DOCUMENT RESUME

ED 411 826 IR 056 653

AUTHOR Voigt, Kristina; Benz, Joachim; Bruggemann, Rainer
TITLE The Internet as an Information Source for Environmental Chemicals--First Results of the Evaluation of the Meta-Database of Internet Resources.
PUB DATE 1996-00-00
NOTE 10p.; In: Online Information 96. Proceedings of the International Online Information Meeting (20th, Olympia 2, London, England, United Kingdom, December 3-5, 1996); see IR 056 631.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; *Databases; *Environmental Research; *Evaluation Criteria; Evaluation Methods; Foreign Countries; Information Sources; *Internet; Online Searching; Online Systems; Resource Allocation; User Needs (Information)
IDENTIFIERS *Environmental Chemistry; University of Kassel (Germany)

ABSTRACT

An evaluation approach using the mathematical method of the Hasse diagram technique is applied on 20 environmental and chemical Internet resources. The data for this evaluation procedure are taken out of a metadatabase called DAIN (Metadatabase of Internet Resources for Environmental Chemicals) which is set up by the GSF Research Centre for Environment and Health and the University of Kassel. The following are chosen as evaluation criteria: search possibilities in Internet resources; quality of resources; number of chemicals in the resource; identification parameters for chemical substances; and information parameters for chemical substances. The 20 Internet resources are ranked with these five different evaluation criteria using a six-figure scoring system. A Hasse diagram is set up and discussed. A further analysis shows that the criterion "information parameters for environmental chemicals" is the most important one in this ranking procedure. (Contains 14 references.) (Author)

The Internet as an information source for environmental chemicals — first results of the evaluation of the meta-database of Internet resources

Kristina Voigt
GSF-Forschungszentrum für Umwelt und Gesundheit, Germany

Joachim Benz
Universität-Gesamthochschule Kassel, Germany

Rainer Brüggemann
Institut für Gewässerökologie und Binnenfischerei im Forschungsverbund Berlin, Germany

Abstract: *An evaluation approach using the mathematical method of the Hasse diagram technique is applied on 20 environmental and chemical Internet resources. The data for this evaluation procedure are taken out of a metadatabase called DAIN (Metadatabase of Internet Resources for Environmental Chemicals) which is set up by the GSF Research Centre for Environment and Health and the University of Kassel. The URL for DAIN is* http://dino.wiz.uni-kassel.de/dain.html. *The following are chosen as evaluation criteria: search possibilities in Internet resources, quality of resources, number of chemicals in the resource, identification parameters for chemical substances, and information parameters for chemical substances. The 20 Internet resources are ranked with these five different evaluation criteria using a six-figure scoring system. A Hasse diagram is set up and discussed. A further analysis shows that the criterion 'information parameters for environmental chemicals' is the most important one in this ranking procedure.*

Keywords: Internet, environmental databases, chemical databases, metadatabases, evaluation, Hasse diagram, ranking

1. Introduction

The Internet provides access to a vast variety of chemistry and environmental information. The quantity and the quality of these information resources continues to improve. However, there is an urgent need to help users of the Internet find the relevant information for a specific subject. A recently published article gives a brief guide to current online resources for environmental professionals (Ref 1). Chemistry resources on the Internet and chemical-specific requirements are given in several articles (Refs 2, 3, 4).

In chemical and environmental sciences several approaches exist to gather the relevant resources. These are lists of Web pages with or without short descriptions of the original site. In Table 1 a few internationally recognised lists of Web sites are given. This is only a small selection of what is available on the Internet. These listed resources are frequently used and recommended by many Webmasters. The table gives not only chemistry and environmental sites but also resources on toxicology and health, as well as sites treating the subject of material safety data sheets.

Table 1: Lists of environmental chemicals' resources on the Internet.

