial-Umwelt-Parameter" weiter differenziert und erhalten dementsprechend (d.h. ein, zwei oder mehrere Spezialparameter) höhere Bewertungsziffern. Es wird hier wieder nach der im Kapitel 6.3 aufgeführten Vorgehensweise verfahren. Die beide zugrundeliegenden nominalen Merkmale Anzahl der Parameter (ja/nein) und Vorhandensein von Spezialparametern (ja/nein) werden durch Zahlenwerte für die Anzahl der Parameter allgemein zusätzlich abgestuft. Danach werden die Merkmale ordinal interpretiert in der Rangordnung Größenordnung der Anzahl der Parameter und Größenordnung der Anzahl der Spezialparameter, wie im folgenden erläutert wird.

Es werden vier Abstufungen von Parametern und vier Abstufungen von Spezialparametern gewählt. Vergibt man dem Parameter 50-70 die Ziffer 3, 30-50 die Ziffer 2, < 30 die Ziffer 1 und < 10 die Ziffer 0 und dementsprechend der Anzahl der Spezialparameter > 2 die Ziffer 3, 2 die

Ziffer 2, 1 die Ziffer 1 und 0 die Ziffer 0, so ergeben sich für die Kombination der beiden Parameter folgende in Tabelle 29 zusammengestellten Ergebnisse (vgl. Teilschritte c,d,e).

**Tabelle 29: Kombinationen der Anzahl der Parameter allgemein mit der Anzahl der Spezialparameter**

Anzahl der Parameter allgemein	Anzahl der Spezialparameter	Kombinationswerte
50-70	> 2	6
50-70	2	5
50-70	1	4
50-70	0	3
30-50	> 2	5
30-50	2	4
30-50	1	3
30-50	0	2
< 30	> 2	4
< 30	2	3
< 30	1	2
< 30	0	1
< 10	> 2	3
< 10	2	2
< 10	1	1
< 10	0	0

Die Werte werden in das Bewertungsziffersystem übertragen, indem die Kombinationswerte um jeweils eine Ziffer heruntergestuft werden. Damit würde der Wert sechs die Bewertungsziffer 5 erhalten, der Wert 5 die Bewertungsziffer 4 etc. Der niedrigste Wert 0 kann nach dem Bewertungsziffer-Ansatz nicht unterschritten werden. Die Beibehaltung dieses Wertes ist daher vertretbar und bedeutet, daß mit der Bewertungsziffer 0 die Fälle (<30,0), (<10,1) und (<10,0) zusammengefaßt sind. Allgemein ergeben sich folgende Zuordnungen für die einzelnen Bewertungsziffern (vgl. Teilschritt f).

**Tabelle 30: Bewertungsziffern für das Kriterium Informationsparameter für Umweltchemikalien (IP)**

Anzahl der Parameter	Spezialparameter	Bewertungsziffer
50 - 70	>2	5
50 - 70	2	4
30 - 50	>2	4
50 - 70	1	3
30 - 50	2	3
< 30	> 2	3
50 - 70	0	2
30 - 50	1	2
< 30	2	2
< 10	2	2
30 - 50	0	1
< 30	1	1
< 10	2	1
< 30	0	0
< 10	1	0
< 10	0	0

Dieses Bewertungskriterium würde besser definiert sein, wenn nicht das Vorhandensein des Parameters sondern vielmehr die Belegung des Parameters mit Daten in Betracht gezogen würde. Da diese Information vom Hersteller von Datenbanken nicht erhältlich ist und Recher-

chen zu aufwendig und teuer wären, wird diese Vorgehensweise als sinnvoll jedoch unrealistisch angesehen.

#### 6.4.3.2 Parameter Entwicklung (PE)

Interessant ist neben der in Kapitel 4.1.2 erläuterten Entwicklung des Chemikalien-Testdatensatzes auch die Fortentwicklung der Datenbanken in bezug auf die inhaltlichen Parameter des Umweltschutzes im Laufe der vergangenen 8-10 Jahre. Wie bei der Vorgehensweise beim Chemikalien-Testdatensatz wurden die Recherchen in den Jahren 1994 und 1995 wiederholt. Die Vergleiche der neuen Rechercheergebnisse mit denen der achtziger Jahre führen zu folgender qualitativen Einstufung. Es wird bei dieser Betrachtungsweise darauf Wert gelegt, wieviele Parameter dazugekommen sind, die vor ca. 10 Jahren noch nicht abgedeckt wurden. Der Quotient wird aus dem Wert der vorher vorhandenen Anzahl der Testdatensatz-Parameter zu der neuen Anzahl der Testdatensatz-Parameter gebildet. Datenbanken, die seit Jahren eine umfangreiche Anzahl an den Parametern abdecken, werden dementsprechend wenig zunehmen. Auch können nur Datenbanken in diesem Sinne betrachtet werden, die es schon seit mehreren Jahren auf dem Markt gibt. Auch hier werden wieder die seit vielen Jahren auf dem Markt befindlichen umfangreichen Datenbanken benachteiligt.

Bei diesem Kriterium handelt es sich um ein Rangordnungsmerkmal (ordinales Merkmal).

Die Verteilung der Bewertungsziffern wird analog zu der Vorgehensweise, angewandt auf die Betrachtung des Chemikalien-Testdatensatzes, vorgenommen.

**Tabelle 31: Bewertungsziffern für das Kriterium Parameter Entwicklung (PE)**

Veränderungen	Quotient	Bewertungsziffer
umfangreiche Zunahme	> 1,7	5
größere Zunahme	1,5-1,6	4
deutliche Zunahme	1,3-1,4	3
kleine Zunahme	1,2	2
unbedeutende Zunahme	1,1	1
keine Zunahme		0

#### 6.4.4 Umweltchemikalien-bezogene Bewertungskriterien

Einige Bewertungskriterien behandeln sowohl Chemikalien- als auch Umweltschutzaspekte.

##### 6.4.4.1 Art der Chemikalien (AR)

Es handelt sich bei diesem Kriterium um ein nominales, qualitatives Merkmal.

Unter der Art bzw. Anwendung der Chemikalien sind spezielle Stoffe, die für eine besondere Anwendung produziert werden, zu verstehen. Beispiele hierfür sind Pestizide und Pharmazeutika. Andere Chemikalien sind zwar nicht für einen besonderen Zweck hergestellt, haben sich aber im Laufe der Zeit als gefährlich erwiesen und sind deshalb auch als Gefahrstoffe eingestuft, gekennzeichnet und gesetzlich geregelt worden. Es gibt nun einige wenige Online Datenbanken, die sich mit solchen speziellen Gruppen von Chemikalien befassen. Ebenso ist es für die Thematik der Umweltchemikalien von großem Interesse, ob die Datenbanken nur Einzelstoffe oder auch Gemische, Zubereitungen, Abbauprodukte etc. umfassen. Gemäß der Betrachtungsweise von Umweltchemikalien ist eine Datenbank, die sich vorwiegend mit der

Thematik von Zubereitungen, Abbauprodukten sowie Gefahrstoffen und Stoffen eines speziellen Anwendungsbereiches (z.B. Pestizide) beschäftigt, wichtiger als eine, die die Allgemeinheit von Chemikalien abdeckt. Die Auswertung der entsprechenden Eintragungen im Datenfeld "Anwendung der Chemikalien" der Metadatenbank der Online Datenbanken ergibt, daß 122 der 453 Datenbanken d.h. ca. 25% spezialisiert sind. Hiervon sind nur 12 Datenbanken für den für die Umweltchemikalien Betrachtung wichtigen Bereich "Gemische und Zubereitungen" ausgewiesen. Aus diesem Grund wird diesen die höchste Bewertungsziffer 5 gegeben. Ebenso gibt es nur 22 Online Datenbanken, die auf dem Gebiet der Gefahrstoffe spezialisiert sind. Hierfür wird die Bewertungsziffer 4 vergeben. 88 Datenbanken sind für spezielle Anwendungen angelegt (Bewertungsziffer 3).

Da eine Datenbank, die sich nicht auf besondere Arten bzw. Anwendungen von Chemikalien spezialisiert, nicht als schlecht einzustufen ist, ist in diesem Fall nicht die Bewertung "ungenügend" und "mangelhaft" vergeben worden. Hier können zwar Informationen zu den oben genannten speziellen Anwendungen oder Arten von Chemikalien vorhanden sein, müssen aber nicht. Diese "allgemeinen" Datenbanken sind nicht für Besonderheiten in dem oben diskutierten Zusammenhang ausgewiesen. De facto führt die Vergabe von nur vier Bewertungsziffern zu einer geringeren Gewichtung dieses Kriteriums im Rangfolgeverfahren.

**Tabelle 32: Bewertungsziffern für das Kriterium Art/Anwendung von Chemikalien (AR)**

Art/Anwendung/Verhalten	Bewertungsziffer
Gemische, Zubereitungen, Abbauprodukte	5
Gefahrstoffe	4
spezielle Anwendungen im Umweltbereich	3
allgemein	2

#### 6.4.4.2 Validität/Güte der Datenquellen (QU)

Hierbei handelt es sich um den schwierigsten Parameter, im Sinne der Gewinnung von Hintergrundinformation, nicht jedoch in Sinne der Klassifizierung. Darüber hinaus trägt es sowohl allgemeinen als auch inhaltlichen Charakter, wobei jedoch die inhaltliche Komponente überwiegt.

##### 6.4.4.2.1 Validität/Güte der Online Datenbanken

Anzustreben ist eine Datenquelle, die umfassend das vom Hersteller angegebene Sachgebiet abdeckt und darüber hinaus auch noch fehlerfrei ist. Diese von den Nutzern von Online Datenbanken geforderte Maximale ist natürlich unrealistisch. Es gibt jedoch durchaus wichtige Unterschiede in der Qualität von Online Datenbanken. Beispielsweise geben manche Hersteller ohne weiteres zu, daß ihre Datenbank unevaluierte Daten enthält. Andere Datenbankhersteller hingegen legen auf die Qualität ihrer Daten großen Wert und prüfen die Daten vor der Eingabe in die Datenbank.

Einen anderen Aspekt der Güte der Online Datenbanken stellen die Einheiten der Daten dar. In den meisten Fällen sind die Einheiten nicht genormt, d.h. die Angaben für einen Parameter variieren beispielsweise von mg/l, mol/l, g/l, ng/kg etc. Manchmal fehlen die Einheiten auch vollständig. Für die Auswertung der Daten sind die Angaben in SI Einheiten erstrebenswert. Für sehr wenige Online Datenbanken trifft das jedoch zu. Dies ist besonders bei den betrachteten bibliographischen Datenbanken der Fall, jedoch gibt es auch nur sehr wenige Faktendatenbanken, die ausschließlich SI Einheiten benutzen.

Darüber hinaus ist für die Betrachtung der Güte der Online Datenbanken die Anzahl der gefundenen Umweltchemikalien des Chemikalien-Testdatensatzes von 68 umweltrelevanten Stoffen aussagefähig. Dieser Chemikalien-Testdatensatz wurde ausführlich in Kapitel 4.1.2 besprochen. Hier wurde auch deutlich herausgestellt, daß es sich zwar um potentielle Lebensmittelkontaminanten handelt, die Chemikalien jedoch aufgrund ihrer großen Produktionsmenge einerseits und ihrer Zugehörigkeit zu Gruppen an potentiellen Umweltkontaminaten andererseits per se eine große Umweltrelevanz haben. Das Heranziehen der Anzahl der Testdatensatz Stoffe für das Bewertungskriterium "Validität/ Güte der Datenbank" ist daher vertretbar.

Die Anzahl der Testdatensatz Stoffe ist deutlich von dem Kriterium Chemikalien-Testdatensatz zu differenzieren. Während das hier beschriebene Kriterium die Anzahl der Testdatensatz Stoffe in Betracht zieht, kommt es bei dem Kriterium Chemikalien-Testdatensatz darauf an, welche der Testdatensatz-Chemikalien in Datenbanken zu finden sind.

Auch hier werden wieder die nominalen Merkmale in der unter 6.3 beschriebenen Vorgehensweise aggregiert. Es gibt drei zugrunde liegende nominale Merkmale Chemikalien-Testdatensatz (ja/nein), evaluierte Daten (ja/nein) und Einheiten vereinheitlicht (ja/nein). Das Merkmal Chemikalien-Testdatensatz wird durch Zahlenwerte zusätzlich abgestuft. Es werden folgende fünf Abstufungen für das Merkmal Chemikalien-Testdatensatz vorgenommen: > 50, 25-50, < 25, < 10, < 5

Vergibt man dem Parameter > 50 die Ziffer 4, 25-50 die Ziffer 3, < 25 die Ziffer 2, < 10 die Ziffer 1 und < 5 die Ziffer 0 und den Merkmalen evaluierte Daten und SI Einheiten bei ja/ja die Ziffer 2, bei einem ja die Ziffer 1 und bei nein/nein die Ziffer 0, so ergeben sich für die Kombinationen folgende Ergebnisse (siehe Spalte 4) der Tabelle 33.

**Tabelle 33: Kombinationswerte für Validität/Güte der Datenbank (QU)**

Chemikalien-Testdatensatz	evaluierte Daten	SI Einheiten	Kombinationswerte
> 50	ja	ja	6
> 50	ja	nein	5
> 50	nein	ja	5
> 50	nein	nein	4
25-50	ja	ja	5
25-50	ja	nein	4
25-50	nein	ja	4
25-50	nein	nein	3
< 25	ja	ja	4
< 25	ja	nein	3
< 25	nein	ja	3
< 25	nein	nein	2
< 10	ja	ja	3
< 10	ja	nein	2
< 10	nein	ja	2
< 10	nein	nein	1
< 5	ja	ja	2
< 5	ja	nein	1
< 5	nein	ja	1
< 5	nein	nein	0

Die Werte werden in das Bewertungsziffersystem übertragen, indem die Kombinationswerte um jeweils eine Ziffer heruntergestuft werden. Damit würde der Wert sechs die Bewertungsziffer 5 erhalten, der Wert 5 die Bewertungsziffer 4 etc. Der niedrigste Wert 0 wird für

Datenbanken vergeben, die weniger als 10 Stoffe des Testdatensatzes abdecken und weder über SI Einheiten noch über evaluierte Daten verfügen. Darüber hinaus wird die Bewertungsziffer 0 für alle Datenbanken vergeben, die weniger als fünf Chemikalien des Testdatensatzes abdecken. Damit ergeben sich folgenden Zuordnungen für die einzelnen Bewertungsziffern.

**Tabelle 34: Bewertungsziffern für das Kriterium Validität der Online Datenbanken (QU)**

Chemikalien-Testdatensatz	evaluierte Daten	SI Einheiten	Bewertungsziffer
> 50	ja	ja	5
> 50	ja	nein	4
> 50	nein	ja	4
25 - 50	ja	ja	4
> 50	nein	nein	3
25 - 50	ja	nein	3
25 - 50	nein	ja	3
< 25	ja	ja	3
< 25	ja	nein	2
< 25	nein	ja	2
< 10	ja	ja	2
< 25	nein	nein	1
< 10	ja	nein	1
< 10	nein	ja	1
< 5	ja	ja	1
< 10	nein	nein	0
< 5	ja	nein	0
< 5	nein	ja	0
< 5	nein	nein	0

#### 6.4.4.2.2 Validität/Güte der CD-ROMs

Auch bei CD-ROMs gibt es wichtige Unterschiede in der Qualität der auf der CD-ROM enthaltenen Informationen. In Analogie zu der Vorgehensweise bei den Online Datenbanken werden wieder die Umstände, ob es sich bei den Daten der CD-ROM um evaluierte oder unevaluierte Daten handelt und ob die verwendeten Einheiten SI Einheiten sind oder nicht herangezogen. Da die CD-ROMs - wie oben erwähnt - nicht mit dem Testsatz an Chemikalien geprüft werden konnten, kann für die Betrachtung der Güte der CD-ROM die Anzahl der gefundenen Umweltchemikalien des Chemikalien-Testdatensatzes von 68 umweltrelevanten Stoffen nicht berücksichtigt werden. Es wird daher auf die Anzahl der Chemikalien als Hilfsmittel ausgewichen.

Es gibt drei zugrunde liegende nominale Merkmale Anzahl der Chemikalien (ja/nein), evaluierte Daten (ja/nein) und SI Einheiten (ja/nein).

Das Merkmal Anzahl der Chemikalien wird durch Zahlenwerte zusätzlich abgestuft.

Es werden folgende vier Abstufungen für das Merkmal Anzahl der Chemikalien vorgenommen: >100.000, > 25.000, > 10.000, < 5000.

Das Kriterium Anzahl an Chemikalien geht hier in abgewandelter Form ein, d.h. in der Art, daß nur hohe Werte Berücksichtigung finden.

Vergibt man dem Parameter > 100.000 die Ziffer 3, > 25.000 die Ziffer 2, > 10.000 die Ziffer 1, ≤ 10.000 die Ziffer 0 und den Merkmalen evaluierte Daten und SI Einheiten bei ja/ja die Ziffer 2, bei einem ja die Ziffer 1 und bei nein/nein die Ziffer 0, so ergeben sich für die Kombinationen folgende Ergebnisse der Tabelle 35 (siehe Spalte 4).

Die Werte der Tabelle 35 können in diesem Fall direkt in das Bewertungsziffersystem übertragen werden.

**Tabelle 35: Kombinationswerte für das Kriterium Validität/Güte der CD-ROM (QU)**

Anzahl der Chemikalien	evaluierte Daten	SI Einheiten	Kombinationswerte = Bewertungsziffern
> 100.000	ja	ja	5
> 100.000	ja	nein	4
> 100.000	nein	ja	4
> 100.000	nein	nein	3
> 50.000	ja	ja	4
> 50.000	ja	nein	3
> 50.000	nein	ja	3
> 50.000	nein	nein	2
> 10.000	ja	ja	3
> 10.000	ja	nein	2
> 10.000	nein	ja	2
> 10.000	nein	nein	1
≤ 10.000	ja	ja	2
≤ 10.000	ja	nein	1
≤ 10.000	nein	ja	1
≤ 10.000	nein	nein	0

Damit ergeben sich folgenden Zuordnungen für die einzelnen Bewertungsziffern.

**Tabelle 36: Bewertungsziffern für das Kriterium Validität/Güte der CD-ROM (QU)**

Chemikalien-Testdatensatz	evaluierte Daten	SI Einheiten	Bewertungsziffer
> 100.000	ja	ja	5
> 100.000	ja	nein	4
> 100.000	nein	ja	4
> 50.000	ja	ja	4
> 100.000	nein	nein	3
> 50.000	ja	nein	3
> 50.000	nein	ja	3
> 10.000	ja	ja	3
> 10.000	ja	nein	2
> 10.000	nein	ja	2
≤ 10.000	ja	ja	2
> 10.000	nein	nein	1
≤ 10.000	ja	nein	1
≤ 10.000	nein	ja	1
≤ 10.000	nein	nein	0

## 6.5 Gesamtmaß $\Gamma$ als grobes Einstufungsmaß für Objekte

Wie aus dem Vorangegangenen ersichtlich, wurden zur Ableitung der Bewertungsziffern auch Aggregationen vorgenommen, und zwar dann, wenn die Charakterisierung der Objekte zunächst eine Kombination von nominalen und ordinalen Kriterien erforderte. Man kann einen Schritt weitergehen und ein Gesamtmaß für die Qualität der Datenbanken ableiten wollen, das wie folgt definiert sein kann:

-Sei  $K_{ij}$  die Ausprägung für das  $j$ -te Objekt unter dem  $i$ -ten Einzelkriterium ( $i = 1, \dots, n$ ).

Ein Gesamtmaß  $\Gamma$  wäre dann für das  $j$ -te Objekt wie folgt zu bilden:

$$\Gamma_j = f(K_{1j}, K_{2j}, \dots, K_{nj})$$

$\Gamma$  = Gesamtmaß,  $K_{ij}$  =  $i$ -tes Kriterium für das  $j$ -te Objekt (siehe auch Anlage B1)

Nach Wahl der Kriterien sind die  $K_{ij}$  in ihrer Aussage orientiert (hohe Ausprägung, hohe Güte); daher muß gelten

$$\frac{\partial f}{\partial K_{ij}} > 0$$

Eine übliche Realisierung für  $f$  ist eine Linearform:

$$\Gamma_j = \sum g_i K_{ij} \quad g_i \text{ Gewichtungsfaktoren } 0 < g_i < n \quad (6-1)$$

Durch die zusätzliche (willkürliche) Wahl  $g_i = 1$  für  $i = 1, \dots, n$

erhält man eine spezielle Berechnung für ein Qualitätsgesamtmaß  $\Gamma$ , in dem alle Kriterien gleichgewichtig eingehen. Diese Größe  $\Gamma$  wird zusätzlich in der letzten Spalte der folgenden Tabellen der Bewertungsziffern für die Gruppen eingetragen. Aus der Ableitung wird ersichtlich, daß  $\Gamma$  mit erheblicher Willkür behaftet ist, die ja gerade durch die Hasse-Diagramm-Technik vermieden wird. Hier dient  $\Gamma$  u. a. dazu, die Vorteile der Hasse-Diagramm-Technik evident zu machen. Durch das Gesamtmaß  $\Gamma$  wird eine Einteilung in Äquivalenzklassen induziert: Die dazugehörige Äquivalenzrelation ist "Gleichheit bezüglich  $\Gamma$ ". In der Folge wird von  $\Gamma$ -äquivalenten Objekten gesprochen, wenn diese zu einer Äquivalenzklasse gehören. Eine Nichtübereinstimmung vom Gesamtmaß  $\Gamma$  und der Auswertung mittels der Hasse-Diagramm-Technik kann grundsätzlich sein oder aufgrund der Willkür der Klasseneinteilung zustande kommen. Dies wird anhand der Gruppen diskutiert.

## 6.6 Bedeutung der Bewertungskriterien und denen Ausprägungsmengen

Die Struktur eines Hasse-Diagramms hängt nicht nur von der Gruppe und von der Anzahl der Bewertungskriterien sondern vielmehr auch von den Kriterien selber ab. Bei den Bewertungskriterien ist deren Ausprägungsmenge, d. h. die Verteilung der Bewertungsziffern auf die einzelnen Objekte von großer Bedeutung. Darüber hinaus ist das "Zusammenspiel" der einzelnen Bewertungskriterien miteinander wichtig. Bezüglich der Ausprägungsmengen der einzelnen Bewertungskriterien werden bei den Betrachtungen der folgenden Gruppen die Mittelwerte (MW), Mediane (ME) und die Standardabweichungen (STDEV) gebildet.

Für die Untersuchung der Bedeutung der Ausprägungsmengen werden im folgenden zwei Größen, der Umfangsindex  $U_i$  und der Shannon-Index  $I_i$  betrachtet.

### 6.6.1 Umfangsindex $U_i$

Es ist zwischen der theoretischen Ausprägungsmenge  $A_{\text{theo}, i} := \{0, 1, 2, 3, 4, 5\}$  und der Kriterien- und Gruppen-spezifischen aktuellen Ausprägungsmenge  $A_{\text{aktu}, i}$  zu unterscheiden. Es gilt, die Kardinalitäten der theoretischen Ausprägungsmenge mit denen der aktuellen Ausprägungsmenge zu vergleichen.

Der Umfangsindex  $U_i$  der Bewertungsziffern wird wie folgt definiert:

$$U_i = \text{card } A_{\text{aktu}, i} \quad (6-2)$$

$U_i$  = Umfangsindex,  $A$  = Ausprägungsmenge,  $i$  = Kriterium

### 6.6.2 Shannon-Index $I_i$

Der Shannon-Index beschreibt die Information eines Bewertungskriteriums bezüglich der konkret gegebenen Gruppe [Bock 1974]. Dabei wird davon ausgegangen, daß eine hohe Entropie informativer ist als eine geringe. Daher ist der Shannon-Index ein Maß für den Informationsgehalt eines Bewertungskriteriums.  $P_{ij}$  sind Wahrscheinlichkeiten, daß die  $j$ -te Ausprägung zum  $i$ -ten Kriterium auftritt und wird durch  $n_{ij}/N$  geschätzt.

Es gilt:

$$I_i = - \sum_j p_{ij} \ln p_{ij}$$

$$I_i = - \sum_j \frac{n_{ij}}{N} \ln \frac{n_{ij}}{N} \quad (6-3)$$

Legende:  $I_i$  = Shannon-Index,  $i$  = Bewertungskriterium,  $j$  = Ausprägungsstufe,  $n_{ij}$  = Zahl der Vergabe einer bestimmten Bewertungsziffer Stufe  $j$  über das  $i$ -te Kriterium,  $N$  = Anzahl der Objekte in einer Objektmenge,  $z$  = Zahl der möglichen Ausprägungen (in diesem Fall 0-5, also 6)

Der Shannon-Index wurde 1982 in einer Arbeit von Rao [Rao 1982] kritisiert. Es ist nicht auszuschließen, daß in manchen Fällen die Diversität durch den Shannon-Index überschätzt wird.

### 6.6.3 Interpretation der Indizes

Anzustreben ist eine gut verteilte Bewertungsziffernvergabe. Haben  $U_i$  und  $I_i$  hohe Werte im Vergleich zu den anderen Bewertungskriterien, so ist das betrachtete Kriterium in bezug auf die Verteilung als "gut" anzusehen. Analoges gilt selbstverständlich für niedrige Werte der beiden Indizes. In diesem Fall ist die Bewertungsziffern Vergabe als "schlecht" anzusehen. Wie zu erwarten, liefert der Shannon-Index differenziertere Werte als der Umfangsindex, da ersterer die Anzahl der Objekte einer Gruppe mitberücksichtigt. In der Folge wird von guter, mittlerer und schlechter Verteilung der Bewertungsziffern gesprochen. Als gut gilt, wenn der Umfangsindex  $U_i = 6$  oder  $5$  beträgt, und der Shannon-Index  $I_i$  groß ist im Vergleich zu den Werten der anderen Kriterien. Darüber hinaus müssen die Werte Mittelwert und Median nahe beieinander liegen. Als mittelmäßig wird ein Kriterium bezeichnet, wenn der Umfangsindex  $U_i = 5$  ist, und der Shannon-Index  $I_i$  deutlich kleiner ist als der höchste Wert in der betrachteten Gruppe. Als schlecht ist ein Kriterium anzusehen, wenn der Umfangsindex  $U_i \leq 4$  ist, und der Shannon-Index  $I_i$  kleine Werte aufweist. Hierbei tritt auch eine große Abweichung von Mittelwert und Median auf.

## 7. Hasse-Diagramme für die ausgewählten Gruppen an Datenquellen

### 7.1 Besonderheiten der Interpretationen und Übersicht

Bei den folgenden Analysen von Online Datenbanken und CD-ROMs wird je eine Gruppe (auch Objektmenge genannt) ausführlich dargestellt, das Hasse-Diagramm erläutert, und die Besonderheiten diskutiert. Bei allen weiteren Gruppen werden die Teilschritte nicht mehr gesondert aufgeführt. Die allgemeinen Angaben zu den Hasse-Diagrammen befinden sich im Anhang C. Ebenso werden die Gegenüberstellungen einzelner Objekte im Hasse-Diagramm mit dem Gesamtmaß  $\Gamma$  sowie die Ausprägungsmengen der Bewertungskriterien im Anhang abgehandelt.

Die Datenbanken bzw. CD-ROMs, die sich als maximale Objekte ergeben haben, werden bei der Auswertung der Hasse-Diagramme besonders in den Vordergrund gestellt. Dabei werden die isolierten Objekte - so vorhanden - nicht mit in die Betrachtung der maximalen Objekte respektive minimalen Objekte mit einbezogen. Isolierte Objekte können sowohl den maximalen als auch den minimalen Objekten zugeordnet werden. Die D-Matrix und die W-Matrix werden nur für die Maximalen berechnet und diskutiert.

Als aussagekräftig oder (synonym verwendet) übersichtlich ist erfahrungsgemäß ein Hasse-Diagramm zu bezeichnen, das in der Anzahl der Ebenen und in der Zahl der Elemente in der breitesten Ebene ausgewogen ist. Als schmal wird ein Diagramm bezeichnet, welches eine deutlich größere Anzahl der Ebenen aufweist als Elemente in der breitesten Ebene. Entsprechend gilt ein Hasse-Diagramm als breit, wenn die Anzahl der Objekte in der breitesten Ebene deutlich über der Anzahl der Ebenen liegt. Gut interpretierbar ist ein Hasse-Diagramm, wenn die Zahl der Unvergleichbarkeiten und Vergleichbarkeiten nicht zu stark voneinander abweichen.

Vergleichsweise nur wenige Aussagen können aus einem Hasse-Diagramm gezogen werden, wenn eine Vielzahl von isolierten Objekten vorliegt.

Oftmals werden anhand der Hasse-Diagramme Parallelitäten (Ketten in Hasse-Diagrammen) oder Antiparallelitäten (Antiketten) von Kriterien, genommen über bestimmte Gruppen festgestellt. Würde das Ziel nur die Identifikation solcher paarweisen Abhängigkeiten sein, könnte auch die Rangkorrelation hilfreich sein [Precht 1987]. Hier wird darauf verzichtet, da zwar die Aufdeckung von derartigen Abhängigkeiten wichtig ist, hier jedoch die Untersuchung der Struktur der Ebenen, insbesondere der Maximalen im Vordergrund steht.

Wie oben ausführlich dargelegt, werden je fünf Gruppen an Datenquellen mit einer Auswahl an Bewertungskriterien dem in Kapitel 3.4 dargestellten Bewertungsverfahren mittels der Hasse-Diagramm-Technik unterzogen.

Im folgenden werden drei Kriterien-Teilmengen betrachtet:

- A. vollständiger Satz an Bewertungskriterien (allgemeine und inhaltliche) [IB<sub>all+inh</sub>]
- B. allgemeine Bewertungskriterien [IB<sub>all</sub>]
- C. inhaltliche Bewertungskriterien [IB<sub>inh</sub>]

Da, wie auch schon oben näher ausgeführt worden ist, die CD-ROMs nicht so eingehend getestet werden konnten wie die Online Datenbanken entfallen die Kriterien CT (Chemikalien-Testdatensatz), CE (Chemikalien Entwicklung) und PE (Parameter Entwicklung) hier.

## 7.2 Hasse-Diagramme für Online Datenbanken

### 7.2.1 Bibliographische Online Datenbanken der Chemie-Gruppe (Gruppe 1)

In Tabelle 37 werden die Bewertungsziffern für 12 Bewertungskriterien zu 34 bibliographischen chemierelevanten Online Datenbanken, deren Bezeichnungen und dreistelliges Akronym der Datenbanken aufgeführt. Das Akronym wird auch in den zu erstellenden Hasse-Diagrammen verwendet. In der letzten Spalte wird das in Kapitel 6.5 eingeführte Gesamtmaß  $\Gamma$  aufgelistet.

Wie aus der Tabelle 37 ersichtlich, haben sechs Datenbanken keinen vollständigen Satz an Bewertungskriterien. Es handelt sich hierbei um CAR (CAPREVIEWS), CAP (CAPLUS), EUR (European Chemical News), INS (INSPEC), JPN (Japan News) und KOS (KOSMET). Die fünf letztgenannten Datenbanken wurden erst im Jahr 1995 recherchiert. Aus diesem Grund können die Bewertungskriterien Chemikalien Entwicklung (CE) Parameter Entwicklung (PE) nicht betrachtet werden. Bei der Datenbank CAPREVIEWS handelt es sich um eine Vorschau auf die Dokumente, die in Kürze in die CA Datenbank eingetragen werden. Aus diesem Grund können die Kriterien Chemikalien-Testdatensatz (CT), Chemikalien Entwicklung (CE), Informationsparameter für Umweltchemikalien (IP) und Parameter Entwicklung (PE) nicht analysiert werden. Es werden somit bei der Auswertung nur die **28 Datenbanken** mit vollständigem Satz an Bewertungskriterien berücksichtigt.

**Tabelle 37: Bibliographische Datenbanken der Chemie-Gruppe und deren Bewertungsziffern**

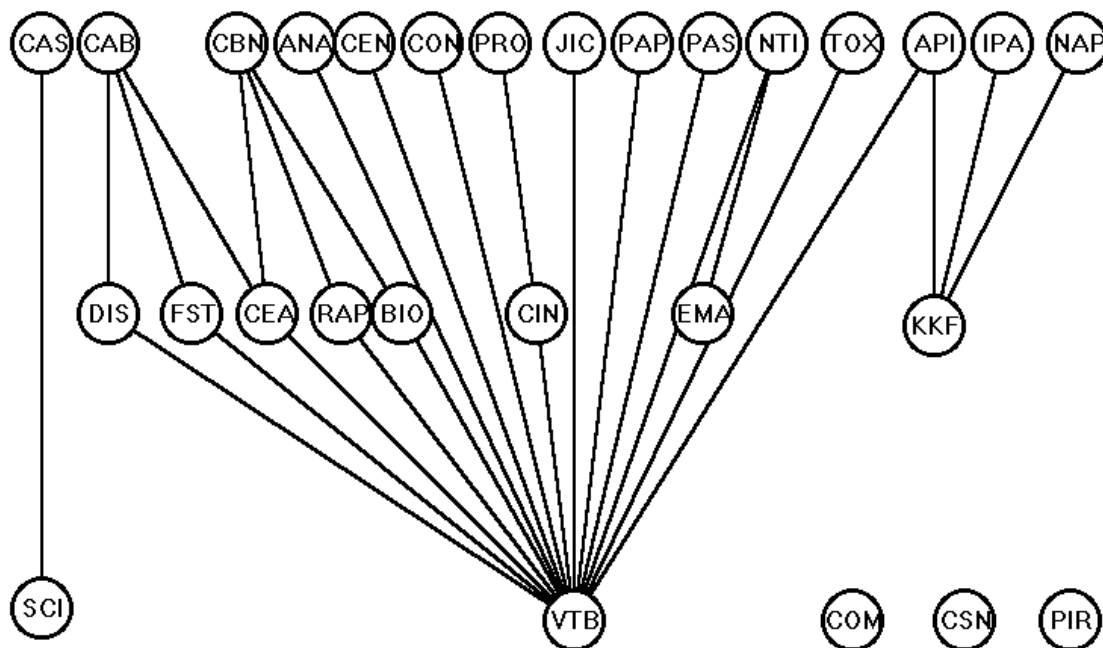
Nr.	Datenbank	Akr.	UM	UP	KO	VE	AN	ID	CT	CE	IP	PE	AR	QU	$\Gamma$
01	ANABST	ANA	2	4	1	3	2	1	4	2	3	5	2	2	31
02	APILIT	API	3	4	0	1	2	0	4	1	4	0	3	3	25
03	BIOTECHABS	BIO	2	4	0	2	3	0	3	1	3	4	2	2	26
04	CA	CAS	5	5	0	3	5	3	5	0	5	0	2	3	36
05	CAB	CAB	4	4	4	3	2	0	4	2	5	5	2	2	37
06	CAPLUS	CAP	5	5	0	0	5	3	5	-	5	-	2	3	-
07	CAPREVIEWS	CAR	2	5	0	0	1	3	-	-	-	-	2	3	-
08	CEABA	CEA	2	4	1	3	0	0	4	1	4	3	2	2	26
09	CBNB	CBN	2	5	1	3	3	0	5	1	4	4	2	2	32
10	CEN	CEN	3	5	2	0	0	1	3	5	4	4	2	2	31
11	CIN	CIN	4	5	0	1	1	1	5	1	4	0	2	2	26
12	COMPENDEX	COM	4	4	1	3	3	0	5	0	4	1	2	3	30
13	CONFSCI	CON	4	4	2	1	1	0	3	1	4	3	2	3	28
14	CSNB	CSN	1	5	0	3	2	1	3	3	4	4	2	2	29
15	DISSABS	DIS	4	4	2	1	2	0	4	2	4	4	2	2	31
16	EMA	EMA	2	4	2	3	1	0	2	3	3	2	2	1	25
17	European C.N.	EUR	2	5	0	0	1	0	3	-	3	-	3	1	-
18	FSTA	FST	2	4	2	3	1	0	3	1	4	3	2	2	27
19	INSPEC	INS	4	5	2	1	3	0	4	-	4	-	2	2	-
20	IPA	IPA	2	4	3	3	0	1	2	0	4	3	3	1	26
21	JICST-E	JIC	4	4	2	2	3	0	4	4	5	0	2	2	32
22	JPNEWS	JPN	5	5	2	0	0	0	2	-	2	-	2	1	-
23	KKF	KKF	2	4	0	0	0	0	2	0	1	0	3	1	13

24	KOSMET	KOS	2	4	1	0	0	1	2	-	3	-	3	1	-
25	NAPRALERT	NAP	2	4	3	0	5	1	4	1	1	2	3	1	27
26	NTIS	NTI	4	4	2	3	1	1	3	3	5	2	2	2	32
27	PAPERCHEM	PAP	2	5	2	1	0	0	4	5	4	4	2	2	31
28	Pascal	PAS	5	4	2	3	3	0	5	2	5	1	2	2	34
29	PIRA	PIR	2	5	0	1	0	1	2	5	5	5	2	2	30
30	PROMT	PRO	4	5	0	1	2	1	5	3	5	1	2	2	31
31	RAPPRA	RAP	2	5	0	3	1	0	4	1	3	3	2	2	26
32	SCISEARCH	SCI	5	5	0	3	3	0	5	0	5	0	2	3	31
33	TOXLINE	TOX	4	4	2	5	4	1	5	1	5	0	2	3	36
34	VTB	VTB	2	3	0	0	0	0	3	1	2	0	2	2	15

Abkürzungen siehe Abkürzungsverzeichnis

### 7.2.1.1 Allgemeine und inhaltliche Bewertungskriterien

Es werden 28 Datenbanken und die Bewertungskriterien UM, UP, KO, VE, AN, ID, CT, CE, IP, PE, AR, QU zur Auswertung mittels der Hasse-Diagramm-Technik herangezogen.



**Abbildung 5:** Hasse-Diagramm Gruppe 1: Online Datenbanken, 12 Bewertungskriterien

Die folgende Tabelle enthält die allgemeinen Angaben zu dem Hasse-Diagramm der vollständigen Satzes an Bewertungskriterien, der auch als Fall 0 bezeichnet wird, sowie zu dem Fall 11, in welchem das Kriterium Art/ Anwendung der Chemikalien ausgelassen wird.

**Tabelle 38: Auswertung der Hasse-Diagramme zu Gruppe 1: Online Datenbanken, 12 Bewertungskriterien**

	Fall 0	Fall 11
<b>Anzahl der Ebenen (NL)</b>	3	3
<b>Elemente in der breitesten Ebene (NEL)</b>	18	17
<b>maximale Objekte</b>	18: {ANA}, {API}, {CAS}, {CAB}, {CBN}, {CEN}, {COM}, {CON}, {CSN}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PIR}, {PRO}, {TOX}	17: {ANA}, {CAS}, {CAB}, {CBN}, {CEN}, {COM}, {CON}, {CSN}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PIR}, {PRO}, {TOX}
<b>maximale Objekte (ohne isolierte Objekte)</b>	15: {ANA}, {API}, {CAS}, {CAB}, {CBN}, {CEN}, {CON}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PRO}, {TOX}	16: {ANA}, {CAS}, {CAB}, {CBN}, {CEN}, {COM}, {CON}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PIR}, {PRO}, {TOX}
<b>minimale Objekte</b>	7: {COM}, {CSN}, {EMA}, {KKF}, {PIR}, {SCI}, {VTB}	3: {CSN}, {KKF}, {VTB}
<b>minimale Objekte (ohne isolierte Objekte)</b>	4: {EMA}, {KKF}, {SCI}, {VTB}	2: {KKF}, {VTB}
<b>isolierte Objekte</b>	3: {COM}, {CSN}, {PIR}	1: {CSN}
<b>isolierte Hierarchien</b>	CAS-SCI	-
<b>Äquivalenzklassen</b>	28	28
<b>nicht-triviale Äquivalenzklassen</b>	-	-
<b>Vergleichbarkeiten V(N)</b>	30	53
<b>Unvergleichbarkeiten U(N)</b>	696	650

Abkürzungen siehe Abkürzungsverzeichnis

### 7.2.1.1.1 Auswertung des Hasse-Diagramms

Es resultiert ein breites Diagramm mit wenigen Ebenen, vielen maximalen, minimalen und isolierten Elementen. Aus dem Diagramm und den dazugehörigen allgemeinen Angaben ist zu entnehmen, daß es keine nicht-trivialen Äquivalenzklassen gibt. In diesem Diagramm befinden sich drei isolierte Elemente, COM (Computerized Engineering Index), CSN (Chemical Safety Newbase) und PIR (Pira International Publishing and Information Services). Diese Datenbanken sind nicht nur mit keiner der anderen Datenbanken der Gruppe vergleichbar, sondern sie sind es auch nicht untereinander. Diese isolierten Elemente kann man sowohl den maximalen als auch den minimalen Objekten zurechnen.

Folgende maximale Objekte sind anzutreffen: ANA (Analytical Abstracts), API (American Petroleum Institute Literature Database), CAS (Chemical Abstract Service Database), CAB (CAB Abstracts), CBN (Chemical Business Newbase), CEN (Chemical Engineering News), CON (Conference Papers Index), IPA (International Pharmaceutical Abstracts), JIC (JICST File on Science, Technology, and Medicine in Japan), NAP (NAPRALERT), NTI (National Technical Information Service), PAP (PAPERCHEM), PAS (PASCAL), PRO (PROMT) und TOX (Toxicology Literature Online). Diese maximalen Objekte sind besser eingestuft in bezug auf die Informationsbasis als die nach unten mit ihnen verbundenen Datenbanken. Beispielsweise ist CAB besser als DIS, FST, CEA und VTB bewertet.

Minimale Objekte sind EMA (Engineered Materials Abstracts), KKF (Kunststoffe Kautschuk Fasern), SCI (Science Citation Search System) und VTB (Verfahrens-technische Berichte).

Diese Datenbanken sind in bezug auf alle 12 Bewertungskriterien schlechter eingestuft als die Datenbanken, mit denen sie nach oben durch Linien verbunden sind.

In diesem Diagramm ist die Zahl der Unvergleichbarkeiten mit 696 sehr hoch, die Zahl der Vergleichbarkeiten hingegen niedrig, d.h. viele Objektpaare unterscheiden sich charakteristisch voneinander. Viele Datenbanken haben wenigstens ein Kriterium unter den 12, das sie zu maximalen Elementen werden läßt. Ein weiteres Kriterium würde nur noch einen unwesentlichen Effekt auf die Struktur des Hasse-Diagramms haben [Brüggemann 1996b]. Die Wegnahme eines Kriteriums könnte jedoch starke Änderungen im Hasse-Diagramm bewirken.

Aus der großen Anzahl der Datenbanken in der breitesten Ebene (NEL = 18) folgt, daß das Hasse-Diagramm als sehr "breit" zu bezeichnen ist, d.h. daß die Informationsbasis Bewertungskriterien enthält, die weitestgehend die "Individualität" jeder Datenbank erfassen. Würde das Hasse-Diagramm nur aus einer Antikette bestehen, also extrem breit sein, dann umfaßte die Informationsbasis alle notwendigen Kriterien, um die Besonderheit jeder Datenbank zu erfassen.

Das Hasse-Diagramm weist zwei Zusammenhangskomponenten auf, die jeweils mehr als ein Objekt umfassen. Dieser Umstand deutet auf eine Besonderheit in der Datenstruktur hin. Tatsächlich findet man anhand der Tabelle 37, daß die beiden Datenbanken CAS (Chemical Abstract Service Database) und SCI (Science Citation Search System) die einzigen Datenbanken sind, die in den Kriterien UM (Umfang der Datenbank) und gleichzeitig in UP (Häufigkeit der Aktualisierung) die höchste Bewertungsziffer haben.

Weiterhin fällt die Datenbank API (American Petroleum Institute Literature Database) auf, die als sog. Artikulationspunkt zwei Teile des Hasse-Diagramms verbindet. Die Besonderheit, die die Datenbanken API, IPA, NAP und KKF vom Rest der Datenbanken trennt, ist die Tatsache, daß nur für diese vier Online Datenbanken für Art/ Anwendung der Chemikalien das Bewertungskriterium drei vergeben wurde.

#### 7.2.1.1.2 Analyse anhand der D-Matrix

Es werden die Maximalen, d.h. ANA, API, CAS, CAB, CBN, CEN, CON, IPA, JIC, NAP, NTI, PAP, PAS, PRO und TOX als Schlüsselemente gewählt. Die isolierten Elemente werden ausgelassen (vgl. hierzu die allgemeinen Ausführungen zur D-Matrix in Kapitel 3.3 und Anhang B.4).

**Tabelle 39: D-Matrix der Maximalen der Gruppe 1: Online Datenbanken**

	ANA	API	CAS	CAB	CBN	CEN	CON	IPA	JIC	NAP	NTI	PAP	PAS	PRO	TOX
ANA	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1
API	1	2	0	1	1	1	1	1	1	1	1	1	1	1	1
CAS	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
CAB	1	1	0	4	2	1	1	0	1	0	1	1	1	1	1
CBN	1	1	0	2	4	1	1	0	1	0	1	1	1	1	1
CEN	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1
CON	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1
IPA	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
JIC	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1
NAP	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
NTI	1	1	0	1	1	1	1	0	1	0	2	1	1	1	1
PAP	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1
PAS	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1

<b>PRO</b>	1	1	0	1	1	1	1	0	1	0	1	1	1	2	1
<b>TOX</b>	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1

Abkürzungen siehe Abkürzungsverzeichnis

Wie aus der Tabelle erkennbar ist, gibt es für die einzelnen Online Datenbanken nur sehr wenige Nachfolgermengen. Die Online Datenbank CBN (Chemical Business Newsbase) und CAB (CAB Abstracts) haben jeweils vier Datenbanken, die vergleichbar schlechter eingestuft werden als sie selbst. Die meisten Datenbanken haben eine oder keine Nachfolger-Datenbank. Die beiden Datenbanken CAB und CBN haben zwei gemeinsame Nachfolger-Datenbanken, nämlich CEA (Chemical Engineering and Biotechnology Abstracts) und VTB (Verfahrens-technische Berichte). Auch die Analyse der Nichtdiagonalelemente macht deutlich, daß es sehr wenige Datenbanken gibt, die miteinander verglichen werden können.

### 7.2.1.1.3 Bedeutung der Bewertungskriterien

In der W-Matrix werden die Einflüsse der Kriterien auf die einzelnen Online Datenbanken beschrieben. In diesem Fall liegen folgende 12 Bewertungskriterien: UM, UP, KO, VE, AN, ID, CT, CE, IP, PE, AR, QU vor. Es ergeben sich daraus n+1 also 13 Fälle. In Fall 0 sind alle Bewertungskriterien vorhanden und in den 12 folgenden Fälle wird jeweils ein Bewertungskriterium ausgelassen. Welches der Kriterien wegfällt, ist aus Tabelle 40 ersichtlich (vgl. hierzu die allgemeinen Ausführungen zur W-Matrix in Kapitel 3.3 und Anhang B.5).

**Tabelle 40: Fälle für 12 Bewertungskriterien**

Fall/Attribut	UM	UP	KO	VE	AN	ID	CT	CE	IP	PE	AR	QU
<b>Fall 0</b>	x	x	x	x	x	x	x	x	x	x	x	x
<b>Fall 1</b>	-	x	x	x	x	x	x	x	x	x	x	x
<b>Fall 2</b>	x	-	x	x	x	x	x	x	x	x	x	x
<b>Fall 3</b>	x	x	-	x	x	x	x	x	x	x	x	x
<b>Fall 4</b>	x	x	x	-	x	x	x	x	x	x	x	x
<b>Fall 5</b>	x	x	x	x	-	x	x	x	x	x	x	x
<b>Fall 6</b>	x	x	x	x	x	-	x	x	x	x	x	x
<b>Fall 7</b>	x	x	x	x	x	x	-	x	x	x	x	x
<b>Fall 8</b>	x	x	x	x	x	x	x	-	x	x	x	x
<b>Fall 9</b>	x	x	x	x	x	x	x	x	-	x	x	x
<b>Fall 10</b>	x	x	x	x	x	x	x	x	x	-	x	x
<b>Fall 11</b>	x	x	x	x	x	x	x	x	x	x	-	x
<b>Fall 12</b>	x	x	x	x	x	x	x	x	x	x	x	-

Abkürzungen siehe Abkürzungsverzeichnis

In Tabelle 41 ist die Auswertung der W-Matrix der 15 maximalen Elemente ANA, API, CAS, CAB, CBN, CEN, CON, IPA, JIC, NAP, NTI, PAP, PAS, PRO und TOX eingetragen. Wenn als Schlüsselemente nur die Maximalen gewählt werden, werden natürlich nicht generelle Veränderungen erfaßt, sondern nur solche, die für die Schlüsselemente gelten. Diese einschränkende Bemerkung gilt auch für alle weiteren Auswertungen der W-Matrix innerhalb dieser Arbeit.

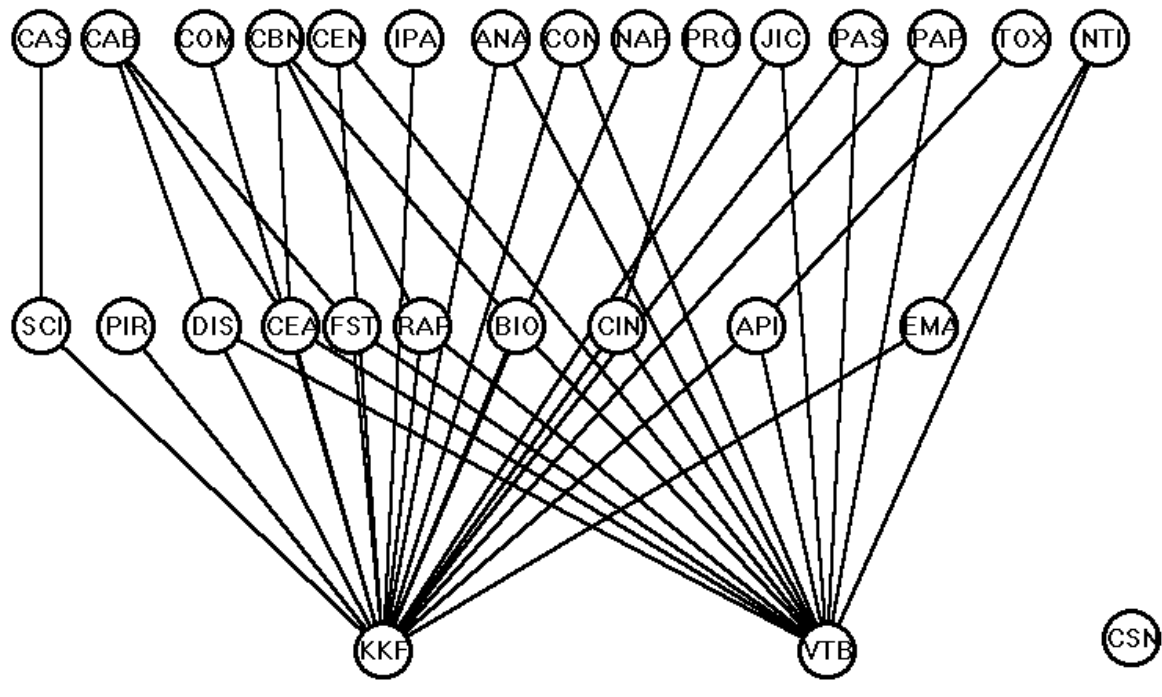
**Tabelle 41: Auswertung der W-Matrix der Maximalen (Gruppe 1: Online Datenbanken)**

Fall 0 / Fall 1	Kriterium UM ausgelassen	0 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 2	Kriterium UP ausgelassen	3 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 3	Kriterium KO ausgelassen	1 Veränderung im Hasse-Diagramm
Fall 0 / Fall 4	Kriterium VE ausgelassen	2 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 5	Kriterium AN ausgelassen	2 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 6	Kriterium ID ausgelassen	1 Veränderung im Hasse-Diagramm
Fall 0 / Fall 7	Kriterium CT ausgelassen	0 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 8	Kriterium CE ausgelassen	5 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 9	Kriterium IP ausgelassen	1 Veränderung im Hasse-Diagramm
Fall 0 / Fall 10	Kriterium PE ausgelassen	12 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 11	Kriterium AR ausgelassen	13 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 12	Kriterium QU ausgelassen	2 Veränderungen im Hasse-Diagramm

Abkürzungen siehe Abkürzungsverzeichnis, Erläuterungen siehe Text

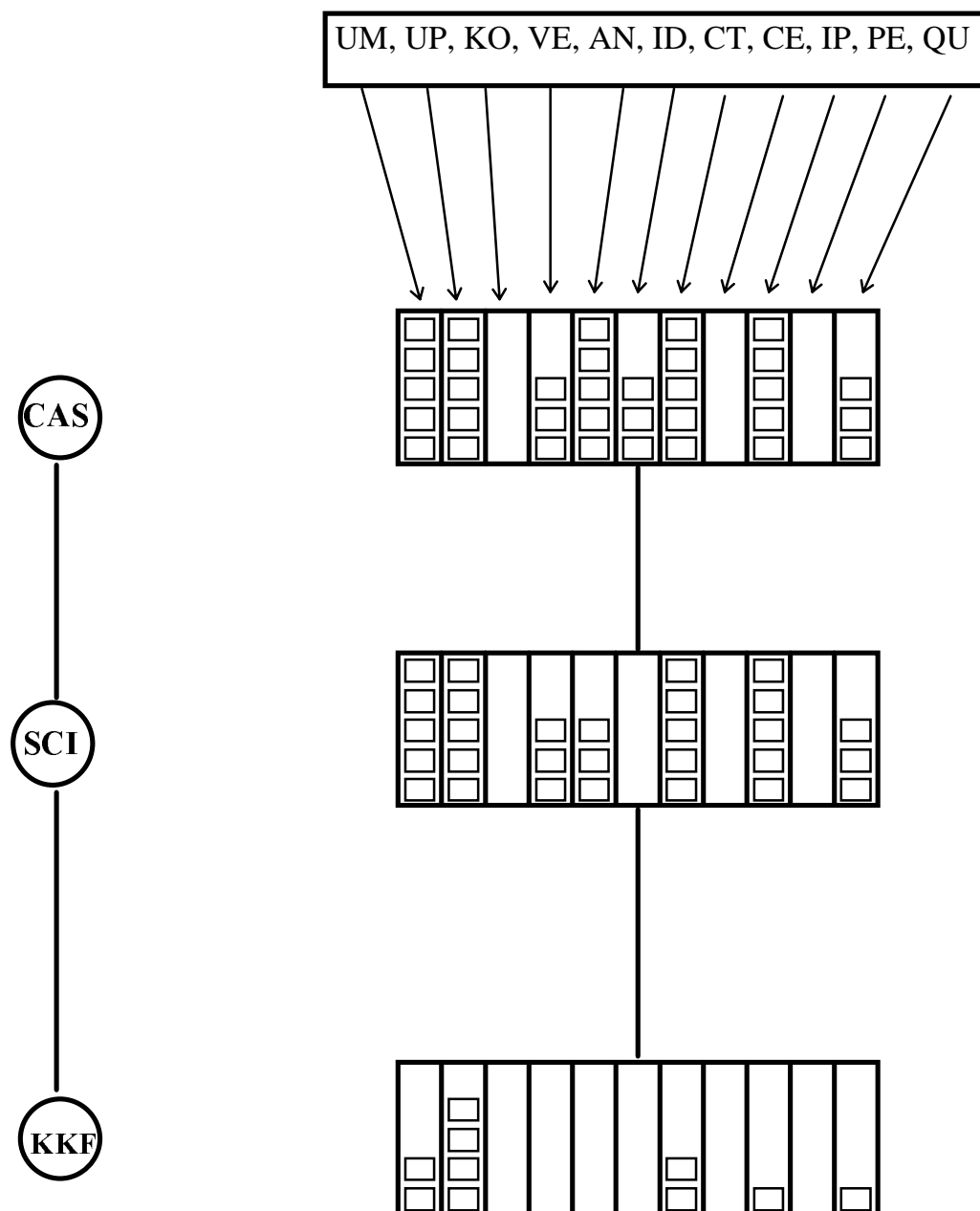
Es gibt bei Auslassung eines Bewertungskriteriums von Null bis 13 Veränderungen im Hasse-Diagramm. Keine Veränderungen gibt es beim Wegfall der Bewertungskriterien UM (Umfang der Datenbank) und CT (Chemikalien-Testdatensatz). Diese beiden Kriterien sind also von keiner Bedeutung in dem vorliegenden Rangfolgeverfahren, wenn als Schlüsselemente die maximalen Datenbanken gewählt werden. Die beiden inhaltlichen Bewertungskriterien PE (Parameter Entwicklung) und AR (Art/ Anwendung der Chemikalien) führen hingegen zu deutlichen Veränderungen im Hasse-Diagramm, sind demnach für das hier durchgeführte Rangfolgeverfahren von großer Bedeutung zumindest was die Beurteilung der Maximalen betrifft. Das Hasse-Diagramm, das bei Auslassung des Kriteriums Art/ Anwendung der Chemikalien entsteht, soll in der folgenden Abbildung 6 nochmals veranschaulicht werden. Die generellen Angaben zum Diagramm sind in Tabelle 38 rechte Spalte zu finden.

Der Wegfall des Bewertungskriteriums Art/ Anwendung der Chemikalien hat einige deutlich sichtbare Veränderungen im Hasse-Diagramm zur Folge. Die Anzahl der Ebenen ist zwar mit drei gleich geblieben und es hat sich auch die Anzahl der Datenbanken in der größten Ebene (NEL) nur geringfügig, d.h. von 18 auf 17 verringert. Jedoch hat sich die Anzahl der Vergleichbarkeiten von 30 auf 53 erhöht. Deutlich wird dieser Umstand u.a. auch dadurch, daß es bei Fall 11 nur noch ein isoliertes Objekt, die Datenbank Chemical Safety Newsbase anstelle von drei isolierten Datenbanken gibt. Auch ist hier keine isolierte Hierarchie mehr vorhanden. Die Datenbank CAS (Chemical Abstracts Service Database) ist jetzt vergleichbar besser eingestuft als SCI (Science Citation Search System) und KKF (Kunststoffe Kautschuk Fasern). Dieser Sachverhalt wird nochmals in der Abbildung 7 deutlich dargestellt. Auf der linken Seite ist das Hasse-Diagramm in der üblichen Kreisdarstellungsweise [Brüggemann 1995a] aufgezeichnet, während auf der rechten Seite das Diagramm der drei Datenbanken in Balkendarstellung aufgezeigt wird.



**Abbildung 6: Hasse-Diagramm Fall 11, Gruppe 1: Online Datenbanken, 12 Bewertungskriterien / - {AR}**

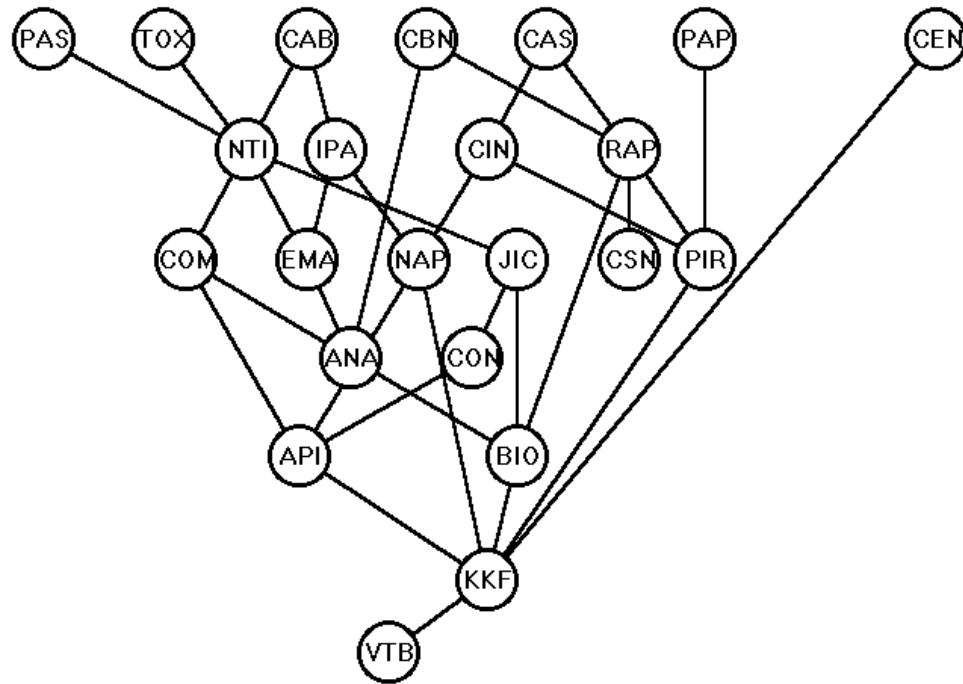
Aus dem Balkendiagramm wird deutlich, daß SCI in allen 11 betrachteten Bewertungskriterien gleich gut oder schlechter eingestuft ist als CAS und KKF wiederum gleiche oder kleinere Bewertungsziffern umfaßt wie bzw. als SCI. Bei dieser betrachteten Kette ist das oben eingeführte Qualitätsgesamtmaß  $\Gamma$ , in dem alle Kriterien gleichgewichtig eingehen, als Bestätigung zum Hasse-Diagramm anzusehen, da CAS den Wert 36, SCI 31 und KKF 13 hat (siehe Tabelle 37). Betrachtet man jedoch andere Datenbanken, z.B. FST und NAP in dem Hasse-Diagramm des Falles 11, so ist NAP ein maximales Objekt, während sich FST in der zweiten Ebene befindet. Beide haben jedoch das Gesamtmaß  $\Gamma = 27$ .



**Abbildung 7:** Ausschnittdiagramme der Kette CAS-SCI-KKF

### 7.2.1.2 Allgemeine Bewertungskriterien

Es werden im folgenden die 28 bibliographischen chemierelevanten Online Datenbanken nur mit den allgemeinen Kriterien UM, UP, KO, VE bewertet.



nicht-triviale Äquivalenzklassen siehe Tabelle 42

**Abbildung 8: Hasse-Diagramm Gruppe 1: Online Datenbanken, 4 allgemeine Bewertungskriterien**

Die dazugehörige Tabelle 42 gibt die Auswertung der Hasse-Diagramme der vollständigen Informationsbasis (Fall 0) sowie des Falls 2 (Auslassung des Bewertungskriteriums UP) bzw. des Falls 3 (Auslassung des Bewertungskriteriums KO), die im folgenden noch erläutert werden.

**Tabelle 42: Auswertung der Hasse-Diagramme zu Gruppe 1: 4 allgemeine Bewertungskriterien**

	Fall 0	Fall 2	Fall 3
Anzahl der Ebenen (NL)	7	8	8
Elemente in der breitesten Ebene (NEL)	7	4	4
maximale Objekte	7: {CAS, SCI}, {CAB}, {CBN}, {CEN}, {PAP}, {PAS}, {TOX}	3: {CAB}, {PAS}, {TOX}	2: {CAS, SCI}, {TOX}
maximale Objekte (ohne isolierte Objekte)	s.o.	s.o.	s.o.
minimale Objekte	2: {CSN}, {VTB}	2: {CSN}, {KKF, VTB}	2: {CSN}, {VTB}
minimale Objekte (ohne isolierte Objekte)	s.o.	s.o.	s.o.
isolierte Objekte	-	-	-
isolierte Hierarchien	-	-	-
Äquivalenzklassen	23	21	16
nicht-triviale Äquivalenzklassen	5: {ANA, CEA}, {CAS, SCI}, {CIN, PRO}, {CON, DIS}, {EMA, FST}	7: {ANA, CEA, CBN}, {CAS, SCI}, {CIN, PRO}, {CON, DIS}, {EMA, FST},	12: {ANA, CEA, EMA, FST, IPA}, {CAS, SCI}, {CAB, COM, NTI}, {CBN, RAP}, {CIN, PRO}, {CON, DIS},

		{KKF,VTB}	{KKF,NAP}, {PAP,PIR}
<b>Vergleichbarkeiten</b> V(N)	158	239	253
<b>Unvergleichbarkeiten</b> U(N)	450	294	288

Abkürzungen siehe Abkürzungsverzeichnis

### 7.2.1.2.1 Auswertung des Hasse-Diagramms

Bei Betrachtung der vollständigen Informationsbasis (IB) treten nicht-triviale Äquivalenzklassen auf, in denen jeweils zwei Objekte die selben Bewertungsziffern für die vier betrachteten allgemeinen Bewertungskriterien aufweisen. Es handelt sich hierbei um {ANA,CEA}, {CAS,SCI}, {CIN,PRO}, {CON,DIS} und {EMA,FST}. Beispielsweise haben beide Datenbanken, CAS und SCI, das Tupel 5,5,0,3 (vgl. hierzu Tabelle 37). Entgegen dem bisher diskutierten Hasse-Diagramm mit einer deutlich größeren Informationsbasis gibt es hier keine isolierten Objekte. Es treten folgende sieben (bzw. acht) maximalen Objekte auf: CAS (Chemical Abstract Service Database) = SCI (Science Citation Search System), CAB (CAB Abstracts), CBN (Chemical Business Newsbase), CEN (Chemical Engineering News), PAP (PAPERCHEM), PAS (PASCAL) und TOX (Toxicology Literature Online). Es gibt nur zwei minimale Objekte, nämlich CSN (Chemical Safety Newsbase) und VTB (Verfahrens-technische Berichte). In dem vorliegenden Fall ist die Zahl der Vergleichbarkeiten deutlich höher als bei dem bisher diskutierten Diagramm. Dementsprechend ist die Anzahl der Unvergleichbarkeiten auch geringer. Das bedeutet, daß es nur wenige Datenbanken gibt, die wenigstens ein Kriterium unter den vier möglichen Bewertungskriterien haben, das sie zu maximalen Objekten werden läßt.

In dem vorliegenden Diagramm ist die Anzahl der Ebenen  $7 \text{ NL} = 7$  (Höhe) und die Anzahl der Objekte in der breitesten Ebene ebenfalls  $7 \text{ NEL} = 7$  (Breite). Hieraus läßt sich ablesen, daß dieses Hasse-Diagramm als "schlank" im Vergleich zu dem bisher diskutierten Diagramm zu nennen ist. Da Breite und Höhe des Diagramms ausgewogen sind, lassen sich gute Schlüsse aus dem Diagramm ziehen. Beispielsweise bilden die Datenbanken CAS (SCI), CIN, NAP, ANA, API, KKF und VTB eine Kette. Dieser Umstand bedeutet, daß z.B. CAS und VTB vergleichbare Objekte darstellen, da ein Pfad von CAS zu VTB besteht. CAS und IPA bilden ein Beispiel für unvergleichbare Objekte. Eine Menge an unvergleichbaren Objekten wird Antikette genannt.

### 7.2.1.2.2 Analyse anhand der D-Matrix

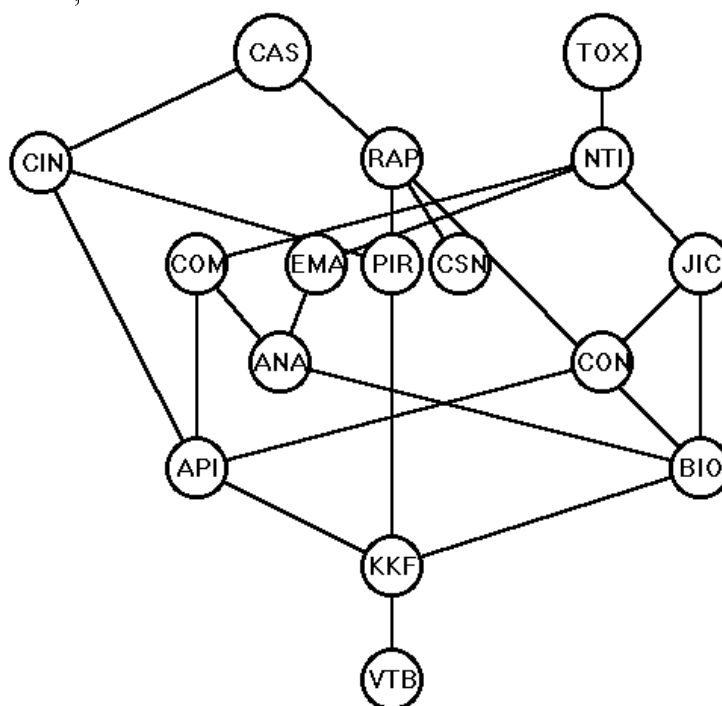
Als Schlüsselemente werden die maximalen Objekte CAS (SCI), CAB, CBN, CEN, PAP, PAS, TOX gewählt. Hier lassen sich die Nachfolger-Datenbanken der einzelnen Maximale der Online Datenbanken und die gemeinsamen Nachfolger-Datenbanken von jeweils zwei Online Datenbanken ablesen. CAB hat 15 Nachfolger-Datenbanken. Die äquivalenten Objekte CAS und SCI haben beispielsweise zehn Datenbanken, die in bezug auf die allgemeinen Bewertungskriterien schlechter eingestuft werden.

**Tabelle 43: D-Matrix der maximalen Online Datenbanken**

	CAS/SCI	CAB	CBN	CEN	PAP	PAS	TOX
CAS/SCI	10	4	6	2	3	4	4
CAB	4	15	5	2	2	13	13
CBN	6	5	8	2	3	5	5
CEN	2	2	2	2	2	2	2
PAP	3	2	3	2	3	2	2
PAS	4	13	5	2	2	13	13
TOX	4	13	5	2	2	13	13

Abkürzungen siehe Abkürzungsverzeichnis, Erläuterungen siehe Text

PAS und TOX haben 13 gemeinsame Nachfolger-Datenbanken. Als Beispiel werden die den beiden maximalen Objekten CAS/SCI und TOX nachfolgenden Datenbanken anhand eines Subgraphen, in welchem also nur die Nachfolger-Datenbanken der beiden genannten Maximalen betrachtet werden, nochmals veranschaulicht.



**Abbildung 9: Subgraph der Datenbanken CAS, TOX der Gruppe 1:  
Online Datenbanken**

Es werden zunächst alle Datenbanken identifiziert, deren Tupel kleiner oder gleich als das Tupel von CAS (5,5,0,3) oder das von TOX (4,4,2,5) sind.

Im Falle von CAS handelt es sich um folgende Datenbanken: RAP (2,5,0,3), CSN (1,5,0,3), CIN/PRO (4,5,0,1), PIR (2,5,0,1), BIO (2,4,0,2), KKF (2,4,0,0), API (3,4,0,1), VTB (2,3,0,0).

Für TOX (4,4,2,5) gibt es folgende 13 Nachfolger-Datenbanken: NTI (4,4,2,3), JIC (4,4,2,2), COM (4,4,1,3), EMA/FST (2,5,0,0), ANA/CEA (2,4,1,3), CON/DIS (4,4,2,1), BIO (2,4,0,2), API (3,4,0,1), KKF (2,4,0,0) und VTB (2,3,0,0).

Betrachtet man z.B. CAS und TOX gemeinsam, so stellt man fest, daß diese beiden Datenbanken vier gemeinsame Nachfolger-Datenbanken aufweisen und zwar BIO (BIOSIS Previews), API (American Petroleum Institute Literature Database), KKF (Kunststoffe Kautschuk Fasern)

und VTB (Verfahrens-technische Berichte). Die Online Datenbank CAS hat die höchste Bewertungsziffer für den Umfang und die Aktualisierung der Datenbank. Die Datenbank verfügt über ca. 12 Millionen Dokumente seit 1967 und wird wöchentlich aktualisiert [STN 1995b]. TOXLINE hingegen liegt mit 1.8 Millionen Dokumenten sowie mit einer monatlichen Aktualisierung unterhalb der Werte von CAS [STN 1995c]. CAS ist derzeit in Subfiles auf CD-ROM verfügbar, wird aber voraussichtlich in naher Zukunft als komplette Version auf CD-ROM vorliegen [CAS 1995a]. TOXLINE hingegen ist nicht nur auf CD-ROM, sondern auch über das Internet auf dem Environmental RouteNet [Cambridge Scientific Abstracts 1996] verfügbar und erhält aus diesem Grunde bei dem Bewertungskriterium VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) die höchste Bewertungsziffer.

### 7.2.1.2.3 Analyse anhand der W-Matrix

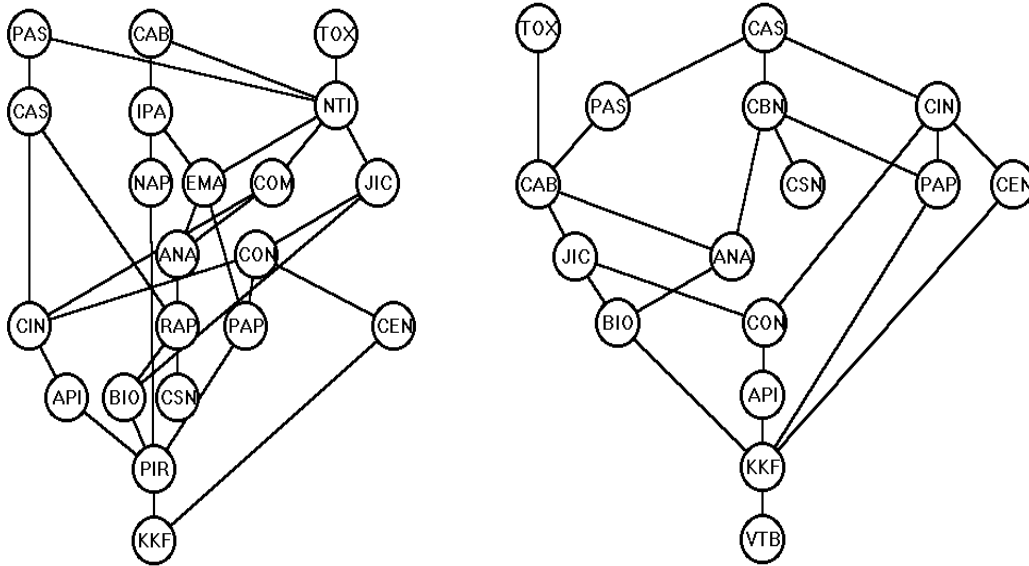
In diesem Fall werden die Einflüsse der vier Bewertungskriterien: UM, UP, KO, VE auf die einzelnen Datenbanken beschrieben. Es ergeben sich daraus  $n+1$  also fünf Fälle.

**Tabelle 44: Auswertung der W-Matrix der Maximalen, 4 allgemeine Bewertungskriterien**

Fall 0 / Fall 1	Kriterium UM ausgelassen	14 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 2	Kriterium UP ausgelassen	26 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 3	Kriterium KO ausgelassen	29 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 4	Kriterium VE ausgelassen	21 Veränderungen im Hasse-Diagramm

Abkürzungen siehe Abkürzungsverzeichnis, Erläuterungen siehe Text

In Tabelle 44 ist die W-Matrix basierend auf den sieben maximalen Objekten CAS, CAB, CBN, CEN, PAP, PAS und TOX dargestellt. Folgende Schlüsse können gezogen werden: Hier gibt es deutlich mehr Veränderungen im Hasse-Diagramm bei der Auslassung eines der allgemeinen Bewertungskriterien als bei dem bisher besprochenen Ansatz. Beim Auslassen der Bewertungskriteriums Kosten der Recherche (KO) gibt es die meisten Veränderungen. Wird das Kriterium Häufigkeit der Aktualisierung (UP) weggelassen, ergeben sich 26 Veränderungen im Diagramm. Diese beiden Bewertungskriterien stellen dementsprechend die wichtigsten allgemeinen Kriterien dar.



nicht-triviale Äquivalenzklassen siehe Tabelle 42

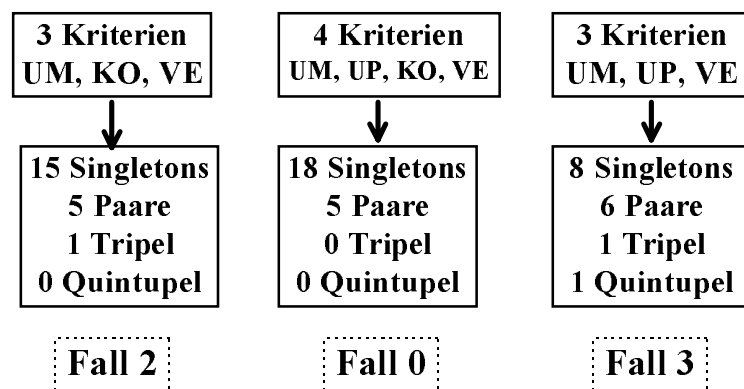
**Fall 2: IB - {UP}**

**Fall 3: IB - {KO}**

**Abbildung 10: Gegenüberstellung von zwei Fällen**

Abkürzungen siehe Abkürzungsverzeichnis, Erläuterungen siehe Text

Vergleicht man die beiden Hasse-Diagramme der Fälle 2 und 3 der Abbildung 10 (jeweils ein Bewertungskriterium ausgelassen) mit dem Ausgangsdiagramm der Abbildung 8 (vollständige Informationsbasis), so ist festzustellen, daß die Anzahl der nicht-trivialen Äquivalenzklassen von fünf auf sieben (Fall 2) bzw. 12 (Fall 3) zugenommen hat (siehe auch Tabelle 42). Es gibt demgemäß mehr äquivalente Objekte. Mit anderen Worten die weggelassenen Kriterien haben zur feineren Klassifizierung geführt, wie in Abbildung 11 deutlich wird.



**Abbildung 11: Zunahme der Tupel bei Auslassung eines Bewertungskriteriums**

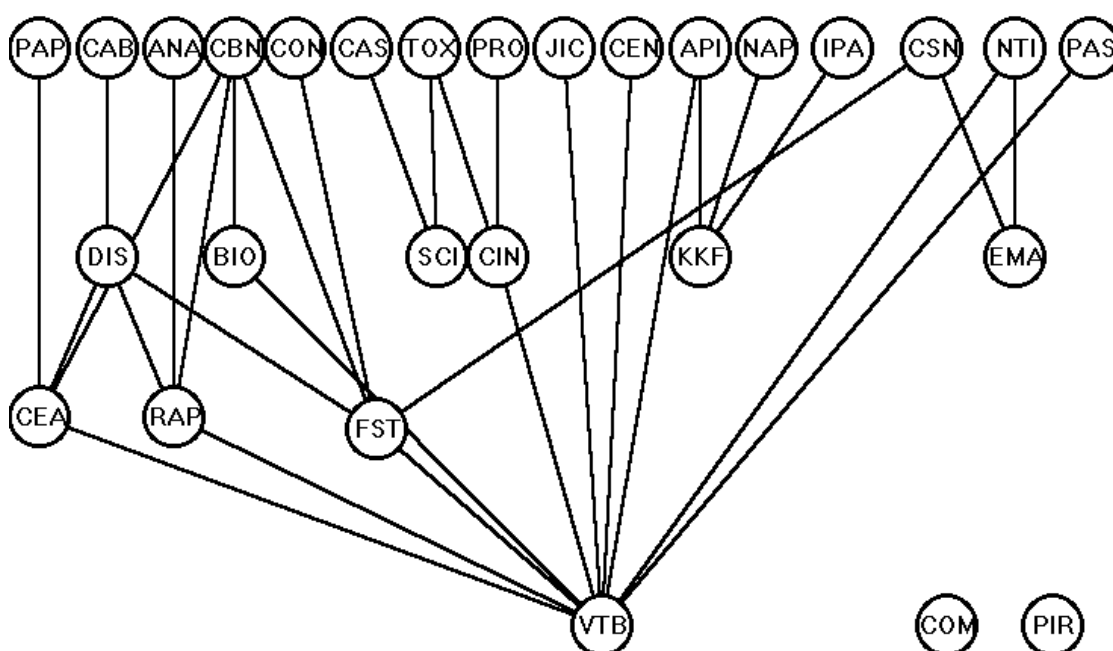
Es ist eine Zunahme der Tupel bei Weglassen eines Bewertungskriteriums festzustellen. Während bei Fall 0 18 Singletons und nur fünf Paare zu verzeichnen sind, erhöht sich die Anzahl der Tupel bei den Fällen 2 und 3.

Die Anzahl der Ebenen hat sich in Abbildung 10 von sieben auf jeweils acht bei den Diagrammen der Fälle 2 und 3 erhöht, die Elemente in der größten Ebene haben sich von sieben im Ausgangsdiagramm auf vier (Fall 2 und 3) verringert. Die Äquivalenzklassen, die die maximalen Objekte darstellen, haben sich von sieben auf drei (Fall 2) bzw. zwei (Fall 3) verringert. Bei den minimalen Objekten haben sich nur Veränderungen bei Fall 2 gegenüber dem Ausgangsdiagramm

gramm ergeben. VTB und KKF bilden eine nicht-triviale Äquivalenzklasse. Ansonsten sind weiterhin VTB und CSN minimale Objekte geblieben. Die Anzahl der Vergleichbarkeiten hat sich von 158 bei voller Informationsbasis auf 239 (Fall 2) bzw. 253 (Fall 3) erhöht, die Anzahl der Unvergleichbarkeiten von 450 (Fall 0) auf 294 (Fall 2) bzw. 288 (Fall 3) erniedrigt. Somit liegt in den beiden Fällen 2 und 3 eine Ausgewogenheit der Vergleichbarkeiten und Unvergleichbarkeiten vor, die die Übersichtlichkeit und die Interpretierbarkeit gegenüber der vollständigen Informationsbasis deutlich erhöht. Beispielsweise werden wieder die beiden Online Datenbanken Chemical Abstract Service Database und Toxicology Literature Online betrachtet. In Fall 2 (Auslassung des Kriteriums Häufigkeit der Aktualisierung) wird CAS mit seinem Tupel (5, 0, 3) schlechter eingestuft als die Online Datenbank PASCAL mit dem Tupel (5, 2, 3). CAS ist kein maximales Element mehr, sondern befindet sich nun in der zweiten Ebene. CAS ist in dem allgemeinen Bewertungskriterium Kosten der Recherche deutlich PAS unterlegen. In Fall 3 hingegen (Auslassung des Bewertungskriteriums Häufigkeit der Aktualisierung) ist CAS ein maximales Objekt wie im Ausgangsdiagramm. Die Datenbank TOX ist in allen beschriebenen Hasse-Diagrammen ein Maximal.

### 7.2.1.3 Inhaltliche Bewertungskriterien

Es werden 28 Datenbanken und die inhaltlichen Bewertungskriterien AN, ID, CT, CE, IP, PE, AR, QU zur Auswertung mittels der Hasse-Diagramm-Technik herangezogen.



**Abbildung 12: Hasse-Diagramm Gruppe 1: Online Datenbanken, 8 inhaltliche Bewertungskriterien**

In der folgenden Tabelle sind die allgemeinen Angaben zu diesem Hasse-Diagramm (Fall 0) sowie zu den weiter unten besprochenen Diagrammen der Fälle 2 und 6 aufgeführt.

**Tabelle 45: Auswertung der Hasse-Diagramme zu Gruppe 1: Online Datenbanken, inhaltliche Bewertungskriterien**

	Fall 0	Fall 2	Fall 6
<b>Anzahl der Ebenen (NL)</b>	4	4	5
<b>Elemente in der breitesten Ebene (NEL)</b>	18	16	11
<b>maximale Objekte</b>	18: {ANA}, {API}, {CAS}, {CAB}, {CBN}, {CEN}, {COM}, {CON}, {CSN}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PIR}, {PRO}, {TOX}	16: {API}, {CAS}, {CAB}, {CBN}, {COM}, {CON}, {CSN}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PIR}, {PRO}, {TOX}	11: {API}, {CAS}, {CEN}, {IPA}, {JIC}, {NAP}, {PAP}, {PAS}, {PIR}, {PRO}, {TOX}
<b>maximale Objekte (ohne isolierte Objekte)</b>	16: {ANA}, {API}, {CAS}, {CAB}, {CBN}, {CEN}, {CON}, {CSN}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PRO}, {TOX}	14: {API}, {CAS}, {CAB}, {CBN}, {CON}, {CSN}, {IPA}, {JIC}, {NAP}, {NTI}, {PAP}, {PAS}, {PRO}, {TOX}	10: {API}, {CAS}, {CEN}, {IPA}, {JIC}, {NAP}, {PAP}, {PAS}, {PRO}, {TOX}
<b>minimale Objekte</b>	6: {COM}, {EMA}, {KKF}, {PIR}, {SCI}, {VTB}	6: {COM}, {EMA}, {KKF}, {PIR}, {SCI}, {VTB}	5: {COM}, {EMA}, {KKF}, {PIR}, {VTB}
<b>minimale Objekte (ohne isolierte Objekte)</b>	4: {EMA}, {KKF}, {SCI}, {VTB}	4: {EMA}, {KKF}, {SCI}, {VTB}	4: {COM}, {EMA}, {KKF}, {VTB}
<b>isolierte Objekte</b>	{COM}, {PIR}	{COM}, {PIR}	{PIR}
<b>isolierte Hierarchien</b>	-	-	-
<b>Äquivalenzklassen</b>	28	28	28
<b>nicht-triviale Äquivalenzklassen</b>	-	-	-
<b>Vergleichbarkeiten V(N)</b>	43	47	83
<b>Unvergleichbarkeiten U(N)</b>	670	662	590

Abkürzungen siehe Abkürzungsverzeichnis

### 7.2.1.3.1 Auswertung des Hasse-Diagramms

Aus dem Hasse-Diagramm und den korrespondierenden allgemeinen Angaben zum Diagramm (Tabelle 45) ist festzustellen, daß es keine nicht-triviale Äquivalenzklassen gibt, d.h. alle Datenbanken haben in bezug auf die acht angewandten inhaltlichen Bewertungskriterien andere Tupel. Es gibt folgende 16 maximalen Objekte: PAP (PAPERCHEM), CAB (CAB Abstracts), ANA (Analytical Abstracts), CBN (Chemical Business Newsbase), CON (Conference Papers Index), CAS (Chemical Abstract Service Database), TOX (Toxicology Literature Online), PRO (PROMT), JIC (JICST File on Science, Technology and Medicine in Japan), CEN (Chemical Engineering News), API (American Petroleum Institute Literature Database), NAP (NAPRALERT), IPA (International Pharmaceutical Abstracts), CSN (Chemical Safety Newsbase), NTI (National Technical Information Service) und PAS (PASCAL). Die beiden Datenbanken COM (Computerized Engineering Index) und PIR (Pira International Publishing and Information Services) sind isolierte Objekte, d.h. sie sind mit keiner der anderen Datenbanken vergleichbar. Sie können sowohl den minimalen und als auch den maximalen Objekten zugeordnet werden. Minimale Objekte, in diesem Fall die Online Datenbanken EMA (Engineered Materials Abstracts), KKF (Kunststoffe Kautschuk Fasern), SCI (Science Citation Search Sy-

stem) und VTB (Verfahrens-technische Berichte) sind vergleichbar schlechter eingestuft als die Datenbanken, mit denen sie nach oben hin im Hasse-Diagramm verbunden sind.

Die Anzahl der Ebenen (NL) ist in diesem Diagramm vier und die Anzahl der Elemente in der größten Ebene (NEL) zehn. Daraus ist abzulesen, daß das Diagramm als sehr "breit" zu bezeichnen ist, d.h. daß die Informationsbasis der inhaltlichen Kriterien umfangreich genug ist, um die "Eigenständigkeit" der Datenbanken zu erfassen.

### 7.2.1.3.2 Analyse anhand der D-Matrix

Auch hier werden wieder alle maximalen Datenbanken (ohne die isolierten Objekte) d.h. ANA, API, CAS, CAB, CBN, CEN, CON, CSN, IPA, JIC, NAP, NTI, PAP, PAS, PRO, und TOX als Schlüsselemente zur Aufstellung der D-Matrix gewählt.

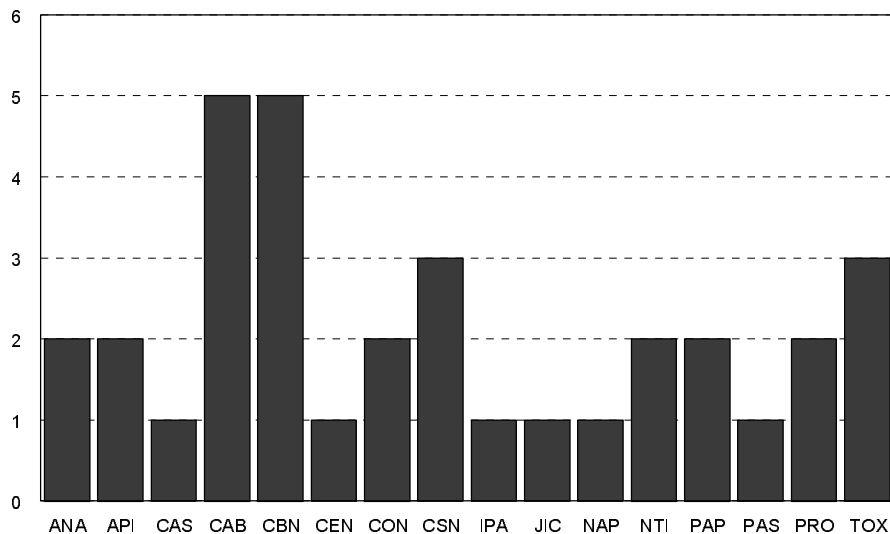
**Tabelle 46: D-Matrix der maximalen Online Datenbanken**

	ANA	API	CAS	CAB	CBN	CEN	CON	CSN	IPA	JIC	NAP	NTI	PAP	PAS	PRO	TOX
ANA	2	1	0	2	2	1	1	1	0	1	0	1	1	1	1	1
API	1	2	0	1	1	1	1	1	1	1	1	1	1	1	1	1
CAS	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
CAB	2	1	0	5	4	1	2	2	0	1	0	1	2	1	1	1
CBN	2	1	0	4	5	1	2	2	0	1	0	1	2	1	1	1
CEN	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1
CON	1	1	0	2	2	1	2	2	0	1	0	1	1	1	1	1
CSN	1	1	0	2	2	1	2	3	0	1	0	2	1	1	1	1
IPA	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0
JIC	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1
NAP	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0
NTI	1	1	0	1	1	1	1	2	0	1	0	2	1	1	1	1
PAP	1	1	0	2	2	1	1	1	0	1	0	1	2	1	1	1
PAS	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1
PRO	1	1	0	1	1	1	1	1	0	1	0	1	1	1	2	2
TOX	1	1	1	1	1	1	1	1	0	1	0	1	1	1	2	3

Abkürzungen siehe Abkürzungsverzeichnis, Erläuterungen siehe Text

Aus der Tabelle läßt sich erkennen, daß es nur äußerst wenig Nachfolger-Datenbanken der einzelnen Datenbanken gibt. Das folgende Histogramm der Abbildung 13 verdeutlicht dieses. So haben beispielsweise die maximalen Objekte CAS, CEN, IPA, JIC, NAP und PAS jeweils nur eine Nachfolger-Datenbank, während die Datenbanken CAB und CBN über jeweils fünf Nachfolger-Datenbanken verfügen. Diese Tatsache läßt den Schluß zu, daß CAB und CBN eine größere Bedeutung bei diesem Rangfolgeansatz zukommt als den Datenbanken, die nur über eine nachfolgende Datenbank verfügen. CAB und CBN haben vier gemeinsame Nachfolger-Datenbanken, nämlich die Datenbanken DIS, CEA, RAP und VTB, d.h. daß CAB und CBN in allen betrachteten inhaltlichen Bewertungskriterien besser eingestuft sind als die vier genannten gemeinsamen Nachfolger-Datenbanken.

Anzahl der Nachfolger-Datenbanken



**Abbildung 13: Histogramm der Nachfolger-Datenbanken der 16 Maximalen**  
Abkürzungen siehe Abkürzungsverzeichnis

### 7.2.1.3.3 Bedeutung der Bewertungskriterien

Hier werden die Einflüsse beschrieben, die die inhaltlichen Bewertungskriterien AN, ID, CT, CE, IP, PE, AR und QU auf das Hasse-Diagramm und damit auch auf die Auswertung haben. Es ergeben sich neun Fälle. In Fall 0 sind alle inhaltlichen Bewertungskriterien vorhanden und in den acht folgenden Fällen ist jeweils ein Bewertungskriterium ausgelassen (siehe auch Kapitel 7.2.1.1.3).

**Tabelle 47: Auswertung der W-Matrix der Maximalen, 8 inhaltliche Bewertungskriterien**

Fall 0 / Fall 1	Kriterium AN ausgelassen	13 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 2	Kriterium ID ausgelassen	4 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 3	Kriterium CT ausgelassen	9 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 4	Kriterium CE ausgelassen	14 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 5	Kriterium IP ausgelassen	5 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 6	Kriterium PE ausgelassen	36 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 7	Kriterium AR ausgelassen	16 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 8	Kriterium QU ausgelassen	6 Veränderungen im Hasse-Diagramm

Abkürzungen siehe Abkürzungsverzeichnis, Erläuterungen siehe Text

In Tabelle 47 ist die Auswertung der W-Matrix der 16 oben aufgeführten maximalen Objekte gegeben. Folgende Schlußfolgerungen können daraus gezogen werden:

Generell ist hier festzustellen, daß bei der Auslassung eines inhaltlichen Bewertungskriteriums es zum Teil zu erheblichen Veränderungen im Hasse-Diagramm kommt. Läßt man beispielsweise das Kriterium PE (Parameter Entwicklung) weg, so kommt es zu 36 Veränderungen im Hasse-Diagramm. Dieses Bewertungskriterium ist somit in diesem Ansatz das Bedeutendste.

Aber auch die Kriterien AR (Art/ Anwendung der Chemikalien) mit 16 Veränderung bei dessen Auslassung, bzw. CE (Chemikalien Entwicklung) mit 14 Veränderungen und AN (Anzahl an Chemikalien) mit 13 Veränderungen bei deren Wegfall sind wichtige Bewertungskriterien in diesem Ansatz. Da es bei der Auslassung von ID (Identifikationsmerkmale) nur zu vier, beim Wegfall von IP (Informationsparameter für Umweltchemikalien) nur zu fünf und bei Wegfall von QU (Validität/Güte der Datenbank) nur zu sechs Veränderungen kommt, können diese Bewertungskriterien in diesem Ansatz als vergleichsweise unbedeutend angesehen werden. Im folgenden soll zur Veranschaulichung das Hasse-Diagramm mit den wenigsten Veränderungen (Fall 2: Auslassung von ID) und das Hasse-Diagramm mit den meisten Veränderungen (Fall 6: Auslassung von PE) nochmals aufgezeigt werden.

Vergleicht man die beiden Hasse-Diagramme der Fälle 2 und 6 der Abbildung 14 (siehe hierzu auch Tabelle 45) mit dem Ausgangsdiagramm der Abbildung 12 (vollständige Informationsbasis), so ist festzustellen, daß es weiterhin keine nicht-triviale Äquivalenzklasse gibt. Das bedeutet, daß sich auch durch die Auslassung eines der beiden diskutierten Kriterien keine gleichen Tupel ergeben. Die Anzahl der Ebenen hat sich bei Fall 6 um eine Ebene von vier auf fünf erhöht. Die Elemente in der breitesten Ebene haben sich in Fall 2 geringfügig, d.h. von 18 bei der vollständigen Informationsbasis auf 16, jedoch deutlich bei Fall 6 auf 11 erniedrigt. In Fall 2 sind noch die beiden isolierten Objekte des Ausgangsdiagramms nämlich COM und PIR vorhanden, während in Fall 6 nur noch die Datenbank PIR (Pira International Publishing and Information Services) ein isoliertes, d.h. ein mit keinem anderen vergleichbares Objekt darstellt. Die Datenbanken EMA, KKF und VTB sind sowohl im Hasse-Diagramm mit der vollen Informationsbasis als auch bei Auslassung des Bewertungskriteriums ID (Identifikationsmerkmale) bzw. der Auslassung von PE (Parameter Entwicklung) minimale Objekte. Die Anzahl der Vergleichbarkeiten hat sich von 43 bei voller Informationsbasis nur geringfügig auf 47 (Fall 2) bzw. erheblich 83 (Fall 6) erhöht, die der Unvergleichbarkeiten von 670 bei voller Informationsbasis auf 662 bei Fall 2 bzw. 590 bei Fall 6 geringfügig verringert. Die Veränderungen, die soeben aufgelistet wurden, können jedoch nicht völlig beliebig den Hasse-Graphen modifizieren:

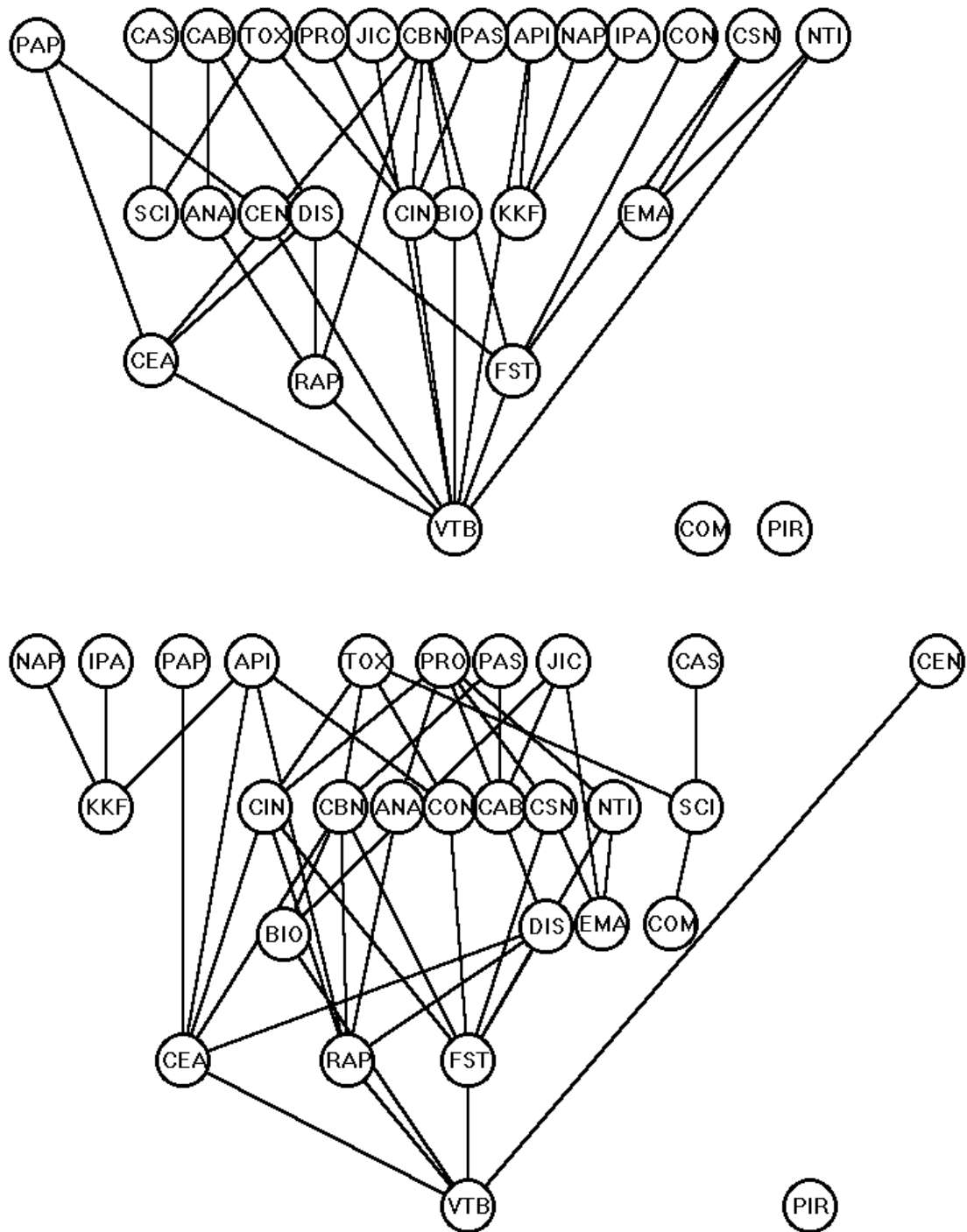
Für Ketten  $K (IB) \subseteq P$  und  $K (IB') \subseteq P$  gilt nämlich:

$$IB \supset IB' \Rightarrow K (IB) \subseteq K (IB')$$

während für Antiketten AK aus

$$IB \supset IB' \Rightarrow AK (IB) \supseteq AK (IB')$$

IB = Informationsbasis, K = Kette, AK = Antikette



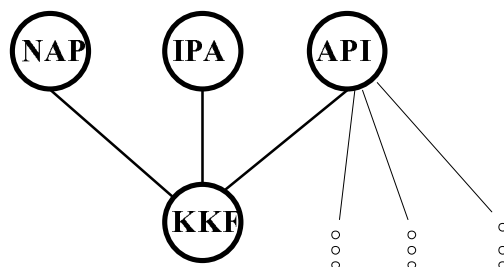
**Abbildung 14: Gegenüberstellung von zwei Fällen / - {ID} (Fall 2: oben) bzw. - {PE} (Fall 6: unten)**

Bei Wegfall eines Bewertungskriteriums (Informationselementes) bleiben die ursprünglich bestehenden Vergleichbarkeiten erhalten, es können jedoch auch neue hinzukommen. Für Antiketten kann entsprechend argumentiert werden.

So gilt z.B.:

$K(\text{IB}) = \{\text{CAB}, \text{DIS}, \text{RAP}, \text{VTB}\}$	IB Fall 0
$K(\text{IB}') = K(\text{IB})$	IB' Fall 2
$K(\text{IB}') = \{\text{PRO}\} \cup K(\text{IB})$	IB' Fall 6 oder
$K(\text{IB}') = \{\text{PAS}\} \cup K(\text{IB})$	IB' Fall 6

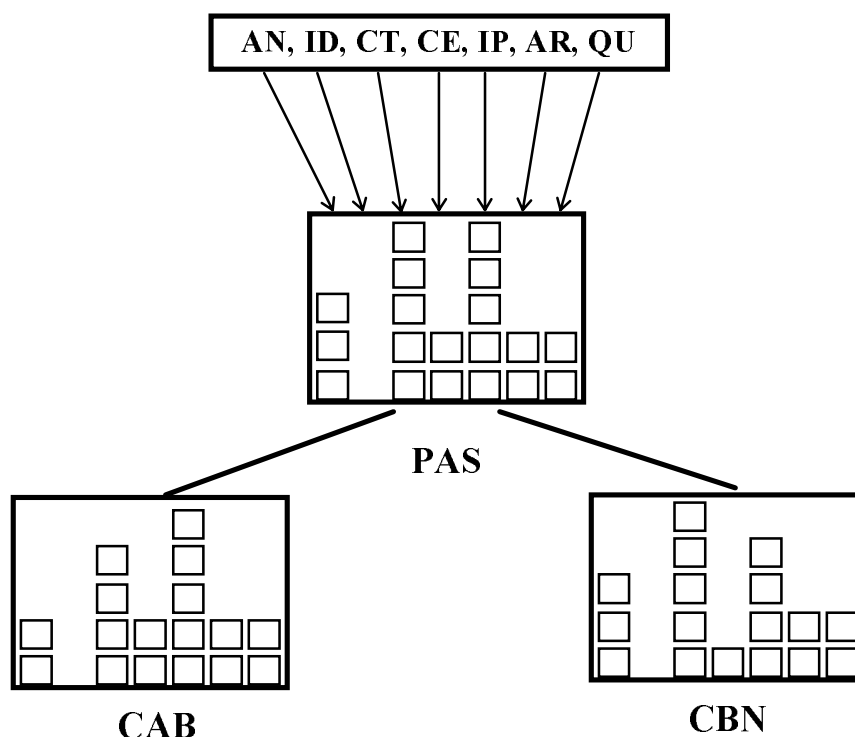
Weiter fällt auf, daß die charakteristische Gruppierung



bestehen bleibt, wenn ID oder PE weggelassen werden. Für diese Gruppierung sind daher andere Kriterien ausschlaggebend und zwar AN und IP. Betrachtet man die Bewertungsziffern in Tabelle 37, so wird deutlich, daß bei dem Bewertungskriterium AN (Anzahl an Chemikalien) NAP sehr hoch eingestuft, API mit 2 bewertet und KKF und IPA mit der Ziffern 0 versehen werden. Für IP (Informationsparameter für Umweltchemikalien) gilt hingegen, daß NAP und KKF nur die Bewertungsziffer 1 erhalten, hingegen API und IPA mit 4 eingestuft werden, d.h. alleine AN und IP bewirken die Struktur dieses Subgraphen.

Zusammenfassend ist festzuhalten, daß die Auslassung des Kriteriums PE (Parameter Entwicklung) zu deutlichen Veränderungen, während die Auslassung des Bewertungskriteriums ID (Identifikationsmerkmale) zu kaum sichtbaren Änderungen im Hasse-Diagramm führt. Wie aus Tabelle 37 ersichtlich, umfaßt die Ausprägungsmenge von ID nur die Bewertungsziffern 0, 1 und 3. Die Bewertungsziffer 3 wird nur einmal, nämlich für die Datenbank CAS vergeben. Die meisten der textbasierten Datenbanken der Chemie-Gruppe verfügen über wenige bis gar keine Chemikalien-bezogenen Merkmale. Bei vielen Datenbanken wird nur die Bewertungsziffer 1 (CAS-Nummer vorhanden) und bei den meisten Datenbanken die Bewertungsziffer 0 vergeben.

An dieser Stelle werden erneut die beiden Online Datenbanken CAB ( CAB Abstracts) und CBN (Chemical Business Newbase) betrachtet. CAB hat bei der Informationsbasis der inhaltlichen Bewertungskriterien  $IB_{inh}$  das Tupel (2, 0, 4, 2, 5, **5**, 2, 2) und CBN (3, 0, 5, 1, 4, **4**, 2, 2). Im Fall 2, d.h. bei der Auslassung des Bewertungskriteriums ID (Identifikationsmerkmale) [kursiv] verändert sich die Position der beiden Objekte im Hasse-Diagramm gegenüber dem Ausgangsdiagramm in keinsten Weise, da beide Objekte dieselbe Bewertungsziffer für ID, nämlich 0 aufweisen. In Fall 6 hingegen, bei dem das Bewertungskriterium PE (Parameter Entwicklung) [fett] ausgelassen wird, verändern sich die Positionen der beiden Datenbanken im Diagramm deutlich. CAB hat mit dem Bewertungsziffer 5 eine sehr große Zunahme bei den Parametern zu verzeichnen; auch bei der Datenbank CBN ist mit dem Bewertungsziffer 4 eine gute Parameter Entwicklung vorzuweisen. Im Fall 6 fällt PE weg und aus diesem Grund verschlechtern sich die Positionen der beiden Datenbanken. Sie sind jetzt keine maximalen Objekte mehr, sondern befinden sich in der zweiten Ebene. Beispielsweise sind jetzt beide Datenbanken Nachfolger der Datenbank PAS (PASCAL), deren vollständiges Tupel der inhaltlichen Bewertungskriterien (3, 0, 5, 2, 5, 1, 2, 2) ist. In dem folgenden Subgraphen wird dieser Sachverhalt für Fall 6 veranschaulicht.



**Abbildung 15:** Teile des Subgraphen von PAS (Erläuterungen siehe Text)

Es sind hier nur sieben inhaltliche Bewertungskriterien dargestellt, nämlich AN, ID, CT, CE, IP, AR, QU. Wie schon oben ausgeführt, ist das Bewertungskriterium Identifikationsmerkmale bei textbasierten Datenbanken sehr schlecht bewertet. Die betrachteten drei Online Datenbanken PAS, CAB und CBN verfügen über keine Identifikationsmerkmale für Chemikalien, nicht einmal über die CAS-Nummer. Auf der anderen Seite fällt auf, daß für die Informationsparameter für Umweltchemikalien sehr hohe Bewertungsziffern vergeben werden, d.h. es gibt bei den betrachteten Online Datenbanken viele umweltrelevante Parameter.

#### 7.2.1.4 Zusammenfassende Analyse der Gruppe 1: Online Datenbanken

Die Gruppe 1 d.h. die textbasierten Datenbanken der Chemie-Gruppe stellt die umfangreichste Gruppe mit über 28 Datenbanken im Vergleich zu den noch zu Analysierenden dar. Auch umfaßt der Bewertungsziffer-Ansatz für diese Gruppe die meisten Bewertungskriterien.

Diese Gruppe soll nun gemäß den Ausführungen in Kapitel 6 vergleichend analysiert werden. Dabei wird sowohl auf die Besonderheiten der Auswertungen der Hasse-Diagramme als auch auf die Gegenüberstellung der Positionen der Objekte im Hasse-Diagramm mit dem Gesamtmaß  $\Gamma$  eingegangen. Ebenso sollen die einzelnen Bewertungskriterien nebst ihrer Ausprägungsmengen näher betrachtet werden.

##### 7.2.1.4.1 Vergleichende Interpretation der Hasse-Diagramme

Der vollständige Satz an Bewertungskriterien umfaßt 12 Kriterien. Die vier allgemeinen und die acht inhaltlichen Bewertungskriterien werden getrennt betrachtet. Wichtige Ergebnisse der einzelnen Rangfolgeverfahren der Gruppe 1 werden nochmals übersichtlich dargestellt.

**Tabelle 48: Zusammenfassung der Rangfolgeverfahren der Gruppe 1:  
Online Datenbanken**

OB/BWK	Art	Abb. Nr.	NL	NEL	MAX	MIN	ÄQ	Iso.E	V(N)	U(N)
28/12	all+inh	6	3	18	18 (15): ANA, API, CAS, CAB, CBN, CEN, CON, IPA, JIC, NAP, NTI, PAP, PAS, PRO, TOX)	7 (4) : EMA, KKF, SCI, VTB)	0	3: COM CSN, PIR	30	696
28/ 4	all	9	7	7	CAS/SCI, CAB, CBN, CEN, PAP, PAS, TOX	2: CSN, VTB	5	0	158	450
28/ 8	inh	13	4	18	18 (16): ANA, API, CAS, CAB, CBN, CEN, CON, CSN, IPA, JIC, NAP, NTI, PAP, PAS, PRO, TOX)	6 (4): EMA, KKF, SCI, VTB	0	2: COM PIR	43	670

Legende:

OB	=	Anzahl der Objekte
BWK	=	Anzahl der Bewertungskriterien
Art	=	Art der Kriterien, allgemein bzw. inhaltlich
Abb. Nr.	=	Nummer der Abbildung
NL	=	Anzahl der Ebenen
NEL	=	Elemente in der breitesten Ebene
ÄQ	=	Anzahl der nicht-trivialen Äquivalenzklassen
MAX	=	Anzahl der maximalen Objekte
MIN	=	Anzahl der minimalen Objekte
Iso E.	=	Anzahl der isolierten Elemente
V(N)	=	Anzahl der Vergleichbarkeiten
U(N)	=	Anzahl der Unvergleichbarkeiten

Bei den Maximalen und Minimalen sind die Zahlen ohne isolierte Objekte in Klammern gesetzt. Am unübersichtlichsten ist das Hasse-Diagramm 28/12, das sowohl die allgemeinen als auch die inhaltlichen Bewertungskriterien berücksichtigt, gefolgt von dem Diagramm mit acht inhaltlichen Bewertungskriterien und demjenigen mit vier allgemeinen.

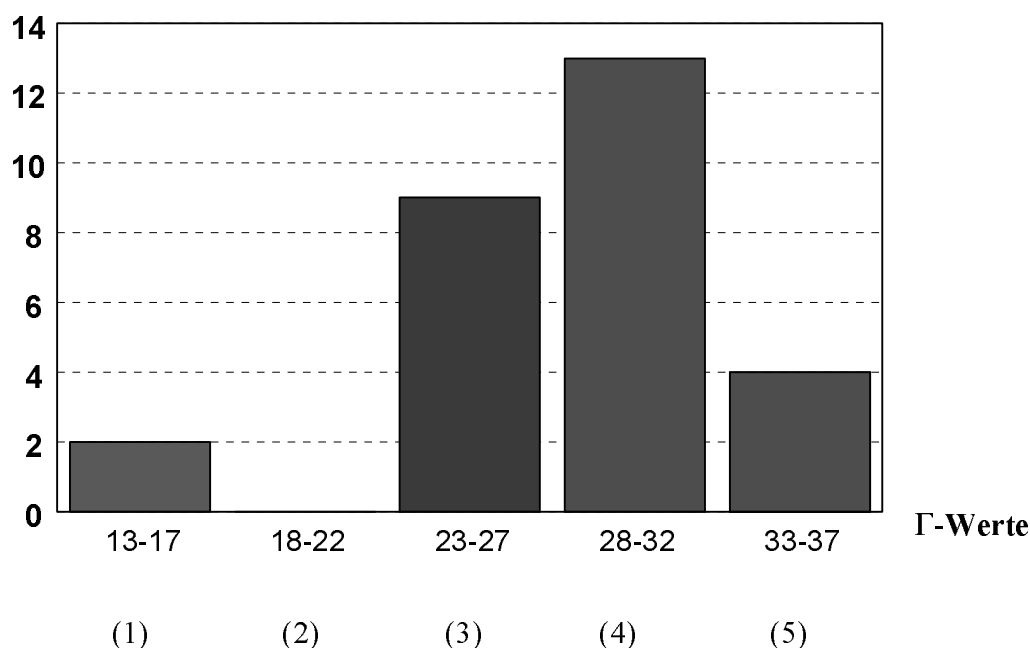
Das am besten interpretierbare Hasse-Diagramm - im Sinne von den Möglichkeiten, Vergleiche durchzuführen - ist jenes mit 28 Datenbanken und vier allgemeinen Bewertungskriterien. Hier ist die Anzahl der Ebenen und die Anzahl der Datenbanken in der breitesten Ebene ausgewogen. Es gibt sieben maximale und nur zwei minimale Objekte. Die Anzahl der Vergleichbarkeiten ist die höchste im Vergleich zu allen anderen diskutierten Diagrammen.

Bei der Betrachtung der Maximalen der drei Gruppen wird deutlich, daß die maximalen Objekte der Gruppe mit allen 12 Bewertungskriterien nicht die Vereinigungsmenge der Maximalen unter den allgemeinen mit denjenigen unter den inhaltlichen Bewertungskriterien ist. Aus diesem Grunde ist nur die Diskussion der einzelnen Informationsbasen, wie sie im vorangegangenen durchgeführt wurde, sinnvoll.

### 7.2.1.4.2 Gegenüberstellung einzelner Datenbanken im Hasse-Diagramm mit dem Gesamtmaß $\Gamma$

Es soll die Position einzelner Objekte im Hasse-Diagramm mit dem Gesamtmaß  $\Gamma$  verglichen werden (vgl. hierzu Tabelle 37 und Abbildung 5). Das Gesamtmaß  $\Gamma$  hat eine Bandbreite von 13-37. Die Länge des Intervalls beträgt 25. Es werden die in Abbildung 16 aufgezeigten fünf Aggregationsklassen gebildet und die Verteilung der  $\Gamma$ -Werte auf die Anzahl der zu den jeweiligen Aggregationsklassen gehörigen Datenbanken dargestellt. Es fällt auf, daß eine Verteilung hin zu den mittleren bis hohen  $\Gamma$ -Werten vorliegt (Aggregationsklassen 3 und 4) und die Aggregationsklasse 2 keine Werte enthält. Bei einem Vergleich der Maximalen mit dem Gesamtmaß  $\Gamma$  fällt auf, daß bei den Datenbanken ANA, CAS, CAB, CBN, CEN, JIC, NTI, PAP, PAS, PRO und TOX eine Übereinstimmung mit hohen  $\Gamma$ -Werten, d.h. Aggregationsklassen (4) und (5) vorliegt. Andere Datenbanken, z.B. API, IPA und NAP sind Maximale, liegen aber in der mittleren Aggregationsklasse (3). Auf der anderen Seite gibt es ein Objekt, nämlich DIS, das ein hohes  $\Gamma$  aufweist und kein Maximal bzw. isoliertes Objekt darstellt. Mit DIS wäre eine Datenbank im Gesamtmaß hoch bewertet, die aber in allen Eigenschaften von anderen noch vorhandenen Datenbanken übertroffen wird. COM und PIR haben hohe  $\Gamma$ -Werte und sind isolierte Objekte.

**Anzahl der Online Datenbanken**



**Abbildung 16: Verteilung der  $\Gamma$ -Werte auf 28 textbasierte Online Datenbanken der Chemie-Gruppe**

Analoges läßt sich für die Minimalen diskutieren. Die Datenbanken KKF und VTB haben ein  $\Gamma < 17$ , d.h. sie sind in der niedrigsten Aggregationsklasse. Gemäß Hasse-Diagramm Auswertung für die Gruppe 1 sind EMA, KKF, SCI und VTB minimale Objekte. EMA hat eine  $\Gamma$  Wert von 25 (Aggregationsklasse 3) und SCI von 31 (Aggregationsklasse 4). Hier liegt somit eine deutliche Abweichung des Gesamtmaßes  $\Gamma$  von der Auswertung mittels der Hasse-Diagramm-Technik vor.

Besonders evident wird der Unterschied zwischen  $\Gamma$  und der Hasse-Diagramm Auswertung bei den Datenbanken API und EMA, die beide über  $\Gamma = 25$  verfügen, also  $\Gamma$ -äquivalent sind. Im Hasse-Diagramm ist jedoch API ein maximales und EMA ein minimales Objekt.

Es ist somit offensichtlich, daß insbesondere die Datenbanken der mittleren Aggregationsklasse 3 deutlich besser durch die Hasse-Diagramm-Technik strukturiert und damit eingestuft sind als mit dem groben Gesamtmaß  $\Gamma$ .

Jedoch bringt auch in anderen Fällen die Hasse-Diagramm-Technik deutlich strukturiertere und genauere Aussagen über die Einstufung der Datenbanken als das Gesamtmaß. CAS und TOX sind  $\Gamma$ -äquivalent. TOX ist sehr gut in den Bewertungskriterien VE, CT und IP eingestuft, während CAS in UM, UP, AN, CE und IP hohe Bewertungsziffern aufweist. In anderen Bewertungskriterien ist CAS schlechter als TOX eingestuft. Diese Tatsache wird durch das Hasse-Diagramm in Balkendarstellung deutlich gemacht.

CAB hat den höchsten Wert für das Gesamtmaß  $\Gamma$ . Diese Datenbank ist aber im Bewertungskriterium ID mit Null eingestuft, also deutlich schlechter als beispielsweise CSN und NAP, bei denen  $\Gamma$  unter 30 liegt. Auch in diesem Fall ist ein Hasse-Diagramm in Balkendarstellung deutlich aussagekräftiger als das Gesamtmaß.

#### 7.2.1.4.3 Bewertungskriterien und deren Ausprägungsmengen

Der Aspekt der Verteilung der Bewertungsziffern auf die Objekte wird im folgenden näher untersucht. Hier wird insbesondere auf die unter Kapitel 6.3 ausgeführten Überlegungen zu den Bewertungskriterien und deren Ausprägungsmengen eingegangen.

Es wird hier die Aufteilung der einzelnen Bewertungskriterien auf die sechs Bewertungsziffern dargestellt. Das Bewertungskriterium AR nimmt eine Sonderstellung ein, da hier generell nur die Ziffern 2, 3, 4 und 5 vergeben werden können (siehe Kapitel 6.4.4.1). Darüber hinaus sind die Mittelwerte, Mediane und Standardabweichungen sowie der Shannon-Index  $I_i$  und der Umfangsindex  $U_i$  verzeichnet.

Am nächsten zu einer symmetrischen Verteilung ist die Verteilung, wenn Mittelwert und Median nur wenig voneinander abweichen. Das ist bei dem Beispiel AN, PE und CE der Fall. Bei IP ist eine Verteilung hin zu den hohen Bewertungsziffern zu beobachten. Eine schlechte Verteilung liegt bei UP, ID, AR und QU vor, bei denen Mittelwerte und Median stark voneinander abweichen, so daß die Betrachtung der Standardabweichung nicht als sinnvoll zu erachten ist. Die beiden eingeführten Indizes, der Shannon-Index und der Umfangsindex dienen ebenfalls zur Beschreibung der Verhältnisse der theoretischen und konkreten Ausprägungsmenge eines Kriteriums, wobei der Umfangsindex nur die Bewertungsziffern, der Shannon-Index hingegen sowohl Bewertungsziffern als auch Anzahl der Objekte berücksichtigt. Aus diesem Grunde ist der Shannon-Index als der aussagekräftigere Index von beiden anzusehen.

**Tabelle 49: Ausprägungsmengen und Indizes der einzelnen Bewertungskriterien der Gruppe 1: Online Datenbanken**

BWK/BWZ	UM	UP	KO	VE	AN	ID	CT	CE	IP	PE	AR	QU
0	0	0	11	4	7	17	0	6	0	8	0	0
1	1	0	4	7	6	10	0	5	2	3	0	4
2	13	0	10	2	6	0	4	2	1	3	24	18
3	2	1	2	14	6	1	7	6	4	5	4	6
4	9	17	1	0	1	0	9	5	12	6	0	0
5	3	10	0	1	2	0	8	4	9	3	0	0
MW	3,00	4,32	1,21	2,07	1,79	0,46	3,75	2,39	3,89	2,25	2,14	2,07
ME	2,5	4	1	3	2	0	4	3	4	2,5	2	2
STDEV	1,186	0,548	1,166	1,303	1,500	0,693	1,041	1,792	1,133	1,818	0,356	0,604
I <sub>i</sub>	1,27	0,79	1,32	1,28	1,65	0,79	1,35	1,74	1,31	1,71	0,41	0,89
U <sub>i</sub>	5	3	5	5	6	3	4	6	5	6	2	3

**Legende:**

BWZ = Bewertungsziffer

STDEV = Standardabweichung MW = Mittelwert

ME = Median,

BWK = Bewertungskriterium I<sub>i</sub> = Shannon-IndexU<sub>i</sub> = Umfangsindex

Für den Umfangsindex liegen Werte von 3-6 vor. Alle sechs möglichen Bewertungsziffern werden nur bei den Bewertungskriterien AN, PE und CE vergeben. Der Wert 3, d.h. nur drei von sechs Ziffern liegen für UP, ID und QU vor. Bei dem Kriterium AR werden zwei von vier aufgeführt. Hier ist die Vergleichbarkeit mit den anderen Bewertungskriterien deutlich eingeschränkt. Mit den Werten des Umfangsindex werden die Ergebnisse der Diskussionen, die anhand des Mittelwertes und Medians angestellt worden sind, untermauert.

Bei der Betrachtung des Shannon-Indexes wird deutlich, daß die Werte für diesen Index differenzierter ausfallen als für den Umfangsindex. Aus den Werten für den Shannon-Index folgt, daß CE das am besten verteilte Bewertungskriterium darstellt, gefolgt von PE und AN. Die am schlechtesten verteilten Bewertungskriterien sind UP und ID mit einem Wert von 0,79, gefolgt von QU. AR wird aus den oben genannten Gründen nicht näher diskutiert. Es soll noch auf die einzelnen Kriterien speziell eingegangen werden. Beispielsweise hat das Kriterium UP (Häufigkeit der Aktualisierung) eine theoretische Ausprägungsmenge von {0,1,2,3,4,5}, in dem diskutierten Ansatz 28/12 jedoch nur {3,4,5}. Das bedeutet, daß sich die Ausprägungsmenge um die Hälfte reduziert. Neben der Ausprägungsmenge ist jedoch auch die Feststellung, wie oft die einzelnen Bewertungsziffern und welche d.h. gute oder schlechte Ziffern vergeben werden, von Bedeutung. Beim Bewertungskriterium UP werden hier nur die hohen, d.h. guten Ziffern vergeben, da textbasierte Online Datenbanken häufig aktualisiert werden. Das allgemeine Bewertungskriterium VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) hat z.B. die Ausprägungsmenge {0,1,2,3,5}, jedoch wird die Bewertungsziffer 5 nur einmal vergeben. Auch bei den inhaltlichen Bewertungskriterien gibt es aktuelle Ausprägungsmengen, deren Umfänge deutlich unter denjenigen der vollständigen Ausprägungsmenge mit den Ziffern 0 bis 5 liegen. Besonders auffällig ist diese Tatsache bei den Identifikationsmerkmalen von Chemikalien, bei denen die Ausprägungsmenge bei {0,1,3} liegt und noch auffälliger bei der Art/Anwendung der Chemikalien. Hier ist die aktuelle Ausprägungsmenge nur {2,3}. Das bedeutet, daß chemiebezogene textbasierte Datenbanken in der Regel nicht über wesentliche Identifikationsmerkmale verfügen und zudem auch in ihren Inhalten sich nicht deutlich auf besondere Arten von Chemikalien spezialisiert haben.

### 7.2.2 Faktendatenbanken der Chemie-Gruppe (Gruppe 2)

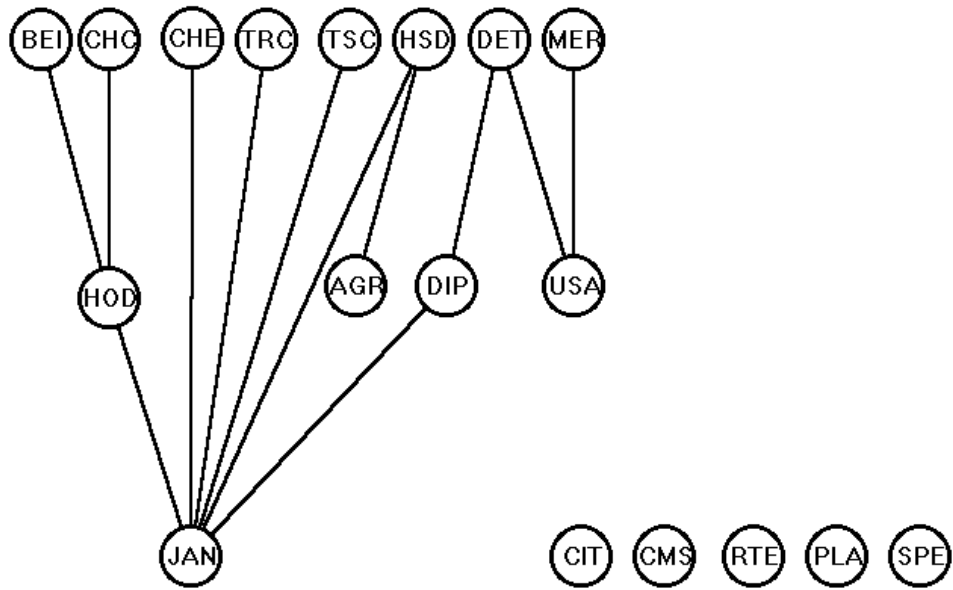
In Tabelle 50 wird der vollständige Datensatz für 18 numerische chemierelevante Online Datenbanken mit der Bezeichnung und dem dreistelligen Akronym der Datenbank aufgelistet. Es gibt im Gegensatz zu den textbasierten Datenbanken bei den Faktendatenbanken keine Parameter Entwicklung. Bei den Faktendatenbanken wird in der Regel an der einmal festgelegten Strukturierung an Datenfeldern nichts mehr verändert. Aus diesem Grund ist keine Aussage über das Hinzukommen neuer Parameter zu treffen. Es liegen hier somit 11 Bewertungskriterien, davon vier allgemeine und sieben inhaltliche vor.

**Tabelle 50: Faktendatenbanken der Chemie-Gruppe und deren Bewertungsziffern**

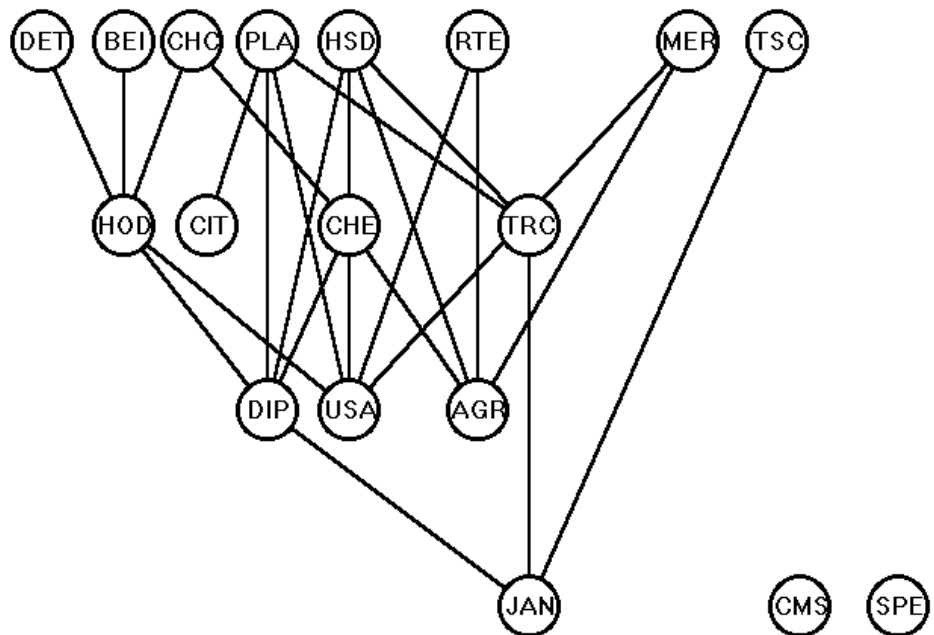
Nr	Datenbank	Akr.	UM	UP	KO	VE	AN	ID	CT	CE	IP	AR	QU	Γ
01	Agrochemicals Hb.	AGR	1	2	0	2	0	2	0	0	2	3	1	13
02	BEILSTEIN	BEI	5	1	3	3	5	5	4	0	3	2	5	36
03	CHCD	CHC	4	2	4	2	4	2	5	1	3	2	4	33
04	Chem-Intell Trade	CIT	0	4	1	1	0	0	1	2	1	2	2	14
05	CHEMSAFE	CHE	3	2	3	2	0	2	4	4	1	5	3	29
06	CHEMSEARCH	CMS	5	4	0	0	5	5	5	0	0	2	5	31
07	DETERM	DET	5	2	3	1	3	2	5	2	2	5	4	34
08	DIPPR	DIP	1	1	2	1	0	2	4	2	1	2	4	20
09	HODOC	HOD	4	1	2	1	4	2	4	0	1	2	4	25
10	HSDB	HSD	2	3	4	3	1	3	4	1	5	4	4	34
11	JANAF	JAN	1	0	2	0	0	2	0	0	0	2	0	7
12	MERCK	MER	3	2	1	3	3	3	3	0	1	3	2	24
13	PLASPEC	PLA	3	4	3	1	3	1	0	0	1	3	0	19
14	RTECS	RTE	5	3	0	3	5	3	5	0	2	2	3	31
15	SPECINFO	SPE	5	3	1	0	5	3	4	1	1	2	4	29
16	TRCTHERMO	TRC	2	3	2	0	2	2	4	1	1	2	4	23
17	TSCA	TSC	4	0	5	2	4	2	5	1	1	2	4	30
18	USAN	USA	2	1	0	1	2	2	0	0	1	3	0	12

Abkürzungen siehe Abkürzungsverzeichnis

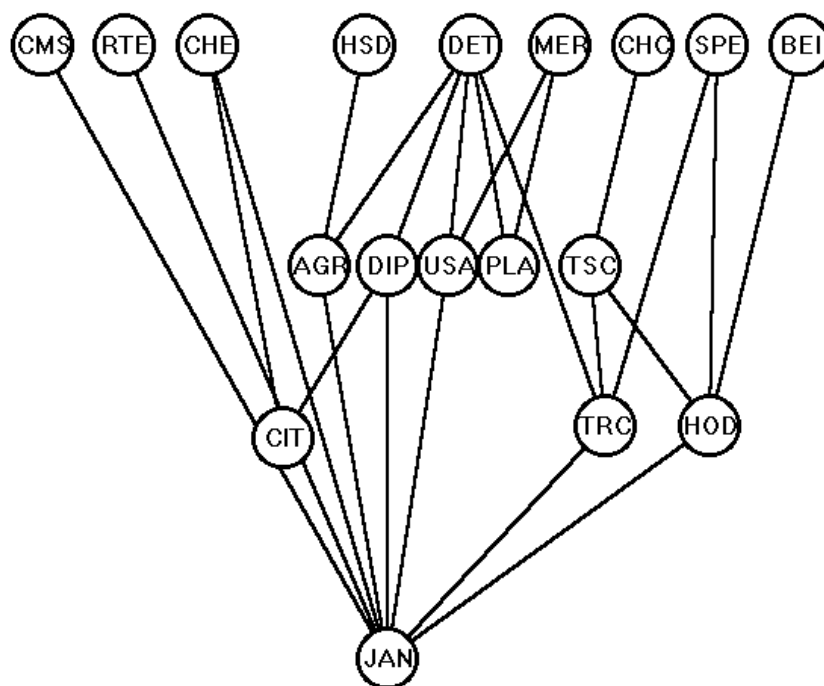
Analog zu der bei den chemierelevanten textbasierten Datenbanken angewandten Vorgehensweise werden drei Ansätze diskutiert. Die 18 Online Datenbanken werden mit den allgemeinen und den inhaltlichen Bewertungskriterien gemeinsam analysiert und dann getrennt mit den vier allgemeinen bzw. sieben inhaltlichen einer Bewertung unterzogen. Die allgemeinen Angaben zu den Hasse-Diagrammen der drei Ansätze sind im Anhang C.2 zu finden. Bei der D-Matrix werden die maximalen Objekte mit der jeweiligen Anzahl an Nachfolger-Datenbanken, bei der W-Matrix werden die Zahlen der Veränderungen, die sich im Hasse-Diagramm bei der Auslassung des aufgelisteten Bewertungskriteriums ergeben, aufgeführt (vgl. hierzu die Kapitel 7.2.1.2.2 und 7.2.1.2.3).