Name of the Resource	URL
Chemistry Information on the Internet	http://hackberry.chem.niu.edu/cheminf.html
Computer Chemistry Centre Erlangen	http://www.organik.uni-erlangen.de/
Directory of Environmental Resources on the Net	http://www.envirosw.com/
Environmental Journalism	http://www.sej.org/
Environmental Sites on the Internet, Stockholm	http://www.lib.kth.se/(lg/eindex.htm
Fachgruppe Informatik im Umweltschutz	http://www.iai.fzk.de/Fachgruppe/GI/
FU CHEMnet WWW Entry Point	http://www.chemie.fu-berlin.de/index.html
Internet Chemistry Resources	http://www.rpi.edu/dept/chem/cheminfo/chemres.html
List of Environmental Toxicology Resources	http://pitt.edu/(martint/pages/envtox.htm
RSC Links to Other Chemistry Web Sites	http://www.worldserver.pipex.com/rsc/wwwsites.htm
Sheffield Chemdex	http://www.shef.ac.uk/(chem/chemdex/
Where to Find MSDS on the Internet	http://www.chem.uky.edu/resources/msds.html
WWW Resources Chemistry	http://www.uky.edu/Subject/chemistry
WWW Resources Environment	http://www.uky.edu/Subject/environment
WWW Virtual Library, Chemistry	http://www.chem.uda.edu/chempointers.html
WWW Virtual Library, Environment	http://ecosys.drdr.virginia.edu/All.html#b

2. DAIN — Metadatabase of Internet Resources for Environmental Chemicals

An approach for a metadatabase called DAIN (Metadatabase of Internet Resources) was established in late 1995. DAIN can be found at the URL *http://dino.wiz.uni-kassel.de/dain.html*.

The set-up of the indexing-file system was presented at Online Information '95 (Ref 5). In this current paper the evaluation of the metadatabase which contains approximately 100 entries will be presented. During the evaluation approach those Internet resources where chemical substances are covered will be looked at. The data-field 'type of database' will be focused on. For the time being the following types of databases are covered: bibliographic, chemical catalogues, full text, inventories, metadatabases, numerical and structural databases. Numerical databases and metadatabases comprise the biggest proportion.

3. Relevant Internet resources for environmental chemicals and evaluation criteria

3.1. Selection of relevant resources

As the current number of resources is still quite easily comprehensible we selected 20 sites, mainly factual data sources. These are given in Table 6.

We searched all these resources again for the evaluation process and found out that only two sites had changed their location but clearly indicated their new address. This means that the chosen set of sites (also called objects in our evaluation approach) are well established data sources and will probably be available under their URLs for a longer period of time in the future.

3.2. Evaluation criteria for environmental chemicals Internet resources

Taking into account the different structure of Internet resources in comparison to online and CD-ROM databases (Refs 6,7) we developed two new evaluation criteria: 'search possibilities' and 'quality of Internet resources'.

3.2.1. Search possibilities

A variety of search possibilities exists for Internet resources. In this field of interest we came across the following: a simple listing of chemicals not even ordered alphabetically, an alphabetically ordered list of chemicals, and a search index for chemicals. Referring to the possibility of searching with the aid of a provided search form, some categories are encountered. The easiest form is one single search form, where the user types in for example the name of the chemical substance. More advanced systems provide several search forms with or without the possibilities of using Boolean operators. The most sophisticated search systems allow all three operators AND/OR/NOT to be used whereas most systems only allow AND/OR facilities.

According to our formerly introduced evaluation system for databases (Refs 8,9) we have worked out a six-figure scoring system where 0 is the worst score and 5 is the best one. Applying this scoring system to the current research activities concerning the evaluation of Internet resources, we set up Table 2.

Table 2: Scores for the criterion 'search possibilities in Internet resources' SE.

Criteria	Score
list	0
alphabetical list	1
index	2
search form, one field	3
search form, several fields	4
search form, several fields, Boolean operators	5

3.2.2. Quality of Internet resources

Data sources can be judged by many quality criteria. We have already published the relevant criteria in connection with the evaluation of online databases (Refs 7–9). However, in this paper we want to concentrate on quality criteria which we encountered while working with Internet resources. One important fact to know is whether the resource contains evaluated data. Expert knowledge is included in very few data sources. In contrast to other data sources, in the Internet world you find a number of awards which are given to some exceptionally good sites: for example the Magellan 3 and 4 StarSite, c/net Best of the Web and MetalWorker's Finalist TOP10 WebSite are mentioned here. The feature 'help' is not found very often within Internet resources. Therefore we also count this 'help function' as a quality criterion. Unfortunately not every site gives a short explanation on what the source is about, so the 'description of the resource' is another quality criterion.

It is quite obvious that more than one criterion may apply for one site. This is why we speak of a nominal criterion with respect to 'quality of the Internet resource'. Detailed explanations of nominal and ordinal criteria are given by Belke (Ref 10).