**Abbildung 17: Hasse-Diagramm Gruppe 2: Online Datenbanken, 11 Bewertungskriterien**



**Abbildung 18: Hasse-Diagramm Gruppe 2: Online Datenbanken, 4 allgemeine Bewertungskriterien**



**Abbildung 19: Hasse-Diagramm Gruppe 2: Online Datenbanken 7 inhaltliche Bewertungskriterien**

Aus den Hasse-Diagrammen sowie der Aufstellung im Anhang C.2 lassen sich folgende generelle Schlüsse ziehen:

- Es gibt keine nicht-trivialen Äquivalenzklassen, d.h. keine Datenbank hat dasselbe Tupel wie eine andere.
- Hier liegen keine isolierten Hierarchien vor.
- In zwei Ansätzen sind isolierte Datenbanken vorhanden, bei den inhaltlichen Bewertungskriterien gibt es kein isoliertes Objekt.

Am unübersichtlichsten ist das Hasse-Diagramm mit 11 Bewertungskriterien, d.h. jenes, welches sowohl die allgemeinen als auch die inhaltlichen Bewertungskriterien berücksichtigt. Vergleicht man das Hasse-Diagramm der vier allgemeinen Bewertungskriterien mit dem der sieben inhaltlichen Bewertungskriterien, so ist keine deutliche Unterscheidung herauszuarbeiten.

Das maximale Objekt BEI, die BEILSTEIN Database [STN International 1995d] hat in vier von 11 Bewertungskriterien die beste Bewertungsziffer (vgl. Tabelle 50). Es handelt sich hierbei um den Umfang der Datenbank, die Anzahl an Chemikalien, Identifikationsmerkmale und die Validität/Güte der Datenbank. Bei den allgemeinen Kriterien ist BEI mit Ausnahme der Anzahl der Dokumente, die im Juli 1995 über 6,5 Millionen Substanzen lag [STN International 1995d], mittel bis schlecht eingestuft. Insbesondere fällt diese Tatsache bei der Häufigkeit der Aktualisierung der Datenbank auf, die z.Zt. unter der Frequenz vieler anderer Online Datenbanken liegt. Die Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien ist mit der Bewertungsziffer drei belegt. Denn es gibt sowohl das Beilstein Handbuch der Organischen Chemie als auch einen aktuellen Teil der Datenbank auf CD-ROM [Jochum 1991] und das inhouse System CrossFire [Lawson 1995].

Die BEILSTEIN Database und deren Inhalte insbesondere in bezug auf die angesprochenen inhaltlichen Bewertungskriterien werden in mehreren neueren Publikationen betrachtet [Bucher 1994], [Ernst 1994]. Die Bewertungsziffer Null wurde für die Chemikalien Entwicklung gegeben. Schon bei Einführung der Datenbank im Jahr 1990 [Jochum 1991] war die Anzahl der

Chemikalien extrem hoch, so daß bei dem hier diskutierten Kriterium der Zunahme der umweltrelevanten Chemikalien keine Steigerung mehr möglich war.

Ein anderes maximales Objekt CHC, die Datenbank Chapman and Hall Chemical Database ist unvergleichbar mit BEI bei der Betrachtung aller 11 Bewertungskriterien als Gesamtheit. CHC hat z.B. eine höhere Bewertungsziffer bei der Häufigkeit der Aktualisierung und eine bessere Bewertung bei den Kosten der Recherche. Ebenso hat CHC bei der Chemikalien Entwicklung positiver abgeschnitten als BEI. In den sonstigen Bewertungskriterien insbesondere den inhaltlichen ist CHC gleich oder vergleichbar schlechter als BEI eingestuft.

Die Datenbank DET (DETERM) enthält physikalisch-chemische und thermophysikalische Daten zu Reinstoffen und Gemischen von über 17.000 anorganischen und organischen Chemikalien. In der Datenbank kann mit der CAS-Nummer recherchiert werden [STN 1995g]. Niedrig bewertet ist sie hingegen in UP (Häufigkeit der Aktualisierung), VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) und bei IP (Informationsparameter für Umweltchemikalien).

Die Faktendatenbank HSD (Datenbank Hazardous Substances Databank) verfügt nur über ca. 4.500 Chemikalien, die als gefährliche Stoffe identifiziert worden sind. Die Chemikalien sind mit umfangreichen Identifikationsmerkmalen versehen. So sind CAS-Nummern, RTECS-Nummern, NIOSH-Nummern, EG-Nummern, aber auch Summenformeln und Strukturformeln in der Datenbank vorhanden. Nach den Strukturformeln kann allerdings nicht gesucht werden. Die Datenbank verfügt über umfangreiche Informationsparameter für Umweltchemikalien. Hervorzuheben ist bei HSD, daß die Einträge durch Experten evaluiert werden [DIMDI 1995]. Eine weitere maximale Datenbank MER (The Merck Index Online) verfügt mit über 10.000 Chemikalien über eine beachtliche Anzahl von Stoffen und hält ebenfalls die Identifikationsmerkmale Strukturformel, Summenformel, Molekulargewicht, CAS-Nummer vor [CIS 1995]. Somit wird sie in diesen beiden Kriterien hoch eingestuft.

RTE (Registry of Toxic Effects of Chemical Substances) umfaßt über 130.000 Chemikalien, deren Datensätze jedoch nicht kritisch geprüft werden. Die Faktendatenbank verfügt über umfangreiche Identifikationsmerkmale, z.B. CAS-Nummern, RTECS-Nummern, NIOSH-Nummern, EG-Nummern, aber auch Summenformeln und Strukturformeln. RTE weist nur einige Informationsparameter für Umweltchemikalien auf [DIMDI 1995].

Das minimale Objekt, die Online Datenbank JAN (JANAF), hat in sieben von 11 Bewertungskriterien die Bewertungsziffer Null erhalten. JAN verfügt über wenige organische Chemikalien und nur über Stoffe mit einem oder zwei Kohlenstoff-Atomen. Somit ist diese Datenbank auch schlecht bewertet beispielsweise bei den Kriterien Chemikalien-Testdatensatz, Chemikalien Entwicklung, Informationsparameter für Umweltchemikalien. Obwohl die Datenbank über durch Experten evaluierte Daten verfügt, ist sie im Bewertungskriterium Validität/Güte der Datenbank als ungenügend eingestuft worden [STN 1995e]. Der Grund hierfür ist, daß in dieser Datenbank keine einzige Chemikalie des Chemikalien-Testdatensatzes zu finden ist.

### **7.2.3 Gruppe 1 und Gruppe 2 der Online Datenbanken im Vergleich**

Es gibt einige Aussagen und Schlußfolgerungen, die aus den beiden Ansätzen mit chemierelevanten Datenbanken gezogen werden können. Für diese Betrachtungen sei auf die korrespondierenden Tabellen 37 und 50 verwiesen.

Vorteile bei numerischen Datenbanken sind:

- Die Ausprägungsmenge des Kriteriums Identifikationsmerkmale ist bei numerischen Datenbanken mit {0,2,3,5} deutlich besser als bei textbasierten Datenbanken mit {0,1,3}.
- Dasselbe gilt für das Bewertungskriterium Art/ Anwendung der Chemikalien, das ebenfalls bei numerischen Datenbanken besser in bezug auf die Ausprägungsmenge mit {2,3,4,5} als bei textbasierten Datenbanken mit {2,3} ausfällt.

- Auch ist die Ausprägungsmenge für Validität/Güte der Datenbank bei numerischen Datenbanken mit {0,1,2,3,4,5} deutlich umfangreicher als bei textbasierten Datenbanken mit {1,2,3}.

Demgegenüber stehen die folgenden Vorteile der textbasierten Datenbanken:

- Die Ausprägungsmenge des Bewertungskriteriums Informationsparameter für Umweltchemikalien ist bei textbasierten Datenbanken mit {1,2,3,4,5} gegenüber der Ausprägungsmenge von numerischen Datenbanken mit {0,1,2,3,5} besser, insbesondere bei der Betrachtung der Verteilung der höheren Bewertungsziffern bei den textbasierten Datenbanken. Hier werden die Bewertungsziffern 4 und 5 sehr oft vergeben.
- Auch bei dem Bewertungskriterium Chemikalien-Testdatensatz liegt bei den textbasierten Datenbanken mit {0,1,2,3,4,5} gegenüber {0,1,3,4,5} bei den numerischen Datenbanken eine bessere Verteilung vor. Darüber hinaus verfügen die textbasierten Datenbanken über deutlich mehr höhere d.h. bessere Bewertungsziffern.
- Ähnlich ist der Sachverhalt bei dem Bewertungskriterium Chemikalien Entwicklung. Hier haben die textbasierten Datenbanken auch mit {1,2,3,4,5} gegenüber {0,1,2,4} über eine bessere Ausprägungsmenge.

Fazit aus den Betrachtungen ist, daß Faktendatenbanken (numerische Datenbanken) besser in den Bewertungskriterien Identifikationsmerkmale, Art/ Anwendung der Chemikalien und Validität /Güte der Datenbank,

während textbasierte Datenbanken besser in Informationsparametern für Umweltchemikalien, Chemikalien-Testdatensatz und Chemikalien Entwicklung sind.

### **7.2.4 Textbasierte Datenbanken der Umwelt-Gruppe (Gruppe 3)**

In Tabelle 51 sind 19 textbasierte umweltrelevante Online Datenbanken mit ihren Bewertungsziffern aufgelistet.

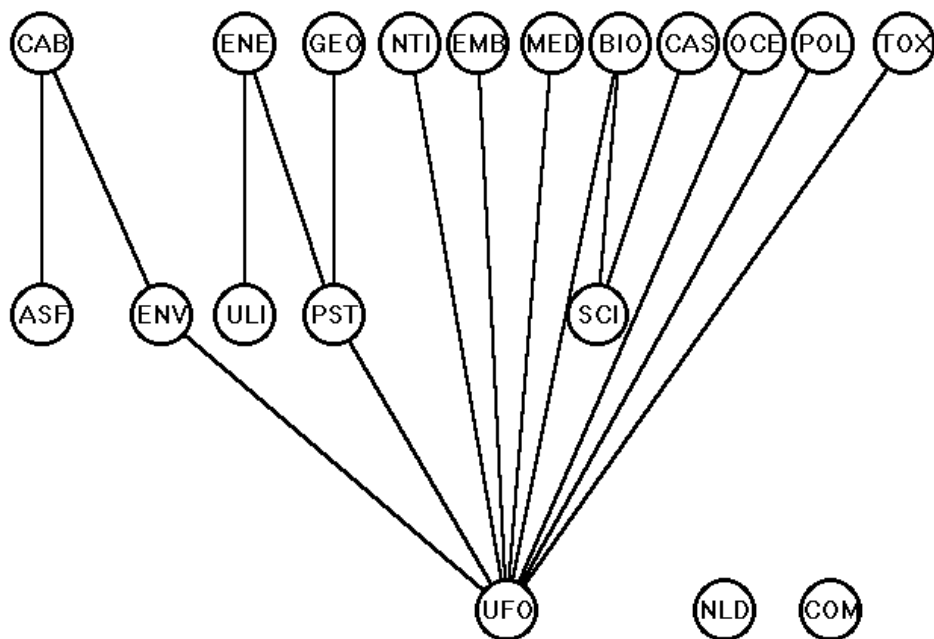
**Tabelle 51: Textbasierte Datenbanken des Umwelt-Gruppe und deren Bewertungsziffern**

Nr.	Name	Akr.	UM	UP	KO	VE	AN	ID	CT	CE	IP	PE	AR	QU	Γ
01	ASFA	ASF	3	4	4	3	1	0	4	0	3	3	2	2	29
02	BIOSIS	BIO	5	5	1	3	3	1	5	1	5	0	2	3	34
03	CA	CAS	5	5	0	3	5	3	5	0	5	0	2	3	36
04	CAB	CAB	4	4	4	3	2	0	4	2	5	5	2	2	37
05	COMPENDEX	COM	4	4	1	3	3	0	5	0	4	1	2	3	30
06	EMBASE	EMB	5	5	1	3	3	1	3	1	5	1	2	3	33
07	ENERGY	ENE	4	4	2	3	2	0	4	5	5	3	2	2	36
08	ENVIROLINE	ENV	2	4	2	3	1	0	3	1	4	4	2	2	28
09	GEOREF	GEO	4	4	1	3	3	0	4	4	3	3	2	1	32
10	MEDLINE	MED	5	5	2	3	2	1	5	1	4	0	2	2	31
11	NLDB	NLD	5	5	0	1	2	0	3	0	4	4	2	3	29
12	NTIS	NTI	4	4	2	3	1	1	3	3	5	2	2	2	32
13	OCEAN	OCE	2	4	1	4	1	0	3	1	3	5	2	2	28
14	POLLUTION	POL	2	4	2	5	2	0	3	5	5	1	2	2	33
15	PSTA	PST	2	4	1	2	0	0	3	3	3	3	2	1	24
16	SCISEARCH	SCI	5	5	0	3	3	0	5	0	5	0	2	3	31

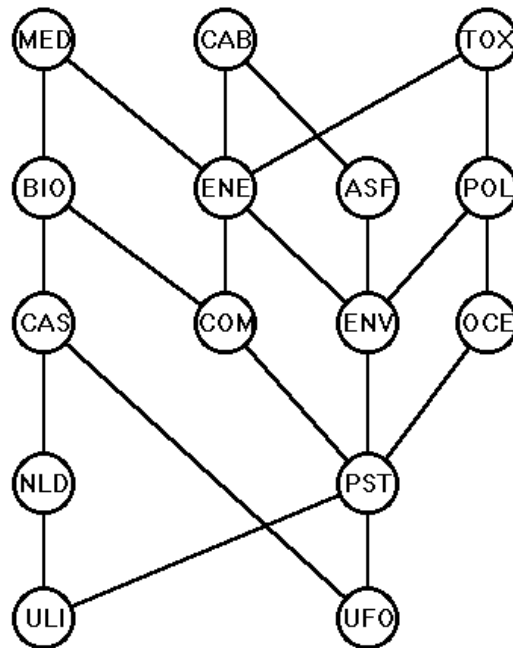
17	TOXLINE	TOX	4	4	2	5	4	1	5	1	5	0	2	3	36
18	UFOR	UFO	1	2	0	2	0	0	1	1	2	0	2	1	12
19	ULIT	ULI	2	4	0	1	1	0	4	4	3	3	2	2	26

Abkürzungen siehe Abkürzungsverzeichnis

Alle 19 aufgeführten Datenbanken verfügen über einen vollständigen Satz an vorhandenen Bewertungskriterien. Hier sind alle verfügbaren, d.h. vier allgemeine und acht inhaltliche Bewertungskriterien komplett anwendbar. Danach wird die Gruppe 3 getrennt mit den vier allgemeinen bzw. acht inhaltlichen einer Bewertung unterzogen. Die allgemeinen Angaben der drei Ansätze sind im Anhang C.3 zu finden.

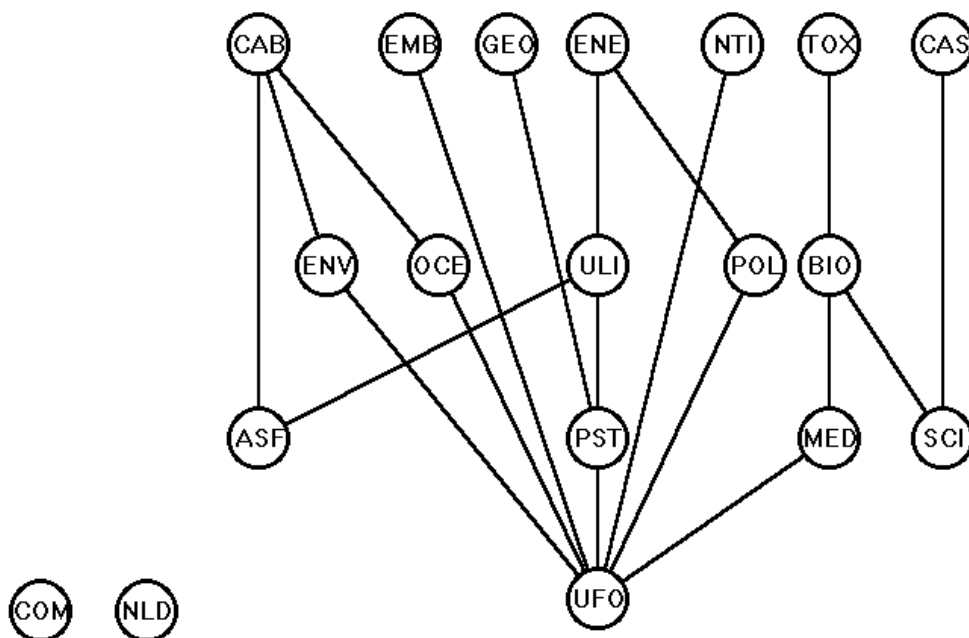


**Abbildung 20: Hasse-Diagramm Gruppe 3: Online Datenbanken, 12 Bewertungskriterien**



nicht-triviale Äquivalenzklassen: {BIO,EMB}; {CAS,SCI}; {COM,GEO}; {ENE,NTI}

**Abbildung 21:** Hasse-Diagramm Gruppe 3: Online Datenbanken 4 allgemeine Bewertungskriterien



**Abbildung 22:** Hasse-Diagramm Gruppe 3: Online Datenbanken, 8 inhaltliche Bewertungskriterien

Aus der Aufstellung der Gruppe 3 (Anhang C.3) und den dazugehörigen Hasse-Diagrammen lassen sich folgende weitere Schlüsse ziehen:

- Es gibt nur bei der Betrachtung der allgemeinen Bewertungskriterien nicht-trivialen Äquivalenzklassen.
- Isolierte nicht-triviale Hierarchien sind nicht vorhanden; isolierte Objekte gibt es jedoch.
- COM und NLD sind isolierte Datenbanken bei der Betrachtung aller 12 Bewertungskriterien und bei den inhaltlichen Kriterien.

Die beiden großen deutschen Umweltdatenbanken "Umweltliteraturdatenbank - ULI" und "Umweltforschungsdatenbank - UFO" werden niedrig bewertet. Der Grund hierfür liegt sowohl in den allgemeinen als auch in den inhaltlichen Kriterien. Bei den allgemeinen Kriterien z.B. den hohen Kosten der Recherche (Bewertungsziffer 0) [Knight-Ridder Information 1995b] als auch bei der Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien (Bewertungsziffer 1) werden die Datenbanken niedrig bewertet [Voigt 1995d]. Die Datenbanken sind zum Zeitpunkt der Auswertung noch nicht auf CD-ROM verfügbar, obwohl seit einiger Zeit angekündigt [Voigt 1995c]. Darüber hinaus verfügen weder ULI noch UFO über eine Unterstützung der Recherche nach Chemikalien, d.h. die Datenbanken haben keine CAS-Nummern. Bei den Kriterien Chemikalien-Testdatensatz und Informationsparameter für Umweltchemikalien verfügt die Datenbank ULI über umfangreiche Informationen [Schultz-Krutisch 1992], [Batschi 1993], [Lohse 1994] und ist dementsprechend hier auch relativ gut eingestuft. Bei UFO läßt sich die schlechte Bewertung auch durch die Art der Datenbank erklären. Während die meisten anderen betrachteten Objekte bibliographische Datenbanken darstellen, handelt es sich bei der Umweltforschungsdatenbank um eine Forschungsdatenbank, die von ihrer Konzeption her deutliche Unterschiede zu bibliographischen Datenbanken aufweist (siehe Anhang A).

### **7.2.5 Umweltrelevante Faktendatenbanken (Gruppe 4)**

Bei den Faktendatenbanken der Umwelt-Gruppe wird das Bewertungskriterium Parameter Entwicklung PE hier generell nicht berücksichtigt. Bei Faktendatenbanken (numerische Datenbanken) werden - wie schon erwähnt - die Datenfelder nur in einigen Ausnahmefällen erweitert bzw. verändert, ansonsten bleiben sie konstant. Die Gruppe 4 umfaßt 14 Online Datenbanken. Die Informationsbasis (IB) der vergleichenden Bewertung beinhaltet folgende 11 Kriterien:

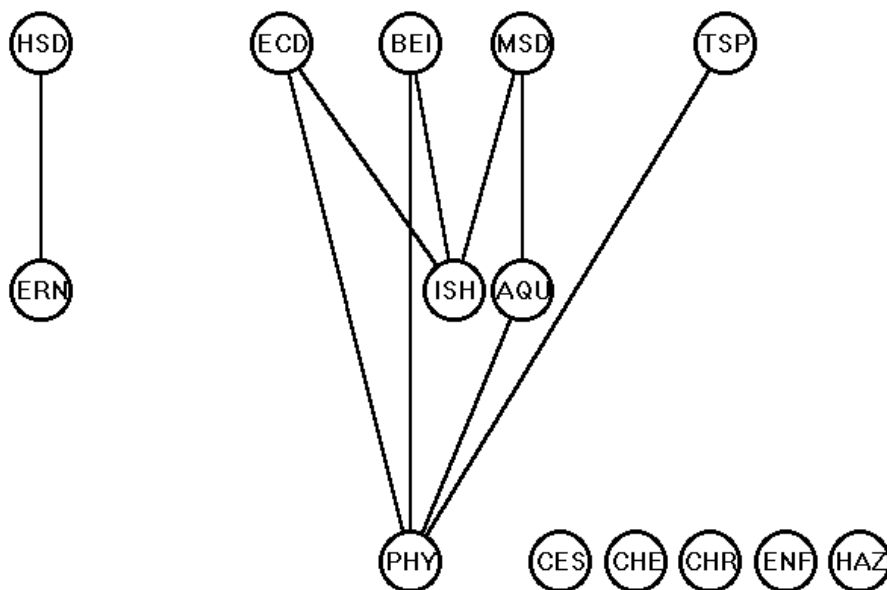
IB: {UM, UP, KO, VE, AN, ID, CT, CE, IP, AR, QU}.

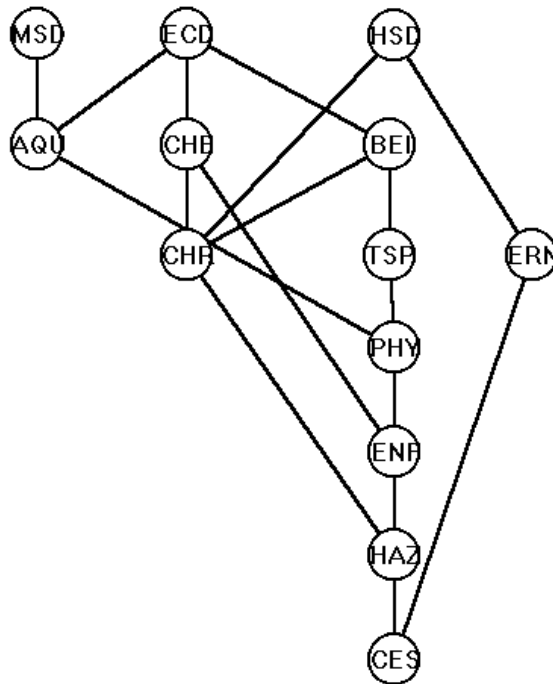
Die 14 Faktendatenbanken der Umwelt-Gruppe sind mit ihren einzelnen Bewertungsziffern und dem Gesamtmaß in der Tabelle 52 zusammengestellt.

**Tabelle 52: Umweltrelevante Faktendatenbanken und deren Bewertungsziffern**

Nr.	Name	Akr.	UM	UP	KO	VE	AN	ID	CT	CE	IP	AR	QU	Γ
01	AQUIRE	AQU	4	2	2	0	2	1	4	0	1	2	4	22
02	BEILSTEIN	BEI	5	1	3	3	5	5	4	0	3	2	5	36
03	CESARS	CES	0	0	0	0	0	2	5	2	4	4	1	18
04	CHEMSAFE	CHE	3	2	3	2	0	2	4	4	1	5	3	29
05	CHRIS	CHR	1	0	2	2	0	2	3	4	2	2	1	19
06	ECDIN	ECD	5	2	5	5	5	3	5	0	4	2	3	39
07	ENVIROFATE	ENF	3	0	2	0	1	2	4	5	4	2	2	25
08	ERNS	ERN	2	3	0	2	1	1	2	0	1	4	1	17
09	HAZINF	HAZ	0	0	1	0	0	1	1	5	1	4	0	13
10	HSDB	HSD	2	3	4	3	1	3	4	1	5	4	4	34
11	ISHOW	ISH	3	0	2	0	3	2	3	0	2	2	3	20
12	MSDS-OHS	MSD	4	5	3	0	4	2	4	1	3	5	4	35
13	PHYTOTOX	PHY	4	0	2	0	0	1	3	0	1	2	1	14
14	TSCAPP	TSP	5	0	3	0	4	1	5	0	2	2	4	26

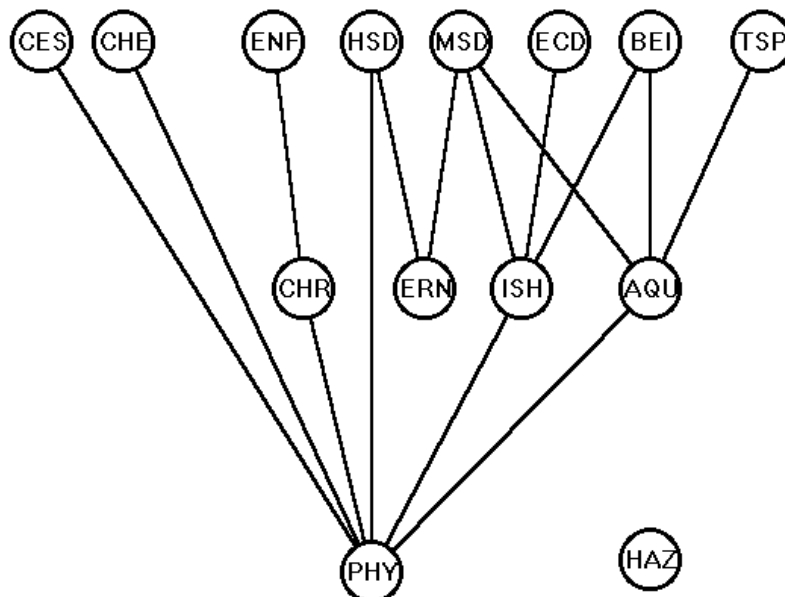
Abkürzungen siehe Abkürzungsverzeichnis

**Abbildung 23: Hasse-Diagramm Gruppe 4: Online Datenbanken, 11 Bewertungskriterien**



nicht-triviale Äquivalenzklasse: {ENF,ISH}

**Abbildung 24:** Hasse-Diagramm Gruppe 4: Online Datenbanken, 4 allgemeine Bewertungskriterien



**Abbildung 25:** Hasse-Diagramm Gruppe 4: Online Datenbanken, 9 inhaltliche Bewertungskriterien

Die allgemeinen Angaben sind im Anhang C.4 zu finden. Am unübersichtlichsten ist das Hasse-Diagramm 14/11, das sowohl die allgemeinen als auch die inhaltlichen Bewertungskriterien berücksichtigt. Vergleicht man das Hasse-Diagramm der allgemeinen (14/4) mit dem der inhaltlichen Bewertungskriterien (14/7), so ist das Diagramm mit den vier allgemeinen Bewertungs-

kriterien besser interpretierbar als dasjenige mit den sieben inhaltlichen Kriterien, da bei den allgemeinen Bewertungskriterien eine gute Ausgewogenheit zwischen Vergleichbarkeiten und Unvergleichbarkeiten vorliegt.

### **7.2.6 Gruppe 3 und Gruppe 4 der Online Datenbanken im Vergleich**

Es können einige Aussagen und Schlußfolgerungen aus den beiden Ansätzen mit umweltrelevanten Datenbanken gezogen werden. Für diese Betrachtungen sei auf die korrespondierenden Tabellen 51 und 52 verwiesen.

Vorteile bei numerischen Datenbanken sind folgende:

- Das Kriterium ID (Identifikationsmerkmale) hat bei den bei numerischen Datenbanken (Gruppe 4) mit der Ausprägungsmenge {1,2,3,5} eine deutlich umfangreichere und somit bessere als bei textbasierten Datenbanken mit der Menge {0,1,3}. Bei letztgenannten wird bei 13 von 19 Datenbanken der Gruppe die Bewertungsziffer Null vergeben. Die Bewertungsziffern 2, 4 und 5 kommen bei textbasierten umweltrelevanten Datenbanken überhaupt nicht vor, d.h. es gibt keine Datenbank in dieser Gruppe, bei der eine Strukturformel suchbar ist. Darüber hinaus verfügt nur die Datenbank Chemical Abstract Service Database über CAS-Nummern und über Strukturformeln (die jedoch nur ausgegeben werden können).
- Das Bewertungskriterium AR (Art/ Anwendung der Chemikalien) ist ebenfalls bei numerischen Datenbanken besser repräsentiert mit den Bewertungsziffern {2,4,5} als bei textbasierten Datenbanken mit der Ziffer 2. Das bedeutet, daß numerische umweltrelevante Datenbanken häufig auf spezielle Anwendungsgebiete ausgerichtet sind.
- Auch ist die Ausprägungsmenge für QU (Validität/Güte der Datenbank) bei numerischen Datenbanken mit {0,1,2,3,4,5} deutlich umfangreicher als bei textbasierten Datenbanken mit {1,2,3}.

Demgegenüber stehen die folgenden Vorteile der textbasierten Datenbanken:

- Diese Art von Datenbanken wird in der Regel häufiger aktualisiert.
- Die Ausprägungsmenge des Bewertungskriteriums IP (Informationsparameter für Umweltchemikalien) ist zwar bei numerischen Datenbanken mit {1,2,3,4,5} umfangreicher als bei textbasierten Datenbanken mit {2,3,4,5}, jedoch wird bei numerischen Datenbanken nur einmal die Ziffer 5 vergeben, bei den textbasierten Datenbanken hingegen siebenmal. Aus diesem Grunde sind die textbasierten Datenbanken in diesem Kriterium stärker als die Faktendatenbanken.
- Ähnlich ist der Sachverhalt bei dem Bewertungskriterium CE (Chemikalien Entwicklung). Hier verfügen die textbasierten Datenbanken ebenfalls mit {0,1,2,3,4,5} gegenüber {0,1,2,4,5} über eine bessere Ausprägungsmenge. Hervorzuheben ist darüber hinaus, daß in ca. 50% der numerischen Datenbanken die Bewertungsziffer 0 vergeben wird, bei textbasierten Datenbanken wird diese schlechte Bewertung nur in ca. 25 % der Datenbanken vergeben.

Fazit aus den Betrachtungen ist, daß Faktendatenbanken (numerische Datenbanken) besser verteilt sind in den Bewertungskriterien ID (Identifikationsmerkmale), AR(Art/ Anwendung der Chemikalien) und QU (Validität/Güte der Datenbank), während textbasierte Datenbanken besser in UP (Häufigkeit der Aktualisierung), IP (Informationsparameter für Umweltchemikalien) und CE (Chemikalien Entwicklung) sind.

### **7.2.7 Umweltchemikalien-relevante Datenbanken der Chemie/Umwelt-Gruppe (Gruppe 5)**

In Tabelle 53 werden die 21 Online Datenbanken, die sowohl zur Chemie- als auch zum Umwelt-Gruppe gehören, mit der Bezeichnung und dem dreistelligen Akronym der Datenbank aufgeführt.

Wie aus der Tabelle ersichtlich, haben drei Datenbanken keinen vollständigen Satz an Bewertungskriterien. Es sind die Datenbanken CAR (CAPREVIEWS), CAP (CAPLUS) und HSD (HSDB). Aus diesem Grund können die Kriterien Chemikalien-Testdatensatz (CT), Chemikalien Entwicklung (CE), Informationsparameter für Umweltchemikalien (IP) und Parameter Entwicklung (PE) nicht analysiert werden.

Analog der Vorgehensweise bei der Gruppe 1 werden auch hier nur diejenigen Datenbanken, in diesem Fall also 18 Datenbanken, die über einen vollständigen Satz an Bewertungskriterien verfügen, bewertet.

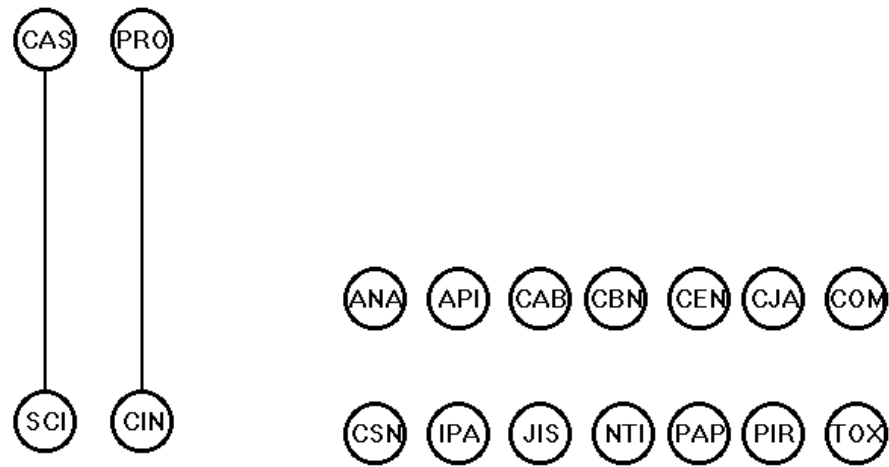
**Tabelle 53: Umweltchemikalien-relevante Datenbanken der Chemie-/Umwelt-Gruppe und deren Bewertungsziffern**

Nr.	Name	Akr.	UM	UP	KO	VE	AN	ID	CT	CE	IP	PE	AR	QU	Γ
01	ANABST	ANA	2	4	1	3	2	1	4	2	3	5	2	2	31
02	APILIT	API	3	4	0	1	2	0	4	1	4	0	3	3	25
03	CA	CAS	5	5	0	3	5	3	5	0	5	0	2	3	36
04	CAB	CAB	4	4	4	3	2	0	4	2	5	5	2	2	37
05	CAPLUS	CAP	5	5	0	0	5	3	5	-	5	-	2	3	-
06	CAPREVIEWS	CAR	2	5	0	0	1	3	-	-	-	-	2	3	-
07	CBNB	CBN	2	5	1	3	3	0	5	1	4	4	2	2	32
08	CEN	CEN	3	5	2	0	0	1	3	5	4	4	2	2	31
09	CIN	CIN	4	5	0	1	1	1	5	1	4	0	2	2	26
10	CJACS	CJA	2	5	4	0	3	1	5	1	5	3	2	3	34
11	COMPENDEX	COM	4	4	1	3	3	0	5	0	4	1	2	3	30
12	CSNB	CSN	1	5	0	3	2	1	3	3	4	4	2	2	29
13	HSDB	HSD	2	3	4	3	1	3	4	1	5	-	4	4	-
14	IPA	IPA	2	4	3	3	0	1	2	0	4	3	3	1	26
15	JIST-E	JIS	4	4	2	2	3	0	4	4	5	0	2	2	32
16	NTIS	NTI	4	4	2	3	1	1	3	3	5	2	2	2	32
17	PAPERCHEM	PAP	2	5	2	1	0	0	4	5	4	4	2	2	31
18	PIRA	PIR	2	5	0	1	0	1	2	5	5	5	2	2	30
19	PROMT	PRO	4	5	0	1	2	1	5	3	5	1	2	2	31
20	SCISEARCH	SCI	5	5	0	3	3	0	5	0	5	0	2	3	31
21	TOXLINE	TOX	4	4	2	5	4	1	5	1	5	0	2	3	36

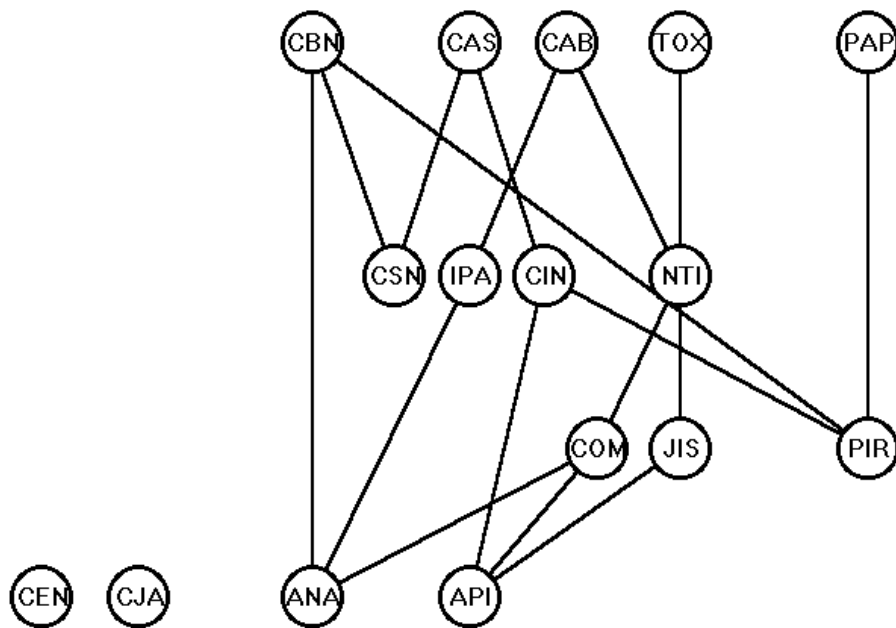
Abkürzungen siehe Abkürzungsverzeichnis

Die 18 ausgewählten Datenbanken werden zunächst wie der mit allen 12 Bewertungskriterien und dann getrennt mit den vier allgemeinen bzw. acht inhaltlichen einer Bewertung unterzogen. Die allgemeinen Angaben der drei Ansätze sind im Anhang C.5 aufgeführt.

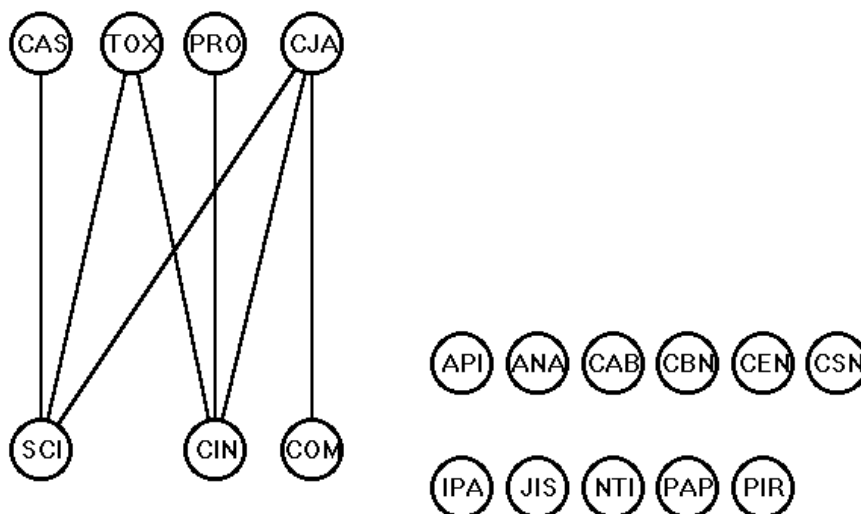
Die Gruppe 5 d.h. die Umweltchemikalien-relevanten Datenbanken, die sowohl der Chemie- als auch der Umwelt-Gruppe angehören, verfügt nur über textbasierte Datenbanken.



**Abbildung 26:** Hasse-Diagramm Gruppe 5: Online Datenbanken, 12 Bewertungskriterien



**Abbildung 27:** Hasse-Diagramm Gruppe 5: Online Datenbanken, 4 allgemeine Bewertungskriterien



**Abbildung 28: Hasse-Diagramm Gruppe 5: Online Datenbanken, 8 inhaltliche Bewertungskriterien**

Am unübersichtlichsten ist das Hasse-Diagramm 18/12, das sowohl die allgemeinen als auch die inhaltlichen Bewertungskriterien berücksichtigt. Das Diagramm verfügt nur über zwei Ebenen und ist mit 16 Datenbanken in der breitesten Ebene als sehr "breit" anzusehen. Es sind bei diesem Ansatz nur zwei Vergleichbarkeiten zu verzeichnen. Darüber hinaus sind 14 von 18 Datenbanken isolierte Objekte. Das Hasse-Diagramm der inhaltlichen Bewertungskriterien verfügt wie das Diagramm mit den 12 Bewertungskriterien über zwei Ebenen, jedoch sind nur sieben Objekte anstelle von 16 in der breitesten Ebene anzutreffen. Ebenso hat sich die Anzahl der isolierten Objekte von 14 auf 11 verringert. Die Anzahl der Vergleichbarkeiten hat zugenommen. Damit ist das Diagramm, welches nur die inhaltlichen Bewertungskriterien berücksichtigt, interpretierbarer und aussagekräftiger als dasjenige mit allen 12 Bewertungskriterien. Das Diagramm der vier allgemeinen Bewertungskriterien ist mit einer Anzahl von vier Ebenen und sieben Objekten in der breitesten Ebene ausgewogener als die beiden oben diskutierten Diagramme. Auch liegen hier nur zwei isolierte Objekte vor, und die Anzahl der Vergleichbarkeiten hat sich auf 41 vergrößert. Aus diesem Diagramm der Gruppe 5 lassen sich somit die meisten Schlüsse ziehen.

### 7.3 Hasse-Diagramme für CD-ROMs

#### 7.3.1 Textbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 1)

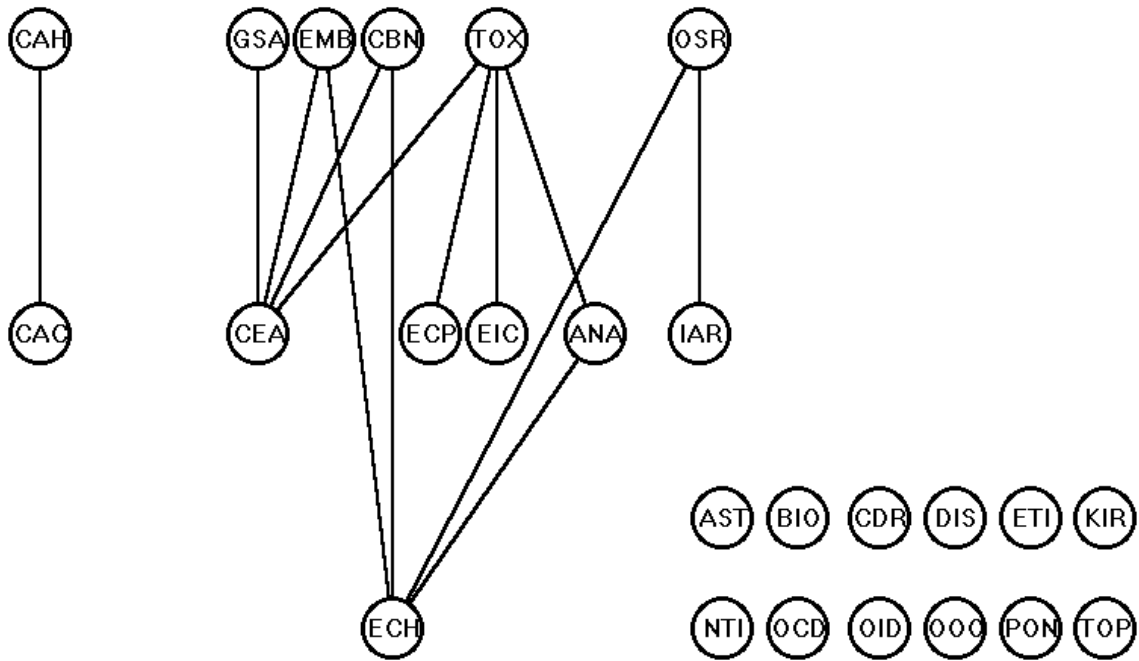
In der folgenden Tabelle 54 werden 26 textbasierte CD-ROMs mit der Bezeichnung der CD-ROM und deren Bewertungsziffern aufgeführt. Darüber hinaus wird das in Kapitel 6.5 eingeführte Gesamtmaß  $\Gamma$  aufgelistet.

**Tabelle 54: Textbasierte CD-ROMs des Chemie-Spezialgebietes und deren Bewertungsziffern**

Nr.	CD-ROM	Akr.	UM	UP	KO	VE	AN	ID	IP	AR	QU	$\Gamma$
01	Analytical Abstracts	ANA	2	3	2	2	2	1	3	2	0	17
02	ASTI	AST	3	4	2	5	0	0	3	2	0	19
03	Biological Abstracts	BIO	4	3	0	1	3	1	5	2	1	20
04	CASurveyor Cancer	CAC	3	4	3	0	4	3	0	2	1	20
05	CASurveyor Food	CAF	3	4	3	0	4	3	0	2	1	20
06	CASurveyor Haz. Mat.	CAH	3	4	3	0	4	3	3	4	1	25
07	CBNB	CBN	2	4	2	1	3	1	4	2	0	19
08	CEABA	CEA	2	3	2	1	0	0	4	2	0	14
09	CD-RÖMPP	CDR	4	1	3	1	3	3	0	2	3	20
10	Dissertation Abstracts	DIS	4	4	0	5	2	0	4	2	0	21
11	Ei CHEM Disc	EIC	4	3	1	3	2	0	4	2	0	19
12	Ei Compendex Plus	ECP	4	3	0	3	3	0	4	2	1	20
13	Ei Tech Index	ETI	1	5	2	3	2	0	5	2	0	20
14	EMBASE-CD:PollTox	EMB	2	3	3	1	1	1	5	2	0	18
15	Environmental Chemistry	ECH	2	3	2	1	0	1	2	2	0	13
16	General Science Abstracts	GSA	2	4	2	5	1	0	5	2	0	21
17	IARCancer Disk	IAR	1	1	2	1	0	0	3	3	0	11
18	Kirk-Othmer	KIR	4	1	0	2	3	1	1	3	2	17
19	NTIS	NTI	3	3	1	5	1	1	5	2	0	21
20	OSH-CD	OCD	2	5	2	0	0	1	3	4	0	17
21	OSH-OFFSHORE bundle	OOO	2	5	1	0	1	1	3	4	0	17
22	OSH-ROM	OSR	5	3	3	1	3	1	3	4	1	24
23	OHS INTERDATA	OID	1	3	4	1	3	2	1	4	0	19
24	PolTox I: NLM, CSA, IFIS	PON	3	3	1	3	3	1	2	2	2	20
25	TOXLINE on SP	TOX	4	3	2	3	3	1	4	2	1	23
26	TOXLINE Plus	TOP	4	3	1	1	4	1	4	2	1	21

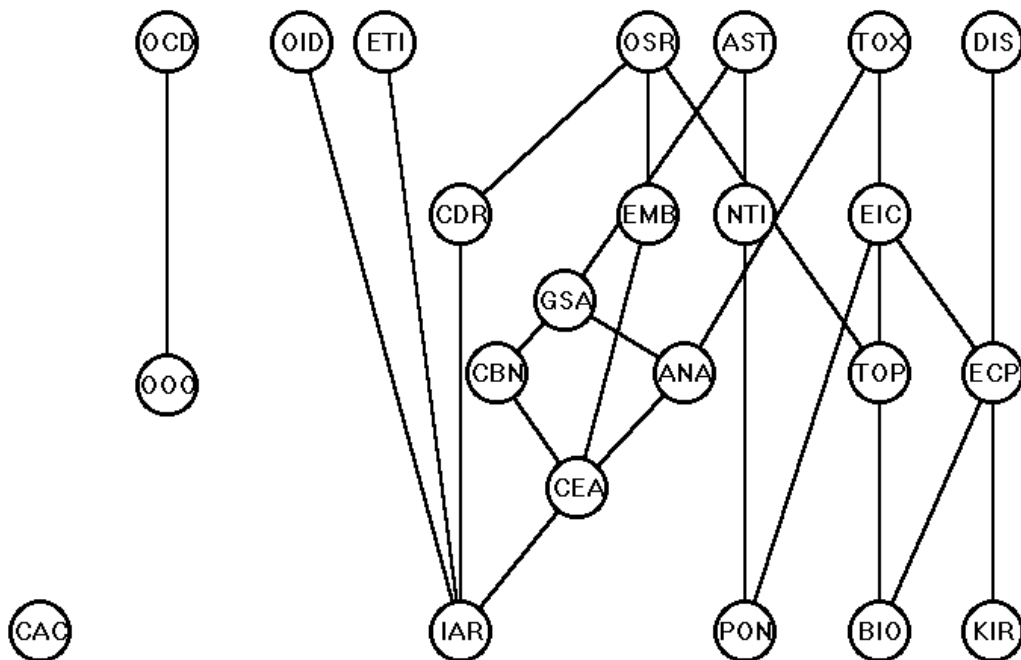
Abkürzungen siehe Abkürzungsverzeichnis

Wie schon bei den Online Datenbanken ausgeführt, werden auch bei den CD-ROMs drei Ansätze diskutiert. Die 26 chemierelevanten textbasierten CD-ROMs werden mit den allgemeinen und inhaltlichen Bewertungskriterien gemeinsam analysiert und dann getrennt mit den vier allgemeinen bzw. mit den fünf inhaltlichen Kriterien einer Bewertung unterzogen. Die tabellarische Auswertung ist im Anhang C.6 aufgeführt. Es treten in allen drei Ansätzen nicht-triviale Äquivalenzklassen und isolierte Objekte auf. Bei der Betrachtung aller neun Bewertungskriterien sowie bei den allgemeinen Bewertungskriterien sind isolierte Hierarchien zu verzeichnen. Hier ist die Anzahl der Ebenen gleich der Anzahl der Elemente in der breitesten Ebene. Am übersichtlichsten ist das Hasse-Diagramm, welches die fünf inhaltlichen Bewertungskriterien berücksichtigt.



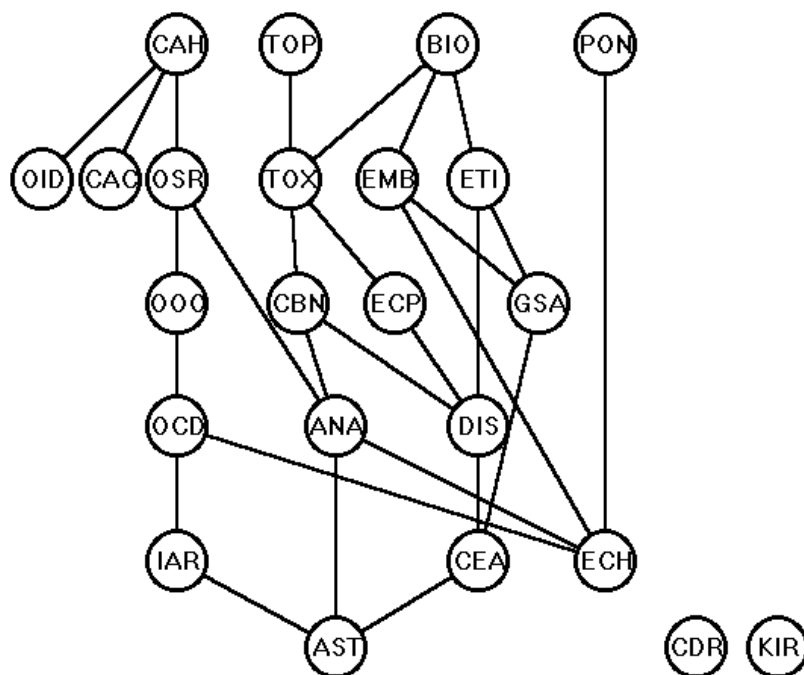
nicht-triviale Äquivalenzklasse: {CAC, CAF}

**Abbildung 29:** Hasse-Diagramm Gruppe 1: CD-ROMs, 9 Bewertungskriterien



nicht-triviale Äquivalenzklasse: {CAC, CAF, CAH}

**Abbildung 30:** Hasse-Diagramm Gruppe 1: CD-ROMs, 4 allgemeine Bewertungskriterien



nicht-triviale Äquivalenzklassen: {CAC,CAF}; {DIS,EIC}

**Abbildung 31:** Hasse-Diagramm Gruppe 1: CD-ROMs, 5 inhaltliche Bewertungskriterien

### 7.3.2 Faktenbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 2)

In der Tabelle 55 die Bewertungsziffern zu neun Bewertungskriterien für 19 faktenbasierte und integrierte CD-ROMs mit der Bezeichnung und Abkürzung der CD-ROM aufgeführt.

**Tabelle 55:** Faktenbasierte und integrierte CD-ROMs des Chemie-Spezialgebietes und deren Bewertungsziffern

Nr	CD-ROM	Akk.	UM	UP	KO	VE	AN	ID	IP	AR	QU	$\Gamma$
01	CANDATA	CAN	5	3	4	1	4	2	1	4	2	26
02	CHEM-BANK	CHB	5	3	1	2	4	3	2	2	2	24
03	ChemReact 10	CR1	5	2	1	1	5	4	0	2	5	25
04	ChemReact 32	CR3	5	2	0	1	5	4	0	2	5	24
05	CHEM Source	CHS	3	3	4	1	3	2	4	4	1	25
06	Current Facts	CUR	5	3	4	0	5	5	1	2	4	29
07	Dictionary of Drugs	DID	4	2	2	1	4	5	2	3	4	27
08	Dictionary of Natural Pro	DIN	3	2	0	1	3	5	2	3	3	22
09	Dictionary of Org. Comp.	DOC	5	2	0	2	5	5	2	2	5	28
10	EINECS-Plus	EIN	5	1	2	2	5	2	0	2	4	23
11	Gefahrgut-CD-ROM	GEF	5	1	0	0	5	2	0	4	3	20
12	Giftliste	GIF	3	2	5	1	3	2	1	4	2	23
13	MSDS	MSD	3	3	4	1	3	2	2	4	1	23
14	NIOSHTIC/DIDS	NIO	5	3	4	1	5	2	2	4	3	29

15	PEST-BANK	PES	4	3	2	1	4	1	0	3	2	20
16	RTECS	RTE	5	3	4	2	5	3	2	2	2	28
17	Sax, Hawley	SAH	4	1	4	1	4	3	1	2	2	22
18	SIGEDA	SIG	2	2	2	1	1	2	2	5	1	18
19	SORBE	SOR	3	1	4	1	3	2	1	4	2	21

Abkürzungen siehe Abkürzungsverzeichnis

### 7.3.2.1 Allgemeine und inhaltliche Bewertungskriterien

Die Interpretation kann dem Hasse-Diagramm und der Tabelle 56, Spalte 2 entnommen werden.

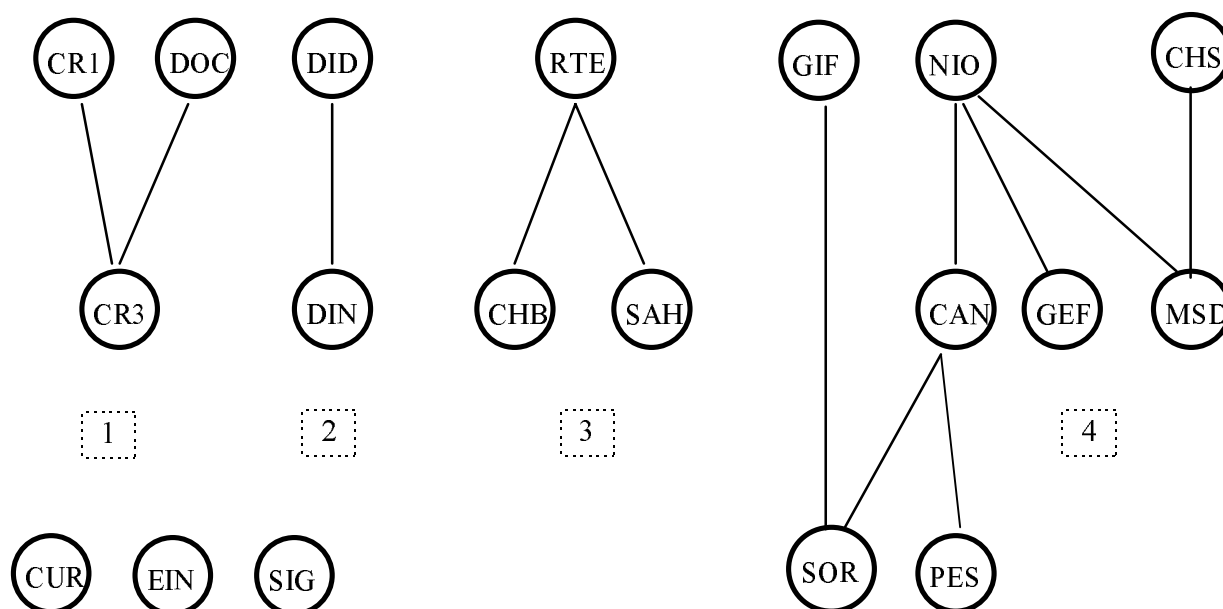
**Tabelle 56:** Auswertung der Hasse-Diagramme zu Gruppe 2: CD-ROMs

	IB <sub>all+inh</sub> Fall 0	IB <sub>all</sub> Fall 0	IB <sub>inh</sub> Fall 0
<b>Anzahl der Bewertungskriterien</b>	9	4	5
<b>Anzahl der Ebenen</b>	3	5	4
<b>Elemente in der breitesten Ebene</b>	10	6	7
<b>maximale Objekte</b>	10: {CR1}, {CHS}, {CUR}, {DID}, {DOC}, {EIN}, {GIF}, {NIO}, {RTE}, {SIG}	2: {GIF}, {RTE}	5: {CHS}, {DID}, {DOC}, {NIO}, {SIG}
<b>maximale Objekte (ohne isolierte Objekte)</b>	7: {CR1}, {CHS}, {DID}, {DOC}, {GIF}, {NIO}, {RTE}	s.o.	4: {CHS}, {DID}, {DOC}, {NIO}
<b>minimale Objekte</b>	11: {CHB}, {CR3}, {CUR}, {DIN}, {EIN}, {GEF}, {MSD}, {PES}, {SAH}, {SIG}, {SOR}	4: {DIN}, {GEF}, {SIG}, {SOR}	7: {DIN}, {EIN}, {GIF}, {SOR}, {MSD}, {PES}, {SAH}, {SIG}
<b>minimale Objekte (ohne isolierte Objekte)</b>	8: {CHB}, {CR3}, {DIN}, {GEF}, {MSD}, {PES}, {SAH}, {SOR}	s.o.	6: {DIN}, {EIN}, {GIF}, {SOR}, {MSD}, {PES}, {SAH}
<b>isolierte Objekte</b>	{CUR}, {EIN}, {SIG}	-	{SIG}
<b>isolierte Hierarchien</b>	3	-	-
<b>Äquivalenzklassen</b>	19	17	17
<b>nicht-triviale Äquivalenzklassen</b>	-	2: {CAN,NIO}, {CHS,MSD}	2: {CR1,CR3}, {GIF,SOR}
<b>Vergleichbarkeiten V(N)</b>	14	75	33
<b>Unvergleichbarkeiten U(N)</b>	314	196	280
<b>D-Matrix der maximalen Objekte (Auswertung)</b>	CR1: 1; CHS: 1; DID: 1; DOC: 1; GIF: 1; NIO: 5; RTE: 2	GIF: 3; RTE: 17	CHS: 1; DID: 4; DOC: 7; NIO: 6
<b>W-Matrix der maximalen Objekte (Auswertung)</b>	UM: 0; UP: 1; KO: 5; VE: 0; AN: 0; ID: 2; IP: 1; AR: 11; QU: 2	UM: 39; UP: 21; KO: 49; VE: 26	AN: 3; ID: 4; IP: 1; AR: 17; QU: 3

Abkürzungen siehe Abkürzungsverzeichnis

Betrachtet wird zunächst das Hasse-Diagramm, welches alle neun Bewertungskriterien gemeinsam darstellt (siehe auch 2. Spalte der Tabelle 56). Wie aus dem Hasse-Diagramm der Abbildung 32 abgelesen werden kann, gibt es viele maximale und minimale Objekte; die Zahl der Unvergleichbarkeiten ist groß während die Anzahl der Vergleichbarkeiten gering ist. Die CD-ROMs ChemReact 10 und ChemReact 32 bilden bei der Betrachtung der inhaltlichen Bewertungskriterien eine Äquivalenzklasse. Beide CD-ROMs unterscheiden sich nur in den Ko-

sten. Die beiden Reaktionsdatenbanken werden auch in einem die Entwicklung dieses Typs von Datenbanken in den vergangenen Jahren zusammenfassenden Artikel beschrieben [Hayward 1995]. "ChemReact 10" umfaßt 10.000 Reaktionen und "ChemReact 32" 32.000 Reaktionen. Da das vorliegende Einstufungsverfahren nicht auf Reaktionen sondern auf Chemikalien und deren Informationsparameter zugeschnitten ist, wird diese Unterscheidung nicht erfaßt. Die CD-ROM CUR (Beilstein Current Facts in Chemistry), auf deren Struktur und Inhalte ausführlich von Warr eingegangen wird [Warr 1993], verfügt über chemische Strukturen, numerische Daten sowie bibliographische Hinweise zu ca. 300.000 Verbindungen. Die CD-ROM enthält jedoch nur Informationen über den aktuellsten 12 monatigen Datenbestand; sie ist nur auf diesem Medium erhältlich. Auch verfügt CUR nur über wenige umweltrelevante Informationsparameter. Aus dem Gesagten folgt, daß diese CD-ROM in vielen Informationsparametern deutlich besser, in anderen jedoch niedriger eingestuft wird als die anderen CD-ROMs. Dies begründet die "Isoliertheit" des Objektes im Hasse-Diagramm mit allen neun Bewertungskriterien.



**Abbildung 32: Hasse-Diagramm Gruppe 2: CD-ROMs, 9 Bewertungskriterien**

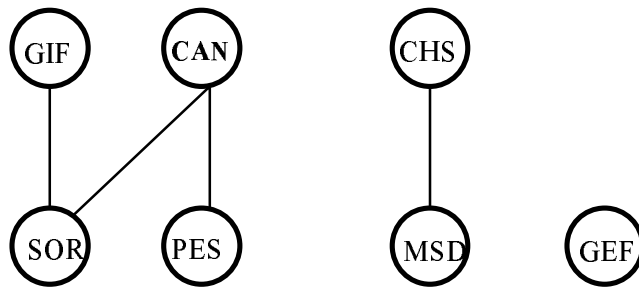
Als strukturelle Besonderheit ist die Tatsache zu vermerken, daß es vier isolierte nicht-triviale Hierarchien gibt. Das heißt, es existieren insgesamt ein "zentrales" Hasse-Diagramm auf der rechten Seite (4) und drei "Nebendiagramme" (1,2,3) auf der linken Seite. Es läßt sich zeigen, daß eine spezielle Datenstruktur hierfür verantwortlich ist. Die Bewertungskriterien ID (Identifikationsmerkmale) und AR (Art/ Anwendung der Chemikalien) führen zum Auftreten dieser Hierarchien. Die CD-ROMs auf der linken Seite des Hasse-Diagramms (1,2,3) verfügen über  $ID = 3$  und  $AR = 2$ , während alle Objekte der rechten Hierarchie (4) Werte für  $ID \leq 2$  und  $AR \geq 3$  aufweisen.

AR, ID sind jedoch nicht das Kriteriumspaar, das die Hierarchie auf der rechten Seite des Diagramms (4) von der Hierarchie, gebildet aus den Objekten DID-DIN (2), trennt.

Die Bewertungskriterien KO, UP trennen zwar DID von CHB und SAH und DIN von SAH, aber nicht DIN von CHB. Unter dem Begriff "trennen" versteht man Ausprägungen, die gewissen Objekten gemein sind, anderen jedoch nicht.

Zwei getrennte Hierarchien können, müssen aber nicht auf Gemeinsamkeiten von Datenstrukturen zurückzuführen sein.

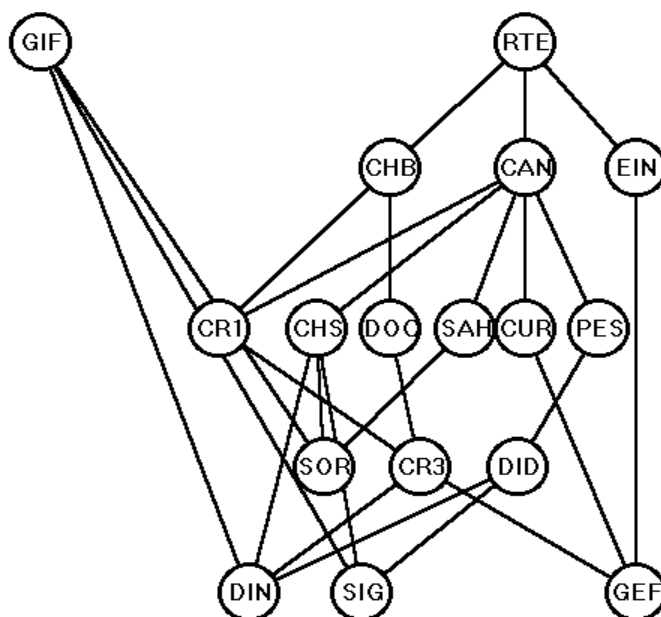
Betrachtet man das Diagramm, welches die meisten Objekte einschließt (rechter Teil der Abbildung 32, also Teildiagramm 4), so fällt auf, daß es Objekte mit einer Sonderstellung gibt. Es handelt sich dabei um Objekte, die sog. Artikulationspunkte bilden. Hier stellt z.B. die CD-ROM NIO einen solchen Artikulationspunkt dar. Beim Weglassen dieses Objektes zerfällt das Hasse-Diagramm erneut in mehrere Teile bestehend aus den CD-ROMs der Abbildung 33.



**Abbildung 33:** Hasse-Diagramm induziert durch Wegfall des Artikulationspunktes NIO

### 7.3.2.2 Allgemeine Bewertungskriterien

Das Hasse-Diagramm, welches vier allgemeine Bewertungskriterien berücksichtigt, zeigt, daß es deutlich übersichtlicher ist als dasjenige mit allen neun Bewertungskriterien. Hier sind alle CD-ROMs in einem zusammenhängenden Diagramm angeordnet. Die Zahl der vergleichbaren CD-ROMs beträgt 75 (siehe Tabelle 56, 3. Spalte).

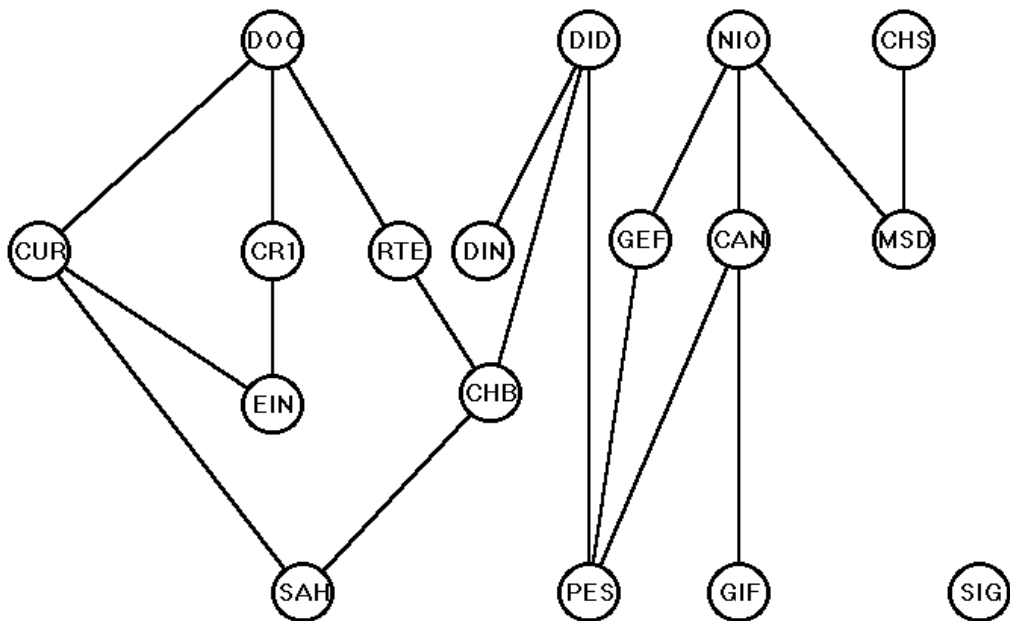


nicht-triviale Äquivalenzklassen: {CAN,NIO}; {CHS,MSD}

**Abbildung 34:** Hasse-Diagramm Gruppe 2: CD-ROMs, 4 allgemeine Bewertungskriterien

### 7.3.2.3 Inhaltliche Bewertungskriterien

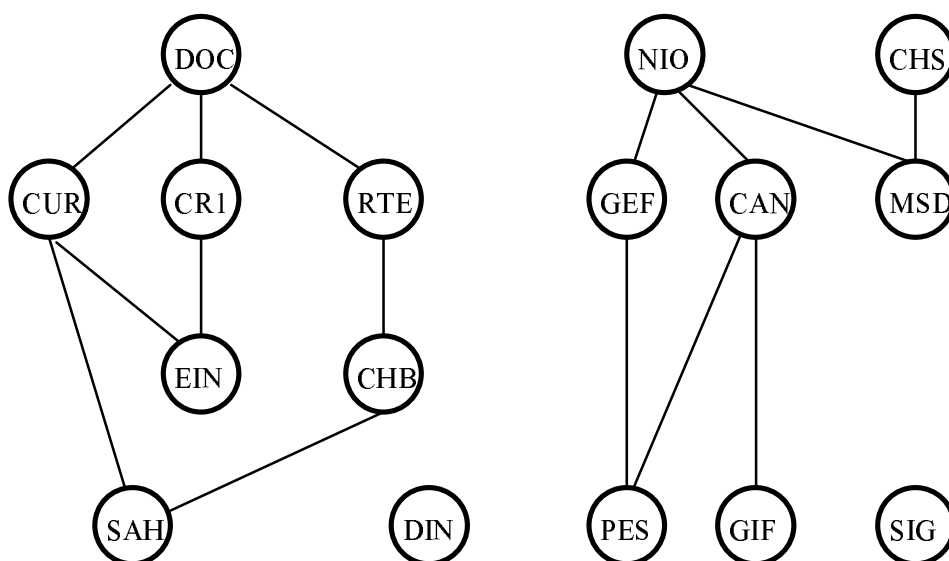
Strukturelle Besonderheiten weist auch das Hasse-Diagramm der fünf inhaltlichen Bewertungskriterien auf (Abbildung 35, Tabelle 56, rechte Spalte). In diesem Hasse-Diagramm gibt es keine isolierten Hierarchien; auch tritt nur ein isoliertes Objekt auf, wie in Abbildung 35 zu erkennen ist.



nicht-triviale Äquivalenzklassen: {CR1, CR3}, {GIF, SOR}

**Abbildung 35:** Hasse-Diagramm Gruppe 2: CD-ROMs, 5 inhaltliche Bewertungskriterien

In diesem Diagramm sind NIO und DID Artikulationspunkte. Beim Wegfall von DID entstehen folgende Teildiagramme.



**Abbildung 36:** Teildiagramme der Gruppe 2 CD-ROMs: inhaltliche Bewertungskriterien

Die Datenstruktur im linken Teildiagramm unterscheidet sich von derjenigen im rechten Teildiagramm. Die Findung dieser speziellen Datenstruktur wird durch das Vorhandensein von Artikulationspunkten, in diesem Beispiel dem Artikulationspunkt DID, durch den das Diagramm in zwei Hierarchien geteilt wird, motiviert. Das Bewertungskriterium AR (Art/ Anwendung der Chemikalien) ist ausschlaggebend für diese Struktur des Diagramms. Im linken Diagramm ist  $AR \leq 2$  und im rechten Teil ist  $AR \geq 3$ . Darüber hinaus hat das Kriterium ID (Identifikationsmerkmale) einen Einfluß auf diese Struktur des Hasse-Diagramms. In dem linken Diagramm hat ID große Werte, während es im rechten Diagramm kleine Werte aufweist.

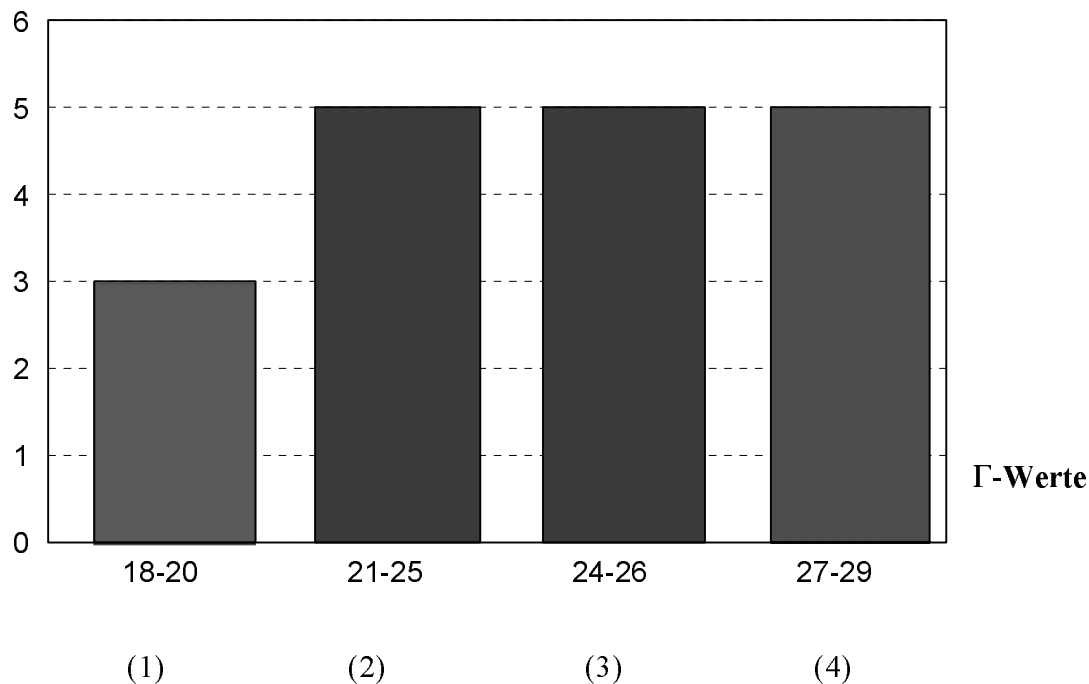
#### 7.3.2.4 Gegenüberstellung einzelner CD-ROMs im Hasse-Diagramm mit dem Gesamtmaß $\Gamma$

Aus Tabelle 55 ist der Intervall für die  $\Gamma$ -Werte abzulesen.

$$18 \leq \Gamma \leq 29.$$

Die Verteilung der  $\Gamma$ -Werte auf die Anzahl der CD-ROMs ist in Abbildung 37 dargestellt.

#### Anzahl der CD-ROMs



**Abbildung 37:** Verteilung der  $\Gamma$ -Werte auf 19 faktenbasierte CD-ROMs des Chemie-Spezialgebietes

Da das Intervall bei dieser Gruppe nur 12 beträgt, werden hier nur vier gleichgroße Aggregationsklassen gebildet.

Bei dieser Gruppe gibt es  $\Gamma$ -äquivalente Objekte, z.B. CUR und NIO oder DOC und RTE. In diesen Beispielen zeigt die Hasse-Diagramm-Technik deutliche Unterschiede der beiden CD-ROMs CUR und NIO mit gleichen  $\Gamma$ -Werten. So ist CUR beispielsweise in den Bewertungskriterien ID und QU höher eingestuft als NIO. Auf der anderen Seite ist NIO besser bewertet als CUR in den Bewertungskriterien VE, IP und AR.

Bei der Gegenüberstellung der höchsten Aggregationsklasse (4), die die CD-ROMs CUR, DID, DOC, NIO und RTE umfaßt, mit den maximalen Objekten CR1, CHS, DID, DOC, GIF, NIO und RTE liegt auch wieder eine gewisse Übereinstimmung vor. CUR ist ein isoliertes

Objekt in der Hasse-Diagramm-Technik, d.h. es kann entweder den maximalen oder den minimalen Objekten zugerechnet werden. Die Maximalen CR1 und CHS liegen in der zweithöchsten Aggregationsstufe, während GIF sogar einen niedrigen  $\Gamma$ -Wert aufweist.

In der untersten Aggregationklasse sind GEF, SIG und SOR aufgeführt. Vergleicht man diese mit den acht Minimalen CHB, CR3, DIN, GEF, MSD, PES, SAH und SOR, so fallen deutliche Unterschiede zwischen den groben  $\Gamma$ -Werten und dem spezifischeren Hasse-Diagramm-Technik Verfahren auf. SIG ist ein isoliertes Objekt im Hasse-Diagramm, CHB, CR3 und PES sind in der Aggregationsstufe 3, d.h. sie haben einen hohen  $\Gamma$ -Wert; DIN, MSD, SAH sind in der Stufe (2).

### 7.3.2.5 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 2 (CD-ROMs)

Wie auch bei den vorangegangenen Gruppen an Online Datenbanken werden hier die Bewertungskriterien analysiert. Die Ergebnisse sind in Tabelle 57 zusammenfassend dargestellt.

**Tabelle 57: Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 2: CD-ROMs, 19 Objekte, 9 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	IP	AR	QU
0	0	0	4	2	0	0	5	0	0
1	0	4	2	13	1	1	5	0	3
2	1	7	4	4	0	9	8	8	7
3	5	8	0	0	5	3	0	3	3
4	3	0	8	0	5	2	1	7	3
5	10	0	1	0	8	4	0	1	3
MW	4,16	2,21	2,42	1,11	4,00	2,95	1,32	3,05	2,79
ME	5	2	2	1	4	2	1	3	2
STDEV	1,015	0,787	1,805	0,567	1,100	1,311	1,057	1,026	1,357
$I_i$	1,34	1,06	1,41	0,83	1,22	1,37	1,22	1,18	1,53
$U_i$	4	3	5	3	4	5	4	4	5

Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3, Abkürzungen siehe Abkürzungsverzeichnis

Der Umfangsindex reicht von drei bis fünf, d.h. es gibt bei der Gruppe 2 kein Kriterium, welches über die vollständige Ausprägungsmenge von sechs Ziffern verfügt. Als sehr gut in bezug auf die Verteilung ist das Bewertungskriterium QU anzusehen. Das wird sowohl durch die beiden Indizes, den Umfangsindex und den Shannon-Index als auch durch den Mittelwert und Median, die fast übereinstimmen, belegt. QU (Validität/Güte der CD-ROM) hat also eine Ausprägungsmenge von  $\{1,2,3,4,5\}$ , d.h., daß viele der betrachteten CD-ROMs evaluierte Daten enthalten. Gut verteilt sind die Kriterien KO und ID. Die Kosten der CD-ROM werden mit einer Ausprägungsmenge von  $\{0,1,2,4,5\}$ , wobei die guten Bewertungsziffern häufig vergeben werden, als insgesamt positiv gesehen. CD-ROMs sind zum großen Teil preiswert zu erwerben. Das Bewertungskriterium ID (Identifikationsmerkmale) hat eine Ausprägungsmenge von  $\{1,2,3,4,5\}$ . Die Bewertungsziffer 0 wird nicht vergeben. Das bedeutet, daß alle betrachteten CD-ROMs mindestens über die CAS-Nummer verfügen, viele sogar noch weitere Identifikationsmerkmale besitzen und sogar eine große Anzahl über suchbare Strukturformeln verfügt. Hier ist eine deutliche Verbesserung gegenüber der Betrachtungen der faktenbasierten Online Datenbanken festzustellen. Schlecht verteilt sind die Kriterien UP und VE. Beispielsweise hat das Kriterium UP (Häufigkeit der Aktualisierung) in dem vorliegenden Ansatz nur die Ausprä-

gungsmenge {1,2,3}. Das bedeutet, daß diese Art der CD-ROMs nur selten aktualisiert wird. Bei der Verfügbarkeit der CD-ROM bei anderen Hosts bzw. auf anderen Medien ist die Situation mit einer Ausprägungsmenge von {0,1,2} als schlecht anzusehen. Daraus läßt sich der Schluß ziehen, daß viele CD-ROMs ausschließlich auf diesem Datenträger erhältlich sind, also Eigenentwicklungen darstellen. AR verfügt über alle möglichen vier Bewertungsziffern, d.h., daß die Art/ Anwendung der Chemikalien bei der Gruppe 2 mit einer Ausprägungsmenge von {2,3,4,5} gut repräsentiert ist. UM, AN und IP sind mittelmäßig verteilt. Bei den Kriterien UM (Umfang der CD-ROM) und AN (Anzahl an Chemikalien) ist zwar nur eine mäßige Verteilung zu verzeichnen, jedoch kommen die hohen Bewertungsziffern häufiger vor als die niedrigen. Mäßig hingegen sind faktenbasierte chemierelevante CD-ROMs im Bewertungskriterium IP (Informationsparameter für Umweltchemikalien) anzusehen. Diese Aussage bezieht sich sowohl auf die Verteilung als auch auf die Tatsache, daß hier die niedrigen Bewertungsziffern öfter vorkommen als die hohen.

### **7.3.3 Gruppe 1 und Gruppe 2 der CD-ROMs im Vergleich**

Bei den textbasierten chemierelevanten CD-ROMs (Gruppe 1) sind die Bewertungskriterien VE, AN und IP als gut herausgearbeitet worden, währenden UP, ID und QU als schlecht verteilt identifiziert werden (siehe Anlage C.6). Demgegenüber sind die faktenbasierten chemierelevanten CD-ROMs (Gruppe 2) gut in den Kriterien KO, ID, AR und QU, schlecht hingegen in UP und VE.

Folgende generelle Schlüsse lassen sich daraus ziehen: Die chemierelevanten CD-ROMs werden in der Regel nicht häufig aktualisiert. Die Verfügbarkeit der CD-ROMs bei anderen Hosts bzw. auf anderen Medien ist bei den textbasierten CD-ROMs gut verteilt, bei den faktenbasierten hingegen schlecht. Die Anzahl der Chemikalien ist gemäß Verteilung bei den textbasierten CD-ROMs besser. Bei der Betrachtung der hohen Werte hingegen, sind die faktenbasierten den textbasierten überlegen. Faktenbasierte CD-ROMs verfügen über viele Einträge bei den hohen Bewertungsziffern. Das Bewertungskriterium IP (Informationsparameter für Umweltchemikalien) ist bei den textbasierten CD-ROMs gut verteilt und auch hoch bewertet, bei den faktenbasierten hingegen mittelmäßig verteilt und schlecht bewertet, da eine Tendenz zu den niedrigen Bewertungsziffern hin besteht. Die Bewertungskriterien ID (Identifikationsmerkmale) und QU (Validität/Güte der CD-ROM) sind gegenläufig, d.h. bei den textbasierten chemierelevanten CD-ROMs ist ID schlecht verteilt und schlecht eingestuft, bei den faktenbasierten hingegen gut. Bei dem Kriterium AR sind die faktenbasierten CD-ROMs eindeutig besser als die textbasierten. Bei der Gruppe 2 gibt es viele auf besondere Chemikalien spezialisierte CD-ROMs.

### **7.3.4 Textbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 3)**

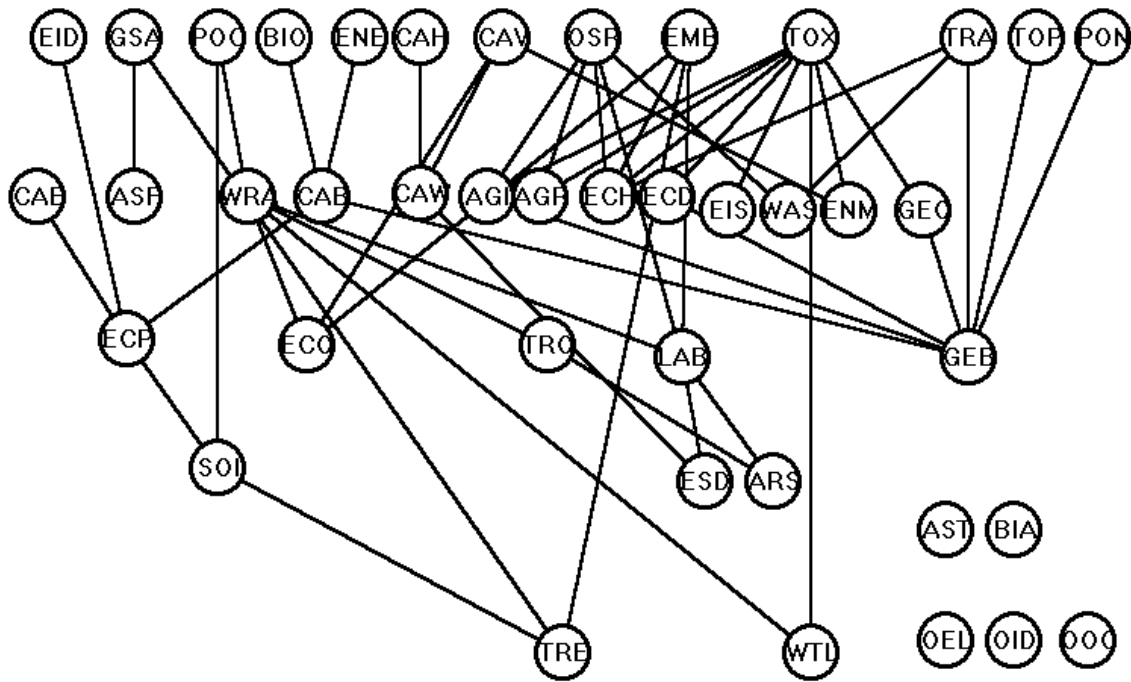
In der Tabelle 58 werden 42 textbasierte CD-ROMs aufgeführt. Es handelt sich dabei um die umfangreichste Gruppe, die bewertet wird.

**Tabelle 58: Textbasierte CD-ROMs des Umwelt-Spezialgebietes und deren Bewertungsziffern**

Nr.	CD-ROM	Akr.	UM	UP	KO	VE	AN	ID	IP	AR	QU	Γ
01	AGRICOLA on SP	AGR	4	3	2	1	1	0	3	2	0	16
02	AGRIS	AGI	4	3	2	1	2	0	2	2	0	16
03	AGRSEARCH	ARS	1	1	3	1	1	0	2	2	0	11
04	ASFA	ASF	2	3	1	4	1	0	4	2	0	17
05	ASTI	AST	3	4	2	5	0	0	3	2	0	19
06	Biolo.& Agri. Index	BIA	3	4	2	4	0	0	4	2	0	19
07	Biological Abstracts	BIO	4	3	0	1	3	1	5	2	1	20
08	CABCD	CAB	4	3	0	1	2	0	5	2	0	17
09	CABPESTCD	CAE	3	3	1	0	2	0	5	3	0	17
10	CASurveyor Haz. Mat.	CAH	3	4	3	0	4	3	3	4	1	25
11	CASurveyor Pollution	CAV	3	4	3	0	4	3	4	2	1	24
12	CASurveyor Waste	CAW	3	4	3	0	4	3	1	2	1	21
13	E-CD	ECD	3	3	0	0	2	0	5	2	0	15
14	Earth Science Disc	ESD	2	3	3	0	0	0	1	2	0	11
15	ECODISC	ECO	1	3	3	0	0	0	4	2	0	13
16	Ei Compendex Plus	ECD	4	3	0	3	3	0	4	2	1	20
17	Ei Energy & Environ.	EID	4	3	2	0	2	0	5	2	0	18
18	EMBASE CD: PolTox	EMB	2	3	3	1	1	1	5	2	0	18
19	Environ. Chemistry	ECH	2	3	2	1	0	1	2	2	0	13
20	Environmental Impact	EIS	0	3	2	2	0	1	4	2	0	14
21	Environmental Manag.	ENM	3	3	2	0	2	1	4	2	0	17
22	ETDE Energy Database	ENE	4	3	1	1	2	0	5	2	0	18
23	General Science Abst.	GSA	2	4	2	5	1	0	5	2	0	21
24	GEOBASE	GEB	3	3	0	1	2	0	2	2	0	13
25	GeoRef	GEO	4	3	1	3	3	0	3	2	0	19
26	LABORDOC	LAB	2	3	3	1	1	0	2	2	0	14
27	LifeScience Collection	LSI	4	3	1	3	3	0	3	2	0	19
28	OCLC Environment	OEL	3	1	4	1	0	0	2	2	0	13
29	OSH- OFFSHORE	OOO	2	5	1	0	1	1	3	4	0	17
30	OSH-ROM	OSR	5	3	3	1	3	1	3	4	1	24
31	OHS INTERDATA	OID	1	3	4	1	3	2	1	4	0	19
32	PolToxI:NLM,CSA, IFI	PON	3	3	1	3	3	1	2	2	2	20
33	PolTox III: CAB	POC	2	3	3	3	2	0	5	2	0	20
34	SOILCD	SOI	2	3	0	0	2	0	5	2	0	14
35	TOXLINE on SP	TOX	4	3	2	3	3	1	4	2	1	23
36	TOXLINE Plus	TOP	4	3	1	1	4	1	4	2	1	21
37	TRANSPORT	TRA	3	3	2	3	3	1	3	4	1	23
38	TREECD	TRE	2	3	0	0	1	0	5	2	0	13
39	TROPAC & RURAL	TRO	2	2	3	2	1	0	3	2	0	15
40	WasteInfo	WAS	1	3	2	1	1	0	3	4	0	15
41	Water Resources Abst.	WRA	2	3	3	3	2	0	4	2	0	19
42	WATERLIT	WTL	2	3	2	1	2	0	4	2	0	16

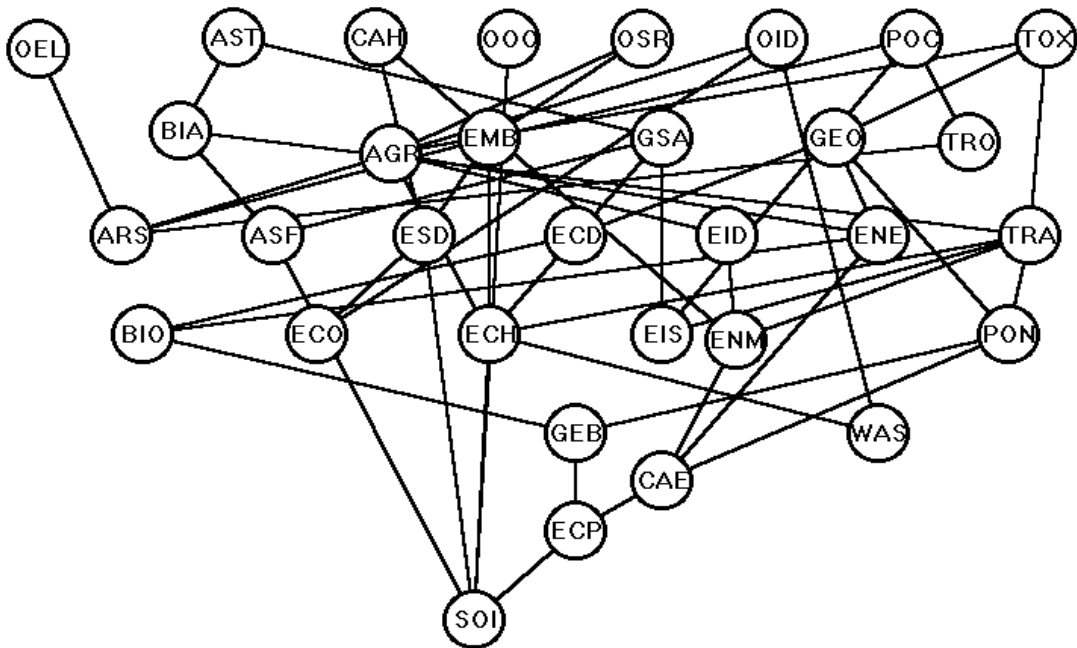
Abkürzungen siehe Abkürzungsverzeichnis

Die tabellarische Auswertung ist im Anhang C.8 zu finden. Die Diagramme der textbasierten umweltrelevanten CD-ROMs lassen erkennen, daß die inhaltlichen Bewertungskriterien zu einem "schmalen" Diagramm führen, da die Anzahl der Ebenen deutlich über der Anzahl der Objekte in der breitesten Ebenen liegen. Auch sind die Anzahl der Vergleichbarkeiten und die Anzahl der Unvergleichbarkeiten ähnlich umfangreich.



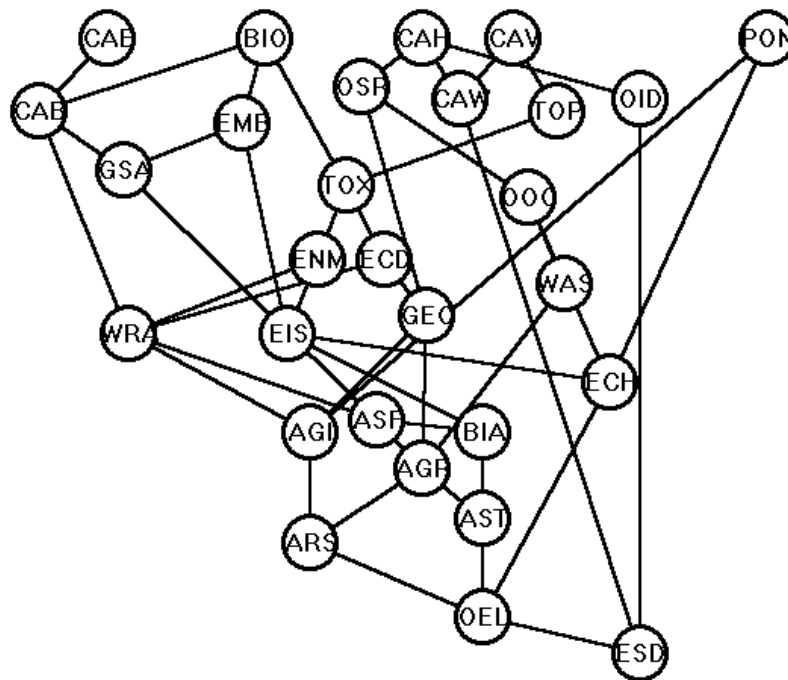
nicht-triviale Äquivalenzklasse: {GEO,LSI}

**Abbildung 38:** Hasse-Diagramm Gruppe 3: CD-ROMs, 9 Bewertungskriterien



nicht-triviale Äquivalenzklassen siehe Anhang C.8, Tabelle C11

**Abbildung 39:** Hasse-Diagramm Gruppe 3: CD-ROMs, 4 allgemeine Bewertungskriterien



nicht-triviale Äquivalenzklassen siehe Anhang C.8, Tabelle C11

**Abbildung 40: Hasse-Diagramm Gruppe 3: CD-ROMs, 5 inhaltliche Bewertungskriterien**

Am unübersichtlichsten ist das Hasse-Diagramm 42/9, welches sowohl inhaltliche als auch allgemeine Kriterien berücksichtigt. Im Diagramm, welches alle neun Bewertungskriterien umfasst, gibt es fünf Ebenen und 19 Objekte in der breitesten Ebene. Dieses Diagramm ist dementsprechend als "breit" anzusehen. Auch die Anzahl der Unvergleichbarkeiten ist um den Faktor 20 größer als die Anzahl der Vergleichbarkeiten.

**7.3.5 Faktenbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 4)**

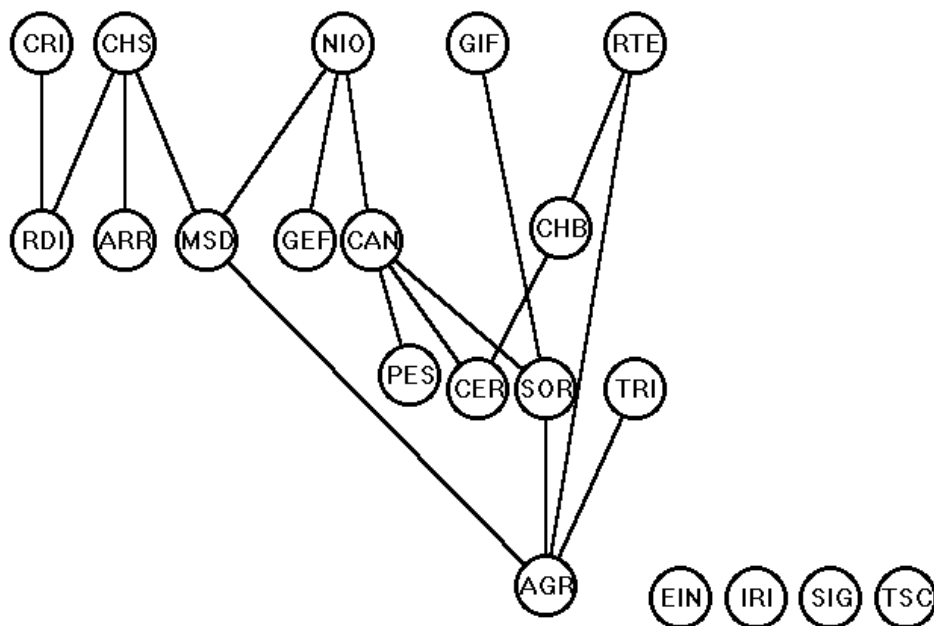
Tabelle 59 zeigt die Bewertungsziffern und  $\Gamma$ -Werte für 20 faktenbasierte CD-ROMs.

**Tabelle 59: Faktenbasierte CD-ROMs des Umwelt-Spezialgebietes und deren Bewertungsziffern**

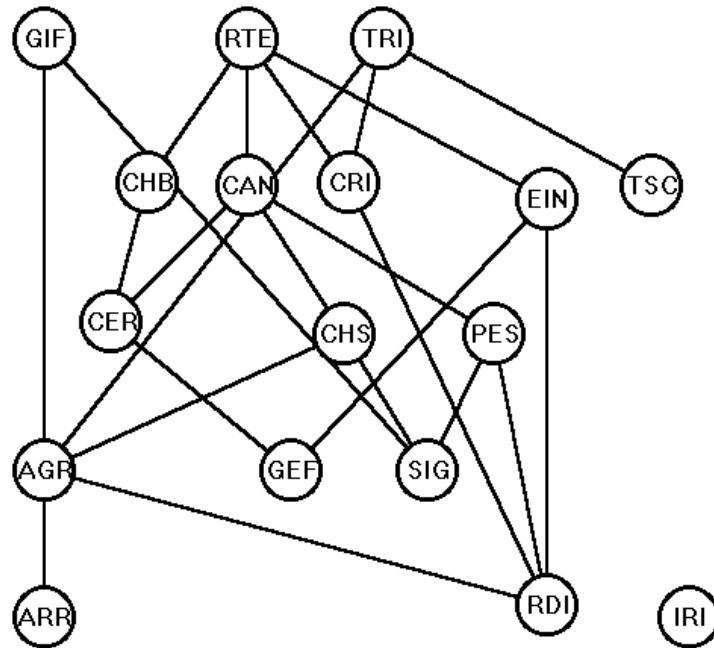
Nr	CD-ROM	Akr.	UM	UP	KO	VE	AN	ID	IP	AR	QU	$\Gamma$
01	AGREP	AGR	3	1	4	1	0	0	0	2	0	11
02	ARRIP	ARR	2	1	4	0	0	0	3	2	0	12
03	CANDATA	CAN	5	3	4	1	4	2	1	4	2	26
04	CERCLA & RCRA	CER	5	3	0	1	1	0	0	2	0	12
05	CHEM-BANK	CHB	5	3	1	2	4	3	2	2	2	24
06	CHEM Source	CHS	3	3	4	1	3	2	4	4	1	25
07	CRIS	CRI	4	1	3	2	1	0	4	2	0	17
08	EINECS plus	EIN	5	1	2	2	5	2	0	2	4	23
09	Gefährgut	GEF	5	1	0	0	5	2	0	4	3	20
10	Giftliste	GIF	3	2	5	1	3	2	1	4	2	23
11	IRIS	IRI	1	1	4	3	0	1	0	4	0	14

12	MSDS	MSD	3	3	4	1	3	2	2	4	1	23
13	NIOSHTIC/DIDS	NIO	5	3	4	1	5	2	2	4	3	29
14	PEST-BANK	PES	4	3	2	1	4	1	0	3	2	20
15	RDIR	RDI	3	1	2	1	0	0	3	2	0	12
16	RTECS	RTE	5	3	4	2	5	3	2	2	2	28
17	SIGEDA	SIG	2	2	2	1	1	2	2	5	1	18
18	SORBE	SOR	3	1	4	1	3	2	1	4	2	21
19	TRI	TRI	4	1	5	2	0	1	0	4	0	17
20	TSCA	TSC	4	0	5	2	4	2	1	2	3	23

Abkürzungen siehe Abkürzungsverzeichnis

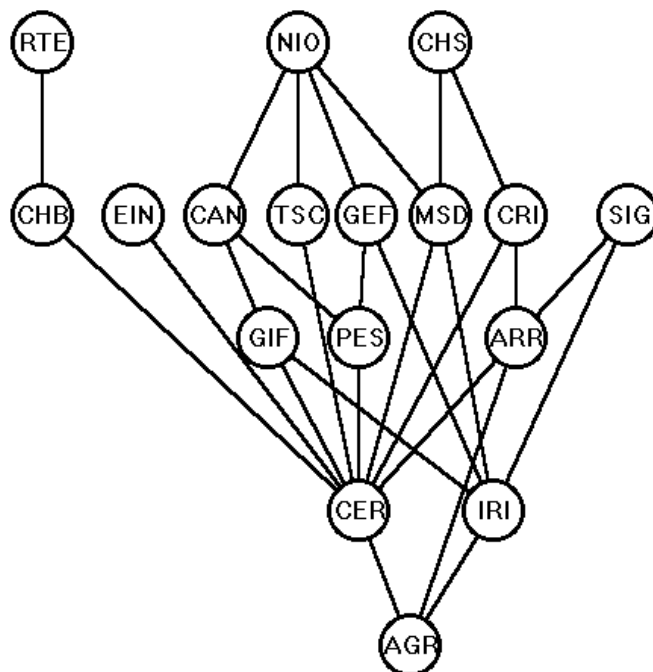


**Abbildung 41:** Hasse-Diagramm Gruppe 4: CD-ROMs, 9 Bewertungskriterien



nicht-triviale Äquivalenzklassen: {ARG, SOR}; {CAN, NIO}; {CHS, MSD}

**Abbildung 42: Hasse-Diagramm Gruppe 4: CD-ROMs, 4 allgemeine Bewertungskriterien**



nicht-triviale Äquivalenzklassen: {ARG, DDI}; {GIF, SOR}; {IRI, TRI}

**Abbildung 43: Hasse-Diagramm Gruppe 4: CD-ROMs, 5 inhaltliche Bewertungskriterien**

Auch bei dieser Gruppe sind die Auswertungen der Hasse-Diagramme im Anhang (C.9) zu finden. Sowohl das Hasse-Diagramm der allgemeinen als auch das der inhaltlichen Bewer-

tungskriterien ist als aussagekräftig zu bezeichnen, da die Anzahl der Ebenen mit der Anzahl der Objekte in der breitesten Ebene ausgewogen ist. Auch liegt bei der Betrachtung der vier allgemeinen Bewertungskriterien nur ein isoliertes Objekt, bei den inhaltlichen Bewertungskriterien keines vor. Das Hasse-Diagramm, welches alle Bewertungskriterien berücksichtigt, ist als "breit" zu bezeichnen, da die Anzahl der Objekte in der breitesten Ebene deutlich über der Anzahl der Ebenen liegt.

### **7.3.6 Gruppe 3 und Gruppe 4 der CD-ROMs im Vergleich**

Wie in den Anhängen C.8 und C.9 nachzulesen ist, sind die Bewertungskriterien UM, KO, AN und IP bei den umweltrelevanten Gruppen der CD-ROMs, d.h. sowohl den textbasierten als auch den faktenbasierten gut verteilt.

Bei einer weiterführenden Auswertung wird jedoch deutlich, daß die textbasierten CD-ROMs über eine geringere Anzahl von Chemikalien verfügen, da die hohen Bewertungsziffern gering oder gar nicht belegt sind. Bei den faktenbasierten CD-ROMs des Umwelt-Spezialgebietes hingegen sind die hohen Bewertungsziffern für dieses Bewertungskriterium gut repräsentiert. Hier gibt es somit eine höhere Anzahl von Chemikalien als in den textbasierten.

Genau gegensätzlich ist der Sachverhalt bei dem Kriterium IP (Informationsparameter für Umweltchemikalien). Hier sind bei den textbasierten CD-ROMs die hohen Bewertungsziffern besser belegt als die niedrigen; bei den faktenbasierten CD-ROMs liegen mehr Einträge in den niedrigen als in den höheren Bewertungsziffern vor.

Das Kriterium VE (Verfügbarkeit der CD-ROM bei anderen Hosts bzw. auf anderen Medien) ist bei der textbasierten Gruppe gut verteilt. QU (Validität/Güte der CD-ROM) ist bei den faktenbasierten gut, bei den textbasierten CD-ROMs schlecht verteilt. In faktenbasierten CD-ROMs gibt es demnach eher evaluierte Daten als in textbasierten.

Sowohl bei den faktenbasierten als auch bei den textbasierten CD-ROMs des Umwelt-Spezialgebietes ist das Bewertungskriterium ID (Identifikationsmerkmale) schlecht verteilt, und die hohen Bewertungsziffern darüber hinaus nicht belegt. Bezüglich der Ausprägungsmenge ist jedoch die Gruppe 3 schlechter als die Gruppe 4, da die Werte 0 und 1 im ersten Fall am häufigsten vergeben werden. Bei den faktenbasierten CD-ROMs sind die Kriterien UP (Häufigkeit der Aktualisierung) und VE (Verfügbarkeit der CD-ROM bei anderen Hosts bzw. auf anderen Medien) schlecht verteilt. Das bedeutet, daß diese Art von CD-ROMs selten aktualisiert wird und darüber hinaus auch fast nur auf diesem Medium verfügbar ist.

### **7.3.7 Umweltchemikalien-relevante CD-ROMs des Chemie/Umwelt-Spezialgebietes (Gruppe 5)**

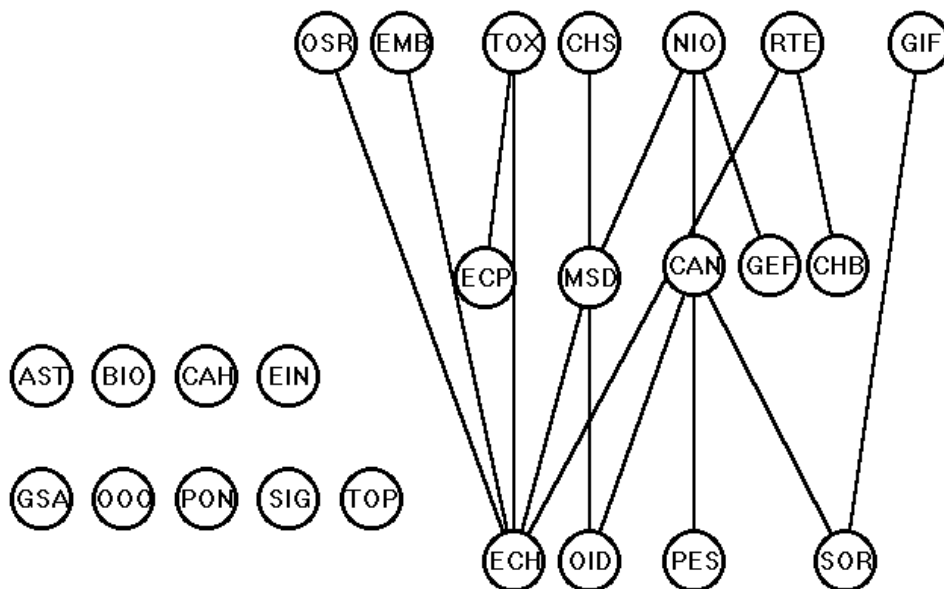
**Tabelle 60: Umweltchemikalien-relevante CD-ROMs des Chemie-/Umwelt-Spezialgebietes und deren Bewertungsziffern**

Nr.	CD-ROM	Akk.	UM	UP	KO	VE	AN	ID	IP	AR	QU	Γ
01	ASTI	AST	3	4	2	5	0	0	3	2	0	19
02	Biological Abstracts	BIO	4	3	0	1	3	1	5	2	1	20
03	CANDATA	CAN	5	3	4	1	4	2	1	4	2	26
04	CASurveyor Haz.Mat.	CAH	3	4	3	0	4	3	3	4	1	25
05	CHEM-BANK	CHB	5	3	1	2	4	3	2	2	2	24
06	CHEM Source	CHS	3	3	4	1	3	2	4	4	1	25
07	Ei Compendex Plus	ECP	4	3	0	3	3	0	4	2	1	20
08	EINECS Plus-CD	EIN	5	1	2	2	5	2	0	2	4	23
09	EMBASE-CD:PoITox	EMB	2	3	3	1	1	1	5	2	0	18

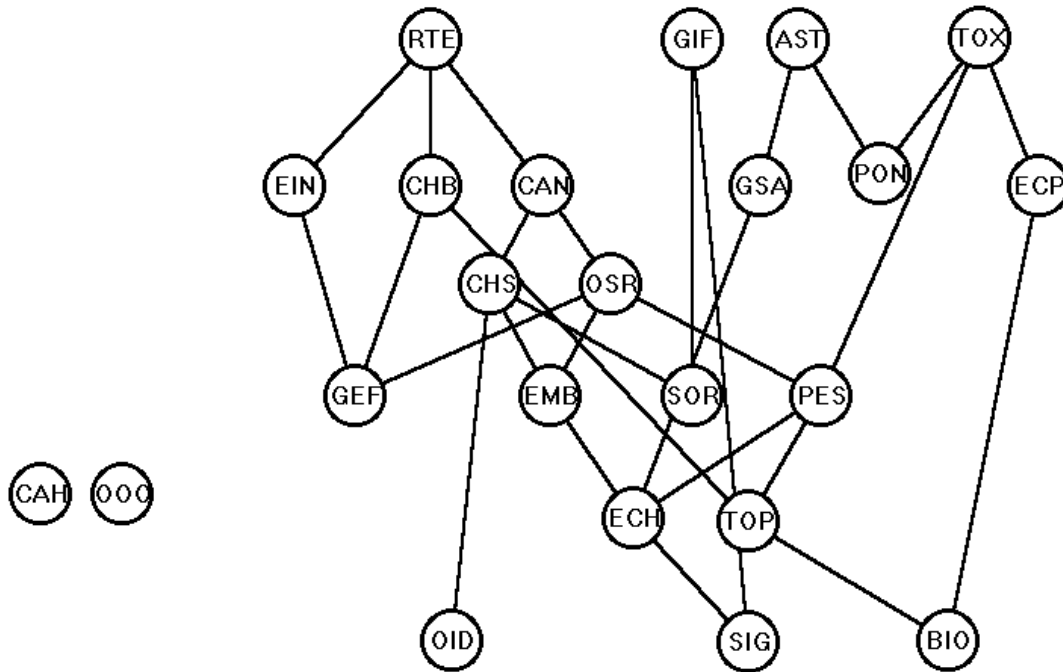
10	Environ. Chemistry	ECH	2	3	2	1	0	1	2	2	0	13
11	Gefahrgut-CD-ROM	GEF	5	1	0	0	5	2	0	4	3	20
12	General Science	GSA	2	4	2	5	1	0	5	2	0	21
13	Giftliste	GIF	3	2	5	1	3	2	1	4	2	23
14	MSDS	MSD	3	3	4	1	3	2	2	4	1	23
15	NIOSHTIC/DIDS	NIO	5	3	4	1	5	2	2	4	3	29
16	OSH-OFFSHORE	OOO	2	5	1	0	1	1	3	4	0	17
17	OSH-ROM	OSR	5	3	3	1	3	1	3	4	1	24
18	OHS INTERDATA	OID	1	3	4	1	3	2	1	4	0	19
19	PEST-BANK	PES	4	3	2	1	4	1	0	3	2	20
20	PolTox I:NLM,CSA, IFI	PON	3	3	1	3	3	1	2	2	2	20
21	RTECS	RTE	5	3	4	2	5	3	2	2	2	28
22	SIGEDA	SIG	2	2	2	1	1	2	2	5	1	18
23	SORBE	SOR	3	1	4	1	3	2	1	4	2	21
24	TOXLINE on SP	TOX	4	3	2	3	3	1	4	2	1	23
25	TOXLINE Plus	TOP	4	3	1	1	4	1	4	2	1	21

Abkürzungen siehe Abkürzungsverzeichnis

In dieser Gruppe sind diejenigen CD-ROMs erfasst, die sowohl dem Chemie- als auch dem Umweltspezialgebiet zuzurechnen sind. Die Erläuterungen zu den Hasse-Diagrammen sind im Anhang C.10 aufgeführt.

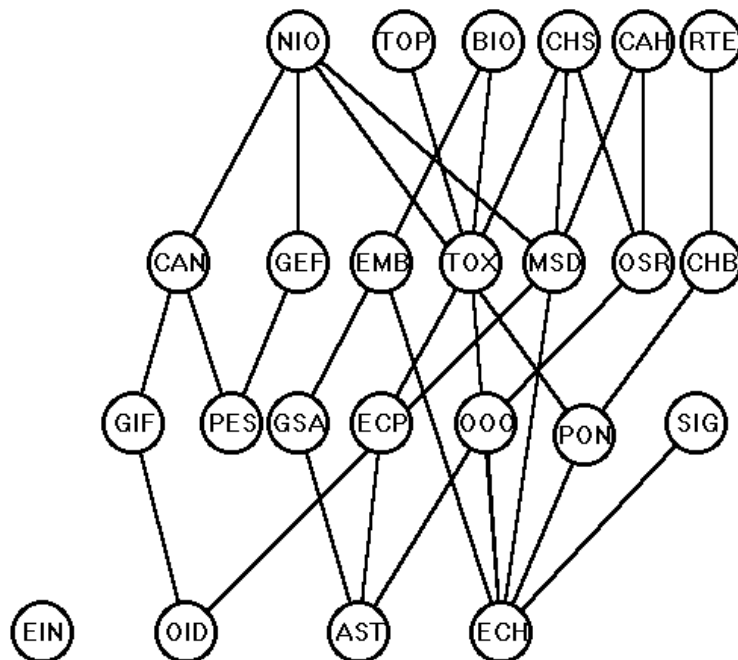


**Abbildung 44: Hasse-Diagramm Gruppe 5: CD-ROMs, 9 Bewertungskriterien**



nicht-triviale Äquivalenzklassen: {CAN,NIO}; {CHS,MSD}

**Abbildung 45:** Hasse-Diagramm Gruppe 5: CD-ROMs, 4 allgemeine Bewertungskriterien



nicht-triviale Äquivalenzklasse: {GIF, SOR}

**Abbildung 46:** Hasse-Diagramm Gruppe 5: CD-ROMs, 5 inhaltliche Bewertungskriterien

Wie auch die vorangegangenen Gruppen werden auch die Umweltchemikalien-relevanten CD-ROMs wieder mit neun Bewertungskriterien, dann mit den vier allgemeinen und mit den fünf inhaltlichen Kriterien bewertet. Das Hasse-Diagramm der vier allgemeinen Bewertungskriterien ist als aussagekräftig einzustufen, da die Anzahl der Ebenen mit der Anzahl der Objekte in der breitesten Ebene übereinstimmt. Das Hasse-Diagramm, welches alle neun möglichen Bewertungskriterien berücksichtigt, ist hingegen als sehr "breit" anzusehen, aber auch das Diagramm der inhaltlichen Kriterien ist als "breit" zu bezeichnen. Bei beiden Diagrammen übersteigt die Anzahl der Objekte in der breitesten Ebene deutlich die Anzahl der Ebenen. Bei dem neun Kriterien umfassenden Ansatz liegen neun isolierte Objekte vor, die demnach mit keiner anderen CD-ROM verglichen werden können. Das Verhältnis Unvergleichbarkeiten zu Vergleichbarkeiten ist bei dem Ansatz der allgemeine Bewertungskriterien am günstigsten.

## 8. Diskussion der Ergebnisse der Bewertung

### 8.1 Analyse der Ergebnisse der Bewertung der Online Datenbanken

Bei der Bewertung der Online Datenbanken werden fünf verschiedenen Gruppen mit 12 bzw. bei faktenbasierten Datenbanken 11 Bewertungskriterien untersucht (siehe auch Anhang C.1 - C.5). Die Tabelle 61 soll nochmals vergleichend die Verteilung der Bewertungskriterien veranschaulichen.

**Tabelle 61: Gegenüberstellung der Online Gruppen und deren Bewertungskriterien**

Gruppe	Bezeichnung der Gruppe	Anzahl der Objekte	Anzahl der Bewertungskriterien (BWK)	gute BWK (gemäß Verteilung)	schlechte BWK
1 ON	chemierelevante textbasierte Datenbanken	28	12	AN, CE, PE	UP, ID, QU
2 ON	chemierelevante Faktendatenbanken	18	11	UM, KO, AN, QU	VE, ID, CE
3 ON	umweltrelevante textbasierte Datenbanken	19	12	AN, CE, PE	UP, ID, QU
4 ON	umweltrelevante Faktendatenbanken	14	11	UM, KO, AN, QU	VE, ID
5 ON	Umweltchemikalienrelevante Datenbanken	18	12	AN, CE, PE	UP, ID, IP, QU

1 ON = Gruppe 1 Online Datenbanken, etc., Abkürzungen siehe Abkürzungsverzeichnis

Das Bewertungskriterium AR wird aus den oben genannten Gründen nicht berücksichtigt.

Folgende Schlüsse lassen sich aus der Aufstellung ziehen:

- AN ist in allen Ansätzen ein gut verteiltes Bewertungskriterium.
- ID ist in allen Gruppen ein schlecht verteiltes Bewertungskriterium.

Auf die Bedeutung der einzelnen Online Datenbanken wurde bereits in den vorangegangenen Kapiteln eingegangen. Es hat sich gezeigt, daß die großen, umfangreichen Datenbanken wie z.B. BEI (BEILSTEIN Database), CAS (Chemical Abstracts), CAB (CAB Abstracts), RTE (Registry of Toxic Effects of Chemical Substances) auch Maximale darstellen. Die Auswertung hat darüber hinaus auch Aufschluß darüber gegeben, daß die Datenbanken, die nicht so bekannt sind, wie z.B. CBN (Chemical Business News Database), CSN (Chemical Safety Newsbase), HSD (Hazardous Substances Databank), MER (The Merck Index Online), MSD (Material Safety Data Sheets from Occupational Health Service) und NTI (National Technical Information Service) auch hoch bewertet werden.

Demgegenüber ist festzuhalten, daß die allgemein als Umweltdatenbanken anerkannten Datenquellen ULI (Umweltliteraturdatenbank), UFO (Umweltforschungsdatenbank) und POL (Pollution Abstracts) nicht als Maximale, sondern im Falle von UFO sogar als Minimale eingestuft werden. Das bedeutet, daß diese Datenbanken nicht über die für Umweltchemikalien relevanten Informationsparameter verfügen.

## 8.2 Analyse der Ergebnisse der Bewertung der CD-ROMs

Bei der Bewertung der CD-ROMs werden wie auch bei den Online Datenbanken wieder fünf verschiedene Gruppen untersucht. Die Anzahl der Bewertungskriterien ist bei allen Gruppen neun, vier allgemeine und fünf inhaltliche. Die folgende Tabelle soll nochmals vergleichend die Verteilung der Bewertungskriterien veranschaulichen.

Neben der Bezeichnung der Gruppen und der Anzahl der CD-ROMs, zugehörig zu der jeweiligen Gruppe werden die aktuellen Ausprägungsmengen ohne das Bewertungskriterium AR ausgewiesen. Das Bewertungskriterium AR wird aus den schon oben genannten Gründen nicht berücksichtigt. Die als "gut" bzw. "schlecht" herausgearbeiteten (vgl. hierzu Kapitel 6.6 und den Anhang C.6 - C.10) Kriterien sind ebenfalls vermerkt.

**Tabelle 62: Gegenüberstellung der CD-ROM Gruppen und deren Bewertungskriterien**

Gruppe	Bezeichnung der Objektmenge	Anzahl der Objekte	Ausprägungsmenge (ohne AR)	gute BWK	schlechte BWK
1 CD	chemierelevante textbasierte CD-ROMs	26	4-6	IP	UP, ID, QU
2 CD	chemierelevante Fakten CD-ROMs	19	3-5	QU, KO, ID	UP, VE
3 CD	umweltrelevante textbasierte CD-ROMs	42	3-6	AN, IP, VE, UM	ID, QU
4 CD	umweltrelevante Faktendaten CD-ROMs	20	4-6	AN IP, KO, QU	UP, VE
5 CD	Umweltchemikalien-relevante CD-ROMs	25	4-6	IP, KO	ID, UP

1 CD = Gruppe 1 CD-ROMs, etc., Abkürzungen siehe Abkürzungsverzeichnis

Folgende Schlüsse lassen sich aus der Aufstellung ziehen:

- IP ist in vier von fünf Ansätzen ein gut verteiltes Bewertungskriterium.
- ID ist in vier von fünf Gruppen ein schlecht verteiltes Bewertungskriterium.

Auf die hoch und niedrig eingestuft CD-ROMs wurde im Kapitel 7.3 im einzelnen verwiesen. Anzumerken ist in diesem Zusammenhang, daß auch die CD-ROM Versionen der großen Datenbanken z.B. BIOSIS (BIOSIS Previews CD-ROM), RTECS (Registry of Toxic Effects of Chemical Substances) und TOX (Toxicology Literature Online CD-ROM) als Maximale eingestuft wurden.

CAH (CASurveyor Hazardous Materials) und CAV (CASurveyor Pollution Control and the Environment), die beiden thematischen Auszüge aus der Chemical Abstracts Online Datenbank werden sehr gut bewertet. Hierbei ist jedoch hervorzuheben, daß andere Auszüge aus CAS, wie z.B. CAF (CASurveyor Food and Feed Chemistry) und CAC (CASurveyor Cancer Chemical Research) als Minimale eingestuft werden. Sie passen nicht genau in das Spektrum der Bewertung von Umweltchemikalien.

Die CD-ROMs DID (Dictionary of Drugs on CD-ROM), DOC (Dictionary of Organic Compounds on CD-ROM), NIO (CCINFODisc Series C1, NIOSHTIC/DIDS) und PON (Pollution and Toxicology Database on CD-ROM), die nicht in der Form online vorliegen, sind gut bewertet worden.

Auffällig ist, daß die wenigen deutschen bzw. deutschsprachigen CD-ROMs, wie beispielsweise GEF (Gefahrgut CD-ROM), GIF (Giftliste, krebserregende, gesundheitsschädliche und reizende Stoffe), SIG (Siemens Gefahrstoff-Datenbank CD-ROM) niedrig bewertet werden. Eine Ausnahme bildet der CDR (CD-Römpf), der entweder eine isolierte CD-ROM darstellt oder in einer mittleren Ebene im Hasse-Diagramm zu finden ist.

### **8.3 Auswertung und Gegenüberstellung der Medien Online/CD-ROM**

Im vorangegangenen wurden zunächst fünf Gruppen an Online Datenbanken und fünf Gruppen an CD-ROMs mittels der Hasse-Diagramm-Technik analysiert. Es wurde insbesondere auf die sich im positiven (Maximale) bzw. negativen Sinn (Minimale) herauskristallisierenden Datenquellen eingegangen. Es wurden auch die textbasierten mit den faktenbasierten Datenquellen verglichen. Darüber hinaus wurde die Bedeutung der einzelnen Bewertungskriterien herausgestellt. Es verbleibt nun die Gegenüberstellung und der Vergleich der einzelnen Gruppen. Als Nebenergebnis ist der Vergleich der Bewertungskriterien anzuschließen. Hierbei bedarf es darüber hinaus jedoch weiterer Forschungsarbeiten, um die Kriterien aus einem übergeordneten Standpunkt heraus zu analysieren. Die Bedeutung der Bewertungskriterien wird hier auf der Grundlage der Auswertung der Gruppen an Datenquellen erarbeitet. Eine übergeordnete Vorgehensweise wäre demnach, die Kriterien unabhängig von den Realisierungen der Datenquellen zu analysieren. Dann könnten beispielsweise die Trennschärfe und der Aussageumfang der Kriterien in Betracht gezogen werden.

#### **8.3.1 Bedeutung der einzelnen Bewertungskriterien in zehn Gruppen**

Die insgesamt zehn betrachteten Gruppen, davon fünf Gruppen an Online Datenbanken und fünf an CD-ROMs sollen im folgenden nun vergleichend gegenübergestellt, und die Bedeutung der Bewertungskriterien herausgearbeitet werden.

Aus der Tabelle 63 sind die in den vorangegangenen Kapiteln und dem Anhang C als "gut" oder als "schlecht" herausgearbeiteten Bewertungskriterien aufgeführt. Die Kriterien CE (Chemikalien Entwicklung) PE (Parameter Entwicklung) und CT (Chemikalien-Testdatensatz) werden in der Folge nicht mehr berücksichtigt. CE und PE sind nur in Online Datenbanken untersucht worden. Das Kriterium CT, das weder als "gut" noch als "schlecht" auffällt, wird nur bei Online Datenbanken eingesetzt. Dennoch läßt die Tabelle bereits einige Tendenzen erkennen für CE und PE erkennen. Sie sind in drei von fünf Online Datenbanken Gruppen gut zu nennen. Die Chemikalien Entwicklung ist bei den chemierelevanten Faktendatenbanken hingegen als schlechtes Bewertungskriterium anzusehen, da diese Datenbanken in den vergangenen Jahren **nicht** in bezug auf die Anzahl an Umweltchemikalien stark zugenommen haben.

**Tabelle 63: Gegenüberstellung der Gruppen an Online Datenbanken/CD-ROMs**

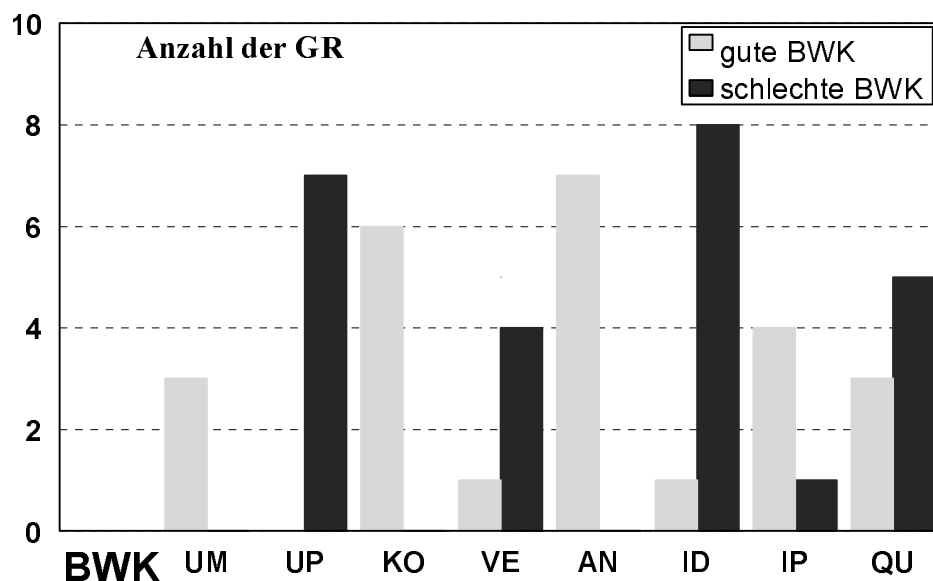
Gruppe	Bezeichnung der Gruppe	Anzahl der Objekte	Ausprägungsmenge (ohne AR)	gute BWK	schlechte BWK
1 ON	chemierelevante textbasierte Datenbanken	28	3-6	AN, CE, PE,	UP, ID, QU
1 CD	chemierelevante textbasierte CD-ROMs	26	4-6	IP	UP, ID, QU
2 ON	chemierelevante Faktendatenbanken	18	4-6	UM, KO, AN, QU	ID, CE, VE
2 CD	chemierelevante Fakten CD-ROMs	19	3-5	KO, ID, QU	UP, VE
3 ON	umweltrelevante textbasierte CD-ROMs	19	3-6	AN, CE, PE	UP, ID, QU
3 CD	umweltrelevante textbasierte CD-ROMs	42	3-6	UM, KO, VE, AN, IP	ID, QU
4 ON	umweltrelevante Faktendatenbanken	14	4-6	UM, KO, AN, QU,	VE, ID
4 CD	umweltrelevante Fakten CD-ROMs	20	4-6	UM, KO, AN, IP, QU	UP, VE
5 ON	Umweltchemikalien-relevante Datenbanken	18	2-6	AN, CE, PE	UP, QU, ID, IP
5 CD	Umweltchemikalien-relevante CD-ROMs	25	4-6	KO, IP	UP, ID

1 ON = Gruppe 1 Online Datenbanken, etc., Abkürzungen siehe Abkürzungsverzeichnis

Bei Auslassung der Kriterien CE, PE und CT sowie AR bleiben folgende acht Bewertungskriterien für eine weitere Auswertung ihrer Relevanz bezüglich der dargestellten Gruppen bestehen:

UM (Umfang der Datenbank bzw. CD-ROM), UP (Häufigkeit der Aktualisierung), KO (Kosten der Recherche bzw. CD-ROM), VE (Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien), AN (Anzahl an Chemikalien), ID (Identifikationsmerkmale für Chemikalien), IP (Informationsparameter für Umweltchemikalien) und QU (Validität/Güte der Datenquelle). Diese werden in der Folge weiter diskutiert. Da die Beurteilung der Kriterienqualität anhand der Datenquellen erfolgt, auf die sie angewandt wurden, könnte parallel auch ein Rückschluß auf die Gruppierungen der Datenquellen erfolgen. Dies wird - trotz der Betonung auf die Beurteilung der Kriterien - bewußt in Kauf genommen.

Die weitere Aufschlüsselung der "guten" und "schlechten" Bewertungskriterien ist in folgender Abbildung 47 verdeutlicht. Die Erläuterung zu "guten" und "schlechten" Bewertungskriterien wird in Kapitel 6.6 gegeben.



**Abbildung 47: "Gute" und "Schlechte" Bewertungskriterien in zehn Gruppen**

Abkürzungen siehe Abkürzungsverzeichnis

Als "gut" bzw. "schlecht" werden die Gruppen bezüglich ihrer Verteilung und den ausgewerteten Indizes eingestuft. Diese "gut/schlecht" Einteilung ist selbstverständlich als eine äußerst grobe anzusehen. Dennoch lassen sich folgende Schlüsse aus der Abbildung ziehen:

Es gibt kein Bewertungskriterium, welches sich nicht in wenigstens einer Gruppe als "gut" oder "schlecht" abhebt. Diese Tatsache läßt den Schluß zu, daß alle Kriterien eine gewisse Relevanz für das hier erarbeitete Bewertungsverfahren für Datenquellen haben.

Die allgemeinen Kriterien UM, UP und KO treten entweder als "gut" oder "schlecht" hervor. UM und KO sind als positiv anzusehen, während UP in sieben von zehn Gruppen "schlecht" bewertet wird. Die Kosten (KO) hingegen wurden für sechs Gruppen als "gut" klassifiziert. Die Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien (VE) ist nur bei den umweltrelevanten textbasierten CD-ROMs als "gut" eingestuft und in vier weiteren Gruppen "schlecht". Es handelt sich dabei um chemierelevanten Faktendatenbanken und CD-ROMs sowie um umweltrelevante Faktendatenbanken und CD-ROMs.

Das Kriterium AN (Anzahl an Chemikalien) ist bei sieben Gruppen als "gut" eingestuft und in keiner anderen Gruppe als "schlecht" bewertet. Es handelt sich damit über das beste Bewertungskriterium (im Sinne der Verteilung) für die Gesamtheit der betrachteten Datenquellen.

Das Bewertungskriterium Identifikationsmerkmale für Chemikalien hingegen ist bei acht Gruppen "schlecht" klassifiziert. Nur bei den faktenbasierten chemierelevanten CD-ROMs sind die Identifikationsmerkmale für Chemikalien in ausreichender Weise berücksichtigt, bei den umweltrelevanten faktenbasierten CD-ROMs ist das Kriterium ID "mittelmäßig" eingestuft.

Das Kriterium IP (Informationsparameter für Umweltchemikalien) ist in vier Gruppen positiv eingestuft, d.h. in allen Gruppen an CD-ROMs mit Ausnahme der chemierelevanten faktenbasierten CD-ROMs. Fast alle CD-ROMs verfügen demnach über ausreichende Umweltchemikalien-relevante Parameter. "Schlecht" ist IP nur bei den Umweltchemikalien-relevanten Online Datenbanken eingestuft. Diese Tatsache läßt den Schluß zu, daß diese Gruppe sich zwar in Chemie- bzw. Chemikalien-relevanten Kriterien ausweist (AN, CE sind gut bewertet), jedoch sowohl bei den Identifikationsmerkmalen (ID) als auch in den umweltrelevanten Parametern eher negativ auffällt. Die Parameter Entwicklung (PE) hingegen ist als positiv zu vermerken, so daß ein Trend zur Verbesserung der Umweltparameter-Situation zu erkennen ist.