In order to come to a table of scores like that given in Table 2, four nominal criteria have to be interpreted in an ordinal manner which is now shown.

First the sequence of the criteria is defined:

evaluated data > award > help feature > description

Following the sequence the value 4 is given for evaluated data, 3 for award, 2 for help feature and 1 for description.

Then all possible combinations of these four criteria are calculated. The results are given in Table 3.

Table 3: Combinations for the criterion 'quality of Internet resource' QU.

Combinations	evaluated data	award	help feature	description	combination values
4 combination	x	x	x	x	10
3 combination	x	x	x		9
	x	x		x	8
	x		x	x	7
		x	x	x	6
2 combination	x	x			7
	x		x		6
	x			x	5
		x	x		5
		x		x	4
			x	x	3
1 combination	x				4
		x			3
			x		2
				x	1

Applying this combination method it so happens that for example the combination of three criteria (award, help feature, description) gives the same number as the combination of two criteria (evaluated data, help feature).

In order to come to the six figure method the combination values are divided by two. The scores are given in the Table 4. It has to be noted that this aggregation is not really necessary: it is performed in order to receive a homogenous treatment of criteria and to keep the list of criteria small.

Table 4: Scores for the criterion 'quality of Internet resource' QU.

Criteria	Score
none of the criteria	0
help function	1
description	1
help function, description	2
evaluated data	2
award	2
award, help function, description	3
evaluated data, help function, description	3
evaluated data, description	3
award, help function	3
evaluated data, help function, description	4
evaluated data, award, description	4
evaluated data, award	4
evaluated data, award, help function	5
evaluated data, award, help function, description	5

3.2.3. Other evaluation criteria

As we have been working in the field of evaluating online databases and CD-ROMs for some years, we have developed some criteria which are of great importance concerning the evaluation of environmental and chemical data sources. These are the 'number of chemicals' (NU), 'identification parameters for chemical substances' (ID)

and 'information parameters for environmental chemicals' (IP) which are also relevant for Internet resources.

The 'identification parameters' for environmental chemicals are structural formula, molecular formula, molecular weight , CAS number, other registry numbers, synonyms and so on. For the evaluation of environmental data sources, the criterion 'information parameters for environmental chemicals' is of great importance. It indicates what kind of parameters are covered in the resource in question. All these parameters have already been published (Refs 6,8,9). Whereas the already established scores for the criteria 'identification parameters' and 'information parameters for environmental chemicals' can be applied for Internet resources in the same way as for online databases and CD-ROMs, the 'number of chemicals' has to be adjusted to the topic of Internet resources. The number of chemicals in most of the Internet resources covered by DAIN is considerably lower than in online databases and CD-ROMs. Therefore we had to change the scores given for this criterion. The scores for 'number of chemicals in Internet resources' is given in Table 5.

Table 5: Scores for the criterion 'number of chemicals in Internet resources' NU.

Number of chemicals	Score
0–99	0
100–249	1
250–499	2
500–999	3
1000–9999	4
> 10,000	5

4. Evaluation approach using the Hasse diagram technique

Hasse diagrams have been described in several publications (Refs 11–13) and the repetition of the theoretical background and of other applications is beyond the scope of this article. However, for the sake of convenience to the reader some helpful remarks should be given: The basis of the Hasse diagram technique is the assumption that ranking can be performed while avoiding the use of an ordering index (Ref 13). Hasse diagrams visualise partially ordered sets. Partially ordered sets arise when an additional order relation is established between elements of a set. If for example Internet resources form a set then they are ordered if and only if all the properties of resource 'A' are equal or better than those of resource 'B'. In that case we write:

$A \geq B$.

The aforementioned order relation is an example among many possible realisations. By definition an order relation must be reflexive, asymmetric and transitive. An Internet resource may be compared with itself (reflexivity); if resource 'A' is 'better' than 'B' then the reverse is not true (asymmetry); and if resource 'A' is better than 'B' and 'B' is better than 'C' then it also follows that 'A' is better than 'C' (transitivity).

A graphical representation of such partially ordered sets draws a pair of ordered elements in the plane such that the better element is located above the worse one, and both are connected by a line. Lines which correspond to transitivity are omitted to keep a simple graphic.

Hasse diagrams are extremely useful if several criteria are given to decide which objects are priority objects. In this approach we look at 20 Internet resources (objects) which are evaluated with five different criteria using a six-figure scoring system (see Table 6).