Das Kriterium QU (Validität/Güte der Datenquelle) ist in drei Gruppen "gut", in fünf hingegen als "schlecht" klassifiziert. Die Validität/Güte der Datenquelle ist bei allen Faktendatenquellen, seien sie auf Umweltschutz oder auf Chemie spezialisiert, als gut zu erachten. Als "schlecht" ist das Bewertungskriterium bei allen textbasierten Datenquellen und bei den Umweltchemikalien-relevanten Online Datenbanken anzusehen.

### **8.3.2 Kriterienmaß $\vartheta$ als grobes Einstufungsmaß für Kriterien**

Beantwortet werden soll nun die Frage, wie gut die einzelnen Gruppen im Vergleich zueinander für Fragestellungen auf dem Gebiet der Chemie und des Umweltschutzes gemessen an den acht oben genannten Bewertungskriterien sind. Wie bereits in Kapitel 8.3.1 dargelegt, liegen zu zehn Gruppen (10 Objekte) und acht Bewertungskriterien Informationen vor.

In Analogie zur Vorgehensweise im Zusammenhang mit dem in Kapitel 6.5 eingeführten Gesamtmaß  $\Gamma$  für Objekte läßt sich auch hier ein **Gesamtmaß  $\vartheta$**  für Kriterien definieren. Seien die Gruppen 1ON (1. Gruppe Online Datenbanken) bis 5CD (5. Gruppe CD-ROMs) mit dem Index  $j$  numeriert und die Bewertungskriterien UM bis QU mit dem Index  $i$ .

In Tabelle 63 wurde bereits eine Einteilung der Bewertungskriterien in "gut" und "schlecht" vorgenommen. In Erweiterung dieser Vorgehensweise ist die zusätzliche Abstufung in "mittel" sinnvoll, damit auch diejenigen Kriterien berücksichtigt werden können, die in Tabelle 63 weder als "gut" noch als "schlecht" bezeichnet werden. Somit ergeben sich drei verschiedene Einstufungen für Bewertungskriterien  $C_{ij}$  d.h. gut, mittel und schlecht. Gut erhält die Bewertungsziffer = 2, mittel = 1, schlecht = 0, d.h. die Ausprägungsmenge für  $C_{ij}$  ist  $\{0,1,2\}$ . Zur Bewertung des  $i$ -ten Kriteriums müssen die Bewertungen  $C_{ij}$  aggregiert werden. Naheliegend ist eine additive Verknüpfung gemäß

$$\vartheta_i = \sum_j g_j C_{ij} \quad g_j = 1 \text{ für alle } j$$

$g_j$  sind Gewichtungsfaktoren, die der Einfachheit halber gleich eins gesetzt werden.

Im Gegensatz zum Gesamtmaß  $\Gamma$  für Objekte, wo in der entsprechenden Datenmatrix (z.B. Tabelle 37) zeilenweise addiert wird, wird bei dem Gesamtmaß  $\vartheta$  für Kriterien in der Datenmatrix der Tabelle 64 spaltenweise addiert.

Es folgt:

$$\vartheta_i = \sum C_{ij} \quad (8-1)$$

$C_{ij}$  = Ausprägung (0,1,2) für das  $i$ -te Bewertungskriterium und die  $j$ -te Gruppierung

### **8.3.3 Hasse-Diagramm für zehn Gruppen an Datenquellen**

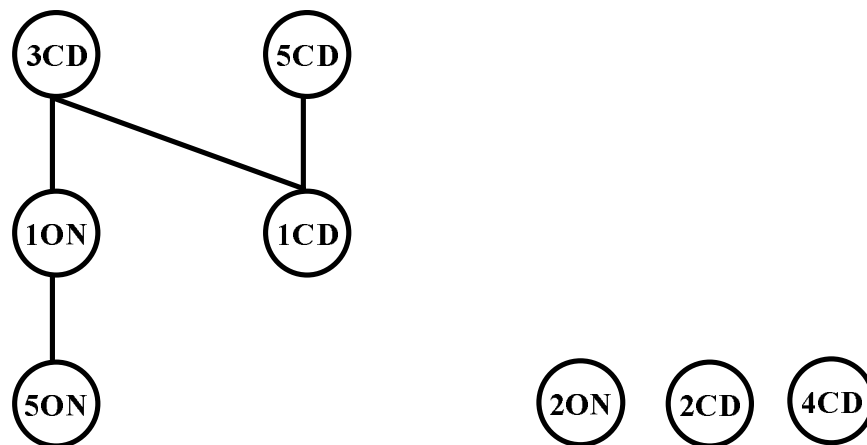
Wie oben dargestellt, liegen zehn Gruppen und acht Bewertungskriterien vor. Darüber hinaus sind drei Einstufungen in gut, mittel und schlecht vorgenommen worden. Unter dieser Voraussetzung kann wieder eine Einstufung der Gruppen mit anschließender Bewertung durchgeführt werden, die in folgender Tabelle 64 aufgeführt wird. Die letzte Spalte enthält wieder das Gesamtmaß  $\Gamma$  (vgl. Kapitel 6.5) und die unterste Reihe führt das in Kapitel 8.3.2 eingeführte Kriterienmaß  $\vartheta$  auf.

**Tabelle 64: Bewertung der zehn Gruppen an Datenquellen mit acht Bewertungskriterien**

Gruppe/BWK	UM	KO	UP	VE	AN	ID	IP	QU	$\Gamma$
1ON	1	1	0	1	2	0	1	0	6
1CD	1	1	0	1	1	0	2	0	6
2ON	2	2	1	0	2	0	1	2	10
2CD	1	2	0	0	1	2	1	2	9
3ON	1	1	0	1	2	0	1	0	6
3CD	2	2	1	2	2	0	2	0	11
4ON	2	2	1	0	2	0	1	2	10
4CD	2	2	0	0	2	1	2	2	11
5ON	1	1	0	1	2	0	0	0	5
5CD	1	2	0	1	1	0	2	1	8
<b>9</b>	<b>13</b>	<b>17</b>	<b>3</b>	<b>7</b>	<b>17</b>	<b>3</b>	<b>13</b>	<b>9</b>	

1 ON = Gruppe 1 Online Datenbanken, etc., Abkürzungen siehe Abkürzungsverzeichnis

Das daraus resultierende Hasse-Diagramm für die partiell geordnete Menge der Gruppen, in welchem alle acht Bewertungskriterien gleichzeitig betrachtet werden, ist in Abbildung 48 aufgezeichnet.



**Äquivalente Objekte:** 1ON = 3ON, 2ON = 4ON

**Abbildung 48: Hasse-Diagramm für zehn Gruppen und acht Bewertungskriterien**

Das Hasse-Diagramm umfaßt drei Ebenen. Die Gruppen 1ON und 3ON sowie 2ON und 4ON sind nicht-triviale Äquivalenzklassen. Die Gruppen 2CD und 4CD sowie 2ON (und 4ON) sind isolierte Objekte. Es handelt sich dabei um die chemierelevanten faktenbasierten Online Datenbanken und CD-ROMs sowie um die umweltrelevanten faktenbasierten Online Datenbanken und CD-ROMs. Daraus folgt, daß diese aufgeführten Gruppen nicht miteinander verglichen werden können, wenn alle acht Bewertungskriterien gleichzeitig zur Einstufung herangezogen werden. Die isolierten Objekte sind beispielsweise alle in den Bewertungskriterien KO und QU gut eingestuft, während sie in dem Kriterium VE schlecht bewertet werden. Ansonsten sind die vier Gruppen weder untereinander noch mit den restlichen sechs Gruppen vergleichbar.

Maximale Gruppen sind 3CD (textbasierte CD-ROMs der Umwelt-Gruppe) und 5CD (Umweltchemikalien-relevante CD-ROMs). Diese beiden Gruppen scheinen am besten geeignet zu sein, die Thematik der "Chemie und Umweltschutz" abzudecken und geeignete Informationen hierüber zu liefern. Die Gruppe 3CD umfaßt 42 textbasierte CD-ROMs des Umwelt-

Spezialgebietes. Sie ist in fünf von acht Bewertungskriterien als gut eingestuft. Das bedeutet, daß die Verteilung der Bewertungskriterien UM, KO, VE, AN, IP als gut anzusehen ist. Schlecht ist diese Gruppe an CD-ROMs in bezug auf die Kriterien ID und QU zu nennen. Die Gruppe 3CD ist ein Maximal, d.h. die Verteilung aller acht Bewertungskriterien ist genommen über 3CD gut. Das bedeutet jedoch auch, daß diese Gruppe CD-ROMs enthält, die ihrerseits maximal oder auch minimal sind. Die Bezeichnung "gut" weist lediglich eine gute Verteilung aus, nicht zwangsweise eine "Güte" schlechthin.

Die minimale Gruppen sind hingegen 5ON, d.h. die Umweltchemikalien-relevanten Online Datenbanken, und 1CD (chemierelevante textbasierte CD-ROMs), die am ungeeignetsten für die Gesamthematik sind.

In Analogie an die Auswertung der zehn einzelnen Gruppen mittels der Hasse-Diagramm-Technik (vgl. Kapitel 7.2 und 7.3 ) sollen auch bei der Betrachtung der Gruppen wieder die Einflüsse der einzelnen Bewertungskriterien mit Hilfe der W-Matrix Analyse untersucht werden.

Für acht Kriterien ergeben sich n+1, d.h. neun Fälle. Die W-Matrix Analyse wurde bereits in Kapitel 7.2.1.2.3 eingehend erläutert. An dieser Stelle wird jedoch die W-Matrix für **alle** Objekte und nicht nur für die maximalen Objekte dargestellt.

**Tabelle 65: W-Matrix Auswertung für zehn Gruppen**

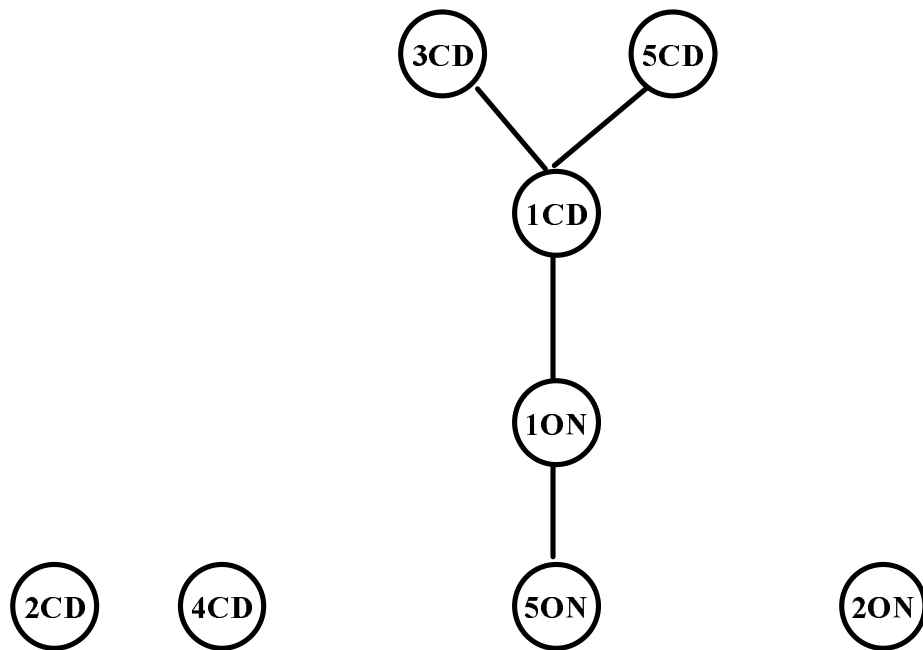
Fall 0 / Fall 1	Kriterium UM ausgelassen	0 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 2	Kriterium UP ausgelassen	0 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 3	Kriterium KO ausgelassen	2 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 4	Kriterium VE ausgelassen	11 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 5	Kriterium AN ausgelassen	6 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 6	Kriterium ID ausgelassen	3 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 7	Kriterium IP ausgelassen	5 Veränderungen im Hasse-Diagramm
Fall 0 / Fall 8	Kriterium QU ausgelassen	3 Veränderungen im Hasse-Diagramm

Abkürzungen siehe Abkürzungsverzeichnis

Aus der Tabelle wird deutlich, daß die Bewertungskriterien VE und AN die größten Einflüsse auf das Hasse-Diagramm und damit auf die Bewertung ausüben, während UM und UP keine, und KO, ID und QU nur eine geringe Bedeutung haben.

Beispielhaft soll hier der Fall 5, d.h. Auslassung des chemierelevanten Bewertungskriteriums "Anzahl an Chemikalien" in Form eines Hasse-Diagramms in Abbildung 49 dargestellt werden.

In diesem Hasse-Diagramm, welches die sieben Kriterien UM, UP, KO, VE, ID, IP und QU berücksichtigt und das Kriterium AN ausläßt, gibt es drei bzw. vier isolierte Objekte. Es handelt sich dabei um die Gruppen 2CD, 4CD und 2ON (=4ON), alles faktenbasierte Datenquellen umfassende Gruppen. Die Situation der isolierten Objekte ist unverändert geblieben, verglichen mit dem Hasse-Diagramm der vollständigen Informationsbasis. Maximale Objekte sind nach wie vor zwei CD-ROM Gruppen, 3CD (umweltrelevante textbasierte CD-ROMs) und 5CD (Umweltchemikalien-relevante CD-ROMs). Beide Objekte sind besser bewertet als die Objekte 1CD (chemierelevante textbasierte CD-ROMs), 1ON (chemierelevante textbasierte Datenbanken) und 5ON (Umweltchemikalien-relevante Datenbanken). Als Minimal ist nach wie vor die Gruppe 5ON (Umweltchemikalien-relevante Datenbanken) eingestuft. Die Anzahl der Ebenen im Diagramm des Falls 5 hat im Vergleich zur Abbildung 48 von drei auf vier zugenommen.

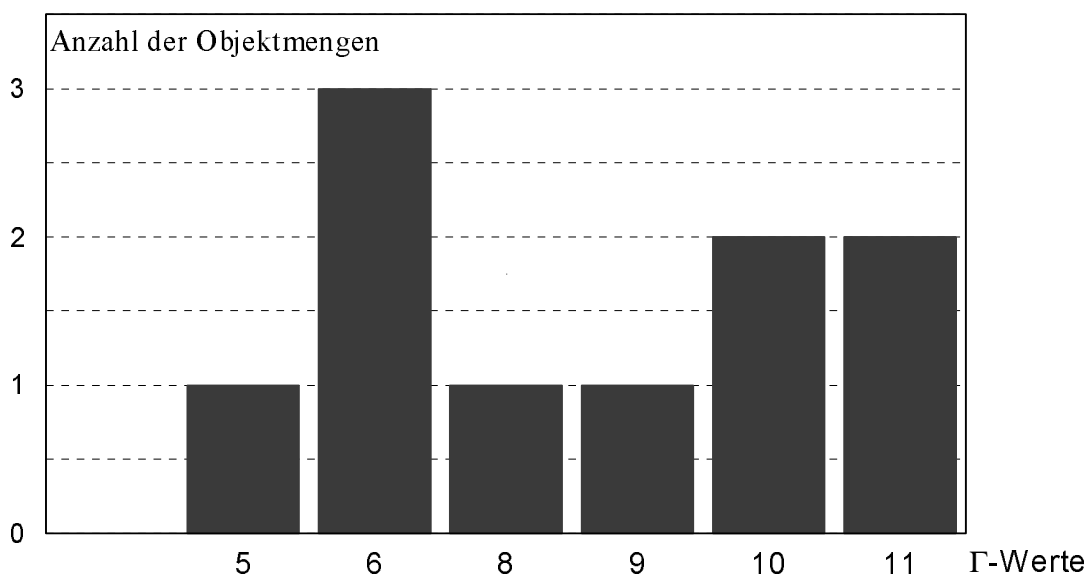


Äquivalente Objekte:  $1ON = 3ON$ ,  $2ON = 4ON$

**Abbildung 49:** Hasse-Diagramm für Fall 5: Zehn Gruppen und acht Bewertungskriterien -{AN}

### 8.3.4 Vergleich des Hasse-Diagramms mit dem Gesamtmaß $\Gamma$

Im folgenden soll eine Gegenüberstellung der Positionen der Gruppen im Hasse-Diagramm mit dem in Kapitel 6.5 eingeführten Gesamtmaß  $\Gamma$  für Objekte vorgenommen werden. Die  $\Gamma$ -Werte sind in Tabelle 64 aufgelistet. Sie reichen von 5 bis 11. Diese  $\Gamma$ -Werte werden im Balkendiagramm der Abbildung 50 nochmals dargestellt.



**Abbildung 50:** Verteilung der  $\Gamma$ -Werte auf 10 Gruppen

Über die höchsten  $\Gamma$ -Werte verfügen die Gruppen 3CD und 4CD. Sie sind  $\Gamma$ -äquivalent. Während 3CD auch ein Maximal im Hasse-Diagramm darstellt, ist die Gruppe 4CD ein isoliertes Objekt. Die Unterschiede zwischen 3CD und 4CD, so z.B. die Tatsache, daß das Objekt 3CD besser in den Kriterien UP und VE eingestuft ist, schlechter hingegen in ID und QU als die Gruppe 4CD, kommen nur bei der Auswertung mittels der Hasse-Diagramm-Technik deutlich zum Vorschein. Weiterhin haben 2ON und 4ON hohe Werte im Gesamtmaß  $\Gamma$ . Diese beiden Objekte gehören in der Hasse-Diagramm-Technik einer nicht-trivialen Äquivalenzklasse an und sind gleichzeitig isolierte Objekte.

Den niedrigsten  $\Gamma$ -Wert hat die Gruppe 5ON, die im Hasse-Diagramm auch ein Minimal darstellt. Die Objekte 1ON, 1CD und 3ON haben einen niedrigen  $\Gamma$ -Wert und sind darüber hinaus auch  $\Gamma$ -äquivalent. Im Hasse-Diagramm sind diese Objekte deutlich differenzierter plaziert. Während 1CD ein Minimal darstellt, sind 1ON (=3ON) in der Mitte des Diagramms angesiedelt. 1CD und 1ON unterscheiden sich in der Bewertung für die Kriterien AN und IP. Während 1ON besser in der "Anzahl an Chemikalien" ist, verfügt 1CD über einen höheren Wert bei den "Informationsparametern für Umweltchemikalien".

### **8.3.5 Auswertung der Ergebnisse des Kriterienmaßes $\vartheta$ für acht Bewertungskriterien**

In der letzten Zeile der Tabelle 64 ist das Kriterienmaß  $\vartheta$  aufgeführt.

Es gilt:  $3 \leq \vartheta \leq 17$ .

Dieses Kriterienmaß unterstreicht die in Kapitel 8.3.2 ausgeführten Bemerkungen zu den Qualitäten der einzelnen Bewertungskriterien. Anhand der Werte für  $\vartheta$  ergeben sich:

Die Kriterien KO (Kosten der Recherche bzw. CD-ROM) und AN (Anzahl an Chemikalien) sind die besten Kriterien, während UP (Häufigkeit der Aktualisierung) und ID (Identifikationsmerkmale für Chemikalien) die schlechtesten Kriterien darstellen. Als zufriedenstellend ist das allgemeine Bewertungskriterium UM (Umfang der Datenquelle) und das inhaltliche IP (Informationsparameter für Umweltchemikalien) anzusehen.

## **8.4 Reflexion der Bewertungskriterien**

In Kapitel 6.4 werden insgesamt 12 Bewertungskriterien eingeführt, die in der Arbeit betrachtet werden. Diese Kriterien lassen sich thematisch nach allgemeinen und inhaltlichen Fragestellungen einordnen. Bei den inhaltlichen Bewertungskriterien wird thematisch nach Chemie und Umweltschutz unterschieden. Die Bewertungskriterien umfassen demnach ein weites Spektrum und decken viele Aspekte in der Thematik "Chemie und Umweltschutz" ab. Im einzelnen wurden die allgemeinen Kriterien: Umfang der Datenbank / CD-ROM (UM), Häufigkeit der Aktualisierung (UP), Kosten der Recherche / der CD-ROM (KO), Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien (VE) sowie die inhaltlichen Kriterien: Anzahl an Chemikalien (AN), Identifikationsmerkmale für Chemikalien (ID), Chemikalien-Testdatensatz (CT), Chemikalien Entwicklung (CE), Informationsparameter für Umweltchemikalien (IP), Parameter Entwicklung (PE), Art/ Anwendung der Chemikalien (AR) und Validität/Güte der Datenquelle (QU) berücksichtigt.

### **8.4.1 Nicht-berücksichtigte Bewertungskriterien**

Neben den oben genannten Kriterien lassen sich selbstverständlich auch noch weitere definieren. Bei den allgemeinen Kriterien wäre in diesem Zusammenhang die Sprache, in der die Da-

tenbank geschrieben ist, beispielhaft zu nennen. Diese deutet auf Landes- bzw. regionale Besonderheiten hin, die beispielsweise in der Gesetzgebung von großer Bedeutung sind. Da nur ca. 5 % der wissenschaftlich-technischen Datenbanken in deutsch vorliegen, wäre das Kriterium "Sprache deutsch" von Bedeutung. In den deutschen Datenbanken ist beispielsweise zu erwarten, daß deutsche Umweltgesetze, Verordnungen und damit im Zusammenhang stehende Grenzwerte (z.B. MAK-, BAT-, TRK-Werte) zu finden sind. Diese Informationen gibt es nicht oder nur selten und unvollständig bzw. veraltet in internationalen englischsprachigen Online Datenbanken und CD-ROMs. Somit würden hier zwei nominale Merkmale "Sprache deutsch ja/nein" und "vorhanden sein von Grenzwerten ja/nein" vorliegen. Wie bei dem Bewertungskriterium IP (Informationsparameter für Umweltchemikalien) könnte die Anzahl an verschiedenen Grenzwerten noch zu einer zusätzlichen Einstufung führen. Anschließend werden die Merkmale in ein ordinales Schema integriert, wie beispielsweise beim Kriterium IP in Kapitel 6.4.3.1 ausgeführt.

Auch für die Retrievalsprache könnte ein allgemeines Kriterium definiert werden. Die Retrievalsprachen der einzelnen Hosts sind unterschiedlich in ihren Funktionalitäten. Es gibt vielfach die Möglichkeit, der statistischen Aufbereitung von Rechercheergebnissen, der Entfernung von Dubletten, der Zwischenanzeige der bisherigen Recherche-Kosten etc. Diese Funktionalitäten könnten ebenfalls zur Einstufung dieses Kriteriums mit herangezogen werden. Ebenso gibt es in einigen Retrievalsystemen zusätzlich eine menügesteuerte Oberfläche, die einfache Recherchen deutlich erleichtert. In diesem Fall könnten diese nominalen Kriterien dann, wie beispielsweise bei dem Kriterium VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) in Kapitel 6.4.1.4 ausgeführt, ordinal interpretiert werden.

Bei den inhaltlichen Bewertungskriterien könnte man insbesondere den Chemikalien-Testdatensatz variieren. Es würde sich eine Auswahl von hochproduzierten Chemikalien anbieten [Kirschner 1996]. Bei den von Kirschner aufgeführten 50 hochproduzierten Chemikalien könnte ein Bewertungskriterium derart definiert werden, daß z.B. das Vorhandensein von 40-50 Chemikalien in der zu bewertenden Datenquelle eine hohe, und das Auffinden von weniger als zehn Stoffen eine niedrige Bewertungsziffer bedeuten würden. Die Interpretation dieses ordinalen Merkmals könnte dann in Anlehnung an die beim Bewertungskriterium AN (Anzahl an Chemikalien) diskutierten Vorgehensweise in Kapitel 6.4.2.1 geschehen. Dabei ist zu beachten, daß im Gegensatz zu der im Zusammenhang mit dem Bewertungskriterium Chemikalien-Testdatensatz (CT) (vgl. Kapitel 6.4.2.3) dargelegten Vorgehensweise dieses Selektionskriterium (hier: hohes Produktionsvolumen) nicht notwendigerweise umweltrelevant ist, da Verhalten und Verbleib einer Chemikalie von den Eigenschaften der Substanz und gleichzeitig der Umwelt abhängen. Mit anderen Worten, es wäre hier als Zwischenschritt die kausale Analyse des Schicksals von Chemikalien in der Umwelt durchzuführen, wie es beispielhaft von Matthies [Matthies 1989] dargestellt wurde. Analoges gilt für die weiter unten aufgeführten spezifischeren Chemikalien-Testdatensätze. Neben dem Einsatz von hochproduzierten Chemikalien als Testdatensatz könnten auch spezielle Gruppen an Umweltchemikalien, z.B. Pestizide oder Lösungsmittel in einem Testdatensatz zusammengefaßt werden. Es würden sich beispielsweise, die als wichtige Umweltchemikalien herausgearbeitete Liste an 81 Lösungsmitteln [Howard 1990] oder die Liste an 71 Pestiziden [Howard 1991] anbieten. Gemäß der Vorgehensweise in Kapitel 6.4.2.1 beim Bewertungskriterium "Anzahl an Chemikalien" könnte auch hier wieder ein ordinales Kriterium definiert werden. Darüber hinaus wäre beispielsweise daran zu denken, spezielle Gruppen an Umweltkontaminanten, z.B., verschiedene Polycyclische Aromatische Kohlenwasserstoffe (PAKs) in einem Testdatensatz zusammenzufassen.

Auch ist eine Ergänzung bzw. Veränderung des Bewertungskriteriums Informationsparameter für Umweltchemikalien in Erwägung zu ziehen. Eine weitere Betonung einzelner Disziplinen, z.B. des Arbeitsschutzes wäre möglich. Hier könnte das Bewertungskriterium IP (Informationsparameter für Umweltchemikalien) beispielsweise um die "Spezialparameter"

Gefahrensymbole, Hinweise auf Schutzmaßnahmen, MAK-, BAT-, TRK-Werte erweitert werden.

#### **8.4.2 Qualität der Bewertungskriterien**

Die Informationsbasen für die einzelnen Gruppen an Datenquellen und deren Ausprägungsmengen werden in Kapitel 7 und dem dazugehörigen Anhang C ausführlich dargestellt. Die Fragestellung wie gut sich die einzelnen Bewertungsziffern auf die zehn behandelten Gruppen an Datenquellen (Objektmengen) abbilden lassen, wird in Kapitel 8.1 für Online Datenbanken und 8.2 für CD-ROMs ausgeführt.

Der Gesichtspunkt der Unabhängigkeit der Bewertungskriterien wird zurückgeführt auf die Frage der Korrelation in den verschiedenen Gruppierungen. Konsistent zur Vorgehensweise im Rahmen der Hasse-Diagramm-Technik wäre die Analyse der paarweisen Rangkorrelation. In dieser Arbeit wird ein anderes Konzept realisiert, nämlich das der W-Matrix (Sensitivitätsanalyse,  $W = 0$ , wenn positive Korrelation), die allerdings nur in Fallbeispielen ausgewertet und in den Tabellen zur Auswertung der Hasse-Diagramme im Anhang aufgelistet wird. Die daraus resultierenden Ergebnisse können im einzelnen in Kapitel zur W-Matrix 7.2.1.2.3 nachgelesen werden. Anhand der Daten in Tabelle 41 des Kapitel 7.2.1.1, in der die Bewertungsergebnisse der bibliographischen Online Datenbanken der Chemie-Gruppe dargestellt werden, läßt sich beispielsweise feststellen, daß in dieser speziellen Gruppe das Bewertungskriterium CT (Chemikalien-Testdatensatz) keine Rolle für die vergleichende Bewertung spielt. Es wird implizit durch die anderen Kriterien mit abgedeckt. Analoges gilt für das Kriterium UM (Umfang der Datenbank). Die W-Matrix wurde allerdings - wie im einzelnen ausgeführt - für die Maximalen berechnet. Damit ist die Unabhängigkeit der Kriterien nur auf die hauptsächlich interessierenden maximalen Datenquellen zu beziehen.

#### **8.4.3 Einteilung und Vergabe der Bewertungsziffern**

Wie in vielen Bewertungsverfahren im Umweltschutz [Hushon 1983], [GDCh 1986] [Walker 1993] ist die Interpretation der Rohdaten und die Konvertierung dieser in Bewertungsziffern in einem hohen Maße auf die Beurteilung von Experten zurückzuführen. In der vorliegenden Arbeit werden die Auswertungen der Metadatenbanken für Umweltschutz und Chemikalien DADB - Metadatenbank der Online Datenbanken und DACD - Metadatenbank der CD-ROMs für die Beurteilung herangezogen. Es wird somit auch das Expertenwissen einbezogen, d.h. die Anzahl der Bewertungsziffern und die Einstufungen enthalten ein gewisses Maß an Subjektivität.

Demnach wäre auch eine andere Einteilung z.B. in eine größere Anzahl von Bewertungsziffern, d.h. feinere Einteilung prinzipiell möglich. Es ist jedoch abzuwägen, ob die hieraus resultierende feinere Differenzierung inhaltlich zu rechtfertigen ist und nicht etwa eine Trennschärfe vorspiegelt, die nicht begründbar ist. Gleichwohl wäre nicht auszuschließen, daß eine gröbere Einteilung zu einer stellenweise anderen Einschätzung der Datenquellen führen könnte. Generell führt jedoch eine Einstufung in eine geringere Klassenzahl zu einem niedrigeren Gewicht des Bewertungskriteriums, was in der hier dargestellten Arbeit zu vermeiden war.

## 9. Zusammenfassung

Da eine ungeheure Informationsflut auf dem Gebiet der Chemie und des Umweltschutzes existiert und die Anzahl der Publikationen auch in naher Zukunft noch ständig anwachsen wird, ist ein wissenschaftlicher Ansatz zur Aufbereitung der Informationsfülle notwendig. Die genannte Proliferation hat zur Etablierung einer großen Anzahl an Datenquellen geführt. Die Zusammenstellungen der Auswahl dieser Datenquellen finden sich getrennt nach den Medien Online Datenbanken und CD-ROMs in zwei sog. **Metadatenbanken** (Datenbanken der Datenquellen), der DADB - Metadatenbank der Online Datenbanken und der DACD - Metadatenbank der CD-ROMs. In diesen Datenbanken der Datenquellen werden Informationen über Online Datenbanken bzw. CD-ROMs auf dem Gebiet der Chemie und des Umweltschutzes bezüglich der Inhalte der betreffenden Datenquelle(n), Datenbanktypen (z.B. bibliographisch oder numerisch) sowie pragmatischen Gesichtspunkten (Anbieter, Hersteller, Kosten, Aktualisierung, Medium etc.) aufbereitet vorgehalten. Sowohl der Datenbankaufbau dieser beiden Metadatenbanken als auch die Auswertung der DADB und DACD gemäß der vorkommenden Datenbanktypen werden dargestellt. Es zeigt sich beispielsweise deutlich, daß textbasierte Datenbanken nach wie vor den faktenbasierten Datenbanken von ihrem Umfang her weit überlegen sind.

Für die **vergleichende Bewertung von Datenquellen** (Online Datenbanken und CD-ROMs) auf dem Gebiet der Chemie und des Umweltschutzes, die den Kern dieser Arbeit bildet, sind Metadatenbanken die entscheidende Voraussetzung.

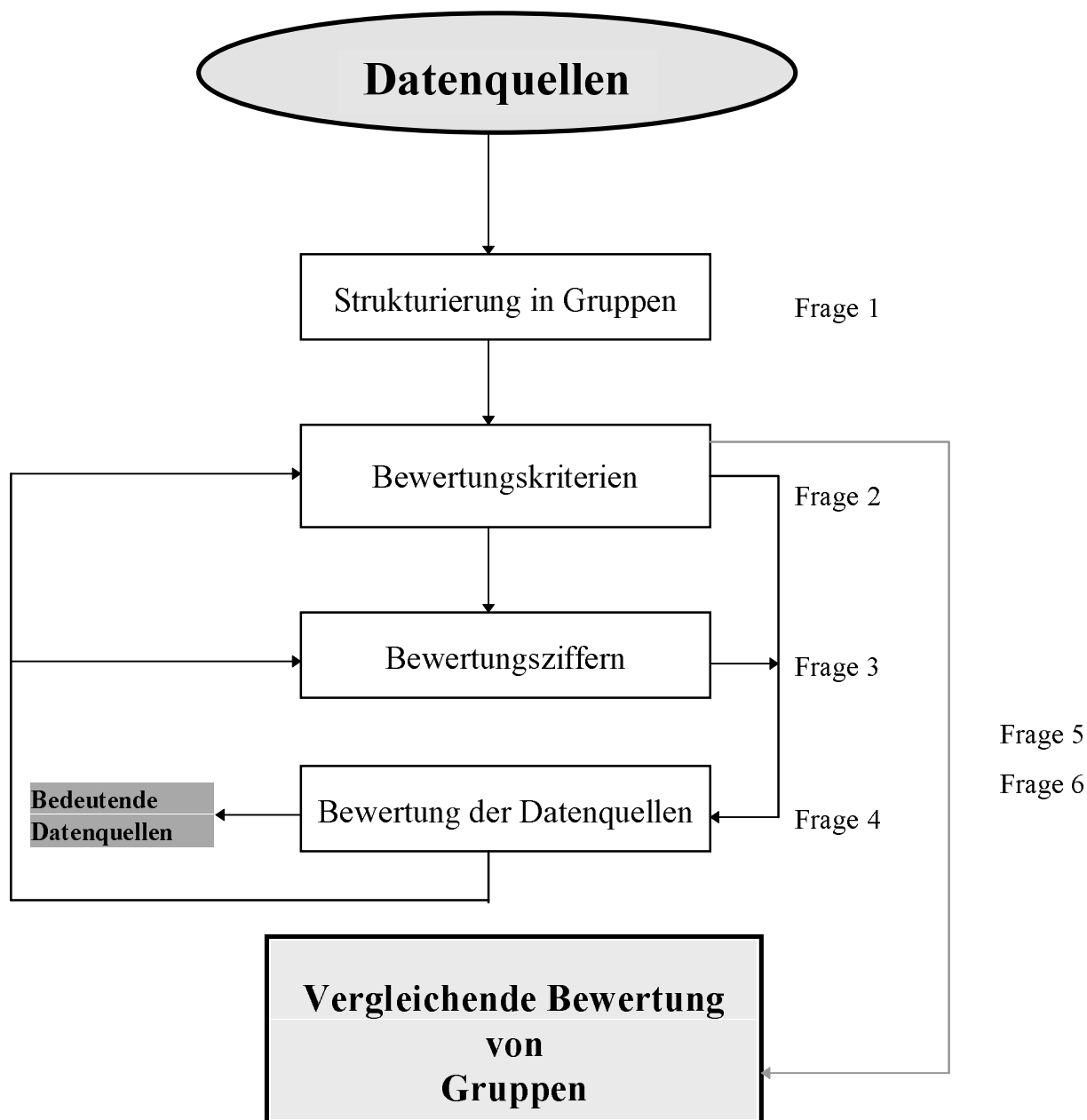
Im einzelnen sind in Beantwortung der Fragestellungen 1-6 aus dem Kapitel 1.4 "Vorgehensweise bei der Bewertung von Datenquellen" die im folgenden dargelegten Ergebnisse erzielt worden. Diese werden in Abbildung 51 schematisch dargestellt.

### 9.1 Gruppierung von Datenquellen

Da in der Metadatenbank der Online Datenbanken 453 und in Metadatenbank der CD-ROMs 347 Quellen vorgehalten werden, müssen repräsentative **Gruppen**, d.h. diejenigen Datenquellen, die einer Bewertung unterzogen werden sollen, gefunden werden. Als Gruppierungsmerkmale werden inhaltliche Themen, nämlich Chemie/Umwelt und Strukturvorgaben Text/Fakten herangezogen. Damit können fünf Gruppen an **Online Datenbanken** und fünf Gruppen an **CD-ROMs** definiert werden. Es handelt sich um chemierelevante textbasierte Datenquellen, chemierelevante Faktendatenquellen, umweltrelevante textbasierte Datenquellen, umweltrelevante Faktendatenquellen und Umweltchemikalien-relevante Datenquellen.

### 9.2 Auswahl von Bewertungskriterien

Die insgesamt 10 Gruppen an Datenquellen werden mit einer Auswahl an Bewertungskriterien eingestuft, die sowohl den Inhalt der Datenquelle betrachten als auch wichtige pragmatische Überlegungen bezüglich des Zugriffs und der Nutzungsmöglichkeiten der Datenquellen berücksichtigen. Als **Bewertungskriterien** werden vier allgemeine, nämlich **UM** (Umfang der Datenquelle), **KO** (Kosten der Recherche bzw. CD-ROM), **UP** (Häufigkeit der Aktualisierung), **VE** (Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien)



**Abbildung 51:** Ablaufschema der vergleichenden Bewertung von Datenquellen in Chemie und Umweltschutz (Erläuterungen siehe Text)

und fünf inhaltliche Bewertungskriterien, und zwar *AN* (Anzahl an Chemikalien), *ID* (Identifikationsmerkmale für Chemikalien), *IP* (Informationsparameter für Umweltchemikalien), *AR* (Art/Anwendung der Chemikalien) und *QU* (Validität/Güte der Datenquelle) angewandt. Darüber hinaus sind die Kriterien *CT* (Chemikalien-Testdatensatz), *PE* (Parameter Entwicklung) und *CE* (Chemikalien Entwicklung) für Online Datenbanken bedeutsam, da hier eine tieferegreifende Auswertung vorgenommen wird als bei den CD-ROMs.

### 9.3 Bewertungsziffern

Mit der Auswahl an Bewertungskriterien alleine ist es nicht getan. Vielmehr muß das Ausmaß, in welcher Weise eine spezielle Datenquelle einem Kriterium entspricht, durch geeignete quantitative Maße beschrieben werden. In vielen Fällen (z.B. bei den Bewertungskriterien UM = Umfang der Datenbank / CD-ROM, UP = Häufigkeit der Aktualisierung und AN = Anzahl an Chemikalien) ist eine Kategorisierung in eine diskrete Einzelbewertung sehr naheliegend, dann nämlich, wenn hinter dem Kriterium ein Kontinuum an Größen wie bei den genannten Beispielen steht. Man spricht dann von Rangordnungsmerkmalen oder **ordinalen Kriterien**. In anderen Fällen (z.B. bei den Bewertungskriterien VE = Verfügbarkeit der Datenquelle bei anderen Hosts bzw. auf anderen Medien oder ID = Identifikationsmerkmale für Chemikalien) sind Qualitäten zu beurteilen, die in geeigneter Form in ein quantitatives, d.h. ein mehrstufiges Bewertungsmaß umgesetzt werden müssen. Man spricht dann von **nominalen Kriterien**. Anhand sorgfältiger Analysen möglicher Kombinationen qualitativer Merkmale wird eine mehrstufige Bewertung erarbeitet, wie beispielsweise für die Identifikationsmerkmale für Chemikalien.

Die einzelnen nominalen und ordinalen Kriterien werden mit einem sechs Ziffern (von Null bis fünf) umfassenden Ansatz eingestuft, wobei hohe Ziffern in bezug auf ein Kriterium eine gute und niedrige Ziffern eine schlechte Datenquelle bedeuten. Mit Ausnahme des Kriteriums AR wird bei allen Bewertungskriterien der volle Umfang an **Bewertungsziffern** vergeben.

### 9.4 Bedeutende Datenquellen

Die Bewertung der Datenquellen wird mittels eines mathematischen Verfahrens aus dem Gebiet der partiell geordneten Mengen, der sog. **Hasse-Diagramm-Technik** durchgeführt. **Maximale** und **Minimale**, d.h. die Datenquellen (Objekte), die in einem gewissen Sinn Extreme sind, werden herausgearbeitet, und die Einflüsse der Bewertungskriterien auf die Hasse-Diagramme spezifiziert.

Die umfangreichen und allgemein bekannten **Online Datenbanken** z. B. BEI (BEILSTEIN Database) CAS (Chemical Abstracts), CAB (CAB Abstracts) und RTE (Registry of Toxic Effects of Chemical Substances) stellen beispielsweise Maximale dar. Die Auswertung hat darüber hinaus auch Aufschluß darüber gegeben, daß die Datenbanken, die nicht so bekannt sind, wie z.B. CBNB (Chemical Business News Database), CSN (Chemical Safety Newsbase), HSD (Hazardous Substances Databank), MER (The Merck Index Online), MSD (Material Safety Data Sheets from Occupational Health Service) und NTI (National Technical Information Service) auch hoch bewertet werden. Demgegenüber ist festzuhalten, daß die allgemein als Umweltdatenbanken anerkannten Datenquellen ULI (Umweltliteraturdatenbank), UFO (Umweltforschungsdatenbank) und POL (Pollution Abstracts) nicht als Maximale, sondern im Falle von UFO sogar als ein Minimal eingestuft werden. Die Grund hierfür ist darin zu sehen, daß diese Datenbanken nicht über die für die Umweltchemikalien-relevanten Informationsparameter verfügen.

Bei den **CD-ROMs** sind die CD-ROM Versionen der großen Datenbanken z.B. BIOSIS (BIOSIS Previews CD-ROM), RTECS (Registry of Toxic Effects of Chemical Substances) und TOX (Toxicology Literature Online CD-ROM) als Maximale eingestuft worden. CAH (CASurveyor Hazardous Materials) und CAV (CASurveyor Pollution Control and the Environment), die beiden thematischen Auszüge aus der Chemical Abstracts Online Datenbank werden sehr gut bewertet. Die CD-ROMs DID (Dictionary of Drugs on CD-ROM), DOC (Dictionary of Organic Compounds on CD-ROM), NIO (CCINFODisc Series C1, NIOSHTIC/DIDS) und PON (Pollution and Toxicology Database on CD-ROM), die nicht in dieser Form als Online Datenbank vorliegen, sind gut bewertet worden. Auffällig ist, daß die wenigen deutschen bzw. deutschsprachigen CD-ROMs, wie beispielsweise GEF (Gefahrgut

CD-ROM), GIF (Giftliste, krebserregende, gesundheitsschädliche und reizende Stoffe), SIG (Siemens Gefahrstoff-Datenbank CD-ROM) niedrig bewertet werden. Eine Ausnahme bildet die CDR (CD-Römpf), von der die Hasse-Diagramm-Technik eine Sonderstellung (ein sog. "isoliertes Element") ausweist.

Neben der Hasse-Diagramm-Technik kommt auch noch ein naheliegendes - in ähnlichen Fragestellungen oft herangezogenes - *Gesamtmaß*  $\Gamma$  zur Einstufung der Datenquellen zur Anwendung. Die Vorteile der Hasse-Diagramm-Technik gegenüber  $\Gamma$  werden evident gemacht: Die Hasse-Diagramm-Technik erlaubt eindeutigere Aussagen bezüglich der Ausprägung der einzelnen Bewertungskriterien.

### 9.5 Bedeutung der Bewertungskriterien bei der Bewertung

Anhand von Maßen, die in der Hasse-Diagramm-Technik bereitgestellt werden, zeigt sich, daß einzelne Bewertungskriterien sehr unterschiedliche Bedeutungen bei der Bewertung der verschiedenen Gruppen haben. Dieser Frage wurde in einer vergleichsweise aufwendigen Technik in einzelnen Fallbeispielen genauer nachgegangen (*Sensitivitätsanalyse*).

Für die darüber hinausführende Analyse der Bewertungskriterien werden der *Umfangindex*  $U_i$  und der *Shannon-Index*  $I_i$  eingeführt.

Bei der Auswertung der Gruppen der Online Datenbanken ist das Kriterium AN (Anzahl an Chemikalien) in allen Ansätzen als ein gut verteiltes und das Kriterium ID (Identifikationsmerkmale für Chemikalien) in allen Gruppen als ein schlecht verteiltes Bewertungskriterium herausgearbeitet worden. Die Betrachtung der fünf Gruppen an CD-ROMs ergibt, daß das Kriterium IP (Informationsparameter für Umweltchemikalien) in vier von fünf Ansätzen gut, während "Identifikationsmerkmale für Chemikalien" in vier von fünf Gruppen schlecht verteilt ist. Lediglich die Gruppe der chemierelevanten faktenbasierten CD-ROMs (2CD) liefert gute Ergebnisse für Identifikationsparameter. Hier gibt es nämlich zahlreiche CD-ROMs, die chemische Strukturformeln berücksichtigen.

### 9.6 Vergleichende Bewertung der Medien Online/CD-ROM

Die zehn Gruppen (siehe Kapitel 9.1) werden bezüglich der Güte der einzelnen Bewertungskriterien mittels der Hasse-Diagramm-Technik bewertet. Dabei zeigt sich, daß viele dieser Gruppen einzigartig sind und sich nicht mehr untereinander vergleichen lassen. Es erweisen sich jedoch die CD-ROMs gegenüber den Online Datenbanken als vorteilhafter. Die umweltrelevanten textbasierten CD-ROMs (3CD) und Umweltchemikalien-relevanten CD-ROMs (5CD) sind die beiden optimalen Gruppen; die Umweltchemikalien-relevanten Online Datenbanken (5ON) erhalten eine ungünstige Einstufung.

## 10. Ausblick

In dieser Arbeit wird eine vergleichende Bewertung von Datenquellen in der Chemie und im Umweltschutz dargestellt. Dabei werden die Medien Online Datenbanken und CD-ROMs betrachtet.

Eine zusätzliche Bewertung der *Internet Ressourcen* in Chemie und Umweltschutz ist nach einer umfassenden Aufarbeitung der in Frage kommenden Ressourcen und deren Eintrag in DAIN - Metadatenbank der Internet Ressourcen vordringlich.

Es werden in dieser Arbeit Gruppen an Datenquellen (Online Datenbanken und CD-ROMs) mit verschiedenen Bewertungskriterien bewertet. Andere *Gruppierungen*, z.B. toxikologische Datenbanken, Gefahrstoff-Datenbanken oder gemäß des Datenbanktyps Volltextdatenbanken, Reaktionsdatenbanken sollten zukünftig ins Auge gefaßt werden.

Die *Bewertungskriterien* lassen sich thematisch nach allgemeinen und inhaltlichen Fragestellungen einordnen. Bei der inhaltlichen Bewertungskriterien wird thematisch nach Chemie und Umweltschutz unterschieden. Die Bewertungskriterien umfassen demnach ein weites Spektrum und decken viele Aspekte in der Thematik "Chemie und Umweltschutz" ab. Neben den gewählten Kriterien lassen sich selbstverständlich auch noch weitere definieren. Bei den allgemeinen Kriterien wäre in diesem Zusammenhang die Sprache, in der die Datenbank geschrieben ist, beispielhaft zu nennen. Da nur ca. 5 % der wissenschaftlich-technischen Datenbanken in deutsch vorliegen, wäre das Kriterium "Sprache deutsch" von Bedeutung. Auch für die Retrievalsprache könnte ein allgemeines Kriterium definiert werden. Bei vielen Datenbankanbietern wird neben der Retrievalsprache zusätzlich eine menügesteuerte Benutzeroberfläche angeboten.

Bei den inhaltlichen Bewertungskriterien könnte man insbesondere den Chemikalien-Testdatensatz variieren. Es würde sich eine Auswahl von hochproduzierten Chemikalien anbieten. Ebenso ist bei der Beurteilung von speziellen Chemikaliengruppen z.B. Pestizide und Lösungsmittel an besonders auf diese Chemikalienklassen abgestimmte Testdatensätze zu denken. Darüber hinaus ist eine Ergänzung bzw. Veränderung des Bewertungskriteriums Informationsparameter für Umweltchemikalien in Erwägung zu ziehen. Eine weitere Betonung einzelner Disziplinen, z.B. des Arbeitsschutzes und Gesundheitsschutzes wäre möglich.

Bei den *Bewertungsziffern* könnte einerseits eine umfangreichere oder geringere Anzahl von Bewertungsziffern in Betracht gezogen werden und andererseits könnten die Grenzen der Einteilungen anders gewählt werden. Bei allen Ansätzen wäre aber ein umfassender Einsatz von Expertenwissen notwendig.

Es wäre auch daran zu denken, alle Datenquellen zusammenzufassen, d.h. alle Gruppen an Online Datenbanken und CD-ROMs und mit einer geeigneten Kriterienuntermenge, d.h. Bewertungskriterien, die sowohl für Online Datenbanken als auch für CD-ROMs gelten, eine *Clusteranalyse* durchzuführen. So könnten einander ähnliche Datenquellen auffindig gemacht werden, um so eine Verknüpfung mit der Hasse-Diagramm-Technik zu erreichen. Es würden anstelle der Bewertungsdatensätze für die einzelnen Datenquellen, solche Bewertungsdatensätze herangezogen, die die Cluster hinreichend charakterisieren. Hier bietet es sich auch an, Fuzzy-Techniken heranzuziehen, nämlich um die harte Entscheidung "Zugehörigkeit zu einem Cluster ja/nein" abzumildern.

Schließlich ist auch daran zu denken, *neuronale Netze* anhand gut bekannter Beispiele zu trainieren und die Einordnung in einer Rangskala durch neuronale Netze mit derjenigen durch die Hasse-Diagramm-Technik zu vergleichen.

## Literaturverzeichnis

Barth A. [1992] Datenbanken in den Naturwissenschaften, Eine Einführung in den Umgang mit Online Datenbanken, VCH Verlagsgesellschaft mbH, Weinheim.

Batschi W.-D. [1993] Informationsbestände des Umweltbundesamtes und ihre Nutzungsmöglichkeiten, Umwelt Technologie Aktuell, **4**, S. 354-359.

Baumewerd-Ahlmann A., Zink L. [1994] Umweltdatenbanken - Sachstand und Zukunftsperspektiven in: Page B., Hilty L.M. Umweltinformatik, Informatikanwendungen für den Umweltschutz, S. 79-101, Oldenbourg-Verlag München.

Bawden D. [1990] User-oriented Evaluation of Information Systems and Service, Gower Publishing Company Ltd. Aldershot.

Behrendt H., Brüggemann R. [1993] Modelling the Fate of Organic Chemicals in the Soil Plant Environment: Model Study of Root Uptake of Pesticides, Chemosphere, **27**, 12, S. 2325-2332.

Beilstein Informationssysteme GmbH [1996] Current Facts, [http://www.beilstein.com/productInfor/Current facts.html](http://www.beilstein.com/productInfor/Current%20facts.html).

Belke W., Graichen D., Starruss M. [1979] Nichtmetrische Klassifizierung von Informationen, Theorie und Anwendung, Akademie-Verlag, Berlin.

Benz J., Voigt K., Mücke W. [1989] Strategy for Computer-aided Searches for Information about Chemicals, Online Review, **13**, 5, S. 83-393.

Benz J., Voigt K. [1995] Indexing File System for the Set-Up of Metadatabanks in Environmental Sciences on the Internet, 19th International Online Information Meeting, 5-7 December 1995 London, Proceedings, S. 455-466, Learned Information Ltd., Oxford.

Berger G. [1991] Chemie-Informationssysteme, in: Buder M., Rehfeld W., Seeger T., Grundlagen der praktischen Information und Dokumentation, S. 593-615, K.G.Saur-Verlag, München.

Bock H.H. [1974] Automatische Klassifikation, Vandenhoeck & Ruprecht, Göttingen.

Bowker Saur [1995] The CD-ROM Catalogue 1995, Bowker Saur, East Ginstead.

Brüggemann R. [1989] in: OECD Environmental Monographs No. 27: Compendium for Environmental Exposure Assessment Methods for Chemicals, S.28-31, OECD, Paris.

Brüggemann R., Münzer B., Altschuh J. [1992] Abschätzung von expositionsrelevanten Substanzeigenschaften, GSF-Bericht 5/92, S. 31-33, GSF-Forschungszentrum für Umwelt und Gesundheit, Neuherberg.

Brüggemann R., Voigt K., Münzer B. [1994] The Technique of Hasse-Diagrams Applied in Environmental Online Databanks, Chemosphere, **29**, 4, S. 683-691.

Brüggemann R., Halfon E. [1995a], Theoretical Base of the Program "Hasse", GSF-Bericht 20/95, GSF-Forschungszentrum für Umwelt und Gesundheit, Neuherberg.

Brüggemann R., Voigt K. [1995b] An Evaluation of Online Databases by Methods of Lattice Theory, Chemosphere, **31**, 7, S. 3585-3594.

Brüggemann R., Voigt K., Halfon E. [1996a] Partially Ordered Sets - a Computerized Tool to Compare Environmental Databases, in: Denzer R., Schimak G., Russell D., Environmental Software Systems, Proceedings of the International Symposium on Environmental Software Systems 1995, S. 185-195, Chapman & Hall, London.

Brüggemann R., Voigt K. [1996b] Stability Concept for Online Databases, Chemosphere, **33**, 10, S. 1197-2006.

Bucher R. [1993] Wunsch und Wirklichkeit, Informationsverhalten und -bedarf in der chemischen Forschung - eine Marktuntersuchung, *cogito*, 1, S. 8-15.

Bucher R., Ernst S., Ilchmann G., Menge T. [1994] Stoffkonzepte in chemischen Datenbanken in: Borriars H.W., Brandt J. Demuth R, ECOINFORMA '94, 3. Fachtagung und Ausstellung für Umweltinformation und Umweltkommunikation, Band 8: Datenbanken, Umweltinformationssysteme, S. 95-110, Umweltbundesamt, Wien.

Bundesminister für Umwelt, Naturschutz und Reaktorsicherheit [1991] Was Sie schon immer über Umweltchemikalien wissen wollten, Verlag W. Kohlhammer, Stuttgart.

Cambridge Scientific Abstracts [1996] Environmental Routenet, <http://www.csa.com/routenet>.

CAS [1995] Marketing Division, CA on CD, Brochure, CAS, Columbus, OH.

Canadian Centre for Occupational Health and Safety CCOHS [1995], LIAISON, A Newsletter for the Users of CCINFO, 10, 2.

CD-Römpp [1995] Römpp Chemie Lexikon auf CD, 9. Auflage: Stuttgart: Georg Thieme Verlag (Hersteller). verfügbar: Georg Thieme Verlag.

Chapman & Hall [1995] Chemical Dictionaries CD-ROM Catalogue 1995, Chapman & Hall, London.

ChemG [1994] Chemikaliengesetz vom 16. September 1980, (BGBl. I S.1718, geändert durch Gesetz vom 25. Juli 1994, BGBl. I S.1703, 1440, 1963, 2705.

CIS [1995] A Division of PSI, The CIS News, PSI International Towson, MD, July/August 1995.

Cox J. [1991] Environment Databases 1991, An Aslib Online Guide, Aslib, The Association for Information Management, London.

DATASTAR [1995] Knight-Ridder Information, Database Catalogue 1995, Knight-Ridder Information AG, Bern.

Davey B. A., Priestley H. A. [1990] Introduction to Lattices and Order, Cambridge University Press, Cambridge.

Denzer R., Güttler R. [1994] Integration von Umweltdaten in: Hilty L.M./Page B.: Umweltinformatik, Informatikanwendungen für den Umweltschutz, S. 227-240, Oldenbourg-Verlag, München.

DIMDI [1995] Deutsches Institut für Medizinische Dokumentation und Information, grips-news Nr.4/95, DIMDI, Köln.

ecomед Verlagsgesellschaft [1996], Verlagsverzeichnis ECOBASE 1995/1996, ecomed Verlagsgesellschaft, Landberg.

Eisenbrandt G., Metzler M. [1994] Toxikologie für Chemiker, Stoffe, Mechanismen, Prüfverfahren, Georg Thieme Verlag, Stuttgart.

Elsevier Science [1995] Electronic Products, Elsevier Science Customer Service Department, Amsterdam.

Ernst S., Bucher R., Menge T. [1994] Stoffbegriff, Registrierung, Stoffthesaurus - Zur Stoffidentität in Datenbanken in: Borriars H.W., Brandt J. Demuth, R ECOINFORMA '94, 3. Fachtagung und Ausstellung für Umweltinformation und Umweltkommunikation, Band 8: Datenbanken, Umweltinformationssysteme, S.83-94, Umweltbundesamt, Wien.

Falbe J., Regitz M. [1992] Römpp Chemie Lexikon, Georg Thieme Verlag, Stuttgart.

- FIZ Karlsruhe [1995] STN International, Datenbanken aus Wissenschaft und Technik, FIZ Karlsruhe.
- Friedrichs M., Fränzle O., Salski A. [1996] Fuzzy Clustering of Existing Chemicals According to their Ecotoxicological Properties, *Ecol. Modelling.*, **85**, S. 27-40.
- Gaehtgens F., Müller P. [1995] Wege zur Integration, Internet & Hostanbindungen, *cogito*, **6**, S. 4-10.
- Gasteiger J. [1987] Software-Entwicklung in der Chemie 1, Proceedings des Workshops "Computer in der Chemie" Hochfilzen/Tirol, Springer-Verlag, Heidelberg.
- Gasteiger J., Weiske C. [1989] CHEMINFORM - An Intended Information System on Chemical Reactions, in: Learned Information Ltd., Online Information 1989, S. 147-154, Learned Information Ltd., Oxford.
- Gasteiger J., Weiske C. [1992] ChemInform - Vom gedruckten Referatedienst zur Reaktionsdatenbank, *Nachr. Chem. Tech. Lab.*, **40**, 10, S. 1114-1120.
- Gayle Alston P. [1991] Environment Online, The Greening of Databases, Part 2: Scientific and Technical Databases, *Database*, **14**, 5, S. 34-52.
- Gayle Alston P., Stoss F.W. [1992] Environment Online: The Greening of Databases, Part 3: Business and Regulatory Information, *Database*, **14**, 6, S. 17-35.
- Gefahrgut CD-ROM [1995] [CD-ROM]. Dangerous Goods CD-ROM Heidelberg: Springer-Verlag (Hersteller). verfügbar: Springer-Verlag.
- Gefährstoff CD [1995][CD-ROM]. Sorbe Sicherheitstechnische Kenndaten chemischer Stoffe. Landsberg: Ecomed Verlagsgesellschaft (Hersteller). verfügbar: Ecomed Verlagsgesellschaft.
- Georgy U. [1993] Patentdatenbanken, Angebot und Recherchetips, Hoppenstedt Verlag, cogito, Hoppenstedt Verlag, Darmstadt, 1993, **1**, S. 22-25.
- Gesellschaft Deutscher Chemiker (GDCh) [1986] Beratergremium für umweltrelevante Altstoffe (BUA), Umweltrelevante Alte Stoffe, Auswahlverfahren und Stoffliste, VCH Verlagsgesellschaft, Weinheim.
- Gesellschaft Deutscher Chemiker (GDCh) [1992] Beratergremium für umweltrelevante Altstoffe (BUA), Umweltrelevante Alte Stoffe III, Prioritätensetzung und eingestufte Stoffe der dritten Stoffliste, VCH Verlagsgesellschaft, Weinheim.
- Glock B. [1996] 10. CIC-Jubiläum in Hochfilzen, *Nachr. Chem, Tech. Lab.*, **44**, 1, S. 47-50.
- Green T. [1995] Online Information, Caught on the Web, *Online*, **19**, 1, S. 24-31.
- Graedel T.E., Hawkins D.T., Claxton L.D. [1986] Atmospheric Chemical Compounds, Sources, Occurrences and Bioassays, Academic Press, New York.
- Greve K., Häuslein A. [1994] Metainformation in Umweltinformationssystemen, in: Hilty L.M., Jaeschke A., Page B., Schwabel A., Informatik für den Umweltschutz, Proceedings 8. Symposium, Hamburg 1994, S. 169-178, Metropolis-Verlag, Marburg.
- Greve K. [1995] Metainformationssysteme und Umweltinformationen- Thesen zum Stand der Forschung und Entwicklung - in: Dollinger F., Strobl J., Angewandte Geographische Informationstechnik VII. Beiträge zum GIS-Symposium 5.-7. Juli 1995, Salzburger Geographische Materialien, Salzburg.
- Halfon E. [1983] Is There a Best Model Structure?, II. Comparing the Model Structures of Different Fate Models, *Ecological Modelling*, **20**, S. 153-163 .
- Halfon E., Reggiani M. G. [1986] On Ranking Chemicals for Environmental Hazard, *Environ. Sci. Technol.*, **20**, 11, S. 1173-1179.
- Harary F. [1976] Graphentheorie, R. Oldenbourg Verlag, München.

- Harry V., Oppenheim C. [1993a] Evaluations of Electronic Databases, Part 1: Criteria for Testing CD-ROM Products, *Online & CD-ROM Review*, **17**, 4, S. 211-222.
- Harry V., Oppenheim C. [1993b] Evaluations of Electronic Databases, Part 2: Testing CD-ROM Products, *Online & CD-ROM Review*, **17**, 6, S. 339-368.
- Hasse H., Vorlesungen über Klassenkörpertheorie [1967] Universität Marburg, Physica-Verlag.
- Hayward J. [1995] Chemical Reaction Databases: Progress in Reaction Searching, Learned Information Ltd., 19th International Online Information Meeting, London 5-7 December 1995, S. 143-155, Learned Information, Ltd., Oxford.
- Heller S.R. [1993] Chemical Information Activities: What the Future Holds, *J. Chem. Inf. Comput. Sci.*, **33**, S. 284-291.
- Heller S.R. [1996] Chemistry on the Internet - the Road to Everywhere and Nowhere, *J. Chem. Inf. Comput. Sci.*, **36**, S. 205-213.
- Herther N.K. [1995] CD-ROM at Ten Years, *Online*, **19**, 2, S. 86-93.
- HMSO [1995] CD-ROM & Disk Catalogue, HMSO Publications, Norwich.
- Hoppenstedt Verlag [1992] cogito, Sonderheft für Einsteiger, Hoppenstedt Verlag, Darmstadt.
- Howard P.H. [1990] Handbook of Environmental Fate and Exposure Data for Organic Chemicals, Volume 2: Solvents, Lewis Publishers, Chelsea.
- Howard P.H. [1991] Handbook of Environmental Fate and Exposure Data for Organic Chemicals, Volume 3: Pesticides, Lewis Publishers, Chelsea.
- Hulpke H., Koch H.A., Wagner H. [1993] Römpp Lexikon Umwelt, Georg Thieme Verlag, Stuttgart.
- Hushon J.M., Kornreich M.R. [1984] Scoring Systems for Hazard Assessment in: Saxena J., Hazard Assessment of Chemicals, Current Developments, S. 63-107, Academic Press, Orlando.
- Ihlenfeldt W.-D., Gasteiger J. [1996] Treasure Hunt: On the Track of Reusable Structural Information in the World Wide Web, in: Gasteiger J., Software-Entwicklung in der Chemie 10, Proceedings of the 10th Workshop "Computer in Chemistry" Hochfilzen/Tirol 1995, S. 403-414, Gesellschaft Deutscher Chemiker, Fachgruppe Chemie-Information-Computer, Frankfurt.
- Jacso P. [1990] The Ideal CD-ROM Workstation for the 1990s, in: Learned Information Ltd., Online Information 90, 11-13 December 1990, London, S. 25-32, Learned Information Ltd., Oxford.
- Jochum C., Bucher R [1991] Die neuen Komponenten des Beilstein Informations-System in: Mitteilungsblatt GDCH, Fachgruppe Chem.-Inf.-Comp., **19**, S. 34-46.
- Joint Research Center [1997] Environmental Institute, ECDIN, <http://ulisse.ei.jrc.it/Ecdin/Ecdin.html>.
- Keitz von S., Keitz von W., Gerlach H. [1993] Modernes Online-Retrieval, VCH Verlagsgesellschaft mbH, Weinheim.
- Kind J. [1991] Online-Dienste, in: Buder M., Rehfeld W., Seeger T., Grundlagen der praktischen Information und Dokumentation, S. 366-401, K.G.Saur-Verlag, München.
- Kirschner E.M. [1996] Growth of Top 50 Chemicals Slowed in 1995 From Very High 1994 Rate, *Chemical & Engineering News*, April 1996.

- Klöpffer W. [1994] Environmental Hazard - Assessment of Chemicals and Products, Part 1: General Assessment Principles, *ESPR, Environ. Sci. & Poll. Res.* **1**, 1, S. 47-53.
- Knight-Ridder Information Inc. [1995a] DIALOG Database Catalogue 1995, Knight-Ridder Information, Inc., Mountain View, CA.
- Knight-Ridder Information Inc. [1995b] DataStar SFr Price List April 1995, Knight-Ridder Information, Inc., Mountain View, CA.
- Knight-Ridder Information Inc. [1996] KR Information OnDisc, At a Glance, KROnDisc, Oxford.
- Koch H., Neckel H. [1995] Umweltdatenbanken bei STN International - ihre Nutzung in der Praxis, *Umweltwissenschaften und Schadstoff-Forschung*, **7**, 6, S. 352-358.
- Korte F. [1987] Lehrbuch der Ökologischen Chemie, Grundlagen, Konzepte für die ökologische Beurteilung von Chemikalien, Georg Thieme Verlag, Stuttgart.
- Krieger J. H. [1995] Chemistry Sites Proliferate on the Internet's World Wide Web, *Chemical & Engineering News*.
- Lankenau K. [1992] STN International Germany, Economic Aspects of Database Production, Distribution of Information and Database Usage, in: Learned Information Ltd. Oxford, International Centre for Scientific and Technical Information Moscow, 2nd East-West Online Information Meeting, 28 September - 2 October, 1992, Moscow, Russia.
- Larson R.A., Weber E.J. [1994] Reaction Mechanisms in Environmental Organic Chemistry, Lewis Publishers, Boca Raton.
- Lawson A. [1995], Organic Synthesis Pathways in CrossFire in: 19th International Online Information Meeting, S. 131-142, Learned Information Ltd., Oxford.
- Lohse S., Bussian B.-M. [1994] Öffentlich zugängliche Datenbanken des Umweltbundesamtes, in: 10. Münchner Gefahrstofftage 2.-5.November 1994, S. 249-262, Ecomed-Verlagsgesellschaft, Landsberg.
- Luckenbach R. [1996] Past Perfect, Present Perfect, Future Perfect - Quality Assessment and Quality Control Mechanisms at Beilstein, *J.Chem.Inf.Comput.Sci*, **36**, S.923-929.
- Ludl H., Schöpe K., Mangelsdorf I. [1995] Searching for Information on Chemical Substances in Selected Biomedical Bibliographic Databases, *Chemosphere*, **31**, 2, S. 2611-2628.
- Matthies M., Brüggemann R., Münzer B., Schernewski G., Trapp S. [1989] Exposure and Ecotoxicity Estimation for Environmental Chemicals (E4CHEM): Application of Fate Models for Surface Waters and Soils, *Ecol.Model.*, **47**, S. 115-130.
- Matthies M., Behrendt H., Trapp S. [1991] Modelling and Model Validation for Exposure Assessment of the Terrestrial Environment, in: Frehse H., Pesticide Chemistry Advances in International Research, Development, Legislation, S. 434-444, VCH Verlagsgesellschaft, Weinheim.
- Matthies M., Altschuh J., Brüggemann R., Voigt K. [1992] Data Needs, Data Availability and Data Estimation for Environmental Exposure and Hazard Assessment, in: Steinberg C., Kettrup A. (Eds.), Proceedings International Symposium on Ecotoxicology - Ecotoxicological Relevance of Test Methods, November 19-20, 1990, Neuherberg, GSF-Bericht 1/92, GSF-Forschungszentrum für Umwelt und Gesundheit GmbH, Neuherberg.
- Meckler Managing Information Technology [1995] CD-ROMs in Print 1995: An International Guide to CD-ROM, CD-I, CDTV, Multimedia & Electronic Book Products, Meckler Managing Information Technology, Westport.
- Metcalf R.L., The Problem of Living in a World Contaminated with Chemicals, in: Oak Ridge National Laboratory, Access and Use of Information Resources in Accessing Health Risks From Chemical Exposure, June 27-29, 1990, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1993.

- Möhring [1995] Technische Universität Berlin, Fachbereich 3 Mathematik, private Mitteilung an Dr. Brüggemann.
- Mücke W., Voigt K., Schulze H. [1987], Approach for Identification of Hazardous Environmental Chemicals in Food, Toxicological and Environmental Chemistry, **13**, S. 129-140.
- Mücke W., Voigt K., Benz J. [1989] Früherkennung von Umweltchemikalien in Lebensmitteln anhand spezifischer Stoffkriterien, Z. Lebensm. Unters Forsch, **189**, S. 99-102.
- Online Inc. [1992] Environment Online: The Greening of Databases, The Complete Environmental Series from DATABASE Magazine, Online Inc., Wilton.
- Opphuizer A., Hutzinger O. [1982] Multi-criteria Analysis and Risk Assessment, Chemosphere, **11**, S. 675-678.
- Orton, D. [1995] Online Searching in Science and Technology, The British Library, London.
- Ottahal A. [1989] Umwelt-Datenbank-Führer, Verlag TÜV Rheinland, Köln.
- Page, B., Hilty L.M. [1994] Umweltinformatik als Teilgebiet der Angewandten Informatik in: Page B., Hilty L.M. Umweltinformatik, Informatikanwendungen für den Umweltschutz, S. 13-26, Oldenbourg-Verlag München.
- Page B. [1996] Environmental Informatics, Towards a New Discipline in Applied Computer Science for Environmental Protection and Research in: Denzer R., Schimak G., Russell D., Environmental Software Systems, Proceedings of the International Symposium on Environmental Software Systems 1995, S. 3-22, Chapman & Hall, London.
- Pilkington S., Rhodes R.G. [1992] Online Engineering Databases 1992, An Aslib Online Guide, Aslib, The Association for Information Management, London.
- Pillmann W. [1994] Austausch von Umweltinformationen in: Hilty L.M./Page B.: Umweltinformatik, Informatikanwendungen für den Umweltschutz, S. 37-50, Oldenbourg-Verlag, München.
- Precht M. [1987] Bio-Statistik, Eine Einführung für Studierende der biologischen Wissenschaften, R. Oldenbourg Verlag, München.
- Rao C.R. [1982], Diversity and Dissimilarity Coefficient: A Unified Approach, Theoretical Population Biology, **21**, S. 24-43.
- Reggiani M. G., Marchetti F. E. [1975] On Assessing Model Adequacy, IEEE Trans. Systems Man and Cyber, SMC-5, S. 322-330.
- Reichardt T. [1996] Environment On-Line - a Guide to Internet Resources, Environ. Sci. Technol., **30**, 2, S. 76-81.
- Seeger T. [1991] Zur Entwicklung der Information und Dokumentation, in: Buder M., Rehfeld W., Seeger T., Grundlagen der praktischen Information und Dokumentation, S. 9-53, K.G.Saur-Verlag, München.
- Schinzer D. [1993] Drei neue Reaktionsdatenbanken, Nachr. Chem. Tech. Lab., **41**, 7/8, S. 826-828.
- Schultz-Krutisch T. [1992] Das Informationsangebot des Umweltbundesamtes: Datenbanken und Informationssysteme, in: Dahm M., Schaarschmidt M., Umweltinformationssysteme - Berichte zu Tagungen der COMM 91, S. 11-22, inno-FA Fachveranstaltungen GmbH, Neumünster.
- Scientific Consulting [1995a] Dr. Schulte-Hillen BDU, Handbuch lieferbarer CD-ROMs, 1994, Verlag Hoppenstedt & Co, Darmstadt.

Scientific Consulting [1995b] Dr. Schulte-Hillen BDU, Handbuch der Datenbanken für Naturwissenschaft, Technik, Patente, 1995, Verlag Hoppenstedt & Co, Darmstadt.

SilverPlatter Information GmbH [1996] SilverPlatter Directory, Towards a Worldwide Library, SilverPlatter Information GmbH, Berlin.

Springer Verlag [1995] Software, CD-ROMs, Electronic Journals, Fall '95, Springer Electronic media, Heidelberg.

Staud J. [1991] Online Datenbanken: Aufbau, Struktur, Abfragen, Addison-Wesley-Verlag, München.

STN International [1993] Finding and Verifying CAS Registry Numbers on STN, S. 6, STN International, Columbus.

STN International [1995a] Caplus - When Timeliness Counts, STNews, **11**, 1, S.1-3.

STN International [1995b] CA Database Description Cream Sheet, Chemical Abstracts Service, Columbus, OH, July 1995.

STN International [1995c] TOXLINE Database Description Cream Sheet, National Library of Medicine, Bethesda, MD.

STN International [1995d], BEILSTEIN Database Description Cream Sheet, Beilstein Informationssysteme, Frankfurt.

STN International [1995e], JANAF Database Description Cream Sheet, United States Department of Commerce, Gaithersburg, MD.

STN International [1995g] DETHERM Database Description Cream Sheet, DECHEMA e.V., Frankfurt.

STN International [1996] For Chemical Information It's STN, STNews, **12**, 2, S.2-9.

Stoss F.W. [1991] Environment Online, The Greening of Databases, Part 1.: General Interest Databases, Database, **14**, 4, S. 13-27.

TFPL Publishing [1996] The CD-ROM Directory 1996, TFPL Publishing, London.

The Royal Society of Chemistry [1996] Online, CD-ROM, PC Products, Internet, 1996, The Royal Society of Chemistry, Information Service, Cambridge.

Trotter W.T. [1992] Combinatorics and Partially Ordered Sets, Dimension Theory, The John Hopkins University Press, London.

Umweltbundesamt [1991] UMPLIS Informations- und Dokumentationssystem Umwelt, Umweltthesaurus - Hierarchischer Teil - Umweltbundesamt, Berlin.

University of Kentucky [1997] Department of Chemistry, Where to Find MSDS on the Internet, <http://www.uky.edu/Arts/Sciences/Chemistry/Resources/MSDS.HTML>.

Urrutia J. [1989] Partial Orders and Euclidian Geometry in: Rival I., Algorithms and Order, NATO ASI Series C: Mathematical and Physical Sciences, Kluwer Academic Publishers, Dordrecht, **255**, S. 387-434.

U.S. Department of Commerce [1995] Environment Highlights Summer 1995, EPA Environmental Databases, U.S. Department of Commerce, National Technical Information Service, Springfield, VA.

Vogt S. [1992] CD-ROM and Online - Competitors or Compatriots, in: Learned Information Ltd., Online Information 92, December 1992, London, S. 33-45, Learned Information Ltd., Oxford.

Voigt K., Rohleder H. [1990] Datenquellen für Umweltchemikalien, 2. Auflage, ecomed-Verlagsgesellschaft, Landsberg.

Voigt K., Pepping T. [1992] Informationssystem Umweltchemikalien, Endbericht, GSF-Bericht 10/92, GSF-Forschungszentrum für Umwelt und Gesundheit, Neuherberg.

Voigt K., Brüggemann R [1993] Metadatabases of Data-Sources for Environmental Chemicals, Proceedings 17th International Online Information Meeting, 7-9 December 1993 London, S. 495-506, Learned Information Ltd., Oxford.

Voigt K., Brüggemann R. [1994] Ansatz eines Scoring-Systems für chemische/umweltrelevante Online Datenbanken in: Jochum C., Software-Entwicklung in der Chemie 8, GDCh-Fachgruppe Chemie-Information-ComputerSoftware-Entwicklung in der Chemie 8, Proceedings of the 8th Workshop "Computer in Chemistry" Seeheim-Jugendheim/Darmstadt November 17-19, 1993, S. 1-15, Gesellschaft Deutscher Chemiker (GDCh), Frankfurt/Main.

Voigt K., Brüggemann R. [1995a] Toxicology Databases in the Metadatabank of Online Databases TOXICOLOGY, **100**, S. 225-240.

Voigt K. [1995b] Informationssystem Umweltchemikalien, Endbericht, Bayerisches Staatsministerium für Landesentwicklung und Umweltfragen, München.

Voigt K., Poetzsch E., Reich M., Schultz-Krutisch T., Zass E. [1995c] German Databases for National and International Users, Using Environmental Information as an Example, Online & CD-ROM Review, **19**, 2, S. 71-77.

Voigt K., Gasteiger J., Ihlenfeldt W.-D., Page B., Specht K., Umstätter W. [1996a] German Experts' Views and Ideas about Information on the Internet, Online & CD-ROM Review, **20**, 3, S. 124-132.

Voigt K., Benz J. [1996b] Umweltschutz im Internet, Datenquellen zu ökologischen und chemischen Fragestellungen, cogito, **2**, S. 5-11.

Walker J.D. [1993] The TSCA Interagency Testing Committee's Approaches to Screening and Scoring Chemicals and Chemical Groups: 1977-1983 in: Oak Ridge National Laboratory, Access and Use of Information Resources in Assessing Health Risks From Chemical Exposure, S. 71-93, Oak Ridge National Laboratory, Oak Ridge.

Warr W.A. [1993] New Chemical Databases on CD-ROM, Database, February, S. 59-67.

Warr W.A. [1994a], Chemical Structure Handling on CD-ROM: Recent Developments and Future Prospects in: Learned Information Ltd., Online Information 94, 18th International Online Information Meeting, Online Information 1994, London 6-8 December, S. 131-137, Learned Information, Ltd., Oxford.

Warr W.A. [1994b] Drugs and Chemical Reactions on CD-ROM, Database, June, S. 56-63.

Weckend E. [1994] Patentinformation als Basisinformation für Umweltschutz und Umweltforschung, in: Borriers H.W., Brandt J., Demuth R., ECOINFORMA '94 3. Fachtagung und Ausstellung für Umweltinformation und Umweltkommunikation, Band 8, S. 33-41, Umweltbundesamt, Wien.

WEKA [1992] Software für Wirtschaft und Verwaltung GmbH, LARS Bedienungsanleitung und Handbuch, WEKA, Frankfurt/Main.

Williams, M.E. [1995] Highlights of the Online/CD-ROM Database Industry: The Information Industry and the Role of the Internet, in: Williams M.E., 16th National Online Meeting, Proceedings 1995, New York, May 2-4, 1995, S. 1-6, Learned Information, Inc., Medford NJ.

Williams, M.E. [1996] Highlights of the Online Database Industry and the Internet, in: Williams M.E., 17th National Online Meeting, Proceedings 1996, New York, May 14-16, 1996, S. 1-4, Learned Information, Inc., Medford NJ.

WIPO [1994] World Intellectual Property Organization, International Patent Classification, 6th Edition, Vol. 3, Section C, WIPO, Geneva.

Wittmann A. [1991] Patentedokumentation in: Buder M., Rehfeld W. Seeger T., Grundlagen der praktischen Information und Dokumentation, S. 522-533, KG Saur, München.

Wolman Y. [1995] Chemical Reaction Databases, Learned Information Ltd., 19th International Online Information Meeting, London 5-7 December 1995, S. 157-162, Learned Information, Ltd., Oxford.

Young Marcaccio K. [1996] Gale Directory of Databases 1995, Gale Research International Ltd., Andover.

Yu P.L. [1973] Introduction to Domination Structure and Multicriteria Decision Problems, Multiple Criterion Decision Making, 1973, S. 249-261.

Zass E. [1994] Die Rolle der Chemie-Information in der Chemie-Ausbildung, in: Gesellschaft Deutscher Chemiker (GDCh), Fachgruppe Chemie-Information-Computer, Mitteilungsblatt Nr. 29, GDCh, Frankfurt, Mai 1994.

Zass E. [1995] Chemie-Datenbanken auf CD-ROM, Nachr. Chem. Tech. Lab., **43**, 5, S. 544-546.

Zupan J., Gasteiger J. [1993], Neural Networks for Chemists, An Introduction, VCH Verlagsgesellschaft, Weinheim.

## Anhänge

### Anhang A. Recherchebeispiele zu Datenquellen in Chemie und Umweltschutz

#### A.1 Faktenbasierte Datenbanken

Hier werden mit Hilfe von Datensätzen Objekte wie Chemikalien, Produkte, Firmen, Forschungsvorhaben beschrieben. Es handelt sich um einen Datenbanktyp, der von allen Online Datenbanken dem konventionellen am ähnlichsten ist, da seine Modellierung datensatzorientiert ist. Bewußt ausgenommen werden alle Datenbanken, die sich auf Text beziehen. Diese faktenbasierten Datenbanken beschreiben ihren sog. "Weltausschnitt" vor allem mit Attributen, Merkmalen, Nomenklaturen und kurzen Texten. Ziel ist hier eine einfache Beschreibung durch diese Informationstypen. Meist wird diese nicht durch längere Texte ergänzt.

##### A.1.1 Faktendatenbanken ( Modelldatenbanken der Naturwissenschaften)

Dieser Datenbanktyp wird im Hauptteil der Arbeit beschrieben (Kapitel 2.5.2).

##### A.1.2 Chemikaliennamen-Verzeichnisse

Es werden mit Hilfe von Datensätzen Objekte, in diesem Fall Chemikalien beschrieben. Für die Recherche nach Informationen über Chemikalien spielen diese Art der faktenbasierten Datenbanken über Chemikalien, die im folgenden Beispiel erklärt werden, eine Schlüsselrolle.

Die *Chemikaliennamen-Verzeichnisse* dienen zur Identifikation von Chemikalien, halten Synonyma, Summenformel, Molekulargewicht und z.T. auch Strukturformeln vor. Verweise auf andere Datenbanken, in denen man weitere Informationen über die gewünschten Chemikalien findet, werden angeboten.

Als Recherchebeispiel wird die Datenbank CHEMLINE bei dem Host DIMDI gewählt.

In CHEMLINE wird nach Identifikationsmerkmalen für die Chemikalie Benzo[a]pyrene gesucht. Aus dem Recherchebeispiel kann man ersehen, daß neben dem systematischen Namen der Chemikalie, die CAS-Nummer, die Synonyma, das Molekulargewicht, die Anzahl der Ringe, die Größe der Ringe und vor allem im sog. Locator Feld, die Verweise auf andere Datenbanken bei DIMDI, in denen sich weitere Informationen über diese Chemikalie befinden, angegeben werden.

#### Recherchebeispiel A1: DIMDI File : CHEMLINE Item: 000050328

BASE COMMAND ACCEPTED FOR CH76;CHEMLINE;LAST-UPDATE=8.06.1995  
Copyright 1993 by the American Chemical Society (ACS) and NLM.

1.00/000001 DIMDI: -CHEMLINE /COPYRIGHT 1993 by ACS and NLM  
ND: 000050328 BASE: CH76 EM: 9502 MD: 950529 NU: 1075695

**NAME** : Benzo(a)pyrene (8CI)(9CI)

**CR CAS NUMBER**: 50-32-8

GRC GROUP OF COMPOUND : Mutation data (RTECS)

Reproductive Effect (RTECS)

Skin / Eye Irritant (RTECS)

Tumor data (RTECS)

NOTE : Potent mutagen and carcinogen. (MESH)

**SY SYNONYMS**: AI3-50461 (NLM); Benz(a)pyrene (HSDB:RTECS);

Benz (a)pyrene (NLM); Benzo(d,e,f)chrysene (RTECS); BP (HSDB:RTECS);  
 CCRIS 76 (CCRIS); HSDB 2554 (HSDB); RCRA waste number U022 (RTECS);  
 3,4-Benz(a)pyrene (HSDB:RTECS); 3,4-Benzopyrene (MESH:RTECS);  
 3,4-Benzpyrene (MESH:RTECS); 3,4-Benzpyrene (carcinogen) (CAS);  
 6,7-Benzopyrene (RTECS)

**SY** MESH HEADING: Benzo(a)pyrene

**SY** NAME OF SUBSTANCE : Benzo(a)pyrene (CCRIS:HSDB:MESH)

**MF** MOLECULAR FORMULA : C20-H12

**NOR** NUMBER OF RINGS : 5

**RS** RING SIZE : 6,6,6,6,6

**RE** RING ELEM. ANALYSIS : C6-C6-C6-C6-C6

**LO** LOCATOR: AIDSLINE; CANCERLIT; CCRIS; DART; DSL; EINECS; EMIC;  
 EMICBACK; ETICBACK; GENETOX; HSDB; HSTAR; IRIS;  
 MEDLINE; MED80; MED85; MED90; MESH; RTECS; TOXLINE;  
 TOXLINE65; TOXLIT; TOXLIT65; TSCAINV; MED75

Fast jeder Host hat in seinem Datenbankangebot eine solche faktenorientierte Datenbank. Sie gibt neben der Identifikation der Chemikalie auch Hinweise zum Auffinden weiterer Informationen in anderen Datenbanken.

### A.1.3 Firmenverzeichnisse und andere Verzeichnisse

Hier werden mit Hilfe von Datensätzen Objekte in diesem Fall beispielsweise Hersteller von Chemikalien, Forschungseinrichtungen, Verbänden, Behörden etc. beschrieben. Interessant im Zusammenhang mit der gestellten Thematik "Umweltschutz und Chemikalien" ist die Datenbank CERCLIS (Database of Hazardous Waste Sites), angeboten vom Host CIS.

### Recherchebeispiel A2: CIS File: CERCLIS Item: COD007069974

CIS (Version 5.0) 28-JUL-1995 12:17:02.07  
 CERCLIS/PRPS(SETS) - Version 5.00/1.28 (May, 1995) (\$105/Hr.)  
 Latest Database Update: Oct, 1994 (Hazardous Waste Sites)

CERCLIS Accession Number COD007069974

(EPAID) EPA ID: COD007069974

(REG) Region: 08

**(ID) SITE IDENTIFICATION INFORMATION:**

**Primary Name:**

(NAME) WESTERN TANNING INC

(STREET) 1454 HWY 50

(CITY) DELTA

(STATE) CO

(ZIP) 81416

(COUNTY) DELTA

(LAT) 38 deg. 45 min. 34.0 sec.

(LONG) 108 deg. 05 min. 23.0 sec.

(LLSRC) R

(DESC) TYPES OF HAZ.MAT'LS DISPOSED: REACTIVE SOLID  
 WASTES, 1,2-DI- CHLOROBENZENE, SULFIDES, ACIDS &  
 BASES USED IN TANNING DYES & FINISHERS. 3 PONDS &  
 A WASTE DUMP WERE LOCATED ON-SITE. BEGAN OPRN ON  
 UNK.DATE AS TANNERY FOR HIDE PREPARATION.

(CNTYCD) County Code: 029

(CONGDS) Congressional District: 03

(FED) Federal Facility Indicator: N [No]

(**FEDDOC**) Federal Docket: N [Not Covered by Federal Facility Hazardous Waste Compliance Docket]  
 (**OWNER**) Ownership Indicator: PR [Private]  
 (**SINCAT**) Site Incident Category: L [Landfill]  
 (**CLASS**) Classification: ND [No Determination]  
 (**LUPD**) Last Update Date: 12-11-92  
 (**NAMSRC**) Name Source: H [Hazardous Waste Data Management System]  
 (**STAT**) Status: N [Not Currently or Formally on Proposed or Final NPL]  
 (USGSU) US Geological Survey Hydrologic Unit: 14020006  
 (**RCRA**) RCRA Flag: E [Environmental Priority Initiative]

(**OI**) **OPERATIONS INFORMATION:**  
 (OPUN) Operable Unit: SITE EVALUATION/DISPOSITION  
 (OPDATA) Operable Unit Data:  
 CUR PLAN CUR PLAN ACTUAL CUR PLAN CUR PLAN ACTUAL  
 EVENT EVENT START START START COMPLETE COMPLETE COMPLETE  
 TYPE LEAD DATE QUARTER DATE DATE QUARTER DATE

DS1	F			07-01-80			
	PA1	S				02-01-84	
	No Further Remedial Action Planned						
	PA2	F				09-28-92	

Da zu den beiden Testdatensatz Chemikalien keine Dokumente gefunden wurden, wird hier ein Beispiel gewählt, in welchem eine Sondermülldeponie beschrieben wird, die 1,2-Dichlorbenzol lagert. Neben den Angaben über den Betreiber und den Standort, gibt es das Datenfeld "DESC (Types of Hazardous Materials Stored)". Es wird deutlich, daß es kein Datenfeld für die CAS-Nummer gibt. Es ist in der CERCLIS Datenbank in diesem Zusammenhang nur die Recherche mit dem Chemikaliennamen möglich.

#### A.1.4 Forschungsdatenbanken

##### Recherchebeispiel A3: DataStar File: UFOR Item: 00035754

Welcome to Knight-Ridder Information

D-S/UFOR/UMWELTFORSCHUNGSDATENBANK '74- SESSION 248  
 COPYRIGHT BY Umweltbundesamt, Berlin, Germany.

1 UFOR  
 AN 00035754.

**PL** Dott-W, Prof.Dr.rer.nat.

**PI** Technische Universitaet Berlin, Fachbereich 21 Umwelttechnik,  
 Fachgebiet Hygiene  
 Amrumer Str. 32  
 D-13353 Berlin  
 Berlin  
 Tel: (030) 31427530.

**TI** Bilanzierung des Abbaus von Pentachlorphenol, **Hexachlorbenzol** und dichlorierten Biphenylen unter aeroben und anoxischen Bedingungen mit Nitrat als H-Akzeptor.

**CC** CH20 = Umweltchemikalien/Schadstoffe: Wirkung von Belastungen; WA23 = Wasserbelastungen: Wirkung auf unterirdisches Wasser; WA30 = Methoden der Wasseruntersuchung, Messtechnik, Datensammlung.

**DE** Grundwassergefahrdung; Chlorkohlenwasserstoff; Kohlenwasserstoff- aromatisch; Mineralisation; Stickstoff; Grundwasser; Nitrit; Substratabbau; Distickstoffoxid; Persistenz; Aerobe-Bedingung; Abbau- aerob; Biphenyl; Pentachlorphenol; **Hexachlorbenzol**; Denitrifikation; Grundwasserverunreinigung; Schadstoffwirkung; Schadstoffabbau;

**HC** Ufokat 92\*CH20-007.

**AB** Chloraromatische Verbindungen werden nicht nur vermehrt in Oberflaechengewassern sondern auch im Grundwasser nachgewiesen und bedrohen damit ein lebenswichtiges Reservoir fuer die Trinkwassergewinnung

(1). Ziel des Forschungsvorhabens ist es, zu untersuchen, ob persistente wenig wasserlösliche Chloraromaten wie z.B. **Hexachlorbenzol**, Pentachlorphenol sowie verschiedene Dichlorobiphenyle, die als Grundwasserkontaminanten eine Rolle spielen, mikrobiell abgebaut werden können. In der Literatur wird der Abbau dieser Verbindungen unter aeroben bzw. anaeroben Bedingungen beschrieben. Wenig bekannt ist bisher, ob die genannten Stoffe unter anoxischen Bedingungen mit Nitrat als H-Akzeptor mineralisiert werden können, obwohl dieser Abbaupfad bei Sauerstoffmangel im Untergrund große Bedeutung zukommen könnte. Bei den geplanten Untersuchungen sollen Bakterienmischkulturen aus verschiedenen Umweltbiotopen mit den genannten Verbindungen als einziger C- und Energiequelle unter aeroben und anoxischen Bedingungen auf ihre Abbauleistung untersucht werden. Im Vordergrund der Untersuchungen steht dabei die Bilanzierung des Abbaus über die Erfassung des Substrat- bzw. des Nitrat-Verbrauchs sowie der Nitrit-, N<sub>2</sub>O-, N<sub>2</sub>- und der CO<sub>2</sub>-Entstehung. Ergibt sich aus der Bilanz, dass eine Mineralisierung unvollständig ist, so sollen angereicherte mikrobielle Metabolite isoliert und identifiziert werden.

**DT** 1-3-91 - 28-2-93.

**SI** Deutsche Forschungsgemeinschaft, (FKZ DO 237/8-1).

**ED** 940525.

Es werden Forschungsprojekte beschrieben. Als Beispiel für diesen Typ der faktenbasierten Datenbanken dient die Datenbank UFOR (Umweltforschungsdatenbank), eine Datenbank, die Informationen über Forschungsprojekte auf dem Gebiet der Naturwissenschaften und Technik in Deutschland, Österreich und der Schweiz vorhält. In diesem Fall geht es also um die Fragestellung, gibt es Forschungsprojekte über Hexachlorbenzol. Die Recherche wird bei dem Host DataStar durchgeführt.

Der gezeigte Datensatz umfaßt die Felder Projektleiter (PL), Performing Institution (PI) = durchführende Institution, Titel des Vorhabens (TI). Die Datenfelder Classification Code (CC) und Deskriptors (DE) erschließen den Inhalt des Dokumentes. Sowohl in der Umweltforschungsdatenbank als auch in der Umweltliteraturdatenbank (siehe A.2.1) wird dieselbe Klassifikation verwendet. Die Felder der Klassifikationen sind qualitative Merkmals-/Nomenklaturfelder und enthalten mehrere Eintragungen. Man spricht in diesem Zusammenhang dann von **Mehrfacheinträgen**. Im Abstract (AB) Feld wird eine Kurzbeschreibung des Forschungsvorhabens gegeben.

### **A.1.5 Strukturdatenbanken (Formalismendatenbanken der Naturwissenschaften)**

Jeder Eintrag in ein Feld eines Datensatzes ist geprägt durch ein Modellierungswerkzeug, z.B. die natürliche Sprache, das Attributkonzept, den Zahlenraum (z.B. durch ein quantitatives Merkmal), durch ein mathematisches Modell usw. oder durch eine formale Sprache.

Datenbanken, die durch eine solche formale Sprache bestimmt werden, werden hier kurz angesprochen. Diese sind jedoch nicht zahlreich. Ein sehr bekanntes Beispiel stellt die Datenbank REGISTRY von Chemical Abstracts bei dem Host STN International dar. Faßt man die chemischen Strukturformeln als formale Sprache auf, können die Datenbanken, die chemische Strukturformeln verwalten, als Datenbanken erachtet werden, die Ausprägungen einer formalen Sprache (eines Formalismus) verwalten. Die interne Repräsentation erfolgt durch sog. "connection tables".

Natürlich ist hier auch das Retrieval verändert. Vom Aufbau des Suchbegriffs bis zur Struktur der Treffer liegt hier eine völlig andere Situation vor, als bei allen anderen Bereichen des Information Retrievals. Der Unterschied zu den übrigen satzorientierten Datenbanken ist so deutlich, daß STN International von "alphanumeric files" und "structure- or substance-based files" spricht. Während im ersten Fall mit alphanumerischen Ausdrücken gesucht wird unter Nutzung der Bool'schen und weiterer Operatoren, kann in den structure-based files auch mit Hilfe von **Strukturen** gesucht werden. Datenbanken, die Strukturformeln verwalten, speichern somit nichts anderes als Terme der formalen Sprache "Strukturformel", wenn sie dies auf nicht-

triviale Weise tun. (Also nicht als Graphik oder alphanumerische Zeichenkette sondern unter zumindest teilweiser Berücksichtigung der Eigenschaften des Informationstyps).

### **Recherchebeispiel A4: STN File: REGISTRY**

```

***** Welcome to STN International *****
FILE 'HOME' ENTERED AT 17:37:39 ON 28 JUL 95

FILE 'REGISTRY' ENTERED AT 17:40:14 ON 28 JUL 95 □
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 1995 American Chemical Society (ACS)
L2 ANSWER 1 OF 1 REGISTRY COPYRIGHT 1995 ACS
RN ***50-32-8*** REGISTRY
CN Benzo[a]pyrene (7CI, 8CI, 9CI) (CA INDEX NAME)
OTHER NAMES:
CN 3,4-Benzopyrene, CN 3,4-Benzpyrene, CN 3,4-Benzpyrene (carcinogen), CN 3,4-Benz[a]pyrene
CN 4,5-Benzpyrene, CN 6,7-Benzopyrene, CN Benzopyrene, CN Benzo[d,e,f]chrysene, CN
Benz[a]pyrene
FS 3D CONCORD
MF C20 H12
CI COM, RPS
LC STN Files: AIDSLINE, ANABSTR, APILIT, APILIT2, APIPAT, APIPAT2, BEILSTEIN*,
BIOBUSINESS, BIOSIS, CA, CABA, CANCERLIT, CAOLD, CAPLUS, CAPREVIEWS, CASREACT, CEN,
CHEMINFORMRX, CHEMLIST, CHEMSAFE, CIN, CJACS, CSCHM, CSNB, DETHERM*, DDFU,
DRUGU, EMBASE, GMELIN*, HODOC*, HSDB*, IFICDB, IFIPAT, IFIUDB, IPA, MEDLINE, MRCK*,
MSDS-OHS, MSDS-SUM, NAPRALERT, PIRA, PROMT, RTECS*, SPECINFO, TOXLINE, TOXLIT,
TRCTHERMO*, USPATFULL
(*File contains numerically searchable property data)
Other Sources: DSL**, EINECS**, TSCA**
(**Enter CHEMLIST File for up-to-date regulatory information)
. C : . C:
. C. : C. : C
.
.
.
. C : . C. : C. : C
C. : C. : C : . C:
.
.
.
C. : C. : C. : C
. C: . C: . C:
1 REFERENCES IN FILE CAPREVIEWS
13419 REFERENCES IN FILE CA (1967 TO DATE)
912 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA
13446 REFERENCES IN FILE CAPLUS (1967 TO DATE)

```

Datenbanktechnisch gesprochen, ist das formale System das Datenmodell, die einzelnen Strukturformeln sind die zulässigen Informationseinheiten. Ein solches Datenmodell hat dann natürlich Auswirkungen auf das Datenbanksystem haben, das die Regeln des formalen Systems bei allen Aktionen berücksichtigen muß bzw. sollte [Staud 1991, S. 177-179].

Da die hier eingesetzte Kommunikationssoftware zur Recherche bei allen Host gilt, ist sie nicht speziell für die STN Datenbanken ausgewiesen. Somit werden die Strukturformeln nicht wie z.B. mit der speziell für STN Datenbanken entwickelten Kommunikationssoftware STN Express ausgegeben.

Zwar steht die Strukturformel im Zentrum dieser Datenbank, aber zahlreiche andere Informationstypen finden hier auch noch Verwendung.

Es seien genannt: Chemikalienname (CN = Chemical Name), Synonyma (CN = Chemical Name), CAS-Nummer (RN = Registry Number), Summenformel (MF = Molecular Formula). Auch werden wieder Hinweise auf andere STN Datenbanken gegeben, in denen sich weitere Informationen über die gesuchten Chemikalien befinden.

### **A.1.6 Metadatenbanken**

Ebenfalls Faktencharakter haben die sog. Metadatenbanken, die auch in der Fachliteratur Verweisdatenbanken genannt werden. Unter einer *Metadatenbank* versteht man eine Datenbank, die auf andere Datenbanken oder andere Datenquellen verweist. Hier werden mit Hilfe von Datensätzen Objekte, in diesem Fall Datenbanken bzw. Datenquellen beschrieben. Als Recherchebeispiel wird ein Datensatz aus dem Gale Directory of Databases gewählt.

Hier wurde nach Umweltdatenbanken (environment and databases) recherchiert. Es wird der Datensatz für die ECDIN (Environmental Chemicals Data and Information Network) Datenbank gezeigt. Es werden unter der Feldbezeichnung NA der Datenbankname, unter PT der Datenbanktyp, die Sprache unter LA etc. als bibliographische Felder gegeben. Der Inhalt der Datenbank wird in den Feldern IT = Index Terms und AB = Abstract erfaßt. Anzumerken ist in diesem Zusammenhang, daß weder der Datenbankhersteller (Produzent) noch der Datenbankanbieter (Host) angezeigt werden.

### **Recherchebeispiel A5: ORBIT File: GALE Item: 9997927**

WELCOME TO ORBIT ONLINE SERVICE. (07/28/95 11:28 A.M. CENTRAL TIME)  
 THE DIRECTORY OF DATABASES IS CURRENT THROUGH THE MAY 1995 (9507) ISSUE.-2- (GALE)  
**AN** - 9997927  
**DAN** - D-0000001301  
**NA** - Environmental Chemicals Data and Information Network  
**RT** - Database  
**MD** - Online  
**DT** - Full-text; Numeric; Bibliographic; Properties  
**AB** - Contains information on chemical substances in the environment, providing chemical identification data on approximately 122,400 chemical compounds, acute toxicity data on approximately 20,000 compounds, and additional data for some 3600 chemical substances. Is divided into the following files: Chemical Substances (Registry Numbers), including ECDIN Registry Name and Number, Chemical Abstracts Service Registry Number, European Customs Union Number, and Registry of Toxic Effects of Chemical Substances (see RTECS) Number; International Registry of Potential Toxic Chemicals (IRPTC) Legal; Chemical Synonyms (Names); Physico-Chemical Properties; Chemical Structures, including Wiswesser Line Notation and chemical structure diagram; Chemical Processes; Uses; Production and Trade Statistics; Occupational Safety and Health, covering hazard information, general recommendations for human safety, personal protection, and medical and biological surveillance; Occupational Exposure Limits; Directive 67/548/EC (Dangerous Substances Hazard Classification); Classical Toxicity, including organism, exposure, effects, test conditions, and bibliographic data; Aquatic Toxicity, including organism, exposure, effects, test conditions, and bibliographic data; Effects on Microorganisms, including organism, exposure, effects, test conditions, and bibliographic data; Odor and Taste Threshold Concentrations; Analytical Methods; Carcinogenicity; Mutagenicity; Concentration in Human Media, including range and mean concentration values, biological media, pathology, and morphology; Concentration in Animal Media; Concentration in Environmental Matrices; Metabolism in Soil; and Chemical Producers.  
**LA** - English (EN)  
**GC** - International  
**TS** - 1970 to date  
**UF** - Periodically, as new data become available  
**IT** - Hazardous substances; Chemicals; Environment; Chemical industry; Chemical properties; Chemical structure; Toxicology; Water resources  
**UP** - 9506

## A.2 Textbasierte Datenbanken

### A.2.1. Bibliographische Datenbanken

Im Gegensatz zu Volltextdatenbanken enthalten *bibliographische Datenbanken* nicht den Text selbst, sondern nur seine "Modellierung". Oftmals enthalten sie ein Kurzreferat (Abstract), das eine Zusammenfassung der Information (des Wissens) gibt, die im Text enthalten ist. Von Faktendatenbanken, denen sie ähnlich sind bezüglich der Felderzusammensetzung, müssen sie unterschieden werden, weil sie Informationen über Objekte (in diesem Beispiel eine Chemikalie) beschreiben und nicht die Objekte selbst.

Als Beispiel wird ein Datensatz aus der Datenbank ULIT (Umweltliteraturdatenbank) bei dem Host DataStar dargestellt.

#### **Recherchebeispiel A6: DataStar File: ULIT Item: 00228438**

Welcome to Knight-Ridder Information 28.07.95

D-S/ULIT/UMWELTLITERATUREDATENBANK '76- SESSION 247

COPYRIGHT BY Umweltbundesamt, Berlin, Germany.

D-S - SEARCH MODE - ENTER QUERY

1 ULIT

AN 00228438.

**AU** Caspers-N, Hartmann-P, Kanne-R, Knoop-G.

**IN** Bayer, Institut fuer Umweltanalyse und Bewertungen, Leverkusen;

Bayer, Institut fuer Umweltanalyse und Bewertungen, Leverkusen;

Bayer, Institut fuer Umweltanalyse und Bewertungen, Leverkusen;

Bayer, Institut fuer Umweltanalyse und Bewertungen, Leverkusen.

**TI** Bund/Laender-Arbeitskreis 'Qualitaetsziele' (BLAK-QZ) - Problematik des Konzeptes bei der Festlegung von Zielvorgaben fuer Oberflaechengewaesser am Beispiel von Trichlorethen und *Hexachlorbenzol*.

**TT** Federal/State Working Group 'Quality Goals' - Problem of the Conception in the Determination of Target Ordinances for Surface Waters Exemplified by Trichlorethane and *Hexachlorobenzol*.

**SO** Umweltwissenschaften und Schadstoff-Forschung, 5 (5) S. 265-270 (1993).

**PT** Zeitschrift.

**LG** GE.

**CC** WA40 = Wasserguetekriterien, Richtwerte, Zielvorstellungen;

CH40 = Umweltchemikalien: Schadstoffe: Richtwerte, Hoechstwerte, Grenzwerte und Zielvorstellungen.

**DE** Oberflaechenwasser; Biozoonose; Oekosystem-aquatisch; Trichlorethylen; *Hexachlorbenzol*; Schadstoffwirkung; Oekotoxizitaet; Daphnien; Sterblichkeit; Umweltqualitaetsziel; Oberflaechengewaesser; Guetekriterien; Grenzwertfestsetzung; Oekotoxikologische-Bewertung;

**FT** Zielvorgaben; BLAK-QZ; Reproduzierbarkeit; Daphnia-maga.

**GD** Bundesrepublik-Deutschland.

**AB** Fuer aquatische Lebensgemeinschaften werden auf Grund eines Konzeptes des und/Laender Arbeitskreises 'Qualitaetsziele' die Zielvorgaben als fachlich begruendete Bewertungsmaassstaebe auf der Grundlage der niedrigsten veroeffentlichten Wirkungswerte festgelegt. Die Vorgehensweise wird an Trichlorethen und *Hexachlorbenzol* erlaeutert. Durch Schiedsanalysen mit heute verfuegbarer Technik konnten aus den Jahren 83/84 stammende Wirkungswerte nicht bestaetigt werden. Die Ergebnisse aus den Schiedsanalysen weisen vielmehr darauf hin, dass insbesondere bei gut untersuchten Verbindungen die Zielvorgaben hoeher angesetzt werden muessen. Bei der Festlegung der Zielvorgaben ist die Datenvalidierung unverzichtbar. (8 Tab. ; 12 Lit.) LI TIB ZB 4603.

IP UBA.

ED 940414.

Die beschriebene Information, ein Artikel, ein Buch oder sonst ein "Text" werden durch inhaltliche und formale Kategorien sowie durch ein Kurzreferat (Abstract) erfaßt. Im einzelnen finden sich neben einem identifizierenden Merkmal, der Accession Number (AN) weitere qualitative Merkmale z. B. Publication Type (PT), Language (LG) und einige formatierte Felder z. B.

Titel (TI), Source (SO). Für die Klassifizierung sind die Datenfelder Classification Code (CC) und Deskriptors (DE) vorgesehen. Diese sind mit denen in UFOR (Anhang A.1.4) identisch. Die Klassifizierungsfelder strukturieren den Artikel und erleichtern somit das Wiederauffinden des entsprechenden Dokumentes. Ein weiteres Feld, welches den Inhalt des Artikels beschreibt, ist das Abstract (AB) Feld.

### A.2.2 Volltextdatenbanken

#### Recherchebeispiel A7: STN File: CJRSC Item: RA405008

```

***** STN Karlsruhe *****
FILE 'CJRSC' ENTERED AT 17:31:34 ON 28 JUL 95
COPYRIGHT (C) 1995 The Royal Society Of Chemistry (RSC)
FILE COVERS 1987 - 26 JUL 1995 (950726/ED)

L1 ANSWER 1 of 6 CJRSC COPYRIGHT 1995 RSC
AN 94:2008 CJRSC
DN RA405008
SO The Analyst, (1994), 119(5), 781-789. CODEN: ANALAO. ISSN: 0003-2654.
TI Electrochemistry of Waste Removal * A Review
AU (1) Bockris, J. O'M.; (2) Bhardwaj, R. C.; (3) Tennakoon, C. L. K.
CS (1,2,3) Department of Chemistry, Texas A&M University, College Station, TX 77843, USA * Presented at EIRELEC '93, Electrochemistry to the Year 2000, Adare, Co. Limerick, Ireland, September 11-15, 1993.
PD MAY 1994
MS Received date: 2 NOV 1993
   Accepted for publication date: 17 JAN 1994
DT Article
PB The Royal Society of Chemistry
AB Summary of Contents Introduction Removal of SO2 and Cl2 From Waste Gas Streams Removal of Greenhouse Gases Replacement of SO2 Injection into the Atmosphere and Metal Recovery by an Electrochemical Technique Catalytic Reduction of CO2 From a Hydrogen-driven Fuel Cell Industrial Pollutants, Aquifers and Metal Recoveries Waste Water Treatment Electrochemical Treatment of Soils Ozonolysis Electrochemical Incineration of Sewage Enzymes for Bacteria as a Possible Aid to Low-cost Energy Production Bacteriolysis Reduction of Biotoxins by Electrochemical Methods Biophoto-electrocatalysis
References Keywords: Review; toxic gases; electrochemical; effluent treatment; biotoxins

TX(68) of 72. Reduction of Biotoxins by Electrochemical Methods. Another example of a biotoxin in everyday life is gasoline. It has been proved that the inhalation of gasoline vapours and the products of combustion of diesel oil in diesel engine exhaust (e.g., ***benzopyrene*** in diesel engine exhaust) by animals causes cancer; therefore it is important to replace petroleum fuel by electricity and/or hydrogen as rapidly as possible.
RE RE(1) of 41. 1. Kreysa, G., and Kulps, S. H. J., Chem. Eng. Technol., 1983, 55, 58.
CP CP(1) of 20. Figure 1. Fig. 1 Schematic diagram of the electrochemical-catalytic process for flue gas desulfurization. Reproduced from ref. 2
TT TT(1) of 1. Table 1 Theoretical space-time yields for several electrode systems

```

**Volltextdatenbanken** enthalten den vollen Text einer Dokumentationseinheit. Jedoch wird in allen Fällen der Text, bzw. die in ihm enthaltene Information, durch weitere Felder beschrieben. Vor allem dienen dazu Attribute, Merkmale, Nomenklaturen und Thesauri.

Ein Datensatz der Volltextdatenbank CJRSC (Chemical Journals of the Royal Society of Chemistry), die von dem Host STN angeboten wird, ist hier gewählt. Diese Volltextdatenbank gibt neben den bibliographischen Angaben ein Kurzreferat (Abstract) und erschließt den Text mit Schlüsselworten (Keywords). Der Text wurde in diesem Beispiel stark gekürzt. Die Tabellen und Graphiken werden nur per Titel aufgeführt, jedoch nicht angezeigt.

### A.3 Integrierte Datenbanken

In Zukunft werden solche Datenbanken eine immer wichtigere Rolle spielen, die zur Modellierung ihres Weltausschnitts alle verfügbaren und sinnvollen Informationstypen einsetzen. Ansätze zur Integration finden sich bereits in einfacher Form bei den Datenbanken zu Chemie und Physik, wo die eigentlichen Daten ergänzt werden um Dateien mit bibliographischen Informationen zur Literatur bezüglich der modellierten Chemikalien. Diese Datenbanken werden integrierte Datenbanken genannt. **Integrierte Datenbanken** verwalten gleichzeitig Fakten (im engeren Sinn), Text und Tabellen und Graphiken [Staud 1991, S. 191-211].

Im Bereich der naturwissenschaftlich orientierten Datenbanken liegen zahlreiche Beispiele für eine Integration vor. Am bekanntesten dürfte die Gruppe der Datenbanken REGISTRY, CAPlus, CA, und CAOLD sein. Alle vier werden von Chemical Abstract Service in Columbus, Ohio hergestellt. Sie beziehen sich auf chemische Substanzen und erfassen neben (chemischen) Fakten auch Strukturformeln (in REGISTRY) und bibliographische Angaben in CAPlus, CA und CAOLD zur Literatur über chemische Substanzen. Seit Januar 1995 ist die bibliographische Datenbank CAPlus online. Es handelt sich hierbei um die aktuellste wissenschaftlich-technische bibliographische Datenbank. CAPlus umfaßt die Datenbanken CA und CAPreviews. Darüber hinaus enthält sie noch weitere neuere Einträge und Datensätze aus zusätzlichen Zeitschriften. Sie umfaßt insgesamt über 12 Millionen Datensätze aus 1.300 wissenschaftlich-technischen Fachzeitschriften und wird täglich aktualisiert. 16 % des Datenbankinhaltes stellen Patente dar, d.h. sie beinhaltet ca. 1.8 Millionen Patente [STN International 1996, S.1-3].

Die Integration ist vielfältig. Zum einen wird durch die CAS-Nummer und durch entsprechende Felder in REGISTRY und CAPlus/CA/CAOLD die Verknüpfung von Strukturformeln und Literaturinformationen ermöglicht, zum anderen finden sich in CAPlus/CA/CAOLD Datensätze, die nicht nur unterschiedliche Objekte beschreiben ("journal record, patent record") sondern dadurch auch einen unterschiedlichen Aufbau haben. Patente zu chemischen Stoffen werden auf andere Weise beschrieben als Fachartikel zu Chemikalien.

Die international übliche **CAS-Nummer** (Chemical Abstract Service Registry Number) der American Chemical Society ist ein Identifikationscode, der vom Chemical Abstract Service für eine chemische Substanz vergeben wird, wenn die Substanz in das Registriersystem eingeordnet wird. Nur eine Nummer existiert für jede Verbindung. Registriernummern werden sequentiell vergeben. Generell kann man festhalten, je höher die Nummer ist, desto später wurde sie vergeben. Das Format der CAS-Nummer ist eine drei-geteilte Nummer, **XXXXXX-XX-X**, bestehend aus zwei bis sechs Ziffern, einem Bindestrich, zwei Ziffern, einem weiteren Bindestrich und einer letzten Ziffer. Wenn mehr Registriernummern vergeben werden, so wird sich die Anzahl der Ziffern in der ersten Zahlenkette auf sieben, acht etc. Ziffern erhöhen [STN International 1993 S. 6].

Die CAS-Nummer dient somit der eindeutigen Identifikation chemischer Verbindungen. Sie ist z.Zt. noch das wichtigste und pragmatischste Identifikationsmerkmal für Chemikalien. Es ist daher unverständlich, daß in vielen der Umweltchemikalien-relevanten Datenbanken diese CAS-Nummer noch keine Berücksichtigung findet. In reinen Chemiedatenbanken ist die Situation der Chemikalien Identifikation über die CAS-Nummer besser. So stellt beispielsweise Heller in seinem Artikel "Chemical Information Activities: What the Future Holds" fest, daß es

ca. 20 Jahre gedauert hat, bis die CAS (Chemical Abstract Service) Registry Number eine weite Verbreitung in Chemiedatenbanken erfahren hat. Diese Nummer hat - wie bereits oben ausgeführt - keinen intellektuellen Bezug zu der Chemikalie wie die Strukturformel. Die durch die ständige Verbesserung der Hard- und Software gegebenen Möglichkeiten die Strukturformel in Datenbanken zu integrieren, lassen die Voraussage der Übernahme der heutigen Stellung der CAS-Nummer durch die Strukturformel als Identifikationsmerkmal in Datenbanken in Zukunft vernünftig erscheinen [Heller 1993]. Auf dem Gebiet der Umweltchemikalien-Datenbanken ist von dieser Vision noch nicht viel zu merken.

### A.3.1 Reaktionsdatenbanken

Eine Reaktionsdatenbank ist eine Datenbank, deren Inhalt chemische Reaktionen darstellen. Die chemischen Reaktionen werden i.a. als topologische Daten gespeichert. Diese bestehen aus Verknüpfungstabellen (connection tables) der beteiligten Verbindungen sowie der Informationen über die Reaktionszentren und die Zuordnung der Atome bzw. Atomgruppen in der Reaktion [Barth 1992, S.411].

Es gibt noch nicht viele Reaktionsdatenbanken online; sie gewinnen jedoch zunehmend an Bedeutung. Die Reaktionsdatenbanken ChemSelect, ChemSynth und CHEMREACT wurden kurz von Schinzer erläutert [Schinzer 1993]. Die Umsetzung des gedruckten Referatedienstes ChemInform in die ChemInformRX Datenbank wurde Ende der achtziger Jahre vollzogen [Gasteiger 1989, Gasteiger 1992]. Den Inhalt dieser Datenbank bilden Veröffentlichungen auf dem Gebiet der Chemie. Besonderes Gewicht wird dabei auf die Synthese organischer Verbindungen gelegt, wobei neue Reaktionen und Synthesemethoden im Vordergrund stehen.

Die Erweiterung des Angebotes an Reaktionsdatenbanken und der Fortschritt in der Suche nach chemischen Reaktionen in diesem Datenbanktyp wurde von Hayward dargelegt. Der Autor kommt jedoch zu dem Schluß, daß die Entwicklungen für den Nutzer längst noch nicht ausreichend sind und daß noch viel mehr in der Zukunft für die Unterstützung der Suche in dieser Art von Datenbanken getan werden muß [Hayward 1995]. Auf die drei bei STN aufliegenden Reaktionsdatenbanken: CASREACT, ChemInformRX und CHEMREACT wird in einer weiteren Publikation eingegangen. Keine der diskutierten Reaktionsdatenbanken deckt die Literatur über chemische Reaktionen auch nur annähernd ab. Aus diesem Grunde sollte eine Recherche nach chemischen Reaktionen nicht nur in Reaktionsdatenbanken sondern zusätzlich auch in anderen Chemikalien-orientierten Datenbanken durchgeführt werden [Wolman 1995].

Als Beispiel soll eine Recherche in ChemInformRX durchgeführt werden.

Es handelt sich bei dem gezeigten Beispiel um die Nitrierung von Benzo[a]pyrene mit Salpetersäure. Das Dokument wurde editiert zum Zwecke der Anpassung an das Textverarbeitungssystem. Zum Reaktionsretrieval müssen die Reaktionssequenzen in Einzelreaktionen aufgelöst werden. Es entsteht Benzo[a]pyrene, 6-Nitro (III) mit einer Ausbeute von 44%, Benzo[a]pyrene, 1-Nitro (IV) mit einer Ausbeute von 8%, Benzo[a]pyrene, 3-Nitro (V) mit einer Ausbeute von 4%. Im Anschluß an die Reaktion werden in dem Feld **Reaktionen RX(1)** die CAS-Nummern der einzelnen an der Reaktion teilnehmenden Chemikalien gegeben. RCT bedeutet Reactant CI Registry Number, RGT Reagent CI Registry Number, PRO Product CI Registry Number. Ferner werden Angaben über die Gesamtausbeute (YDS Yield Stage), die Temperatur (T) gegeben. Auch wird das Dokument unter dem Datenfeld **KW = Keywords** inhaltlich erschlossen. Bei dem besprochenen Beispiel handelt es sich um eine Arylierung.

**Recherchebeispiel A8: STN File: CHEMINFORMRX Item: 9123131**

FILE 'CHEMINFORMRX' ENTERED AT 17:46:23 ON 28 JUL 95

COPYRIGHT (C) FIZ-CHEMIE BERLIN

FILE LAST UPDATED: 17 MAY 95 &lt;950517/UP&gt;

L1 ANSWER 1 OF 1 COPYRIGHT 1995 FIZ CHEMIE

AN 9123131 CHEMINFORMRX

TI Synthesis, Chemical Properties, and Mutagenicity of 1,6- and 3,6- Dinitrobenzo(a)pyrenes.

AU FUKUHARA, K.; MIYATA, N.; MATSUI, M.; MATSUI, K.; ISHIDATE, M. JUN.; KAMIYA, S.

CS Natl. Inst. Hygienic Sci., Kamiyoga, Tokyo 158, Japan

SO Chem. Pharm. Bull., 38(11), 3158-3161 (1990), CODEN: CPBTAL ISSN: 0009-2363

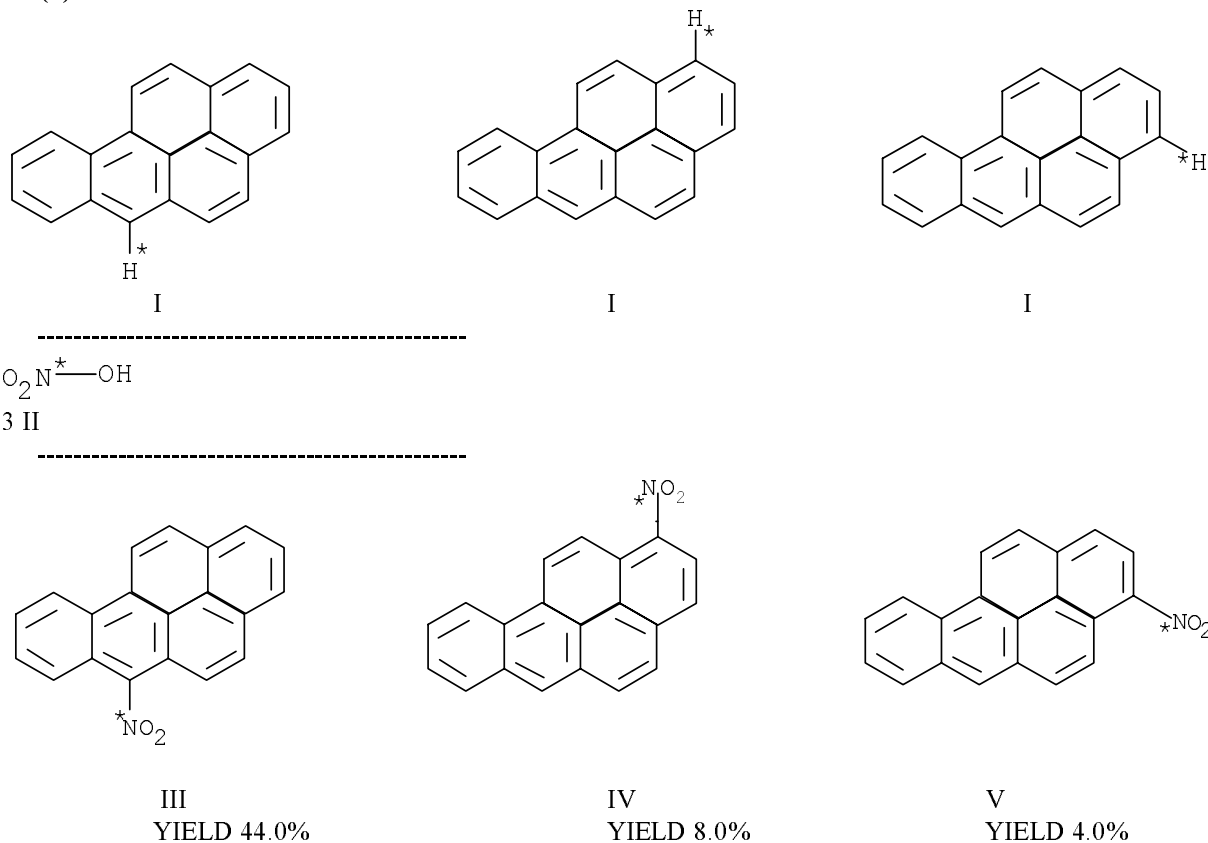
LA English

AB Nitration of the **benzo(a)pyrenes** (I) with nitric acid (II) yields the nitrobenzopyrenes (III) - (V) or (III) - (VII) depending on the reaction conditions. The mutagenic activities of all products are studied and correlated with their redox potentials.

CC Q1150 other polycyclic compounds

DED 19930426

RX(1) OF 2 3 \*\*\*I\*\*\* + 3 II ==&gt; III + IV + V



RX(1) RCT I, \*\*\*157507\*\*\* (50-32-8)

II, 136 (7697-37-2)

RGT 4 (108-24-7), Ac2O

PRO III, 157508 (63041-90-7)

IV, 157509 (70021-99-7)

V, 157510 (70021-98-6)

YDS 56.0 %, T 20.0 Cel, TIM 24 hr, KW Arylation

NTE reaction: I (II) -&gt; III + IV + V

An dem Beispiel läßt sich deutlich erkennen, daß diese Reaktionsdatenbanken der Definition der integrierten Datenbanken entsprechen. Die eigentlichen Daten der Reaktionen (Tabellen, Diagramme) werden ergänzt durch bibliographische Angaben. Während die Felder AU

(Autor), TI (Titel), SO (Quelle), AB (Abstract) zu den bibliographischen Feldern gehören, sind die Datenfelder RX etc. (Reaktionen und deren Diagramme) Fakten im engeren Sinne. Ebenso findet man unter KW (Keywords) ein inhaltserschließendes Feld.

### **A.3.2 Patentdatenbanken**

Als Patentdatenbank wird eine Datenbank bezeichnet, deren Inhalte Patente und allgemeine Patentinformationen sind [Barth 1992, S.409]. Die in den Patentedokumenten enthaltenen Informationen sind im wesentlichen einheitlich strukturiert.

Man unterscheidet:

- die bibliographischen Daten
- die Zusammenfassung, gegebenenfalls ergänzt durch eine schematische Zeichnung
- die Beschreibung, meist in Verbindung mit mehreren schematischen Zeichnungen
- die Patentansprüche.

Damit gehören die Patentdatenbanken per Definition deutlich zu der Klasse der integrierten Datenbanken, die gleichzeitig Fakten (im engeren Sinn), Text und Tabellen und Graphiken speichern.

Neben den in vielen anderen Datenbanktypen ebenfalls vorkommenden Datenfeldern ist hier noch besonders auf das Datenfeld *IC*, das für die Internationale Patentklassifikation (*International Patent Classification*) steht, hinzuweisen. Schon 1877 wurde mit der Einrichtung eines Ordnungssystems für Patente, mit dessen Hilfe die eingereichten Erfindungen einem genau bezeichneten Sachverhalt und damit einer auf diesem Gebiet sachkundigen Prüfstelle zugewiesen werden konnten und können, begonnen. Ein solches Ordnungssystem bezeichnet man als Patentklassifikation. Seit Mitte der siebziger Jahre verwenden die meisten Patentämter die Internationale Patentklassifikation IPC, die rund 70.000 Unterteilungen aufweist. Die für den Schwerpunkt der Erfindung oder für deren wichtigste Bestandteile maßgebenden Klassifikationssymbole werden auf die Patentedokumente aufgedruckt und sind ein wesentliches Hilfsmittel für das Auffinden von Patentedokumenten auf einem bestimmten, eng umgrenzten technischen und wissenschaftlichen Sektor. Die IPC wird entsprechend dem sog. Straßburger Klassifikationsabkommen von 1971 in internationaler Zusammenarbeit regelmäßig überarbeitet und in fünfjährigem Abstand als Gesamtwerk von der Weltorganisation für Geistiges Eigentum (WIPO) neu herausgegeben. Die erste Fassung der Internationalen Patentklassifikation erschien am 1. September 1968 [Wittmann 1991]. Zur Zeit wird mit der sechsten Auflage 1994 gearbeitet [WIPO 1994]. Für die Chemie ist die Section C, Chemistry; Metallurgy, insbesondere für die im besprochenen Chemikalien-Testdatensatz verwendeten Umweltchemikalien die Sektion C07C Acyclische und Carbocyclische Verbindungen wichtig.

Patentdatenbanken werden in erster Linie von Patentabteilungen und Erfindern genutzt. In letzter Zeit werden jedoch auch Patentdatenbanken im Bereich der technisch-naturwissenschaftlichen Informationsgewinnung in Forschung und Entwicklung mit herangezogen [Georgy 1993]. Auch im Bereich des Umweltschutzes und der Umweltforschung ist die Patentinformation von zunehmender Bedeutung. Die deutschsprachige Datenbank PATDPA weist die Besonderheit eines umweltbezogenen Codierungsfeldes auf, wodurch umweltrelevante Patentedokumente besonders leicht und effektiv auffindbar sind [Weckend 1994].

Das Recherchebeispiel ist aus der PATDPA Datenbank. Die Datenbank enthält seit 1983 Abbildungen, chemische Strukturen und mathematische Formeln. Bei der durchgeführten Recherche nach Hexachlorbenzol und Benzo[a]pyrene sind keine graphischen Informationen ausgegeben worden, da die Wiedergabe von Graphiken nur mit der Kommunikationssoftware STN Express erfolgt, die hier nicht eingesetzt wurde. Da es für Hexachlorbenzol explizit eine IPC Nummer gibt, nämlich C07C25/12, wurde die Recherche mit dieser Nummer durchgeführt.

Allerdings wurden nur zwei Dokumente gefunden. Das ausgegebene Dokument ist aus dem Jahre 1981. Erstaunlich ist auch das Fehlen eines Abstracts, denn Kurzreferate sollen bis zum Jahre 1981 zurückreichen.

Ebenfalls wurde nach Polykondensierten Kohlenwasserstoffen C07C15/20 recherchiert und hier eine Trefferquote von 22 erzielt. Für den kontrollierten Begriff Umweltschutz gibt es zwar 7735 Treffer, jedoch ergibt die Verknüpfung mit den beiden Testchemikalien keinen Erfolg.

Es sind im vorliegenden Dokument also nur reine bibliographische Informationen, die das Patent beschreiben, erfaßt.

### **Recherchebeispiel A9: STN File: PATDPA Item: 810902**

(FILE 'HOME' ENTERED AT 15:48:52 ON 01 AUG 95)

FILE 'PATDPA' ENTERED AT 15:49:27 ON 01 AUG 95

**L1** 7735 S UMWELTSCHUTZ/CC  
**L2** 2 S C07C025-12/IC  
**L3** 22 S C07C015-20/IC  
**L2** ANSWER 2 OF 2 COPYRIGHT 1995 DPA/FIZ KA  
**AN** 80(57):244 PATDPA ED 810902 EW 8136  
**SN** EP79300050.6 UP 861001  
**TI** (B1) Verfahren zum Abscheiden von *Hexachlorbenzol* aus Tetrachloraethylen.  
**IN**  
**INC** Blunk, John Ross (\*US Louisville Kentucky 40207)  
 Rozmus, Glenn Frank (\*US Columbia Tennessee 38401)  
**PA**  
**PAC** Stauffer Chemical Co. (\*US Westport, Conn.)  
**PAN** 1005316 US  
**PAT** (CORP) Juristische Person  
**AG** derzeit kein Vertreter bestellt  
**SO** Patentblatt 101 (1981) Heft 36, EP B1 EP-Patentschrift  
**DT** P  
**LA** German  
**NTE** 780111: FPRD (32) Erstes Prioritaetsdatum  
 790111: ADR (86) Anmeldetag d. EP-Anm. m. DE-Benennung  
 790808: AOR (87) EP-Publ. der EP-Anmeldung mit DE-Benennung  
 810902: PGR (87) EP-Patent mit DE-Wirk.  
 XXXXXX: ZEJP H339 Erledigt wegen Nichtzahlg. d. Jahresgebuehr  
**NTL** XXXXXX: ZEJP H339 Erledigt wegen Nichtzahlg. d. Jahresgebuehr  
**PIT** EP PG Europaeische Patentschrift, Einspr.-Frist 9 Mon.  
**PI** EP 3398 B1 810902 PGR OP9 (87) letzte Publ./ EP-Schrift  
**DS** R BE; R DE; R FR; R GB; R NL  
**AI** EP 79-300050 A 790111 ADR (86) EP-Anm. mit DE-Ben.  
 DE 79-2960710 E 790111 ADRN (22) DE-AKZ fuer EP-Patent  
**PRAI** US 78-868652 A 780111 CP (32) Unionsprioritaet  
**FI**  
**FIA** DE 79-2960710 E 790111 ADRN (22) DE2960710  
 EP 79-300050 A 790111 ADR (86) EP3398  
**FIP** EP 3398 A1 790808 AOR LXX (87) EP-Publik. mit DE-Ben. :  
 R BE; R DE; R FR; R GB; R NL  
 DE 2960710 CE 810902 PGRN (45) DE-AKZ fuer EP-Patent  
 EP 3398 B1 810902 PGR (87) EP-Patent mit DE-Wirk. :  
 R BE; R DE; R FR; R GB; R NL  
**OS** CA 92:58412 'CA92(7):58412B'  
 CA 97:215717 'CA97(25):215717C'  
**IC**  
**ICM** (ID3) C07C025-12 (511) IPC-Hauptklasse

ICS (ID2) C07C017-38	(512) IPC-Nebenklasse
(ID2) B01D029-10	(512) IPC-Nebenklasse
PST ABSCHIEDEN; <i>HEXACHLORBENZOL</i> ; TETRACHLORAETHYL; VERFAHREN	
FA INC; PAC; PAN; DS; OS; ICS; ICS; PAS; PAT	

#### A.4 Zusammenfassung

Datenbanken zur Thematik "Chemikalien und Umweltschutz" werden in zehn verschiedene oben beschriebene und in Tabelle A1 zusammengefaßte Typen eingeteilt.

**Tabelle A1: Beispiele für einzelne Datenbanktypen**

Kapitel Nr./ Anhang Nr.	Oberbegriff	Typ	Beispiel Datenbank	Host
2.5.2 A.1.1	faktenbasierte Datenbanken	Faktendatenbanken (numerische Datenbanken)	CHCD	DIALOG
A.1.2	faktenbasierte Datenbanken	Chemikaliennamen-Ver- zeichnisse	CHEMLINE	DIMDI
A.1.3	faktenbasierte Datenbanken	Firmenverzeichnisse und sonstige Verzeichnisse	CERCLIS	CIS
A.1.4	faktenbasierte Datenbanken	Forschungsdatenbanken	UFOR	DataStar
A.1.5	faktenbasierte Datenbanken	Strukturdatenbanken	REGISTRY	STN
A.1.6	faktenbasierte Datenbanken	Metadatenbanken	GALE	ORBIT
A.2.1	textbasierte Da- tenbanken	Bibliographische Daten- banken	ULIT	DataStar
A.2.2	textbasierte Da- tenbanken	Volltextdatenbanken	CJRSC	STN
A.3.1	integrierte Da- tenbanken	Reaktionsdatenbanken	ChemInformRX	STN
A.3.2	integrierte Da- tenbanken	Patentdatenbanken	PATDPA	STN

## Anhang B. Hasse-Diagramme als Werkzeug zur Bewertung

Die Bezeichnung „*Hasse-Diagramm*“ deutet auf den Mathematiker H. Hasse hin, der dieses Diagramm eingeführt und bekannt gemacht hat [Hasse 1967]. Desöfteren wird das Hasse-Diagramm (nur das Diagramm) als technischer Ausdruck für die Visualisierung von partiell geordneten Mengen benutzt. Mitte der siebziger Jahre wurden Hasse-Diagramme erneut im Rahmen funktionalanalytischer Untersuchungen aufgegriffen [Reggiani 1975]. Grundbegriffe zum Einsatz der Hasse-Diagramm-Technik für das Rangfolgeverfahren von Chemikalien und von Deponien wurden von E. Halfon erläutert [Halfon 1983, 1986]. Die Technik der Hasse-Diagramme stellt ein mathematisches Verfahren dar, das die Parallelität von Informationen beibehält. Die Grundlage hierfür ist die Theorie *partiell geordneter Mengen* (im englischen auch *posets* von *partially ordered sets*). Sie wird oft im Rahmen der Verbandstheorie mitabgehandelt, bekommt aber zunehmendes Gewicht als eigenständige Disziplin.