Table 6: 20 selected Internet resources for evaluation, their criteria and scores.

Acc	Name of Internet Resource	SE	QU	NU	ID	IP
ARS	ARS PPD	1	3	2	3	3
ATS	ATSDR ToxFAQs	2	1	0	1	1
BIO	Biocatalysis/Biodegradation	3	4	0	4	2
CHE	Chemikalien Sicherheitsdaten	5	2	4	3	2
ECD	ECDIN	4	2	5	3	5
ENV	ENVIRO-NET	3	0	3	3	4
EPA	EPA CFS	1	0	2	2	4
EXT	EXTOXNET, PIPS	3	1	1	2	4
FIS	Fischer Scientific	4	2	4	3	2
HAC	Hazardous Chemical Db.	5	0	4	3	2
HAD	HazDat	3	3	1	2	1
ICS	ICSC	1	1	0	3	2
OPP	OPPT Chemical Fact Sheets	1	0	0	2	5
OXM	Oxford MSDS	2	0	4	2	4
PES	PESTIS	3	0	2	1	4
SID	SIDS	0	1	0	3	4
SIR	SIRI, MSDS Archive	3	1	4	2	4
STA	Stanford Uni Portfolio	5	0	4	2	1
UTM	Utah MSDS	2	0	4	2	4
WHO	WHO	0	1	0	1	2

4.1. Hasse diagram for 20 Internet resources

The Hasse diagram for this set of 20 objects is given in Figure 1.

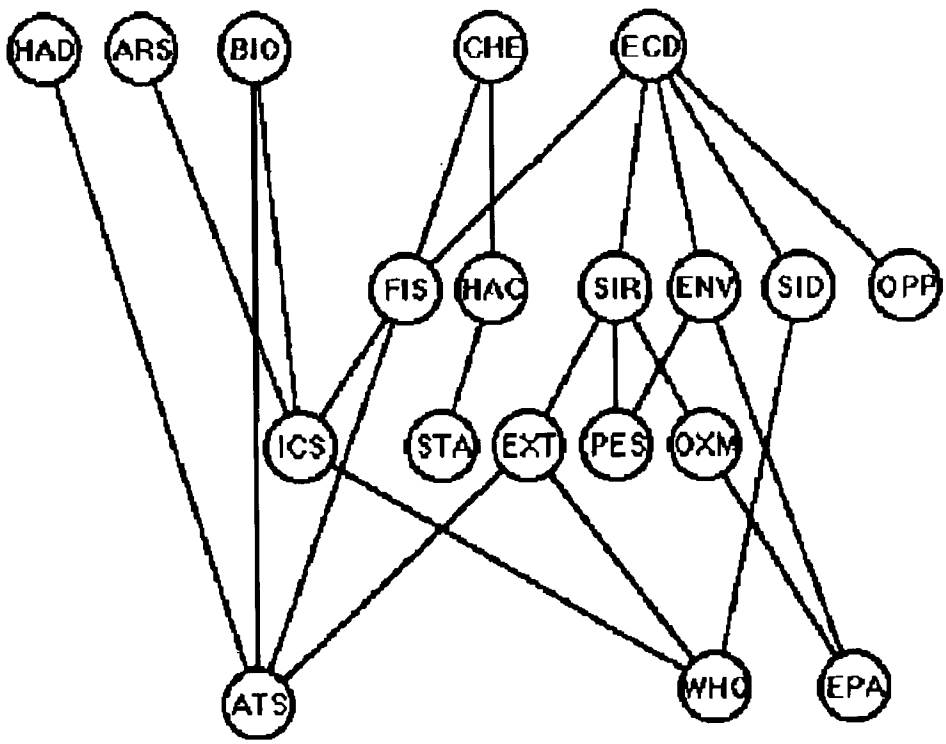

Figure 1: Hasse diagram for 20 Internet resources evaluated with five criteria.

The following conclusions can be drawn from the diagram.

The two resources OXM,UTM have the same scores (see also Table 6): they are members of a so-called non-trivial equivalence class. This means that these two Internet resources which treat the subject of material safety data sheets for chemical substances do not differ from one another if we only regard these five evaluation criteria explained in the above way. Five Internet resources are so-called maximal objects, which means that they are better regarding all five evaluation criteria than those objects which are situated in a lower level. The five objects are: ARS (ARS PPD), BIO (Biocatalysis/Biodegradation), CHE (Chemikalien Sicherheitsdaten), ECD (Environmental Chemicals Data and Information Network) and HAD (HazDat). These Internet resources are the best ones out of the chosen set of 20 sites. However, they are incomparable with each other. Taking, for example, a look at BIO (Biocatalysis/Biodegradation Database) and at ECD (Environmental Chemicals Data and Information Network) this can be demonstrated (see Table 6). BIO has a higher score in 'quality' identification parameters than ECD. The criteria 'search possibilities', 'number of chemicals' and 'information parameters' of BIO are less than those of ECD. Therefore no order relation exists, which means both Internet resources are not comparable. Consequently both resources are not connected by a line.

Minimal objects are those resources which have no successors. In Figure 1 the following six objects are minimals: ATS, EPA, OPP, PES, STA and WHO. These resources are the worst sites in this evaluation approach of a set of Internet resources for environmental chemicals. Taking a look at Table 6 it can be seen that these six minimal resources cannot be compared with each other (see explanation given for maximals).

Figure 1 shows four levels and six elements are found in the largest level.

Successors of a given resource, for example ARS, are those resources which can be reached beginning at ARS and following the lines downwards. This means the successors of ARS are ICS and WHO. With respect to all the given evaluation criteria in Table 6, these sites are worse than ARS.

Taking a look at the successors of the maximal object ECD, it can be seen that ECD has 13 resources which are situated lower in the Hasse diagram. This means that ECD has a considerably higher number of objects which have lower scores than for example CHE, BIO, ARS and HAD. For example HAD is only comparable with ATS.

4.2. Sensitivity analysis of ranking with respect to its evaluation criteria

The ranking of a set of objects does not only depend on the numerical values (scores) but even more on the choice of criteria. In other words the ranking of objects is sensitive to the set of criteria. A matrix called the W-matrix which has been discussed by Brüggemann (Ref 14) quantifies the influence of the evaluation criteria on the ranking. In case 0 all evaluation criteria — that is to say in this example all five evaluation criteria SE, QU, AN, ID and IP — are looked upon. If one criterion out of five criteria is omitted then n+1 cases are possible. For this evaluation six cases result. In case 0 all criteria are evaluated (see Figure 1). In the following five cases one criterion is omitted at a time according to Table 7.

Table 7: Cases for five evaluation criteria.

Case /Criteria	SE	QU	AN	ID	IP
Case 0	x	x	x	x	x
Case 1	-	x	x	x	x
Case 2	x	-	x	x	x
Case 3	x	x	-	x	x
Case 4	x	x	x	-	x
Case 5	x	x	x	x	-

The Hasse software which was developed by the third author (Ref 14) is able to calculate the W-matrix for all objects. This calculation for the five criteria gives the following results:

- Case 1: (SE omitted) 18 changes in the Hasse diagram
- Case 2: (QU omitted) 29 changes in the Hasse diagram
- Case 3: (AN omitted) 9 changes in the Hasse diagram
- Case 4: (ID omitted) 6 changes in the Hasse diagram
- Case 5: (IP omitted) 46 changes in the Hasse diagram

The more changes that are induced by omission of a criterion, the more important the criterion in question is.

This means that in this evaluation approach 'information parameters for environmental chemicals' is the most important criterion followed by 'quality of Internet resources' and 'search possibilities in Internet resources'. The two criteria which are set-up especially for Internet resources have a big influence on the ranking of the 20 chosen Internet sites.