Die Hasse-Diagramme werden u. a. zur vergleichenden Beurteilung von Objekten, die durch ein Tupel, d. h. durch eine Reihe von Zahlen charakterisiert werden, eingesetzt. In dem vorliegenden Fall werden Datenbanken über Chemikalien und Umweltschutz als zu vergleichende Objekte betrachtet.

Innerhalb einer Menge  $E$  ( $x, y, z \in E$ ) ist eine Ordnung (Halbordnung) von  $E$  eine mathematische Relation mit den folgenden Eigenschaften:

A	$x \leq x$	Reflexivität	(1-1)
B	$x \leq y$ und $y \leq x \Rightarrow y = x$	Antisymmetrie	(1-1')
C	$x \leq y$ und $y \leq z \Rightarrow x \leq z$	Transitivität	(1-1'')

Bemerkung:  $x \leq y$  und  $y \leq x$  für alle Paare  $(x, y \in E)$  ist nicht gefordert.

**Reflexivität** bedeutet, daß der Vergleich von identischen Objekten gestattet ist.

**Antisymmetrie** drückt die Tatsache aus, daß z.B.  $3 < 5$  ist, jedoch die Umkehrung falsch ist.

**Transitivität** bringt die Tatsache zum Ausdruck, daß z.B.  $1 < 100$  und  $100 < 1000$ , daraus folgt, daß  $1 < 1000$  [Davey 1990].

Eine Menge  $E$ , die über eine Ordnungsrelation  $\leq$  verfügt, wird als *partiell geordnete Menge* bezeichnet [Davey 1990]. Anzumerken ist in diesem Zusammenhang, daß eine Ordnung, die auf ' $<$ ' basiert, als strikte Ordnung definiert werden kann. Die Bezeichnung 'Ordnung' ist nicht notwendigerweise beschränkt auf die üblichen Vergleiche der Größe von Zahlen. Andere binäre Beziehungen, die 1-1, 1-1' und 1-1'' genügen, können auch mit Hilfe von Hasse-Diagrammen analysiert werden. Aus diesem Grunde ist die Ordnungstheorie ein mächtiges Instrument der vergleichenden Bewertung im naturwissenschaftlichen Bereich. Im folgenden wird nur auf einige wenige grundsätzliche Begriffe, die bei der Auswertung der Hasse-Diagramme der Online Datenbanken und CD-ROMs zum Tragen kommen, eingegangen. Weitergehende Erläuterungen zur Hasse-Diagramm-Technik sind von Brüggemann ausgeführt worden [Brüggemann, 1995a]. In dieser Publikation geht Brüggemann auch ausführlich auf das in der GSF-Forschungszentrum für Umwelt und Gesundheit erarbeitete Programm "Hasse" ein.

### B.1 Grundlagen der Hasse-Diagramm-Technik

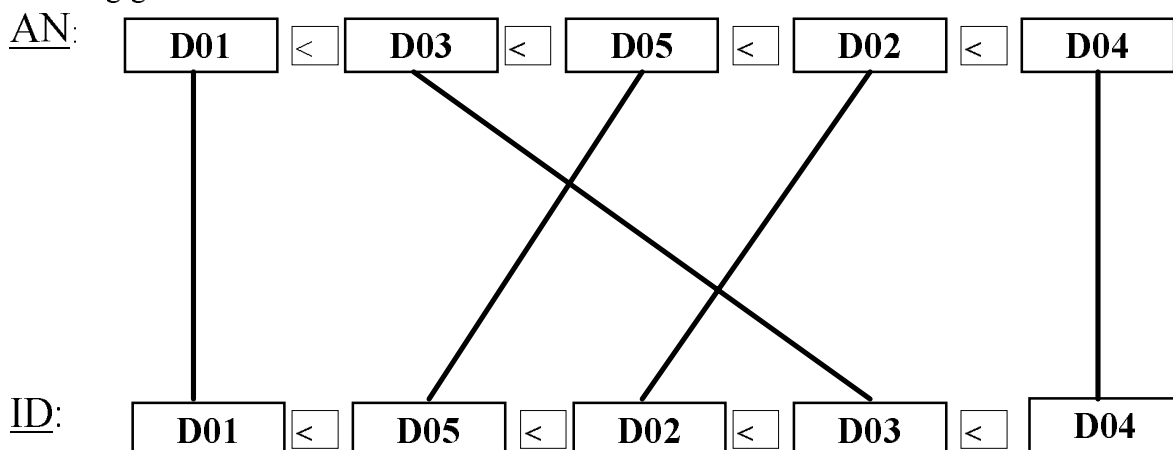
Im folgenden wird erklärt, warum Halbordnungen für das Rangfolgeverfahren von Objekten nützlich sind. Der Ausgangspunkt hierzu ist das Konzept von *Relationen* zwischen Elementen einer Menge. Durch Relationen erhält die Menge von Objekten eine Struktur. Als spezielle Relation wird, da ein "Anordnen", also ein Rangfolgeverfahren, im Hauptinteresse steht, die Ordnungsrelation gewählt (Näheres siehe unten). Im folgenden wird dazu im wesentlichen der Darstellung von Brüggemann und Halfon [Brüggemann 1995a] gefolgt.

### **B.1.1 Permutationsdiagramm versus Hasse-Diagramm**

Objekte, im vorliegenden Fall Datenbanken, werden beispielsweise durch ihre Inhalte d.h. hier durch die Anzahl der in ihnen enthaltenen Chemikalien (AN) charakterisiert. In vielen Fällen ist es jedoch nicht ausreichend, die Datenbanken nur anhand eines Kriteriums zu beschreiben, denn andere Kriterien wie z.B. das Vorhandensein von Identifikationsmerkmalen (ID) [Strukturformel, Summenformel, Molekulargewicht, CAS-Nummer, Synonyma] sind ebenso von Bedeutung. Wie ersichtlich, müssen Objekte aktuell durch mehr als ein Kriterium (auch Attribut genannt) charakterisiert werden. Diese Objekte können dann nicht geordnet werden, weil Konflikte zwischen Attributen bestehen. Als Beispiel seien fünf Objekte (Datenbanken) D01, D02, D03, D04, D05 genannt, die durch die Anzahl an Chemikalien (AN) und die Identifikationsmerkmale (ID) charakterisiert werden. Häufig reagieren die Attribute nicht parallel d.h. eine Datenbank, die viele Chemikalien enthält, umfaßt oft nur wenige Identifikationsmerkmale. Ordnet man die fünf Objekte nun gemäß AN und ID an, so ergibt sich folgende Abbildung B1.

Diese Art von Diagramm wird *Permutationsdiagramm* genannt [Urrutia 1989]. Es zeigt einige Umkehrungen in den zwei Folgen: Einige Objekte tauschen gegenseitig ihre Positionen in Abhängigkeit davon, welches Kriterium zur Definition der Folge angewandt wird (z.B. D03 und D05). Andere Objekte verändern ihre Beziehung zu anderen nicht, wenn das Kriterium wechselt. Dies gilt für D01 und D04 oder D05 und D02. Offensichtlich bleibt ein sog. "Rest an Ordnung", wenn beide Kriterien gleichzeitig betrachtet werden. Diese Tatsache motiviert den Begriff "*halb- bzw. partiell geordnete Mengen*". In dem besprochenen Beispiel der fünf Datenbanken liegt eine partielle Ordnung zugrunde, weil mehr als ein Kriterium zur Charakterisierung der Objekte herangezogen wird.

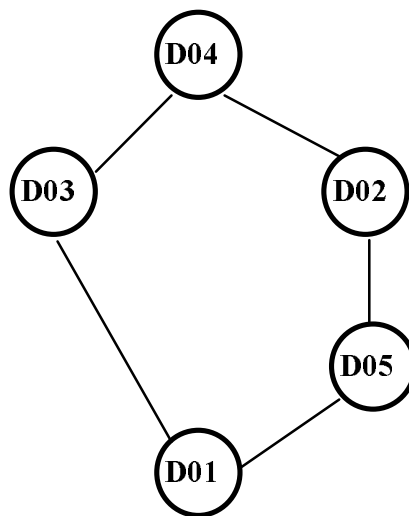
Falls die "normale" Ordnung, d.h. die Ordnung, in welcher jedes Objekt mit einem anderen verglichen werden kann, benutzt wird, wird die Bezeichnung *lineare* Ordnung oder *totale* Ordnung gewählt.



**Abbildung B1: Permutationsdiagramm für fünf Datenbanken**

Permutationsdiagramme werden sehr unübersichtlich, wenn viele Objekte, besonders aber auch viele Kriterien herangezogen werden. In diesem Fall entstünde eine korrespondierende Anzahl von Folgen und für jedes Paar der Folgen würde ein Permutationsdiagramm gezeichnet. Anstelle dieses aufwendigen Verfahrens, welches zu  $n(n-1)/2$  Paaren an Permutationen führt ( $n$  = Anzahl der Attribute) gibt die Hasse-Diagramm-Technik ein nützliches Werkzeug zur Visuali-

sierung. Die partiell geordneten Mengen für die fünf Datenbanken werden im Hasse-Diagramm in der Abbildung B2 dargestellt.



**Abbildung B2: Hasse-Diagramm als Alternative zum Permutationsdiagramm**

Es bleibt eine sog. Restordnung, wenn beide Kriterien gleichzeitig betrachtet werden:

$D04 > D03 > D01$  und

$D04 > D02 > D05 > D01$ .

In der Abbildung wird ebenso deutlich, daß D03 und D05 bzw. D03 und D02 ihre Position verändern, wenn nur ein Attribut betrachtet wird.

### **B.1.2 Definition von Grundbegriffen**

**Kriterien** umfassen sowohl quantitative als auch qualitative Eigenschaften.

**Attribute** sind quantitative, meßbare Daten. Attribute werden mit #1, #2,... #n gekennzeichnet.

Das **Objekt** ist der Gegenstand des Interesses und wird von Attributen charakterisiert. Beziehungen zwischen Objekten können ordnungstheoretisch untersucht und mit Hilfe von Hasse-Diagrammen (vgl. Abbildung B2) dargestellt werden. Objekte werden zur Menge  $E$  zusammengefaßt und werden daher auch desöfteren **Elemente** genannt.

Ein **Element** ist ein Mitglied einer Menge.

**Daten** sind numerische Werte, die zu jedem Kriterium gehören, mit denen Objekte charakterisiert werden. Ist der numerische Wert gemeint, der zu einem Objekt bezüglich eines bestimmten Kriteriums gehört, dann wird auch von der "Ausprägung" gesprochen.

Unter dem Begriff der **äquivalenten Objekte** in einem Hasse-Diagramm versteht man unterschiedliche Objekte, die dieselben Ausprägungen in bezug auf einen bestimmten Satz von Attributen aufweisen. Gleichheit bezüglich einer gegebenen Menge von Attributen definiert eine **Äquivalenzrelation**  $\mathcal{R}$ .

Die **Kardinalität** einer endlichen Menge ist eine Anzahl von Elementen einer Menge, beschrieben durch  $\text{card}(G)$  für die Menge  $G$ .

Ein sog. **Fall** ist eine Untermenge von ausgesuchten Attributen, hervorgehend aus einer Grundmenge von Attributen. Diese ausgewählten Attribute sind spezifisch für eine bestimmte Fragestellung. Jeder Fall korrespondiert mit exakt einem Hasse-Diagramm.

Man kann eine Tabelle aufstellen, in welcher die Reihen Objekte repräsentieren, und die Spalten die Daten für jedes Objekt, korrespondierend zu dem Spalten-definierenden Attribut enthalten. Es ist nützlich, die Tabelle als **Datenmatrix**  $\mathbf{P}$  [ $m \times n$ ,  $m$  Objekte (Reihen),  $n$  Attribute (Spalten)] anzusehen. Der Eintrag  $p_{ij}$  von  $\mathbf{P}$  ist der numerische Wert des  $j$ -ten Attributes vom  $i$ -ten Element.  $p_i$  bezieht sich auf das  $i$ -te Objekt mit allen seinen  $n$  Komponenten gemäß der Attribute.

Die Bezeichnung **Tupel** wird als Synonym für eine Reihe einer Datenmatrix  $\mathbf{P}$  verwendet.

### **B.1.3 Menge der Attribute**

Angenommen man will eine Menge  $E$  von  $m$  Elementen (Objekten) untersuchen. Jedes Element wird beschrieben durch eine Menge von  $n$  Attributen. Diese Menge ist eine Sammlung von Eigenschaften, mit denen die Objekte charakterisiert werden. Ein einzelnes Objekt wird beschrieben durch ein  $n$ -Tupel und die  $i$ -te Komponente ist der numerische Wert der  $i$ -ten Eigenschaft. Die gesamte Menge von Attributen wird beschrieben durch  $A$ , irgendeine Untermenge von Attributen durch  $B$  mit

$B \subseteq A$ .

Ein Rangfolgeverfahren kann durchgeführt werden, indem man entweder alle  $n$  Attribute heranzieht oder mit einer Untermenge  $B$  von nur einigen Attributen. Die Ausgangsbasis des Rangfolgeverfahrens ist die Gesamtinformation, die in der Menge  $A$  zusammengefaßt ist. Aus diesem Grunde wird die Menge  $A$  auch als **Informationsbasis** der vergleichenden Bewertung von Objekten bezeichnet [Brüggemann 1995a]. Bei der Nutzung aller Attribute, z.B. alle Elemente von  $A$  gibt es demnach nur ein mögliches Rangfolgeverfahren. Falls irgendein Attribut ausgelassen wird, - in diesem Fall werden  $n-1$  Attribute betrachtet - kann man  $n$  zusätzliche Rangfolgeverfahren Analysen durchführen. Insgesamt können  $2^n - 1$  Fälle verglichen werden, falls alle möglichen Untermengen von  $A$  außer der leeren Menge in Betracht gezogen werden. Meistens ist jedoch nur der Einfluß eines Attributes auf das Rangfolgeverfahren von Interesse. In diesem Fall wird der Fall mit der vollständigen Menge der Attribute mit jenen Fällen verglichen, bei denen ein Attribut ausgelassen wird. Deshalb sind meist nur  $n+1$  Vergleiche nötig.

## **B.2 Graphen und Hasse-Diagramme**

Eine Menge  $E$  (partiell geordnete Menge) wird beschrieben durch  $(E, \leq)$ ,  $a$  und  $b$  zwei Objekte der Gruppe  $E$ . Jedes Objekt wird anhand der  $n$ -elementigen Attributmenge  $A$  charakterisiert. Die Notation " $a \leq b$ " gilt, wenn alle Komponenten des Tupels von  $a$  kleiner oder gleich sind zu den korrespondierenden Komponenten des Tupels von  $b$ .

Unter Zuhilfenahme der Notation  $p_i$  gelten die folgenden Formeln:

$$a, b \in E: a \leq b: \Leftrightarrow p_{a,j} \leq p_{b,j} \text{ für alle } j \in \{1, \dots, n\} \quad (1-2)$$

Objekte, für die die  $\leq$  - Relation gilt, sind **vergleichbar** miteinander. Objekte, die nicht miteinander verglichen werden können, werden **unvergleichbar** genannt.

Eine Kurzschreibweise, die die Unvergleichbarkeit zweier Objekte  $b$  und  $c$  ausdrückt, ist  $b \parallel c$ .

### **B.2.1 Lesen von Hasse-Diagrammen**

Hasse-Diagramme werden von oben nach unten gelesen. Die Objekte werden in Form von Kreisen dargestellt, wobei das Akronym des Objektnamens im Kreisinnern geschrieben steht. Benachbart sind Objekte, die genau durch eine Linie miteinander verbunden sind.

### **B.2.2 Verbandstheoretische Bezeichnungen**

**Maximale Elemente** oder auch **Maximale** genannt, sind Elemente der Gruppe  $E$  (oder  $E/\mathcal{R}$ ), die keine nach oben benachbarten Elemente aufweisen.

**Minimale Elemente** oder **Minimale** genannt, sind Elemente der Objektmenge  $E$  (oder  $E/\mathcal{R}$ ), die keine nach unten benachbarten Elemente aufweisen.

Man spricht vom **kleinsten Element**, wenn nur ein Minimal und vom **größten Element**, wenn nur ein Maximal vorhanden ist.

In einigen Fällen gibt es keine Pfade zwischen Teilen des sog. **nicht-gerichteten Graphen**, d.h. der Graph weist mehrere Zusammenhangskomponenten auf (siehe Standardliteratur zur Graphentheorie, z.B. [Harary 1976]). Dieses Umstand wird auch umschrieben mit: Das Hasse-Diagramm zerfällt in mehrere **Hierarchien**.

Falls nur ein Objekt (oder eine Äquivalenzklasse) einen Teil des Hasse-Diagramms darstellt, nennt man dieses Objekt ein **isoliertes Element**.

Objekte, die alle untereinander **vergleichbar** sind, werden auch **Ketten** in der Verbandstheorie genannt. Eine Menge an paarweise **unvergleichbaren** Objekten wird **Antikette** genannt.

Als ein **Artikulationspunkt** wird ein Objekt bezeichnet, welches bei dessen Auslassung zu einem Zerfall des Hasse-Diagramms in mehrere Hierarchien führt.

### **B.2.3 Kenngrößen (Maßzahlen) zur Unterstützung der Interpretation von Hasse-Diagrammen**

In Brüggemann et al. [Brüggemann 1995a] wurde eine Reihe von Maßzahlen eingeführt. Hier sollen nur die für die Arbeit wichtigen Kenngrößen besprochen werden.

Die **Anzahl der Äquivalenzklassen** ( $NECA = \text{number of equivalence classes}$ ) bedeutet die Anzahl der Äquivalenzklassen mit mehr als einem Objekt und wird auch als Anzahl der **nicht-trivialen Äquivalenzklassen** bezeichnet.

Die **Anzahl der Ebenen** ( $NL = \text{number of levels}$ ) bedeutet die Höhe des Hasse-Diagramms und wird definiert als  $H(E) = L(E) + 1$ .

Die **Anzahl der Elemente in der breitesten Ebene** ( $NEL = \text{number of elements in level}$ ) ist ein Maß für die Breite des Hasse-Diagramms.

Die Anzahl der **Vergleichbarkeiten**  $V(N)$  und **Unvergleichbarkeiten**  $U(N)$  sind nützliche Hilfsmittel zur Interpretation von Hasse-Diagrammen.

## **B.3 Schlüsselemente und Nachfolgermengen**

Oft interessiert die Zahl der Elemente, die zu einem herausgegriffenen Objekt vergleichbar sind, oder es ist die Frage zu beantworten, welche Elemente gleichzeitig vergleichbar sind zu zwei herausgegriffenen Objekten. Vorzugsweise wird ein maximales Element als Ausgangspunkt für die Analyse gewählt. Das ausgewählte Element wird **Schlüsselement** (key element) genannt. Es können auch mehrere Schlüsselemente gleichzeitig gewählt werden. Alle Schlüsselemente sollen eine Menge  $K (\subseteq E)$  bilden.

Die Analyse eines Schlüsselementes impliziert eine Suche nach allen Elementen, die im Hasse-Diagramm unterhalb des Schlüsselementes lokalisiert sind, d.h. alle Elemente, die von einem Schlüsselement nach unten erreicht werden können. Diese Elemente zusammen mit

den Elementen, die äquivalent aber nicht identisch mit dem Schlüsselement sind, werden *Nachfolger* (successors) genannt. Die Menge aller Nachfolger eines Schlüsselementes  $k$  wird als  $G(k)$  bezeichnet.

**Die Nachfolgermengen (successor sets) und ihre Kardinalitäten sind das Kernstück der Hasse-Analyse.** Auf der einen Seite umfaßt die Analyse der Struktur des Hasse-Diagramms, sowohl die Eigenschaften der Nachfolgermengen  $G(k)$  als auch deren Beziehung mit den Nachfolgermengen anderer Schlüsselemente. Auf der anderen Seite werden die Nachfolgermengen, gefunden für zwei Hasse-Diagramme, resultierend durch zwei Untermengen von Attributen herangezogen, um gewisse Unterschiede zu quantifizieren. Die Kardinalität von Nachfolgermengen [bezeichnet als  $\text{card } G(k)$ ] und deren mengentheoretische Kombination spielen in diesem Zusammenhang eine wichtige Rolle.

#### B.4 D-Matrix - Schnittmengen-Matrix

Die Schnittmengen- oder *D-Matrix* dient der Analyse der Struktur des Hasse-Diagramms. Es sollen Relationen zwischen Schlüsselementen aufgezeigt werden.

Es wird zunächst die Größe  $D$  für eine Menge  $K$  von Schlüsselementen wie folgt eingeführt:

$$D := \text{card} \bigcap_{k \in K} G(k) \quad (1-3)$$

Da die Bildung von Schnittmengen assoziativ ist, ist eine binäre Schnittmenge von Nachfolgermengen der Ausgangspunkt für die Analyse von  $D$ . Deshalb ist anstelle von  $D$  die *D-Matrix* von Interesse.

Das Konzept der *D-Matrix* führt zu vereinfachten Graphen. Man nennt diese Vorgehensweise "*Unfolding*" ("Auffalten" des Diagramms). Das Auffalten der Struktur eines Hasse-Diagramms erfolgt durch die Berechnung der Kardinalitäten der Nachfolgermengen und deren Schnittmengen. Die notwendigen Informationen werden in der Matrix  $\mathbf{D}$  zusammengefaßt, wobei  $m = \text{card}(E)$  ist. Es folgt:

$$D_{ij} := \text{card}(G(i) \cap G(j)), \quad (i, j = \text{Schlüsselemente}) \quad (1-4)$$

#### B.5 Bedeutung der Attribute beim Rangfolgeverfahren

Während die vorher besprochene *D-Matrix* zur Analyse der Struktur von Hasse-Diagrammen dient, zeigt die in diesem Kapitel zu beschreibende *W-Matrix* den Einfluß der Attribute auf das Rangfolgeverfahren.

Um die Bedeutung eines Attributes zu bestimmen, werden die Resultate des Rangfolgeverfahrens unter Nutzung verschiedener Untermengen an Attributen verglichen. Dies impliziert einen Vergleich von Hasse-Diagrammen. Dieser wiederum läuft auf den Vergleich von Nachfolgermengen hinaus. Dazu werden Distanzen zwischen Nachfolgermengen in einer sog. *W-Matrix* gespeichert. Details zur Distanzberechnung sind in [Brüggemann 1996a] zu finden.

Vereinfacht formuliert: Die Matrix  $\mathbf{W}$  enthält die "Entfernungen" zwischen verschiedenen Hasse-Diagrammen, resultierend aus derselben Objektmenge, aber abgeleitet unter verschiedenen Attributmengen. Der Einfluß von Attributen kann durch die Auswahl der Teilmengen der Informationsbasis erfaßt werden.

Bisher wurde nämlich die Unähnlichkeit jeweils zweier Hasse-Diagramme anhand eines Schlüsselementes und verschiedener Attributteilmenen bestimmt. Unterschiedliche Attributteilmenen beeinflussen aber die Position aller Elemente  $\in E$ . Es wird daher wie folgt verallgemeinert:

$$W(K,i,j) := \sum_{k \in K \subset E} W(k,i,j) \quad (1-5)$$

$W(K)$  ist eine symmetrische Matrix.

$W(E)$  ist die totale Unähnlichkeitsmatrix der Menge  $E$ . Angenommen  $m := \text{card } E$ . In der Praxis sind nur  $W(k)$  und  $W(E)$  Matrizen von Bedeutung. Falls die Bedeutung der Attribute untersucht werden soll, reicht es aus, den Fall, in welchem alle Attribute vorhanden sind (Fall 0) mit jenen Fällen zu betrachten, wo exakt ein Attribut ausgelassen wird:

(1) Man vergleicht Hasse-Diagramme, hervorgerufen durch  $A$  mit jenen, die gebildet werden durch Attributmengen mit nur  $n-1$  Attributen.

Die verbleibenden  $n$  Matrix-Elemente der ersten Reihe  $W(k,A,A_1) \dots, W(k,A,A_n)$  werden zusammengefaßt um ein Sensitivitätstupel des Schlüsselementes  $k$ ,  $s(k) := [W[(k,A, A_1), \dots, W(k,A,A_n)]]$  zu bilden.

2) Es wird folgende Numerierung vorgenommen:

$$A_1 = \{\#2, \dots, \#n\} \quad (1-6)$$

$$A_i = \{\#1, \dots, \#i-1, \#i+1, \dots, \#n\}, i \text{ ist ausgelassen}$$

$$A_n = \{\#1, \dots, \#n-1\}$$

(3) Es werden die Matrixelemente  $W(k,A,A_1) \dots W(k,A,A_n)$  gebildet. Je größer  $W(k,A,A_i)$ , desto größer ist die symmetrisierte Differenz zwischen  $G(k,A)$  und  $G(k,A_i)$ , und entsprechend größer ist der Einfluß von Attributen  $\#i$  auf die Position der Schlüsselementes  $k$  im Hasse-Diagramm unter  $A$  verglichen mit dem unter  $A_i$ .

(4) Die Matrix  $W$  hängt von der Auswahl des Schlüsselementes  $k$  ab. Alle Objekte werden als Schlüsselemente ausgewählt. Aus diesem Grund muß anstelle von  $W(k)$ ,  $W(E)$  untersucht werden.  $W(E)$  ist die vollständige Matrix der Menge  $E$ .

(5)  $W(E)$  wird als Sensitivitätsmaß herangezogen. Dementsprechend wird als Sensitivität des Rangfolgeverfahrens  $\sigma(i)$ , gegenüber dem Attribut  $\# i$  definiert:

$$\sigma(i) := W(E,A,A_i) \quad 1 \leq i \leq n \quad (1-7)$$

## Anhang C. Detailliertere Bewertungsergebnisse zu zehn Gruppen an Datenquellen

### C.1 Bibliographische Online Datenbanken der Chemie-Gruppe (Gruppe 1)

Diese Gruppe wird ausführlich im Hauptteil in Kapitel 7.2.1 analysiert.

### C.2 Faktendatenbanken der Chemie-Gruppe (Gruppe 2)

Die Hasse-Diagramme befinden sich im Hauptteil in Kapitel 7.2.2.

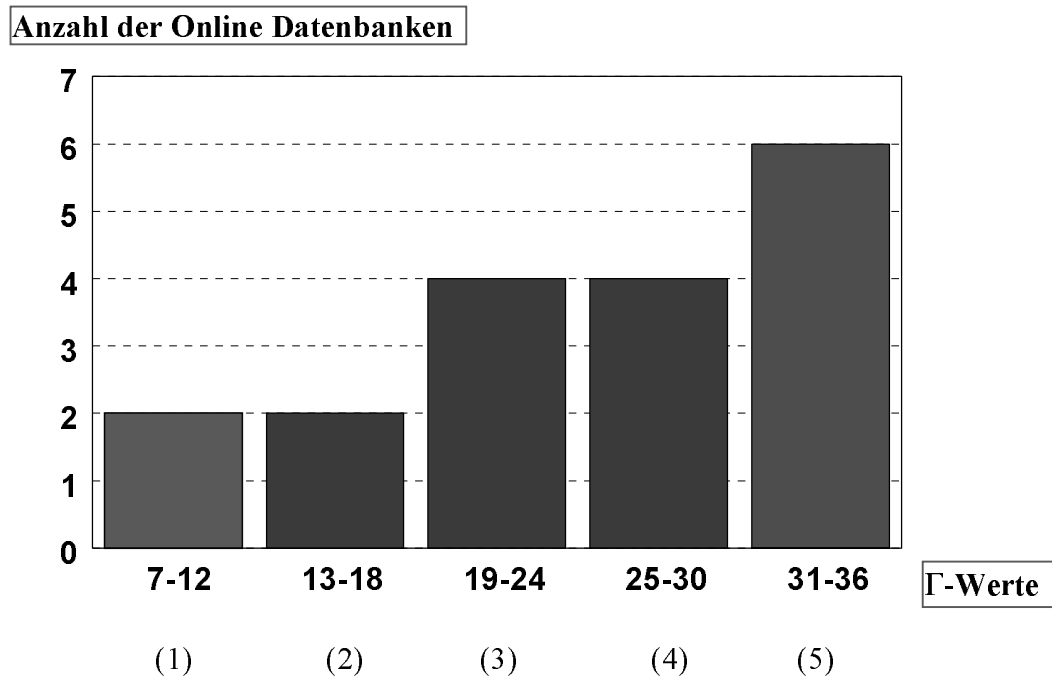
**Tabelle C1: Auswertung der Hasse-Diagramme der Gruppe 2: Online Datenbanken**

	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Bewertungskriterien</b>	11	4	7
<b>Anzahl der Ebenen (NL)</b>	3	4	4
<b>Elemente in der breitesten Ebene (NEL)</b>	13	10	9
<b>maximale Objekte</b>	13: {BEI}, {CHC}, {CIT}, {CHE}, {CMS}, {DET}, {HSD}, {MER}, {PLA}, {RTE}, {SPE}, {TRC}, {TSC}	10: {BEI}, {CHC}, {CMS}, {DET}, {HSD}, {MER}, {PLA}, {RTE}, {SPE}, {TSC}	9: {BEI}, {CHC}, {CHE}, {CMS}, {DET}, {HSD}, {MER}, {RTE}, {SPE}
<b>maximale Objekte (ohne isolierte Objekte)</b>	8: {BEI}, {CHC}, {CHE}, {DET}, {HSD}, {MER}, {TRC}, {TSC}	8: {BEI}, {CHC}, {DET}, {HSD}, {MER}, {PLA}, {RTE}, {TSC}	s.o.
<b>minimale Objekte</b>	8: {AGR}, {CIT}, {CMS}, {JAN}, {PLA}, {RTE}, {SPE}, {USA}	6: {AGR}, {CIT}, {CMS}, {JAN}, {SPE}, {USA}	3: {CIT}, {JAN}, {PLA}
<b>minimale Objekte (ohne isolierte Objekte)</b>	3: {AGR}, {JAN}, {USA}	4: {AGR}, {CIT}, {JAN}, {USA}	s.o.
<b>isolierte Objekte</b>	{CIT}, {CMS}, {PLA}, {RTE}, {SPE}	{CMS}, {SPE}	0
<b>isolierte Hierarchien</b>	0	0	0
<b>Äquivalenzklassen</b>	18	18	18
<b>nicht-triviale Äquivalenzklassen</b>	0	0	0
<b>Vergleichbarkeiten V(N)</b>	15	38	34
<b>Unvergleichbarkeiten U(N)</b>	276	230	238
<b>D-Matrix der maximalen Objekte (Auswertung)</b>	BEI: 2; CHC: 2; CHE: 1; DET: 3; HSD: 2; MER: 1	BEI: 4; CHC: 4; DET: 4; HSD: 5; MER: 2; PLA: 5; RTE: 2; TSC: 1	BEI: 2; CHC: 4; CHE: 1; CMS: 1; DET: 7; HSD: 2; MER: 3; RTE: 1; SPE: 3
<b>W-Matrix der maximalen Objekte (Auswertung)</b>	UM: 0; UP: 7; KO: 2; VE: 1; AN: 4; ID: 0; CT: 0; CE: 3; IP: 2; AR: 3; QU: 1	UM: 13; UP: 30; KO: 17; VE: 8	AN: 10; ID: 1; CT: 3; CE: 13; IP: 5; AR: 14; QU: 2

Abkürzungen siehe Abkürzungsverzeichnis

### **C.2.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 2 (Online Datenbanken)**

Die diskutierten Maximalen und Minimalen werden an dieser Stelle mit dem Gesamtmaß  $\Gamma$  verglichen (siehe auch Tabelle 50 im Hauptteil). Die  $\Gamma$ -Werte reichen von 7-36, die in fünf Aggregationsklassen eingeteilt werden (siehe Abbildung C1).



**Abbildung C1: Verteilung der  $\Gamma$ -Werte auf 18 faktenbasierte Online Datenbanken der Chemie-Gruppe**

Während die beiden unteren Aggregationsklassen (1) und (2) nur jeweils über zwei Datenbanken verfügen, ist eine Zunahme im mittleren Bereich und insbesondere in der höchsten Aggregationsklasse (5) zu verzeichnen. Folgende sechs Datenbanken, die über  $\Gamma \geq 31$  verfügen, sind somit in der obersten Aggregationsklasse (5) zu finden: BEI, CHC, CMS, DET, HSD und RTE. Für die maximalen Objekte BEI, CHC, DET, HSD und RTE liegen Übereinstimmungen zwischen Maximal und hohem Gesamtmaß vor. CMS und RTE sind isolierte Objekte, die man auch den Maximalen zurechnen kann. Somit sind alle Datenbanken der Gruppe 2 mit hohem  $\Gamma$  auch Maximale. Die Datenbanken MER und TRC sind zwar Maximale, verfügen aber nur über ein mittelhohes Gesamtmaß, d.h. sie sind in der Aggregationsklasse (3).

Minimale Datenbanken sind ARG, JAN und USA. Diese drei Datenbanken verfügen auch über ein niedriges Gesamtmaß, sind in die Aggregationsklassen (1) und (2) eingestuft. Das Fazit aus dem Gesagten ist, daß bei den faktenbasierten chemierelevanten Online Datenbanken sowohl bei den minimalen als auch maximalen Objekten eine gute Übereinstimmung zwischen diesen Aussagen und dem Gesamtmaß anzutreffen ist. Wie jedoch schon bei der Analyse der Gruppe 1 der Online Datenbanken gezeigt hat, ist die Aussagefähigkeit des Hasse-Diagramms gegenüber dem Gesamtmaß deutlich höher, da bei der Balkendarstellung des Hasse-Diagramms die Ausprägungen der einzelnen Bewertungskriterien für die Objekte erkennbar sind. So sind z.B. HSD und DET, die  $\Gamma$ -äquivalent sind, in allen Bewertungsziffern mit Ausnahme von QU unterschiedlich eingestuft. Diese Tatsache kommt nur in dem Hasse-Diagramm zum Ausdruck.

### **C.2.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 2 (Online Datenbanken)**

Wie oben ausgeführt, ist die Struktur des Hasse-Diagramms im besonderen abhängig von den Bewertungskriterien und deren Ausprägungsmengen. In Analogie zur Vorgehensweise bei Gruppe 1 (siehe Hauptteil) werden auch hier die Ausprägungsmengen und Indizes für Gruppe 2 aufgelistet.

**Tabelle C2: Ausprägungsmengen und Indizes der einzelnen Bewertungskriterien der Gruppe 2: Online Datenbanken, 18 Objekte, 12 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	CT	CE	IP	AR	QU
0	1	2	5	4	5	2	4	9	2	0	3
1	3	4	2	6	1	0	1	5	10	0	1
2	3	5	4	4	2	10	0	3	3	11	2
3	3	4	4	4	3	4	1	0	2	4	2
4	3	3	2	0	3	0	7	1	0	1	8
5	5	0	1	0	4	2	5	0	1	2	2
MW	3,06	2,11	1,94	1,44	2,56	2,33	3,17	0,83	1,50	2,67	2,94
ME	3	2	2	1	3	2	4	0,5	1	2	4
STDEV	1,662	1,278	1,589	1,097	1,977	1,283	1,978	1,098	1,200	1,029	1,697
$I_i$	1,71	1,57	1,67	1,37	1,69	1,15	1,38	1,16	1,27	1,05	1,55
$U_i$	6	5	6	4	6	4	5	4	5	4	6

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

Auf der Grundlage der bei Gruppe 1 (Kapitel 7.2.1.4.3) eingeführten Erläuterungen in bezug auf Mittelwert, Median, Umfangsindex und Shannon-Index ist hier anzumerken, daß die Bewertungskriterien UM, KO, AN und QU eine gute Verteilung aufweisen. Bei der gezielten Betrachtung des Shannon-Indexes ist UM am besten, gefolgt von AN und KO. Schlecht sind die Kriterien VE, ID und CE verteilt.

Die einzelnen Bewertungskriterien werden nochmals eingehend beleuchtet.

Die allgemeinen Bewertungskriterien UM (Umfang der Datenbank) und KO (Kosten der Recherche) verfügen über alle Bewertungsziffern, d.h. es werden sechs verschiedene Zahlen vergeben. UP (Häufigkeit der Aktualisierung) verfügt über fünf verschiedene Bewertungsziffern und das Kriterium VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) über nur vier. Das letztgenannte allgemeine Kriterium hat somit die geringste Bandbreite und ist das am wenigsten aussagekräftigste.

Auch bei den inhaltlichen Bewertungskriterien gibt es Ausprägungsmengen, die unter der vollständigen Ausprägungsmenge mit den Ziffern 0 bis 5 liegen. Diese Tatsache fällt bei den CE (Chemikalien Entwicklung) auf, bei der die Ausprägungsmenge bei  $\{0,1,2,4\}$  liegt. Auch das Bewertungskriterium IP (Informationsparameter für Umweltchemikalien) verfügt zwar über die Bewertungsziffern 0,1,2,3,5, ist jedoch nur als mittelmäßig anzusehen. Dieser Umstand kann durch die Verteilung der Bewertungsziffern begründet werden. Die Bewertungsziffern 0, 1 und 2 werden im Vergleich zu den hohen Bewertungsziffern häufiger vergeben.

Die Bewertungskriterien AN (Anzahl an Chemikalien) und QU (Validität/Güte der Datenbank) verfügen über alle möglichen Bewertungsziffern. Das Bewertungskriterium ID (Identifikationsmerkmale) mit einer Ausprägungsmenge von  $\{0,2,3,5\}$  ist zwar schlecht in bezug auf die Indizes anzusehen, jedoch ist eine deutliche Verbesserung bei dieser Gruppe der

faktenbasierten Datenbanken hin zu den mittleren und hohen Bewertungsziffern im Vergleich zu den textbasierten chemierelevanten Online Datenbanken zu erkennen.

Das bedeutet, daß chemiebezogene faktenbasierte Datenbanken in der Regel Schwächen bei den umweltbezogenen bzw. Umweltchemikalien-bezogenen Kriterien aufweisen. Stärken dieser Art der Datenbanken sind die Verfügbarkeit der Datenbanken auf anderen Medien und die Anzahl der Chemikalien. Ebenso ist eine Verbesserung der Situation bei den Identifikationsmerkmalen ersichtlich.

### C.3 Textbasierte Datenbanken der Umwelt-Gruppe (Gruppe 3)

**Tabelle C3: Auswertung der Hasse-Diagramme Gruppe 3: Online Datenbanken**

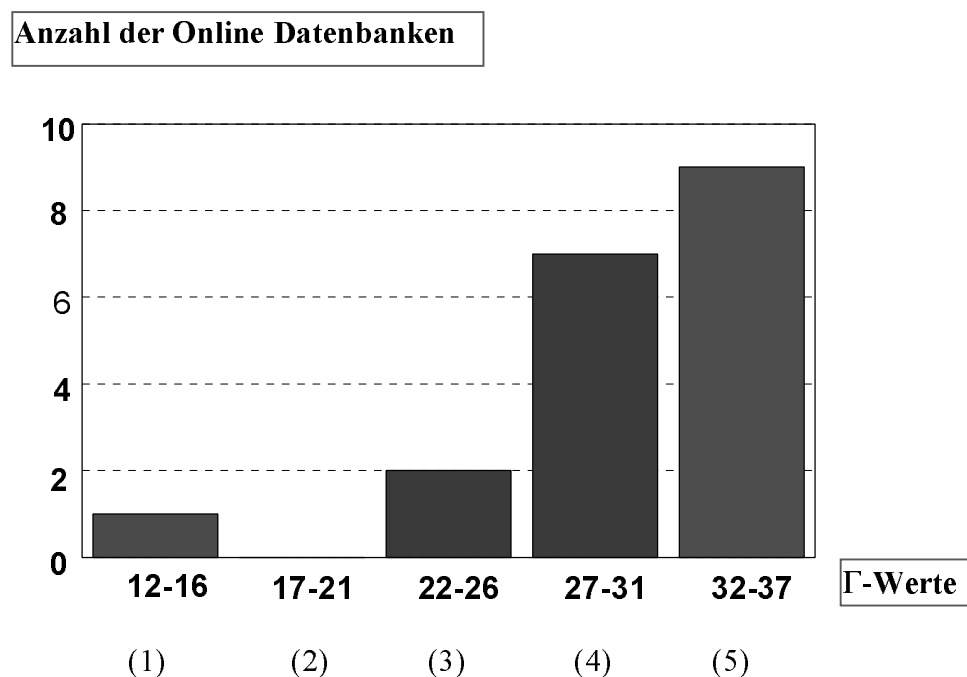
	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Bewertungskriterien</b>	12	4	8
<b>Anzahl der Ebenen (NL)</b>	3	5	4
<b>Elemente in der breitesten Ebene (NEL)</b>	13	4	9
<b>maximale Objekte</b>	13: {BIO}, {CAS}, {CAB}, {COM}, {EMB}, {ENE}, {GEO}, {MED}, {NLD}, {NTI}, {OCE}, {POL}, {TOX}	3: {CAB}, {MED}, {TOX}	9: {CAS}, {CAB}, {COM}, {EMB}, {ENE}, {GEO}, {NLD}, {NTI}, {TOX}
<b>maximale Objekte (ohne isolierte Objekte)</b>	11: {BIO}, {CAS}, {CAB}, {EMB}, {ENE}, {GEO}, {MED}, {NTI}, {OCE}, {POL}, {TOX}	s.o.	7: {CAS}, {CAB}, {EMB}, {ENE}, {GEO}, {NTI}, {TOX}
<b>minimale Objekte</b>	6: {ASF}, {COM}, {NLD}, {SCI}, {UFO}, {ULI}	2: {UFO}, {ULI}	5: {ASF}, {COM}, {NLD}, {SCI}, {UFO}
<b>minimale Objekte (ohne isolierte Objekte)</b>	4 {ASF}, {SCI}, {UFO}, {ULI}	s.o.	3: {ASF}, {SCI}, {UFO}
<b>isolierte Objekte</b>	{COM}, {NLD}	0	{COM}, {NLD}
<b>isolierte Hierarchien</b>	0	0	0
<b>Äquivalenzklassen</b>	19	15	19
<b>nicht-triviale Äquivalenzklassen</b>	0	4: {BIO,EMB}, {CAS,SCI}, {COM,GEO}, {ENE,NTI}	0
<b>Vergleichbarkeiten V(N)</b>	19	98	29
<b>Unvergleichbarkeiten U(N)</b>	304	154	284
<b>D-Matrix der maximalen Objekte (Auswertung)</b>	BIO: 2; CAS: 1; CAB: 3; EMB: 1; ENE: 3; GEO: 2; MED: 1; NTI: 1; OCE: 1; POL: 1; TOX: 1	CAB: 9; MED: 13; TOX: 10	CAS: 1; CAB: 4; EMB: 1; ENE: 5; GEO: 2; NTI: 1; TOX: 4
<b>W-Matrix der maximalen Objekte (Auswertung)</b>	UM: 0; UP: 0; KO: 2; VE: 2; AN: 1; ID: 1; CT: 3; CE: 5; IP: 0; PE: 15; AR: 0; QU: 1	UM: 0; UP: 0; KO: 4; VE: 6	AN: 1; ID: 1; CT: 4; CE: 9; IP: 0; PE: 18; AR: 0; QU: 3

Abkürzungen siehe Abkürzungsverzeichnis

Die Hasse-Diagramme befinden sich im Hauptteil in Kapitel 7.2.4.

### **C.3.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 3 (Online Datenbanken)**

Auch bei der Gruppe 3, d.h. den textbasierten umweltrelevanten Datenbanken soll in Analogie zur Vorgehensweise bei den chemierelevanten textbasierten Datenbanken (Gruppe 1) gezeigt werden, inwieweit die Hasse-Diagramm-Technik mit den Aussagen, die aus dem Gesamtmaß  $\Gamma$  gezogen werden können, übereinstimmen oder voneinander abweichen (vgl. hierzu auch Tabelle 51 im Hauptteil). Die Werte für das Gesamtmaß  $\Gamma$  liegen zwischen 12 und 37. Sie werden in fünf Aggregationsklassen eingeteilt, wobei die vier unteren Stufen jeweils fünf Zahlen und die oberste Stufe sechs Zahlen umfassen.



**Abbildung C2: Verteilung der  $\Gamma$ -Werte auf 19 textbasierte Online Datenbanken der Umwelt-Gruppe**

Die Abbildung macht deutlich, daß wesentlich mehr Datenbanken in den oberen Aggregationsklassen zu finden sind als in den unteren. In die Aggregationsklasse (1) ist nur eine Datenbank eingeteilt und in die zweite Klasse keine. Neun Datenbanken sind in der obersten Aggregationsklasse, d.h. sie verfügen über  $\Gamma$ -Werte  $\geq 32$  ( BIO, CAS, CAB, EMB, ENE, GEO, NTI, POL, TOX) und nur die Datenbank UFO weist einen  $\Gamma$ -Wert  $< 16$  auf. Den höchsten  $\Gamma$ -Wert hat die Datenbank CAB, gefolgt von CAS, ENE und TOX mit jeweils  $\Gamma = 36$ . Bei dieser Gruppe liegt eine gute Übereinstimmung zwischen hohem Gesamtmaß und den maximalen Datenbanken vor. Die Datenbank SCI ( $\Gamma = 31$ ) hingegen ist ein minimales Objekt im Hasse-Ansatz. SCI ist schlechter eingestuft als BIO und CAS (vgl. auch das Hasse-Diagramm in Abbildung 20). OCE ist ein Maximal bei der Hasse-Analyse, liegt jedoch im Gesamtmaß  $\Gamma$  in der Aggregationsklasse 4. Bei den ausgeführten Beispielen läßt sich erkennen, daß die Hasse-Diagramm-Technik der groben Einstufung mittels Gesamtmaß überlegen ist. Besonders deutlich wird auch die bessere Aussagefähigkeit der Hasse-Diagramme bei dem Vergleich der Datenbanken OCE und ENV, die beide einen  $\Gamma$ -Wert 28 haben, somit  $\Gamma$ -äquivalent sind. Während OCE ein maximales Objekt darstellt, ist ENV in der zweiten Ebene des Hasse-Diagramms der

Gruppe 3 zu finden. Darüber hinaus ist die Hasse-Diagramm-Technik auch bei der Betrachtung der Datenbanken CAS, ENE, TOX, die  $\Gamma$ -äquivalent sind, insbesondere in der Balkendiagramm Darstellung aussagekräftiger. Hierbei würden die deutlichen Unterschiede z.B. bei dem Bewertungskriterium ID hervortreten.

### **C.3.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 3 (Online Datenbanken)**

**Tabelle C4: Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 3: Online Datenbanken, 12 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	CT	CE	IP	PE	AR	QU
0	0	0	5	0	2	13	0	5	0	6	0	0
1	1	0	6	2	5	5	1	7	0	3	0	3
2	5	1	6	2	5	0	0	1	1	1	19	9
3	1	0	0	12	5	1	7	2	5	5	0	7
4	6	12	2	1	1	0	5	2	4	2	0	0
5	6	6	0	2	1	0	6	2	9	2	0	0
MW	3,57	4,21	1,37	2,95	2,05	0,42	3,79	1,74	4,11	2,00	2,00	2,21
ME	4	4	1	3	2	0	4	1	4	2	2	2
STDEV	1,346	0,713	1,294	1,026	1,311	0,769	1,084	1,727	0,994	1,795	0,000	0,713
$I_i$	1,39	0,81	1,32	1,16	1,60	0,77	1,24	1,59	1,19	1,64	0	1,01
$U_i$	5	3	4	5	6	3	4	6	4	6	1	3

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

Bei der Betrachtung der Indizes fällt auf, daß bezüglich der Verteilung die Bewertungskriterien AN, CE und PE sehr gut abschneiden. Eine schlechte Verteilung weisen UP, ID und QU auf. AR ist bezüglich der Verteilung als besonders negativ anzusehen, obwohl dieses Kriterium - wie schon oben ausgeführt - nur über vier anstelle von sechs Bewertungsziffern verfügt.

Das Kriterium UP (Häufigkeit der Aktualisierung) hat nur drei verschiedene Bewertungsziffern und die VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) fünf. Die Datenbank Toxicology Literature Online hat die höchste Bewertungsziffer, weil sie neben der Verfügbarkeit bei vielen anderen Hosts und auf CD-ROM auch im Internet verfügbar ist [Cambridge Scientific Abstracts 1996]. UP (Häufigkeit der Aktualisierung der Datenbank) ist unter den allgemeinen Bewertungskriterien am schlechtesten verteilt. Bibliographische Datenbanken werden in der Regel monatlich bis dreimonatlich aktualisiert. Damit werden sie deutlich häufiger gepflegt als Faktendatenbanken (numerische Datenbanken). Am häufigsten, d.h. wöchentlich, wird die Datenbank CAS aktualisiert [STN 1995a].

Bei dem Kriterium ID (Identifikationsparameter für Chemikalien) umfaßt die Ausprägungsmenge nur  $\{0, 1, 3\}$ . Die Bewertungsziffer 3 wird nur einmal, und zwar für CAS vergeben.

Von besonderem Interesse bei umweltbezogenen Datenbanken ist die Betrachtung des Bewertungskriteriums IP (Informationsparameter für Umweltchemikalien), welches zwar in der Verteilung nicht so gut abschneidet, jedoch eine deutliche Verlagerung zu den hohen Bewertungsziffern aufweist.

Vergleicht man die Bewertungskriterien UP, ID und QU, die alle über einen Umfangsindex von 3 verfügen, so läßt die Auswertung des Shannon-Indexes die Interpretation zu, daß QU am besten verteilt ist.

Zusammengefaßt läßt sich feststellen, daß umweltbezogene textbasierte Datenbanken in der Regel Stärken bei den umweltbezogenen bzw. Umweltchemikalien-bezogenen Kriterien aufweisen. Schwächen dieser Art der Datenbanken sind die Identifikationsmerkmale und die Art/Anwendung der Chemikalien.

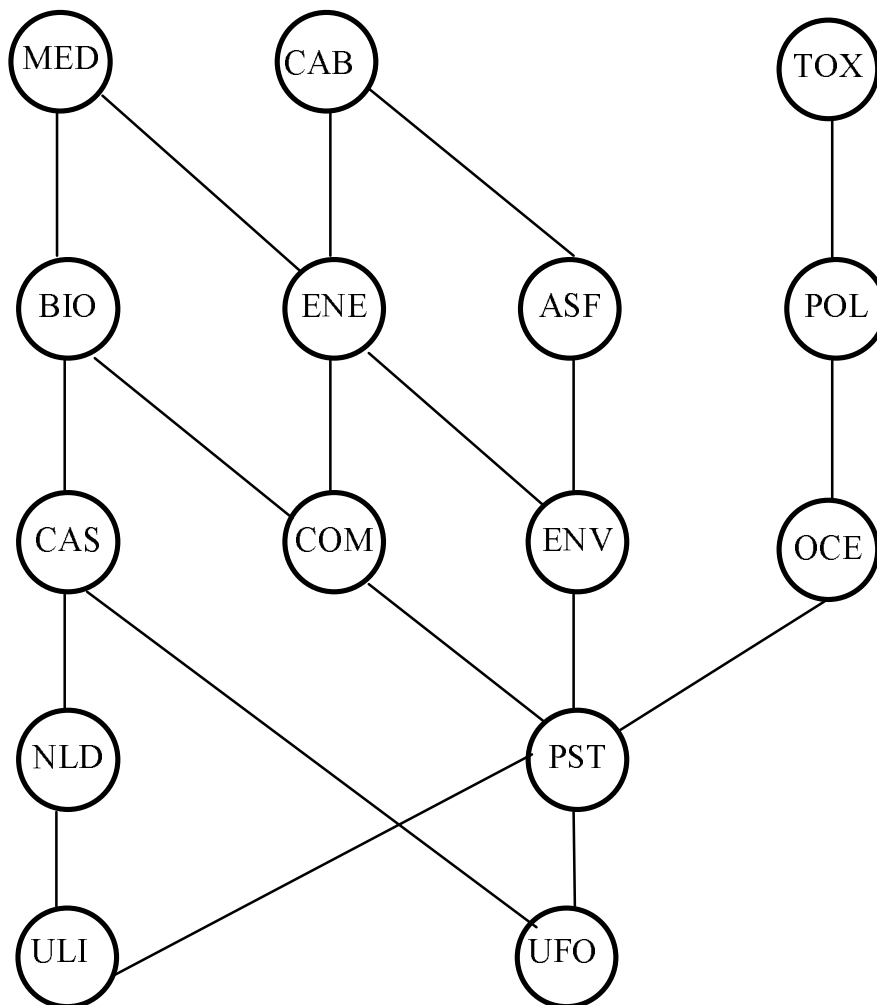
### **C.3.3 Besonderheiten in der Struktur des Hasse-Diagramms der Gruppe 3: allgemeine Bewertungskriterien**

An dieser Stelle soll das Diagramm  $P = "19"$ ,  $IB_{all}$  für eine weitere Interpretation aufgegriffen werden, da sich hierbei einige wichtige Besonderheiten ergeben (siehe Abbildung C3 und Tabelle C3, Spalte 3).

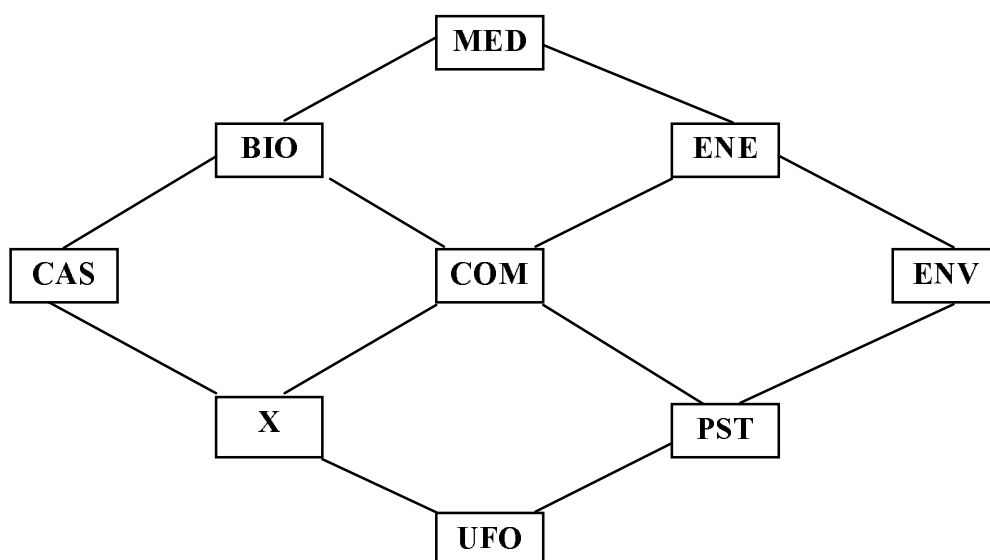
Bei diesem Hasse-Diagramm, in welchem nur vier Bewertungskriterien berücksichtigt werden, kann eine deutlichere Aussage getroffen werden, als bei der Informationsbasis mit 12 Kriterien. Es gibt hier beispielsweise vier nicht-triviale Äquivalenzklassen mit zwei Objekten, d.h. die Datenbanken BIO/EMB, CAS/SCI, COM/GEO und ENE/NTI weisen in bezug auf die vier Kriterien gleiche Tupel auf. Im Diagramm sind keine isolierten Objekte vorhanden. Es ergeben sich folgende drei maximale Objekte CAB (CAB Abstracts), MED (Medical Literature Analysis and Retrieval System Online) und TOX (Toxicology Literature Online).

Minimale Objekte sind UFO (Umweltforschungsdatenbank) und ULI (Umweltliteraturdatenbank). Auffallend ist die regelmäßige Struktur des Diagramms, insbesondere der Datenbanken MED, BIO, ENE, CAS, COM, ENV, PST und UFO, wie in den Abbildungen C3 und C4 gezeigt wird.

Ergänzt man nun die Gruppe  $\{MED, ENE, ENV, BIO, COM, PST, CAS, UFO\}$  um ein fiktives Objekt X, so läßt sich diese Gruppe wie in Abbildung C4 gezeigt, darstellen. Dessen Struktur läßt vermuten, daß auch eine gewisse Regelmäßigkeit in der Datenstruktur vorliegt [Trotter 1992]. So läßt sich zeigen, daß die Dimension der partiell geordneten Menge, die dem Hasse-Diagramm von Abbildung C3 zugrunde liegt, gleich zwei ist [Möhring 1995]. Die Dimension einer partiell geordneten Menge bedeutet die minimale Zahl an Kriterien, durch die ein isomorphes (d.h. jede Vergleichbarkeit in einem Diagramm impliziert eine entsprechende im anderen Diagramm und umgekehrt) Hasse-Diagramm aufgebaut werden kann. Im hier vorliegenden Beispiel genügen die Bewertungskriterien UM und KO, um ein zum Hasse-Diagramm isomorphes Diagramm zu erzeugen (Abbildung C4). Isomorph bedeutet, daß alle Ordnungsrelationen übereinstimmen.



**Abbildung C3:** Gruppe 3: Online Datenbanken, 4 allgemeine Bewertungskriterien



**Abbildung C4:** Gruppe 3: Online Datenbanken, allgemeine Bewertungskriterien  
Strukturelle Besonderheiten

## C.4 Umweltrelevante Faktendatenbanken (Gruppe 4)

Diagramme befinden sich im Hauptteil in Kapitel 7.2.5.

**Tabelle C5: Auswertung der Hasse-Diagramme der Gruppe 4: Online Datenbanken**

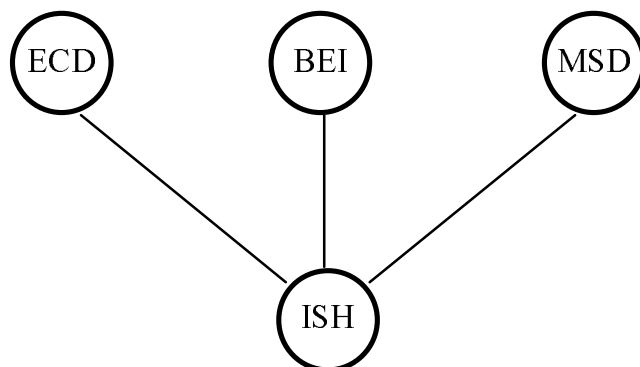
	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Bewertungskriterien</b>	11	4	7
<b>Anzahl der Ebenen (NL)</b>	3	7	3
<b>Elemente in der breitesten Ebene (NEL)</b>	10	4	9
<b>maximale Objekte</b>	10: {BEI}, {CES}, {CHE}, {CHR}, {ECD}, {ENF}, {HAZ}, {HSD}, {MSD}, {TSP}	3: {ECD}, {HSD}, {MSD}	9: {BEI}, {CES}, {CHE}, {ECD}, {ENF}, {HAZ}, {HSD}, {MSD}, {TSP}
<b>maximale Objekte (ohne isolierte Objekte)</b>	5: {BEI}, {ECD}, {HSD}, {MSD}, {TSP}	s.o.	8: {BEI}, {CES}, {CHE}, {ECD}, {ENF}, {HSD}, {MSD}, {TSP}
<b>minimale Objekte</b>	8: {CES}, {CHE}, {CHR}, {ENF}, {ERN}, {HAZ}, {ISH}, {PHY}	1: {CES}	3: {ERN}, {HAZ}, {PHY}
<b>minimale Objekte (ohne isolierte Objekte)</b>	3: {ERN}, {ISH}, {PHY}	s.o.	2: {ERN}, {PHY}
<b>isolierte Objekte</b>	5: {CES}, {CHE}, {CHR}, {ENF}, {HAZ}	0	1: {HAZ}
<b>isolierte Hierarchien</b>	1: HSD-ERN	0	0
<b>Äquivalenzklassen</b>	14	13	14
<b>nicht-triviale Äquivalenzklassen</b>	0	1: {ENF,ISH}	0
<b>Vergleichbarkeiten V(N)</b>	10	56	20
<b>Unvergleichbarkeiten U(N)</b>	162	72	142
<b>D-Matrix der maximalen Objekte (Auswertung)</b>	BEI: 2; ECD: 2; HSD: 1; MSD: 3; TSP: 1	ECD: 10; HSD: 4; MSD: 6	BEI: 3; CES: 1; CHE: 1; ECD: 2; ENF: 2; HSD: 2; MSD: 4; TSP: 2
<b>W-Matrix der maximalen Objekte (Auswertung)</b>	UM: 1; UP: 2; KO: 0; VE: 1; AN: 0; ID: 1; CT: 1; CE: 6; IP: 0; AR: 0; QU: 2	UM: 8; UP: 3; KO: 0; VE: 3	AN: 0; ID: 1; CT: 1; CE: 6; IP: 0; AR: 0; QU: 2

Abkürzungen siehe Abkürzungsverzeichnis

### C.4.1 Besonderheiten in der Struktur des Hasse-Diagramms der Gruppe 4: Online Datenbanken

Es ist auffällig, daß CES bei den allgemeinen Kriterien ein Minimal, bei den inhaltlichen Kriterien hingegen ein Maximal darstellt.

Die folgende Struktur ist im Hasse-Diagramm sowohl bei der Betrachtung der IB<sub>all+inh</sub> als auch bei IB<sub>inh</sub> zu finden. Dieser Umstand macht deutlich, daß die Vergleichbarkeiten und Unvergleichbarkeiten, die für diese Struktur verantwortlich sind, alleine durch IB<sub>inh</sub> induziert werden.



**Abbildung C5: Strukturelle Besonderheit der Gruppe 4: Online Datenbanken**

Die Maximalen BEI, ECD und HSD sollen näher betrachtet werden (vgl. hierzu Tabelle 52).

Die Tupel dieser maximalen Objekte sowie der Datenbank ISH sind:

BEI (5, 1, 3, 3, 5, 5, 4, 0, 3, 2, 5)

ECD (5, 2, 5, 5, 5, 3, 5, 0, 4, 2, 3)

MSD (4, 5, 3, 0, 4, 2, 4, 1, 3, 5, 4)

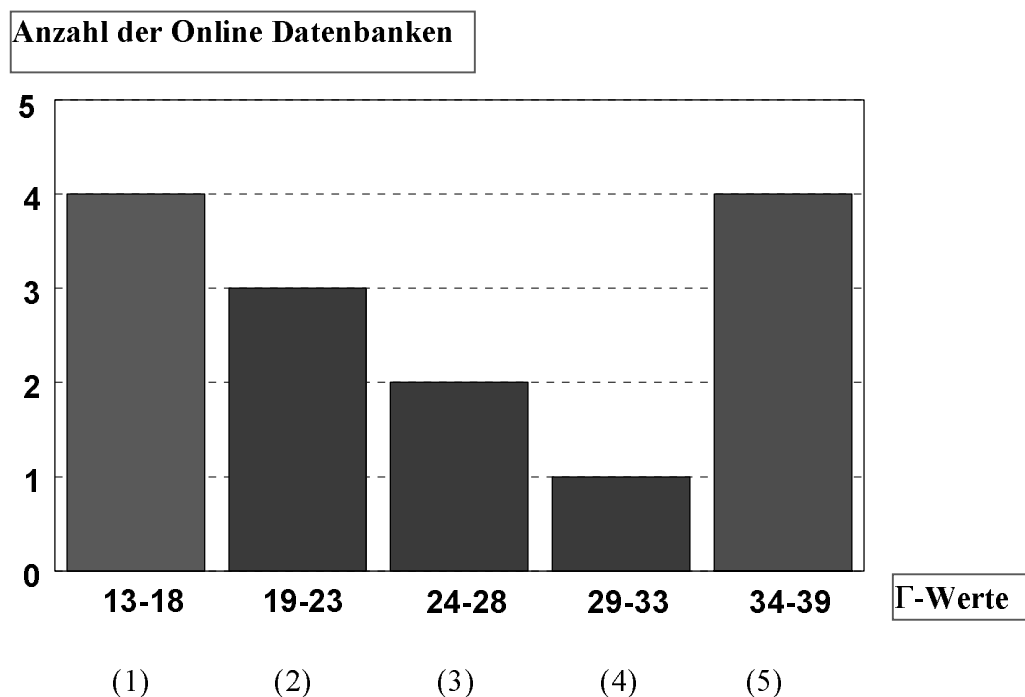
ISH (3, 0, 2, 0, 3, 2, 3, 0, 2, 2, 3)

Die drei maximalen Datenbanken sind unvergleichbar miteinander. Beispielsweise hat die BEILSTEIN Database (BEI) eine höhere Bewertungsziffer für das Kriterium ID (Identifikations-merkmale) als Environmental Chemicals Data and Information Network (ECD) und Material Safety Data Sheets (MSD). Die Datenbank BEI verfügt beispielsweise über eine deutlich bessere Bewertung in den Identifikationsmerkmalen, da sie Strukturformeln, Summenformeln, Molekulargewichte, CAS-Nummern und sonstige Registriernummern vorhält und diese auch recherchierbar sind [STN 1995d], während ECD zwar auch über alle diese Identifikationsmerkmale verfügt, die Strukturformeln jedoch nur ausgegeben werden. In dem Bewertungskriterium QU (Validität/Güte der Datenbank) ist BEI ebenfalls deutlich besser eingestuft als ECD und MSD, da beim BEILSTEIN Institut in der Regel nur durch Experten validierte Daten in die Datenbank aufgenommen werden, was bei ECD und MSD nicht der Fall ist. Auf der anderen Seite verfügt ECD über eine höhere Anzahl an IP (Informationsparametern für Umweltchemikalien) als MSD und BEI. ECD wird deutlich besser bewertet als BEI und MSD bei dem allgemeinen Bewertungskriterium VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) sowie bei den KO (Kosten der Recherche). ECDIN gibt es nicht nur auf CD-ROM sondern auch im Internet [Joint Research Center 1996]. Im Internet ist die Datenbank kostenlos verfügbar. Die BEILSTEIN Database ist bei anderen Hosts verfügbar und auch teilweise auf CD-ROM [Lawson 1995]. Die Datenbank MSD ist nicht im Internet verfügbar [STN 1995f]. Zwar gibt es viele Ressourcen mit Sicherheitsdatenblättern über Chemikalien (Material Safety Data Sheets) im Internet [University of Kentucky 1997], jedoch nicht diese als Online Datenbank zusammengefaßte Sammlung der Occupational Health Services Inc. in Nashville.

Die Datenbank ISH (Information System for Hazardous Organics in Water) ist in allen Bewertungskriterien gleich oder vergleichbar schlechter eingestuft als die diskutierten maximalen Objekte.

### **C.4.2 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 4 (Online Datenbanken)**

Für die umweltrelevanten Faktendatenbanken reicht das Gesamtmaß  $\Gamma$  von 13-39 (siehe Tabelle 52). Auch diese Werte werden in fünf Aggregationsklassen aufgeteilt. In diesem Fall umfassen alle Stufen fünf Zahlen bis auf die erste und die fünfte Klasse, die jeweils sechs Zahlen beinhalten. In folgender Abbildung ist die Verteilung der  $\Gamma$ -Werte auf die Anzahl der Online Datenbanken der umweltrelevanten Faktendatenbanken dargestellt.



**Abbildung C6: Verteilung der  $\Gamma$ -Werte auf 14 faktenbasierte Online Datenbanken der Umwelt-Gruppe**

Es fällt auf, daß sowohl die höchste als auch die niedrigste Aggregationsklasse viele Datenbanken enthält. Ferner ist bei dieser Gruppe die Tatsache zu verzeichnen, daß es keine  $\Gamma$ -äquivalenten Objekte gibt, was aufgrund der niedrigen Anzahl an Objekten auch nicht verwundert. Diese Gruppe ist die kleinste Gruppe überhaupt.

Der höchste  $\Gamma$ -Wert wird für ECD vergeben. Daneben verfügen BEI, HSD und MSD über  $\Gamma$ -Werte der höchsten Aggregationsklasse. Diese vier Datenbanken sind auch maximale Objekte im Hasse-Diagramm. Die mittels der Hasse-Diagramm-Technik als maximal eingestufte Datenbank TSP hat ein Gesamtmaß  $\Gamma = 26$ , ist somit in der dritten Aggregationsklasse.

Bei den Minimalen ERN, ISH, PHY haben alle Datenbanken  $\Gamma$ -Werte  $\leq 20$ , sind demnach in den Aggregationsklassen (1) und (2) zu finden.

Die Datenbanken CES, CHR, HAZ liegen bei dem Gesamtmaß ebenfalls unter dem Wert von 20. Diese Datenbanken sind im Hasse-Diagramm isolierte Objekte. Diese Besonderheit, die auch auf besondere Datenstrukturen hinweist, kann anhand von aggregierten Maßen, wie dem Gesamtmaß  $\Gamma$ , nicht erfaßt werden. Die Hasse-Diagramm-Technik kann insbesondere in der Balkendarstellung die Objekte wesentlich differenzierter darstellen als das Gesamtmaß, z.B. die

deutlichen Unterschiede bei HSD und MSD in den Bewertungskriterien UM, UP, VE, AN und IP herausstellen.

### **C.4.3 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 4**

Zur Untersuchung der Verteilung der 11 Bewertungskriterien der Gruppe 4 wird hier analog zur Vorgehensweise, die bei der Gruppe 1 beschrieben wurde, verfahren.

**Tabelle C6: Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 4: Online Datenbanken, 14 Objekte, 11 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	CT	CE	IP	AR	QU
0	2	7	2	8	5	0	0	7	0	0	1
1	1	1	1	0	3	5	1	2	5	0	4
2	2	3	5	3	1	6	1	1	3	8	1
3	3	2	4	2	1	2	3	0	2	0	3
4	3	0	1	0	2	0	6	2	3	4	4
5	3	1	1	1	2	1	3	2	1	2	1
MW	2,95	1,29	2,29	1,21	1,86	2,00	3,64	1,57	2,43	3,00	2,57
ME	3	2,5	2	0	1	2	4	0,5	2	2	3
STDEV	1,730	1,590	1,383	1,626	1,966	1,109	1,151	2,027	1,399	1,240	1,555
$I_i$	1,74	1,33	1,57	1,12	1,63	1,20	1,40	1,37	1,50	0,96	1,61
$U_i$	6	5	6	4	6	4	5	5	5	3	6

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

Aus der Tabelle wird deutlich, daß das Bewertungskriterium UM (Umfang der Datenbank) am besten verteilt ist, da  $U_i = 6$ , d.h. alle möglichen sechs Bewertungsziffern vergeben sind und der höchste Wert für den Shannon-Index  $I_i$  vorliegt. Auch liegen Mittelwert und Median sehr dicht beieinander, was ebenfalls ein Hinweis für eine gute Verteilung ist. Ebenso weisen eine gute Verteilung die Kriterien AN (Anzahl an Chemikalien), QU (Validität/Güte der Datenbank), KO (Kosten der Recherche) auf, wobei die Reihenfolge gemäß Shannon-Index gewählt wird, da dieser die differenzierteste Angabe bezüglich der Verteilung gibt.

Bei AN ist jedoch eine Neigung hin zu den kleinen Bewertungsziffern zu vermerken. Das bedeutet, daß viele umweltrelevante Faktendatenbanken über nur eine geringe Anzahl an Chemikalien verfügen. Das Bewertungskriterium IP (Informationsparameter für Umweltchemikalien) ist bezüglich der Indizes mittelmäßig verteilt. Auch hier wird deutlich, daß die niedrigen Bewertungsziffern 1 und 2 häufiger vergeben werden als die hohen.

Die schlechteste Verteilung weist das Kriterium AR (Art/ Anwendung der Chemikalien) auf, wobei dieses Kriterium, wie schon oben erwähnt, nur über eine Ausprägungsmenge von vier Ziffern, d.h. {2,3,4,5} verfügt. Anzumerken ist außerdem, daß nur die Ziffern 2, 4 und 5 vergeben werden. Das bedeutet, daß die Gruppe 4 mehrere Datenbanken umfaßt, die sich auf besondere Chemikalien spezialisieren. Schlecht verteilt sind auch die Kriterien VE (Verfügbarkeit der Datenbank bei anderen Hosts bzw. auf anderen Medien) und ID (Identifikationsmerkmale). Bei ID ist jedoch auffällig, daß die schlechteste Bewertungsziffer nicht vergeben wird, d.h. alle Faktendatenbanken neben dem Chemikaliennamen auch über die CAS-Nummer verfügen, einige darüber hinaus auch über Summenformel, Molekulargewicht und Strukturformel. Hier liegt eine wesentlich bessere Situation vor als bei der Gruppe 3, d.h. den textbasierten Daten-

banken. Für die VE werden nur vier verschiedene Bewertungsziffern vergeben, wobei die schlechteste, d. h. Null für acht von 14 Datenbanken, jedoch die beste Bewertungsziffer nur einmalig vergeben wird. Wie schon oben erwähnt, ist die Datenbank ECDIN auch im Internet frei verfügbar [Joint Research Center 1997]. Viele Faktendatenbanken werden nur von einem Host angeboten und sind weder in Buchform noch auf CD-ROM erhältlich. Im Mittelfeld bezüglich der Verteilung liegen die Kriterien UP, CT und CE. Das Kriterium UP (Häufigkeit der Aktualisierung) verfügt hier zwar über die Ausprägungsmenge {0,1,2,3,5}, wobei jedoch die schlechteste Einstufung, d.h. unregelmäßige Aktualisierung der Datenbank sehr häufig auftritt. Bei CE (Chemikalien Entwicklung) ist zwar auch die Ausprägungsmenge mit {0,1,2,4,5} als günstig anzusehen, jedoch wird die Ziffer Null sehr häufig vergeben. Das bedeutet, daß es bei vielen Faktendatenbanken der Umwelt-Gruppe keine Verbesserung in der Anzahl der Umweltchemikalien gegeben hat.

## C.5 Umweltchemikalien-relevante Datenbanken (Gruppe 5)

**Tabelle C7: Auswertung der Hasse-Diagramme der Gruppe 5: Online Datenbanken**

	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Bewertungskriterien</b>	12	4	8
<b>Anzahl der Ebenen (NL)</b>	2	4	2
<b>Elemente in der breitesten Ebene (NEL)</b>	16	7	15
<b>maximale Objekte</b>	16: {ANA}, {API}, {CAS}, {CAB}, {CBN}, {CEN}, {CJA}, {COM}, {CSN}, {IPA}, {JIS}, {NTI}, {PAP}, {PIR}, {PRO}, {TOX}	7: {CAS,SCI}, {CAB}, {CBN}, {CEN}, {CJA}, {PAP}, {TOX}	15: {ANA}, {API}, {CAS}, {CAB}, {CBN}, {CEN}, {CJA}, {CSN}, {IPA}, {JIS}, {NTI}, {PAP}, {PIR}, {PRO}, {TOX}
<b>maximale Objekte (ohne isolierte Objekte)</b>	2: {CAS}, {PRO}	5: {CAS,SCI}, {CAB}, {CBN}, {PAP}, {TOX}	4: {CAS}, {CJA}, {PRO}, {TOX}
<b>minimale Objekte</b>	16: {ANA}, {API}, {CAB}, {CBN}, {CEN}, {CIN}, {CJA}, {COM}, {CSN}, {IPA}, {JIS}, {NTI}, {PAP}, {PIR}, {SCI}, {TOX}	6: {ANA}, {API}, {CEN}, {CJA}, {CSN}, {PIR}	14: {ANA}, {API}, {CAB}, {CBN}, {CEN}, {CIN}, {COM}, {CSN}, {IPA}, {JIS}, {NTI}, {PAP}, {PIR}, {SCI}
<b>minimale Objekte (ohne isolierte Objekte)</b>	2: {CIN}, {SCI}	4: {ANA}, {API}, {CSN}, {PIR}	3: {CIN}, {COM}, {SCI}
<b>isolierte Objekte</b>	14: {ANA}, {API}, {CAB}, {CBN}, {CEN}, {CJA}, {COM}, {CSN}, {IPA}, {JIS}, {NTI}, {PAP}, {PIR}, {TOX}	2: {CEN}, {CJA}	11: {ANA}, {API}, {CAB}, {CBN}, {CEN}, {CSN}, {IPA}, {JIS}, {NTI}, {PAP}, {PIR}
<b>isolierte Hierarchien</b>	2: CAS-SCI; PRO-CIN	0	0
<b>Äquivalenzklassen</b>	18	16	18
<b>nicht-triviale Äquivalenzklassen</b>	0	2: {CAS,SCI}, {CIN,PRO}	0
<b>Vergleichbarkeiten V(N)</b>	2	41	7
<b>Unvergleichbarkeiten U(N)</b>	302	228	292
<b>D-Matrix der maximalen Objekte (Auswertung)</b>	CAS: 1; PRO: 1	CAS/SCI: 6; CAB: 6; CBN: 3; PAP: 1; TOX: 5	CAS: 1; CJA: 3; PRO: 1; TOX: 2
<b>W-Matrix der maximalen Ob-</b>	UM: 0; UP: 0; KO: 0;	UM: 10; UP: 15; KO:	AN: 1; ID: 0; CT: 0;

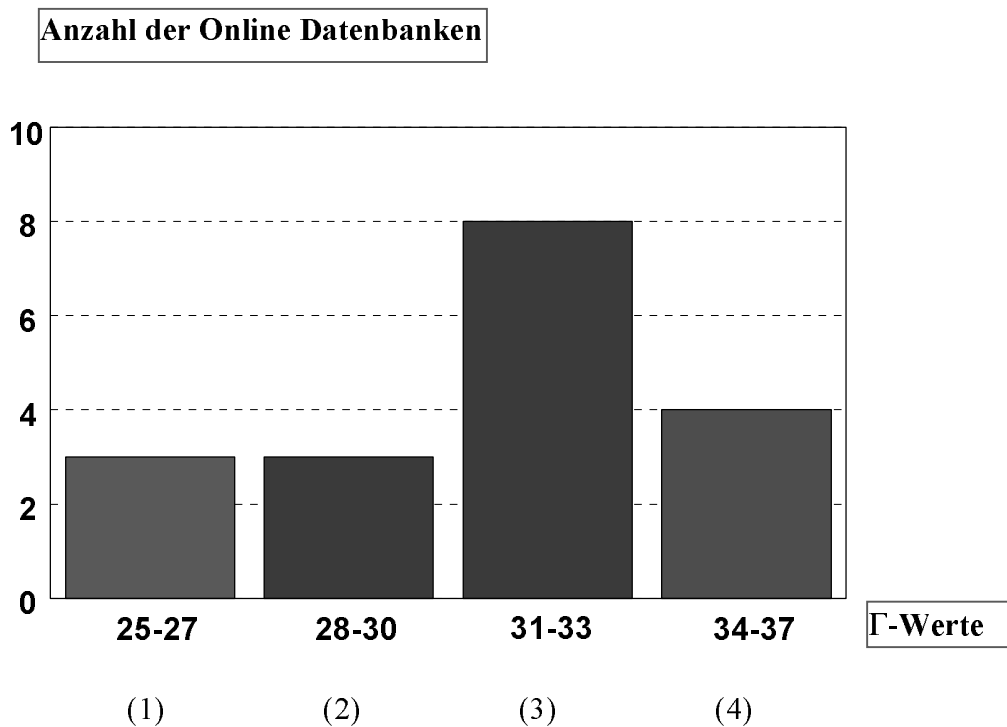
<b>jekte (Auswertung)</b>	VE: 0; AN: 0; ID: 0; CT: 0; CE: 1; IP: 0; PE: 0; AR: 0; QU: 0	16; VE: 4	CE: 7; IP: 0; PE: 9; AR: 3; QU: 0
---------------------------	---	-----------	-----------------------------------

Abkürzungen siehe Abkürzungsverzeichnis

Die Hasse-Diagramme sind in Kapitel 7.2.7 im Hauptteil zu finden.

### **C.5.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 5 (Online Datenbanken)**

Die Werte für das Gesamtmaß  $\Gamma$  liegen zwischen 25 und 37, haben also eine Bandbreite von nur 13 (siehe Tabelle 53). Diese Werte werden in vier Aggregationsklassen eingeteilt, wobei jede Klasse drei bzw. die höchste vier Datenbanken umfaßt. Die Verteilung der Umweltchemikalien-relevanten Datenbanken auf die vier Aggregationsklassen ist in der Abbildung C7 zusammengestellt.



#### **Abbildung C7: Verteilung der $\Gamma$ -Werte auf 18 Umweltchemikalien-relevante Online Datenbanken**

Die Umweltchemikalien-relevanten Datenbanken haben im Durchschnitt höhere  $\Gamma$ -Werte als die vier vorher betrachteten Gruppen. Die höchsten  $\Gamma$ -Werte haben CAB, CAS, CJA und TOX, die auch maximale bzw. isolierte Objekte bei der Betrachtung aller 12 Bewertungskriterien darstellen. Die maximale Datenbank PRO verfügt jedoch nur über ein Gesamtmaß  $\Gamma = 31$ . Die geringsten  $\Gamma$ -Werte (Aggregationsklasse 1) haben API, CIN und IPA, die ebenfalls minimale Objekte oder isolierte Objekte im Hasse-Diagramm darstellen. Das Minimal SCI hingegen verfügt über einen  $\Gamma$ -Wert von 31, ist also in der Aggregationsklasse (3). Daraus läßt sich schließen, daß das "Mittelfeld" der Werte für das Gesamtmaß deutlich besser durch die Hasse-Diagramm-Technik eingeordnet wird. Ebenfalls visualisiert die Hasse-Diagramm-Technik die

Unterschiede, z.B. bei den  $\Gamma$ -äquivalenten Datenbanken CAS und TOX in Form der Balkendarstellung. Die Datenbank CAS ist beispielsweise besser in den Bewertungskriterien UP, AN und ID eingestuft, während TOX der Datenbank CAS in den Kriterien KO, VE und CE überlegen ist. In den Bewertungskriterien CT, IP, PE, AR und QU verfügen beide Datenbanken über dieselben Bewertungsziffern.

### **C.5.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 5**

In der Tabelle wird bei den Umweltchemikalien-relevanten Datenbanken die Verteilung auf die einzelnen Bewertungskriterien untersucht (vgl. hierzu Ausführungen bei der Gruppe 1).

**Tabelle C8: Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 5: Online Datenbanken**

BWK/ BWZ	UM	UP	KO	VE	AN	ID	CT	CE	IP	PE	AR	QU
0	0	0	7	2	4	7	0	4	0	5	0	0
1	1	0	3	5	2	10	0	5	0	2	0	1
2	6	0	5	1	5	0	2	2	0	1	16	11
3	2	0	1	9	5	1	3	3	1	3	2	6
4	7	8	2	0	1	0	5	1	8	4	0	0
5	2	10	0	1	1	0	8	3	9	3	0	0
MW	3,17	4,56	1,33	2,17	2,00	0,72	4,06	2,06	4,44	2,44	2,1	2,28
ME	3,5	5	1	3	2	1	4	1,5	4,5	3	2	2
STDEV	1,200	0,511	1,372	1,339	1,455	0,752	1,056	1,798	0,616	1,947	0,323	0,575
$I_i$	1,38	0,69	1,43	1,27	1,61	0,86	1,26	1,69	0,87	1,69	0,35	0,83
$U_i$	5	2	5	5	6	3	4	6	3	6	2	3

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

Die Verteilung ist bei keinem der allgemeinen Bewertungskriterien als gut anzusehen, hingegen bei den inhaltliche Kriterien AN, CE und PE. Die drei Bewertungskriterien haben einen Umfangsindex  $U_i$  von 6. Der Shannon-Index  $I_i$  ist bei CE und PE am größten. Diese beiden Kriterien, die die Entwicklung der Chemikalien bzw. der Umwelt-Parameter in den Datenbanken widerspiegeln, haben sowohl beim Umfangsindex als auch beim Shannon-Index dieselben Werte. Das Bewertungskriterium AN (Anzahl an Chemikalien) weist zwar eine gute Verteilung auf, jedoch sind die schlechten und mittleren Bewertungsziffern häufiger vergeben als die guten. Die Umweltchemikalien-relevanten Datenbanken verfügen also in der Regel nur über eine geringe bis mittelgroße Anzahl an Chemikalien.

Sehr schlecht sind die Kriterien UP und AR verteilt. Wie schon mehrfach erwähnt, verfügt AR nur über eine Ausprägungsmenge von vier Bewertungsziffern, kann also nur in diesem Zusammenhang gesehen werden. Bei der AR (Art/Anwendung der Chemikalien) liegt die Ausprägungsmenge bei  $\{2,3\}$ , d.h. die Qualität der betrachteten Datenbanken ist mäßig. Das Kriterium UP (Häufigkeit der Aktualisierung) hat eine Ausprägungsmenge von  $\{4,5\}$  gegenüber der vollständigen sechs Ziffern umfassenden Ausprägungsmenge. Das bedeutet, daß diese Gruppe häufig aktualisiert wird. Über eine schlechte Verteilung verfügen auch die Bewertungskriterien ID, IP und QU. Auffällig schlecht ist die Verteilung bei dem Kriterium ID (Identifikationsmerkmale). Hier sind nur die schlechten Bewertungsziffern vergeben. Die Gruppe 5 verfügt nur über eine Datenbank, die Strukturformeln vorhält.

Bei dem Kriterium IP (Informationsparameter für Umweltchemikalien) ist die Verteilung zwar schlecht, jedoch sind die hohen Ziffern oft vergeben. Für dieses Kriterium sind die ausgewählten Datenbanken als günstig zu bezeichnen.

Die Kriterien UM, KO, VE und CT liegen im sog. Mittelfeld in bezug auf ihre Verteilung.

Das allgemeine Bewertungskriterium KO (Kosten der Recherche) hat z.B. die Ausprägungsmenge {0,1,2,3,4}, jedoch wird die Bewertungsziffer 0 bei ca. 50 % der Datenbanken vergeben. Die Kosten der Recherche in Umweltchemikalien-relevanten Datenbanken sind also sehr hoch.

Das Kriterium CT (Chemikalien-Testdatensatz) liegt bezüglich der Verteilung zwar im "Mittelfeld", insbesondere hat CT eine deutlich ausgeprägte Neigung hin zu den hohen Bewertungsziffern. Das bedeutet, daß diese Gruppe über die ausgewählten Umweltchemikalien verfügt, was auch zu erwarten ist.

Zusammengefaßt läßt sich feststellen, daß die Umweltchemikalien-relevanten Datenbanken sich als gut bezüglich ihrer Verteilung in den Bewertungskriterien AN, CE und PE erweisen.

Betrachtet man die Vergabe der hohen im Vergleich zu den niedrigen Bewertungsziffern, so sind bei den inhaltlichen Kriterien CT und IP hervorzuheben.

Schwächen weisen die Datenbanken dieser Gruppe hingegen bei den Kriterien ID, AR und QU, sowohl bezüglich ihrer Verteilung als auch in bezug auf den Vergleich von niedrigen zu hohen Bewertungsziffern auf.

## C.6 Textbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 1)

Die Hasse-Diagramme sind in Kapitel 7.3.1 im Hauptteil zu finden.

**Tabelle C9: Auswertung der Hasse-Diagramme Gruppe 1: CD-ROMs**

	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Ebenen</b>	3	5	6
<b>Elemente in der breitesten Ebene</b>	18	8	6
<b>maximale Objekte</b>	18: {AST}, {BIO}, {CAH}, {CBN}, {CDR}, {DIS}, {ETI}, {EMB}, {GSA}, {KIR}, {NTI}, {OCD}, {OOO}, {OSR}, {OID}, {PON}, {TOX}, {TOP}	8: {AST}, {CAC,CAF,CAH}, {DIS}, {ETI}, {OCD}, {OSR}, {OID}, {TOX}	6: {BIO}, {CAH}, {CDR}, {KIR}, {PON}, {TOP}
<b>maximale Objekte (ohne isolierte Objekte)</b>	6: {CAH}, {CBN}, {EMB}, {GSA}, {OSR}, {TOX}	7: {AST}, {DIS}, {ETI}, {OCD}, {OID}, {OSR}, {TOX}	4: {BIO}, {CAH}, {PON}, {TOP}
<b>minimale Objekte</b>	18: {AST}, {BIO}, {CAC,CAF}, {CEA}, {CDR}, {DIS}, {EIC}, {ECP}, {ETI}, {ECH}, {IAR}, {KIR}, {NTI}, {OCD}, {OOO}, {OID}, {PON}, {TOP}	6: {BIO}, {CAC,CAF,CAH}, {IAR}, {KIR}, {OOO}, {PON}	6: {AST}, {CAC,CAF}, {CDR}, {ECH}, {KIR}, {OID}
<b>minimale Objekte (ohne isolierte Objekte)</b>	6: {CAC}, {CEA}, {ECH}, {ECP}, {EIC}, {IAR}	5: {BIO}, {IAR}, {KIR}, {OOO}, {PON}	4: {AST}, {CAC,CAF}, {ECH}, {OID}
<b>isolierte Objekte</b>	12: {AST}, {BIO}, {CDR}, {DIS}, {ETI}, {KIR}, {NTI}, {OCD}, {OID}, {OOO}, {PON}	{CAC,CAF,CAH}	{CDR}, {KIR}

	{TOP}		
isolierte Hierarchien	1: CAH-CAC	1: OCD-OOO	-
Äquivalenzklassen	25	23	23
nicht-triviale Äquivalenzklassen	{CAC, CAF}	{CAC,CAF,CAH}, {CEA,ECH}	{CAC,CAF}, {DIS,EIC}, {EMB,NTI},
Vergleichbarkeiten V(N)	17	65	93
Unvergleichbarkeiten U(N)	618	528	470

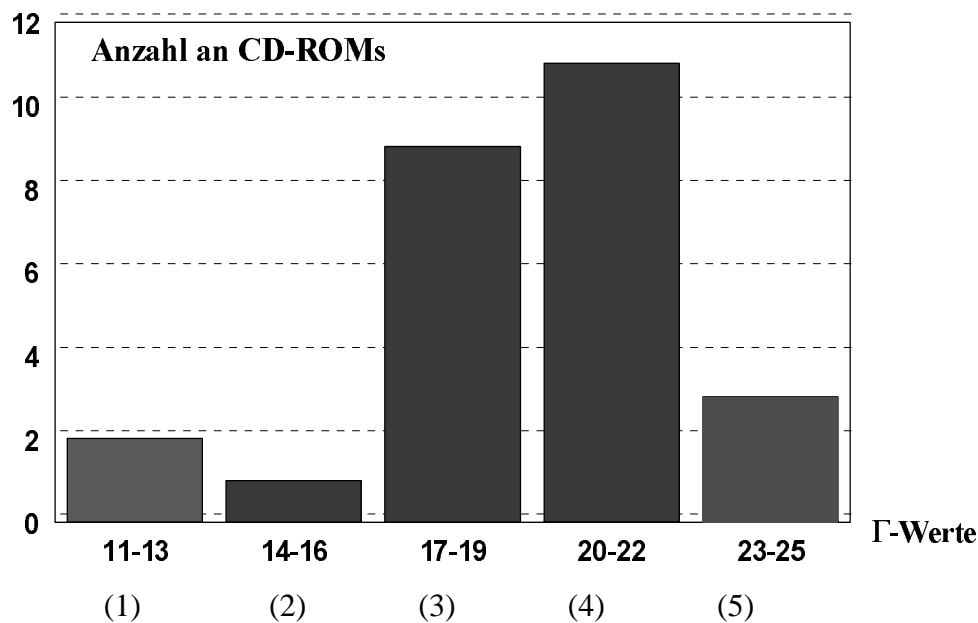
Abkürzungen siehe Abkürzungsverzeichnis

### C.6.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 1 (CD-ROMs)

Die Maximalen und Minimalen sollen an dieser Stelle mit dem Gesamtmaß  $\Gamma$  verglichen werden. Wie aus Tabelle 54 im Hauptteil ersichtlich, ist

$$11 \leq \Gamma \leq 25.$$

Die Verteilung der  $\Gamma$ -Werte auf die Anzahl der CD-ROMs ist in Abbildung C8 dargestellt.



**Abbildung C8:** Verteilung der  $\Gamma$ -Werte auf 26 textbasierte CD-ROMs des Chemie-Spezialgebietes

Die Länge des Intervalls beträgt 15. Es werden fünf gleichartige Aggregationsklassen gebildet. Aus der Graphik läßt sich deutlich erkennen, daß die mittleren Intervalle die größten Anteile an CD-ROMs innehaben. Einige CD-ROMs dieses Bereiches sind  $\Gamma$ -äquivalent. Das trifft beispielsweise für CBN und EIC ( $\Gamma=19$ ) zu. Im Hasse-Diagramm ist jedoch CBN ein Maximal, während EIC ein Minimal darstellt. Jedoch sind beide Objekte unvergleichbar miteinander, wie die Bewertungsziffern in Tabelle 54 im Hauptteil deutlich zeigen. CBN ist höher eingestuft als EIC in den Bewertungskriterien UP, KO, AN und ID; umgekehrt ist EIC in den Kriterien UM und VE der CD-ROM CBN überlegen. Dieser charakteristische Unterschied wird nur durch die Hasse-Diagramm-Technik evident, die Werte des Gesamtmaßes hingegen verstecken diese Tatsache.

In die höchste Aggregationsklasse (5) sind die CD-ROMs CAH, OSR, TOX eingeteilt. Die maximalen Objekte der Hasse-Diagramm-Auswertung sind CAH, CBN, EMB, GSA, OSR und TOX. Es liegt demgemäß eine Übereinstimmung bei den Objekten CAH, OSR und TOX zwi-

schen Maximalen und der höchsten Aggregationsstufe vor. CBN und EMB sind hingegen in der mittleren Aggregationsklasse (3) und GSA in der Aggregationsstufe 4. Auch in diesem Fall ist die Hasse-Diagramm-Technik wieder deutlich aussagekräftiger.

Über niedrige  $\Gamma$ -Werte verfügen die CD-ROMs IAR, ECH und CEA. Diese Objekte sind in den beiden unteren Aggregationsklassen (1) und (2). Minimale sind CAC, CEA, ECH, ECP, EIC und IAR. Auch in diesem Beispiel liefert das Hasse-Diagramm für CAC, ECP und EIC wieder spezifischere Ergebnisse, da diese Objekte in mittleren bis hohen Aggregationsstufen liegen (3 und 4).

### **C.6.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 1 CD-ROMs**

In Anlehnung an die Auswertung der Bewertungskriterien bei Online Datenbanken wird auch bei den CD-ROM Gruppen vorgegangen. Es wird die Verteilung der einzelnen Bewertungskriterien auf die verschiedenen Bewertungsziffern in der folgenden Tabelle dargelegt. Darüber hinaus werden die in Kapitel 6.6 eingeführten Indizes betrachtet.

**Tabelle C10 : Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 1: CD-ROMs, 26 Objekte, 9 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	IP	AR	QU
0	0	0	4	5	5	8	3	0	15
1	3	3	5	10	4	13	2	0	8
2	8	0	10	2	4	1	2	19	2
3	6	13	6	5	9	4	7	2	1
4	8	7	1	0	4	0	7	5	0
5	1	3	0	4	0	0	5	0	0
MW	2,85	3,27	1,08	1,89	2,12	1,04	3,08	2,46	0,58
ME	3	3	2	1	2,5	1	3	2	0
STDEV	1,120	1,080	1,096	1,681	1,395	0,999	1,598	0,811	0,809
$I_i$	1,44	1,20	1,44	1,49	1,55	1,12	1,67	0,74	1,00
$U_i$	5	4	5	5	5	4	6	3	4

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

Der Umfangsindex reicht von 3 bis 6. Nur bei dem Bewertungskriterium IP sind alle sechs möglichen Bewertungsziffern auch mit Daten belegt. Da dieses inhaltliche Kriterium auch den höchsten Wert für den Shannon-Index ausweist, ist es als das am besten verteilte Kriterium anzusehen. Darüber hinaus liegen der Mittelwert und Median dicht beieinander. Als gut verteilte Kriterien sind ferner AN, VE, KO und UM einzustufen. Schlecht verteilt sind QU, ID und UP. Wie schon bei den Online Datenbanken erläutert, sind bei AR (Art/ Anwendung der Chemikalien) nur vier von sechs Bewertungsziffern vergeben, so daß dieses Kriterium nicht mit in den Vergleich der Verteilungen voll mit eingeschlossen werden kann.

Besonders hervorgehoben werden soll die Betrachtung des Bewertungskriteriums ID ( Identifikationsmerkmale von Chemikalien) werden, bei denen die Ausprägungsmenge bei  $\{0,1,2,3\}$  liegt. Das bedeutet, daß chemiebezogene textbasierte CD-ROMs in der Regel nicht über wesentliche Identifikationsmerkmale verfügen. Für die AR (Art/ Anwendung der Chemikalien) ist auch bei den CD-ROMs wie bei den Online Datenbanken nur eine theoretische Ausprägungsmenge von  $\{2,3,4,5\}$  möglich. Bei der vorliegenden Gruppe sind jedoch nur die Bewertungs-

ziffern 2 und 4 vergeben worden. Das bedeutet, daß die textbasierten CD-ROMs des Chemie-Spezialgebietes sich in ihren Inhalten nicht deutlich auf besondere Arten von Chemikalien spezialisiert haben. Diese Aussagen decken sich mit denjenigen, die für textbasierte Online Datenbanken aufgestellt wurden. Bei den CD-ROMs ist jedoch die Situation der Identifikationsmerkmale geringfügig verbessert, wie sich an der häufigeren Vergabe der Bewertungsziffer 3, d.h. u.a. "chemische Strukturformeln vorhanden" zeigt. Das Bewertungskriterium QU (Validität/Güte der CD-ROM) verfügt über die Ausprägungsmenge {0,1,2,3}, d.h. daß nur die schlechten Bewertungsziffern vergeben werden. Das Bewertungskriterium IP (Informationsparameter für Umweltchemikalien) hingegen umfaßt, wie oben erwähnt, die vollständige Ausprägungsmenge. Die hohen Bewertungsziffern werden häufig vergeben. Daraus läßt sich schließen, daß textbasierte CD-ROMs über viele Umweltchemikalien-relevante Parameter verfügen. Dieser Umstand gilt auch für die textbasierte Online Datenbanken.

### C.7 Faktenbasierte CD-ROMs des Chemie-Spezialgebietes (Gruppe 2)

Die Hasse-Diagramme und die ausführliche Darstellung dieser Gruppe sind in Kapitel 7.3.2 im Hauptteil zu finden.

### C.8 Textbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 3)

Die Hasse-Diagramme sind in Kapitel 7.3.4 im Hauptteil zu finden.

**Tabelle C11: Auswertung der Hasse-Diagramme Gruppe 3: CD-ROMs**

	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Ebenen</b>	5	7	10
<b>Elemente in der breitesten Ebene</b>	19	8	6
<b>maximale Objekte</b>	19: {AST}, {BIA}, {BIO}, {CAE}, {CAH}, {CAV}, {EID}, {EMB}, {ENE}, {GSA}, {OEL}, {OOO}, {OSR}, {OID}, {PON}, {POC}, {TOX}, {TOP}, {TRA}	8: {AST}, {CAH,CAV,CAW}, {OEL}, {OOO}, {OSR}, {OID}, {POC,WRA}, {TOX}	5: {BIO}, {CAE}, {CAH}, {CAV}, {PON}
<b>maximale Objekte (ohne isolierte Objekte)</b>	14: {BIO}, {CAE}, {CAH}, {CAV}, {EID}, {EMB}, {ENE}, {GSA}, {OSR}, {PON}, {POC}, {TOX}, {TOP}, {TRA}	s.o.	s.o.
<b>minimale Objekte</b>	17: {AGR}, {ARS}, {ASF}, {AST}, {BIA}, {ESD}, {ECO}, {ECH}, {EIS}, {ENM}, {GEB}, {OEL}, {OOO}, {OID}, {TRE}, {WAS}, {WTL}	5: {ARS}, {ECO}, {EIS}, {SOI,TRE}, {WAS}	1: {ESD}
<b>minimale Objekte (ohne isolierte Objekte)</b>	12: {AGR}, {ARS}, {ASF}, {ESD}, {ECO}, {ECH}, {EIS}, {ENM}, {GEB}, {TRE}, {WAS}, {WTL}	s.o.	s.o.
<b>isolierte Objekte</b>	{AST}, {BIA}, {OEL}, {OID}, {OOO}	-	-
<b>isolierte Hierarchien</b>	-	-	-
<b>Äquivalenzklassen</b>	41	32	29

nicht-triviale Äquivalenzklassen	{GEO,LSI},	{AGR,AGI}, {BIO,CAB}, {CAH,CAV,CAW}, {EMB,LAB}, {ECH,WTL}, {ENE, TOP}, {GEO,LSI}, {POC,WRA}, {SOI, TRE},	{AGR,TRO}, {AGI,GEB}, {ARS,LAB}, {BIA,ECO}, {CAB,ECP,EID,ENE,POC,SOI}, {GSA,TRE}, {GEO,LSI}, {OSR,TRA}, {WRA,WTL}
Vergleichbarkeiten V(N)	85	290	485
Unvergleichbarkeiten U(N)	1554	1164	798

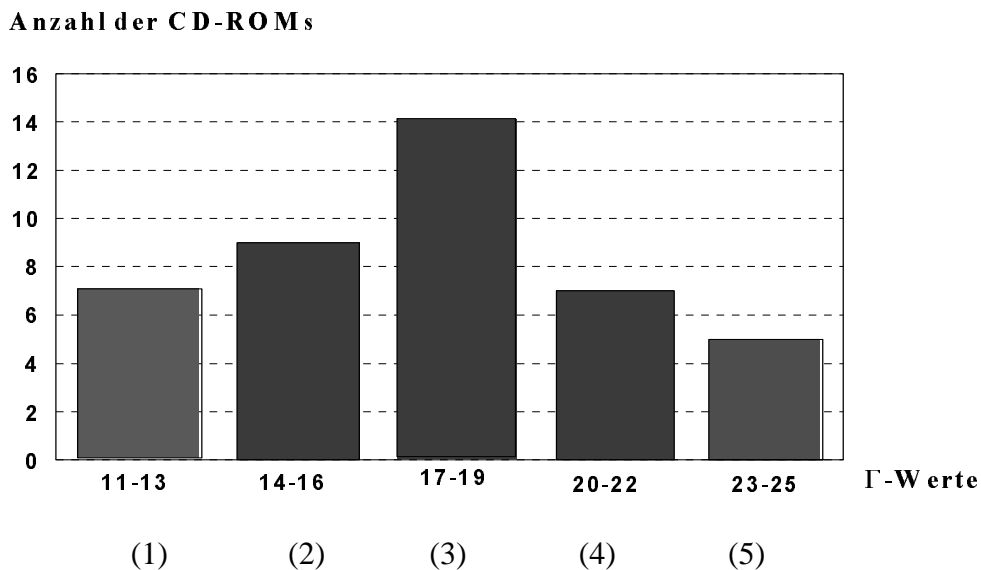
Abkürzungen siehe Abkürzungsverzeichnis

### C.8.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 3 (CD-ROMs)

Analog der bisherigen Vorgehensweise werden zur Erleichterung der Gegenüberstellung der maximalen und minimalen Objekte im Hasse-Diagramm mit den nach dem Gesamtmaß  $\Gamma$  prioritären Objekte in fünf Aggregationsklassen zusammengefaßt. Wie aus Tabelle 58 des Hauptteils hervorgeht, ist

$$11 \leq \Gamma \leq 25.$$

Die Verteilung der  $\Gamma$ -Werte auf die Anzahl der CD-ROMs ist in Abbildung C9 dargestellt.



**Abbildung C9:** Verteilung der  $\Gamma$ -Werte auf 42 textbasierte CD-ROMs des Umwelt-Spezialgebietes

Da auch hier wie bei der Gruppe 1 der CD-ROMs das Intervall 15 beträgt, werden wieder fünf Aggregationsklassen gebildet. Aus der Graphik läßt sich deutlich erkennen, daß eine gute Verteilung der CD-ROMs bezüglich der  $\Gamma$ -Werte vorliegt und die mittleren Aggregationsklassen (2) und (3) die größten Anteile an CD-ROMs innehaben. Auch hier gibt es einige  $\Gamma$ -äquivalente Objekte. Beispielsweise haben ASF, CAB, CAE, ENM und OOO den  $\Gamma$ -Wert = 19, liegen also in einer der beiden mittleren Aggregationsklassen. In der Auswertung mittels der Hasse-Diagramm-Technik sind jedoch ASF und ENM minimale Objekte, während CAE ein maximales Objekt darstellt. Somit ist die Hasse-Diagramm-Technik deutlich aussagekräftiger als das grobe Gesamtmaß.

Vergleicht man die 14 Maximalen (ohne isolierte Objekte) mit den Aggregationsstufen (4) und (5), so ist bei folgenden zehn CD-ROMs eine Übereinstimmung anzutreffen: BIO, CAH, CAV, GSA, OSR, PON, POC, TOX, TOP, TRA.

CAW und ECD haben hohe  $\Gamma$ -Werte, sind jedoch im Hasse-Diagramm nicht als Maximal ausgewiesen sondern in der zweiten Ebene zu finden (vgl. Hasse-Diagramm der Abbildung 38). Auf der anderen Seite sind die CD-ROMs CAE, EID, EMB, ENE, die im Hasse-Diagramm maximale Objekte darstellen, in der Gesamtmaß Betrachtung nur in der mittleren Aggregationsstufe (3) anzutreffen. Auch in diesem Fall liefert die Hasse-Diagramm-Technik wieder fundiertere Bewertungsergebnisse als das Gesamtmaß, da sich beispielsweise die beiden  $\Gamma$ -äquivalenten CD-ROMs EID und EMB deutlich in den Bewertungskriterien UM, KO, VE, AN, ID unterscheiden. Diese Differenzierung kommt nur in der Hasse-Diagramm-Technik zum Ausdruck.

Analog lassen sich die Minimalen den Aggregationsstufen (1) und (2) gegenüberstellen. Hierbei sind nur bei zehn CD-ROMs Übereinstimmungen festzustellen. Es handelt sich um: AGR, ARS, ESD, ECO, ECH, EIS, GEB, TRE, WAS und WTL. Die im Hasse-Diagramm als Minimale ausgewiesenen Objekte ASF und ENM sind bei der Betrachtung des Gesamtmaßes in der mittleren Aggregationsklasse (3) zu finden. Die im  $\Gamma$ -Wert als niedrig eingestuften CD-ROMs sind im Hasse-Diagramm in der zweiten Ebene (AGI, ECD), in der dritten Ebene (LAB und TRO) und SOI ist sogar in der vierten Ebene anzutreffen. OEL ist im Hasse-Diagramm ein isoliertes Objekt.

### **C.8.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 3 CD-ROMs**

Die neun Bewertungskriterien der textbasierten umweltrelevanten CD-ROMs sind mit ihren Ausprägungsmengen in der folgenden Tabelle aufgeführt.

**Tabelle C12: Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 3: CD-ROMs, 42 Objekte, 9 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	IP	AR	QU
0	1	0	7	12	7	27	0	0	32
1	4	2	8	16	10	11	3	0	9
2	13	1	13	2	12	1	7	35	1
3	12	32	12	8	9	3	10	1	0
4	11	6	2	2	4	0	11	6	0
5	1	1	0	2	0	0	11	0	0
MW	2,74	3,07	1,86	1,48	1,83	0,52	3,48	2,31	0,27
ME	3	3	2	1	2	0	4	2	0
STDEV	1,106	0,677	1,160	1,452	1,228	0,862	1,254	0,715	0,497
$I_i$	1,47	0,81	1,48	1,48	1,55	0,91	1,53	0,52	0,63
$U_i$	6	5	5	6	5	4	5	3	3

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

Bei der Gruppe 3 liegen die Ausprägungsmengen der einzelnen Bewertungskriterien zwischen 3 und 6. Die allgemeinen Bewertungskriterien haben alle eine gute Verteilung, d.h. es werden entweder alle sechs oder fünf Bewertungsziffern vergeben.

Für die Kriterien VE und UM treten alle sechs möglichen Bewertungsziffern auf. Sie sind - gemessen an dem Umfangsindex - die am besten verteilten Kriterien. Eine gute Verteilung weist

insbesondere auch das Bewertungskriterium IP auf. Hier ist die Ausprägungsmenge {1,2,3,4,5}, wobei die guten Bewertungsziffern häufig vorkommen.

AN und KO sind ebenso gut verteilt. Auch das Kriterium AN (Anzahl an Chemikalien) ist mit einer Ausprägungsmenge von {0,1,2,3,4} als gut verteilt anzusehen, jedoch sind die schlechten Bewertungsziffern häufiger vergeben als die guten.

Eine schlechte Verteilung haben QU und ID. Das Kriterium ID (Identifikationsmerkmale) hat eine Ausprägungsmenge von {0,1,2,3}. Das bedeutet, daß die umweltrelevanten textbasierten CD-ROMs über wenige bis gar keine Identifikationsmerkmale verfügen. Selbst die CAS-Nummer tritt nicht immer auf. Auch das Bewertungskriterium QU (Validität/Güte der CD-ROM) hat eine Ausprägungsmenge von {0,1,2}. Wenige der betrachteten CD-ROMs enthalten evaluierte Daten.

Auch die AR (Art/ Anwendung der Chemikalien) ist bei der Gruppe 3 sehr schlecht repräsentiert mit einer Ausprägungsmenge von {2,4} gegenüber der vollständigen Menge {2,3,4,5}.

## C.9 Faktenbasierte CD-ROMs des Umwelt-Spezialgebietes (Gruppe 4)

Die Hasse-Diagramme sind in Kapitel 7.3.5 im Hauptteil zu finden.

**Tabelle C13: Auswertung der Hasse-Diagramme Gruppe 4: CD-ROMs**

	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Bewertungskriterien</b>	<b>9</b>	<b>4</b>	<b>5</b>
<b>Anzahl der Ebenen</b>	4	5	5
<b>Elemente in der breitesten Ebene</b>	10	5	6
<b>maximale Objekte</b>	10: {CHS}, {CRI}, {EIN}, {GIF}, {IRI}, {NIO}, {RTE}, {SIG}, {TRI}, {TSC}	4: {GIF}, {IRI}, {RTE}, {TRI}	5: {CHS}, {EIN}, {NIO}, {RTE}, {SIG}
<b>maximale Objekte (ohne isolierte Objekte)</b>	6: {CHS}, {CRI}, {GIF}, {NIO}, {RTE}, {TRI}	3: {GIF}, {RTE}, {TRI}	s.o.
<b>minimale Objekte</b>	10: {AGR}, {ARR}, {CER}, {EIN}, {GEF}, {IRI}, {PES}, {RDI}, {SIG}, {TSC}	6: {ARR}, {GEF}, {IRI}, {RDI}, {SIG}, {TSC}	1: {AGR}
<b>minimale Objekte (ohne isolierte Objekte)</b>	6: {AGR}, {ARR}, {CER}, {GEF}, {PES}, {RDI}	5: {ARR}, {GEF}, {RDI}, {SIG}, {TSC}	s.o.
<b>isolierte Objekte</b>	4: {EIN}, {IRI}, {SIG}, {TSC}	{IRI}	0
<b>isolierte Hierarchien</b>	0	0	0
<b>Äquivalenzklassen</b>	20: {AGR}, {ARR}, {CAN}, {CER}, {CHB}, {CHS}, {CRI}, {EIN}, {GEF}, {GIF}, {IRI}, {MSD}, {NIO}, {PES}, {RDI}, {RTE}, {SIG}, {SOR}, {TRI}, {TSC}	17: {AGR,SOR}, {ARR}, {CAN,NIO}, {CER}, {CHB}, {CHS,MSD}, {CRI}, {EIN}, {GEF}, {GIF}, {IRI}, {PES}, {RDI}, {RTE}, {SIG}, {TRI}, {TSC}	17: {AGR}, {ARR,RDI}, {CAN}, {CER}, {CHB}, {CHS}, {CRI}, {EIN}, {GEF}, {GIF,SOR}, {IRI,TRI}, {MSD}, {NIO}, {PES}, {RTE}, {SIG}, {TSC}
<b>nicht-triviale Äquivalenzklassen</b>	0	3: {AGR,SOR}, {CAN,NIO}, {CHS,MSD}	3: {ARR,RDI}, {GIF,SOR}, {IRI,TRI}
<b>Vergleichbarkeiten V(N)</b>	25	74	73
<b>Unvergleichbarkeiten U(N)</b>	330	238	240
<b>D-Matrix der maximalen Objekte (Auswertung)</b>	CHS: 4; CRI: 1; GIF: 2; NIO: 7; RTE: 3; TRI: 1	GIF: 5; RTE: 15; TRI: 6	CHS: 8; EIN: 2; NIO: 11; RTE: 3; SIG: 4
<b>W-Matrix der maximalen Objekte</b>	UM: 1; UP: 0; KO: 3; VE: 4; AN: 0; ID: 0; IP: 10;	UM: 3; UP: 7; KO: 3; VE: 3	AN: 1; ID: 2; IP: 140; AR: 14; QU: 5

<b>(Auswertung)</b>	AR: 7; QU: 2		
---------------------	--------------	--	--

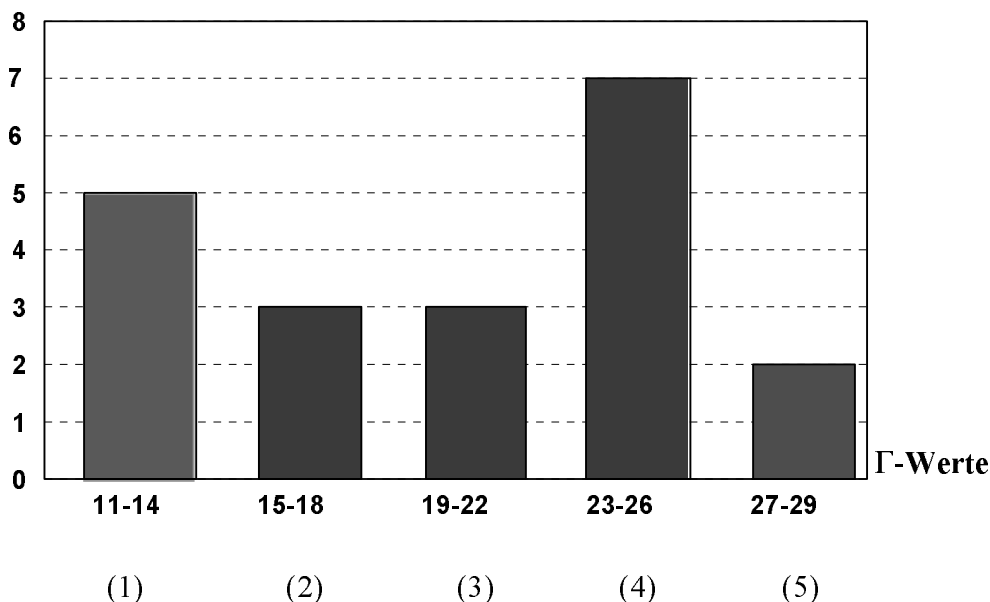
Abkürzungen siehe Abkürzungsverzeichnis

### **C.9.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 4 (CD-ROMs)**

Die  $\Gamma$ -Werte reichen von 11 bis 29 gemäß Tabelle 59 im Hauptteil. Diese Werte werden in fünf Aggregationsklassen zusammengefaßt, wobei die unteren vier Stufen vier, die höchste Stufe nur drei Zahlen enthalten.  $11 \leq \Gamma \leq 29$ .

Die Verteilung der  $\Gamma$ -Werte auf die Anzahl der CD-ROMs ist in Abbildung C10 dargestellt.

**Anzahl der CD-ROMs**



**Abbildung C10: Verteilung der  $\Gamma$ -Werte auf 20 faktenbasierte CD-ROMs des Umwelt-Spezialgebietes**

Aus Tabelle 59 im Hauptteil läßt sich ferner ablesen, daß einige CD-ROMs über dieselben  $\Gamma$ -Werte verfügen. Diese Tatsache trifft beispielsweise für EIN, GIF, MSD und TRS zu, die  $\Gamma$ -Werte von 23 haben. Betrachtet man hingegen die Auswertung mittels der Hasse-Diagramm-Technik, so sind EIN und TSC isolierte Objekte, während GIF ein Maximal darstellt und MSD in der zweiten Ebene liegt. Die Hasse-Diagramm-Technik liefert erneut deutlich differenziertere Ergebnisse als das Gesamtmaß.

Vergleicht man die maximalen ohne isolierte Objekte CHS, CRI, GIF, NIO, RTE, TRI mit der obersten Aggregationsstufe (5) (NIO, RTE), so liegen Übereinstimmungen bei den CD-ROMs NIO und RTE vor. CHS und GIF sind in der Aggregationsstufe (4) anzutreffen, CRI und TRI sind im Hasse-Diagramm maximale Objekte, gemäß der Gesamtmaß Betrachtung jedoch in Aggregationsstufe (2), d.h. einer niedrigen Stufe. Hier ist die Abweichung markant zu nennen, da nahezu entgegengesetzte Ergebnisse erhalten werden. Analog werden die Minimale AGR, ARR, CER, GEF, PES, RDI mit der niedrigsten Aggregationsklasse (1) [AGR, ARR, CER, IRI, RDI] verglichen. Für vier CD-ROMs, d.h. AGR, ARR, CER und RDI liegen Überein-

stimmungen vor. Die Minimale GEF und PES hingegen haben beide  $\Gamma$ -Werte von 20, d.h. sie liegen in der mittleren Aggregationsstufe (3). IRI ist in der niedrigsten Aggregationsstufe (1), im Hasse-Diagramm jedoch ein isoliertes Objekt. Auch bei der Gruppe 4 wird somit wieder die Überlegenheit der Hasse-Diagramm-Technik gegenüber dem Gesamtmaß  $\Gamma$  verdeutlicht.

### **C.9.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 4 CD-ROMs**

Analog zur Vorgehensweise bei den bereits besprochenen Gruppen wird auch hier wieder eine Auswertung der Bewertungskriterien durchgeführt.

**Tabelle C14: Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 4: CD-ROMs, 20 Objekte, 9 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	IP	AR	QU
0	0	1	2	2	5	5	7	0	7
1	1	9	1	11	3	3	4	0	3
2	2	2	4	6	0	10	5	9	6
3	6	8	1	1	4	2	2	1	3
4	4	0	9	0	4	0	2	9	1
5	7	0	3	0	4	0	0	1	0
MW	3,70	1,85	3,15	1,30	2,55	1,45	1,40	3,10	1,40
ME	4	1,5	4	1	3	2	1	3,5	1,5
STDEV	1,218	1,040	1,565	0,733	1,959	0,999	1,353	1,071	1,273
$I_i$	1,43	1,11	1,49	1,07	1,59	1,21	1,50	1,02	1,44
$U_i$	5	4	6	4	5	4	5	4	5

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

Der Umfangsindex reicht von 4 bis 6, d.h. es liegt bei den textbasierten umweltrelevanten CD-ROMs eine gute Verteilung vor, da mindestens vier von sechs Bewertungskriterien vergeben werden. Am besten sind laut Umfangsindex und Shannon-Index die Kriterien KO und AN verteilt. AN hat nur für die Bewertungsziffer 2 keine Eintragung; ansonsten liegt eine gleichmäßige Verteilung vor. Das Kriterium KO (Kosten der CD-ROM) verfügt über hohe Werte bei den guten Bewertungsziffern. Somit sind die umweltrelevanten faktenbasierten CD-ROMs als insgesamt preisgünstig anzusehen. Als gut verteilt sind auch die Bewertungskriterien UM, IP und QU zu betrachten. Auffällig ist jedoch, daß bei IP (Informationsparameter für Umweltchemikalien) die höchste Bewertungsziffer kein einziges Mal vergeben wird. Das bedeutet, daß die als umweltrelevant eingestuften CD-ROMs nicht über alle in dieser Arbeit herausgestellten Parameter verfügen. Eine mäßige Verteilung liegt bei VE, UP und ID vor, wobei bei allen drei Kriterien auffällt, daß die beiden höchsten Bewertungsziffern nicht vergeben sind.

### **C.10 Umweltchemikalien-relevante CD-ROMs (Gruppe 5)**

Die Hasse-Diagramme sind in Kapitel 7.3.7 im Hauptteil zu finden.

**Tabelle C15: Auswertung der Hasse-Diagramme Gruppe 5: CD-ROMs**

	<b>IB<sub>all+inh</sub> Fall 0</b>	<b>IB<sub>all</sub> Fall 0</b>	<b>IB<sub>inh</sub> Fall 0</b>
<b>Anzahl der Bewertungskriterien</b>	<b>9</b>	<b>4</b>	<b>5</b>
<b>Anzahl der Ebenen</b>	3	6	4
<b>Elemente in der breitesten Ebene</b>	16	6	8
<b>maximale Objekte</b>	16: {AST}, {BIO}, {CAH}, {CHS}, {EIN}, {EMB}, {GSA}, {GIF}, {NIO}, {OOO}, {OSR}, {PON}, {RTE}, {SIG}, {TOX}, {TOP}	6: {AST}, {CAH}, {GIF}, {OOO}, {RTE}, {TOX}	8: {BIO}, {CAH}, {CHS}, {EIN}, {NIO}, {RTE}, {SIG}, {TOP}
<b>maximale Objekte (ohne isolierte Objekte)</b>	7: {CHS}, {EMB}, {GIF}, {NIO}, {OSR}, {RTE}, {TOX}	4: {AST}, {GIF}, {RTE}, {TOX}	7: {BIO}, {CAH}, {CHS}, {NIO}, {RTE}, {SIG}, {TOP}
<b>minimale Objekte</b>	16: {AST}, {BIO}, {CAH}, {CHB}, {ECP}, {EIN}, {ECH}, {GEF}, {GSA}, {OOO}, {OID}, {PES}, {PON}, {SIG}, {SOR}, {TOP}	8: {BIO}, {CAH}, {GEF}, {OOO}, {OID}, {PON}, {SIG}, {SOR}	5: {AST}, {ECH}, {EIN}, {OID}, {PES}
<b>minimale Objekte (ohne isolierte Objekte)</b>	7: {CHB}, {ECP}, {ECH}, {GEF}, {OID}, {PES}, {SOR}	6: {BIO}, {GEF}, {OID}, {PON}, {SIG}, {SOR}	4: {AST}, {ECH}, {OID}, {PES}
<b>isolierte Objekte</b>	9: {AST}, {BIO}, {CAH}, {EIN}, {GSA}, {OOO}, {PON}, {SIG}, {TOP}	2: {CAH}, {OOO}	1: {EIN}
<b>isolierte Hierarchien</b>	0	0	0
<b>Äquivalenzklassen</b>	25	23	24
<b>nicht-triviale Äquivalenzklassen</b>	0	2: {CAN,NIO}, {CHS,MSD}	1: {GIF,SOR},
<b>Vergleichbarkeiten V(N)</b>	22	89	64
<b>Unvergleichbarkeiten U(N)</b>	556	426	474
<b>D-Matrix der maximalen Objekte (Auswertung)</b>	CHS: 3; EMB: 1; GIF: 1; NIO: 7; OSR: 1; RTE: 2; TOX: 2	{AST}, {GIF}, {RTE}, {TOX}	BIO: 6; CAH: 6; CHS: 8; NIO: 9; RTE: 3; SIG: 1; TOP: 4
<b>W-Matrix der maximalen Objekte (Auswertung)</b>	UM: 1; UP: 0; KO: 3; VE: 4; AN: 0; ID: 0; IP: 10; AR: 7; QU: 2	UM: 1; UP: 0; KO: 3; VE: 4;	AN: 0; ID: 0; IP: 10; AR: 7; QU: 2

Abkürzungen siehe Abkürzungsverzeichnis

### **C.10.1 Gegenüberstellung Hasse-Diagramm und Gesamtmaß $\Gamma$ : Gruppe 5 (CD-ROMs)**

Die  $\Gamma$ -Werte reichen von 13 bis 29 gemäß Tabelle 60 des Hauptteils, also ist die Bandbreite 17. Diese Werte werden in fünf Aggregationsklassen zusammengefaßt, wobei die oberste und unterste Aggregationsklasse vier Ziffern enthalten, die mittleren Klassen nur drei.

$$13 \leq \Gamma \leq 29.$$

Die Verteilung der  $\Gamma$ -Werte auf die Anzahl der CD-ROMs ist in der folgenden Abbildung dargestellt.

Es wird deutlich, daß die drei mittleren Aggregationsklassen stark repräsentiert sind, während die obere und untere Stufe niedrig sind.

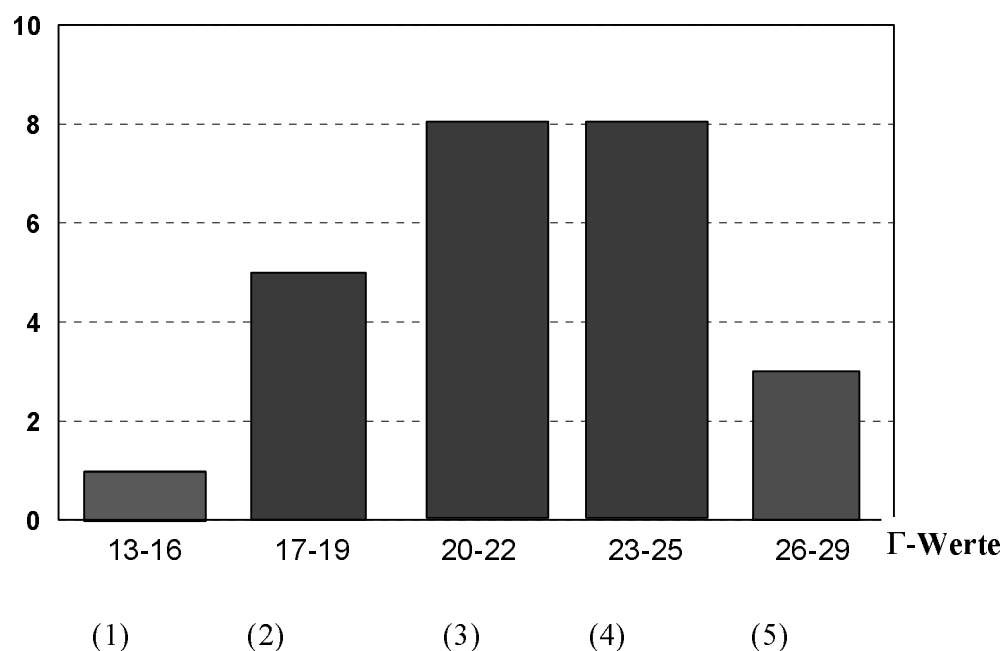
Auch in dieser Gruppe sind wieder einige Objekte  $\Gamma$ -äquivalent. Beispielsweise haben EIN, GIF, MSD, TOX den  $\Gamma$ -Wert von 23. Gemäß der Auswertung dieser Gruppe mittels der

Hasse-Diagramm-Technik ist EIN ein isoliertes Objekt, GIF und TOX maximale Objekte und MSD in der zweiten Ebene des Hasse-Diagramms zu finden. Das unterstreicht erneut die Vorzüge der Hasse-Diagramm-Technik gegenüber der Auswertung des Gesamtmaßes.

Vergleicht man die Maximalen CHS, EMB, GIF, NIO, ORS, RTE, TOX mit den Objekten der obersten Aggregationsstufe (5) (CAN, NIO, RTE), so fällt auf, daß beide Ansätze nur in NIO und RTE übereinstimmen. CAN ist in der höchsten Aggregationsstufe, jedoch im Hasse-Diagramm befindet sich dieses Objekt in der zweiten Ebene. Die Maximalen CHS, GIF, ORS, TOX sind Aggregationsstufe (4) und EMB in Aggregationsklasse (2), also in einer niedrigen Stufe.

Beim Vergleich der Minimalen CHB, ECP, ECH, GEF, OID, PES, SOR mit den beiden niedrigsten Aggregationsstufen (1) und (2), d.h. mit den Objekten AST, EMB, ECH, OOO, OID, SIG fällt auf, daß nur bei ECH und OID eine Übereinstimmung der beiden Ansätze vorliegt. Die Minimalen CHB, ECP, GEF, PES und SOR sind in der Aggregationsklasse (3) oder (4) zu finden. Auch bei der Betrachtung der Gruppe 5 liefert die Hasse-Diagramm-Technik deutlich bessere Ergebnisse als das Gesamtmaß.

### Anzahl der CD-ROMs



**Abbildung C11:** Verteilung der  $\Gamma$ -Werte auf 25 Umweltchemikalien-relevante CD-ROMs

### C.10.2 Bewertungskriterien und deren Ausprägungsmengen: Gruppe 5 CD-ROMs

Bei dieser Gruppe an CD-ROMs hat der Umfangsindex ein Intervall von 4-6. Die Verteilung dieser Gruppe auf die einzelnen Kriterien ist insgesamt als gut anzusehen.

Das Kriterium IP ist am besten gemäß Shannon-Index und Umfangsindex verteilt, gefolgt von KO. Eine gute Verteilung für die Informationsparameter für Umweltchemikalien ist für die Umweltchemikalien-relevanten CD-ROMs auch zu erwarten. Hingegen ist das chemierelevante Kriterium ID (Identifikationsmerkmale) nur mäßig gut verteilt. Insbesondere fällt auf, daß die

hohen Bewertungsziffern nicht belegt sind. Das bedeutet, daß bei dieser für Umweltchemikalien spezifischen Gruppe nur in wenigen CD-ROMs mit der Strukturformel gesucht werden kann. UM und AN sind gut verteilt. Für AR sind alle vier möglichen Bewertungsziffern vergeben.

**Tabelle C16: Ausprägungsmengen der einzelnen Bewertungskriterien der Gruppe 5: CD-ROMs, 25 Objekte, 9 Bewertungskriterien**

BWK/BWZ	UM	UP	KO	VE	AN	ID	IP	AR	QU
0	0	0	3	3	2	3	3	0	6
1	1	3	4	14	4	9	4	0	9
2	5	2	7	3	0	10	7	12	7
3	7	16	3	3	10	3	4	1	2
4	5	3	7	0	5	0	4	11	1
5	7	1	1	2	4	0	3	1	0
MW	3,48	2,88	2,40	1,56	2,96	1,52	2,44	3,04	1,32
ME	3	3	2	1	3	2	2	3	1
STDEV	1,229	0,927	1,472	1,325	1,513	0,872	1,557	1,060	1,069
I <sub>i</sub>	1,49	1,13	1,64	1,29	1,48	1,24	1,74	0,97	1,40
U <sub>i</sub>	5	5	6	5	5	4	6	4	5

Legende und weitere Erläuterungen siehe Tabelle 49 in Kapitel 7.2.1.4.3

**LEBENS LAUF**

Name: Kristina Voigt  
 Geboren am: 03.09.53 in Berlin  
 Wohnort: Claude-Lorrain-Straße 37, 81543 München  
 Nationalität: Deutsch

**Ausbildung:****Schule:**

Renee-Sintenis-Grundschule in Berlin Frohnau	1960	1966
Georg-Herwegh-Gymnasium in Berlin-Hermsdorf	1966	1973
Trimester am Technical College in Kettering, England	04.71	06.71
Abitur	19.01.73	

**Hochschule:**

Technische Universität Berlin	1973	1979
Fachrichtung Lebensmittelchemie		
Staatsexamen Teil A	06.01.78	
Praktikum an der Landesanstalt für Lebensmittel-, Arzneimittel- und Gerichtliche Chemie	01.07.78	30.09.79
Staatsexamen Teil B	05.12.79	

**Bisherige Tätigkeit:**

Teilzeitbeschäftigung bei der Aeroamerica Inc. als Stewardeß	15.08.75	30.09.79
Teilzeitbeschäftigung bei der Air Berlin USA Inc. als Stewardeß	17.03.80	31.03.81
Chemikerin in der Abteilung für Chemie-Dokumentation des Europäischen Patentamtes, Dienststelle Berlin	01.05.80	30.09.80
Wissenschaftliche Angestellte in der Gruppe Umwelt- chemikalien des Umweltbundesamtes in Berlin	01.10.80	30.06.83
Wissenschaftliche Angestellte in der Projektgruppe Um- weltgefährdungspotentiale von Chemikalien der GSF-Forschungszentrum für Umwelt und Gesundheit	01.07.83	
Leitung des Projektes "Informationssystem Umweltchemikalien"	01.01.88	

**Sprachkenntnisse:**

Englisch	9 Jahre Gymnasium, Trimester in England, 3 Semester Englisch für Chemiker an der Technischen Universität Berlin; berufliche Weiterbildung
Französisch	5 Jahre Gymnasium, Weiterbildung durch Kurse, Privatunterricht und Sprachreisen nach Frankreich