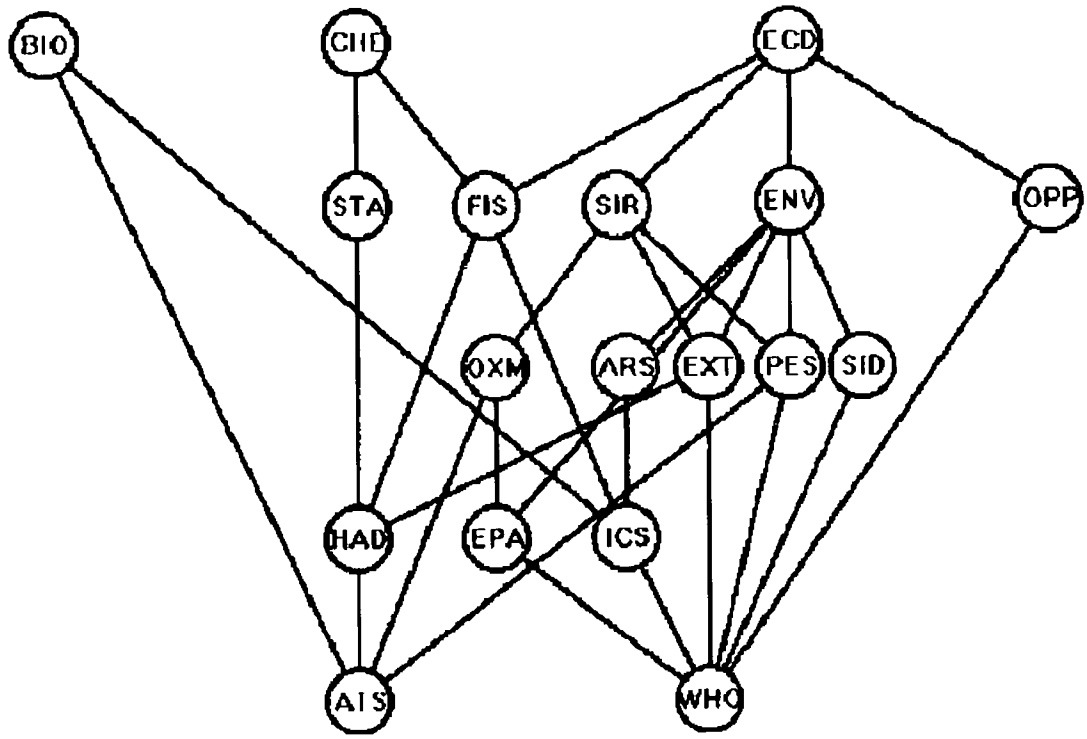

Figure 2: Hasse diagram for Case 2 (Omission of QU): 20 Internet resources.

Comparing this diagram with only four evaluation criteria with the diagram in Figure 1, a few changes can easily be seen. It is obvious that only three instead of five maximal objects are existing now. HAD and ARS are respectively in the fourth and third levels of the diagram. This Case 2 diagram has five levels as opposed to four in the Case 0 diagram. In this diagram only two resources, ATS and WHO, are minimals. CHE/HAC and OXM/UTM are members of non-trivial equivalence classes.

5. Future research

Apart from the content of the Internet resource the update plays an important if not vital role. In this respect two aspects of the update of a resource have to be examined. First it is important to check if the resource can still be found under the same URL and second one has to figure out if and how the resource has been updated. Automated systems to fulfil this urgent need to inspect the update of an Internet resource are under development.

Additionally a great effort will be made to extend the DAIN Metadatabase of Internet Resources for Environmental Chemicals continuously. The authors would be very grateful for proposals and hints in this regard.

Furthermore the Hasse diagram technique will be extended by additional features coming from:

- graph theory (for example identification of articulation points as points of interest for further data analysis);
- multivariate statistical methods, for example integration of cluster analysis;
- establishing more instruments to define the similarities of Hasse diagrams.

Kristina Voigt
GSF-Forschungszentrum für Umwelt und Gesundheit
Projektgruppe Umweltgefährdungspotentiale von Chemikalien
Ingolstaedter Landstrasse 1
85758 Oberschleissheim
Germany
Tel: +49 89/3187 4029
Fax: +49 89/3187 3369
E-mail: voigt@wiz.uni-kassel.de

References

[1] Reichardt, T. (1996) Environment On-Line — a Guide to Internet Resources, *Environ. Sci. Technol.*, **30**(2), 76–81.

[2] Ihlenfeldt, W.-D. and J. Gasteiger (1995) Chemie auf dem Internet, Neue Wege der Vermittlung chemischer Information, *Chemie in unserer Zeit*, **29**(5), 249–259.

[3] Kazic, T. (1995) Chemical information: how do we get it and what do we do with it? in H. Collier (Ed.), *Proceedings of the 1995 International Chemical Information Conference*, Nimes, France, 23–25 October 1995, Infonortics Ltd, Calne, pp. 48–61.

[4] Rzepa, H.S., B.J. Whitaker and M.J. Winter (1995) Chemistry and the WWW, *Chemistry in Britain*, 685.

[5] Benz J. and K. Voigt (1995) Indexing file system for the set-up of metadatabanks in environmental sciences on the Internet, in D. Raitt and B. Jeapes (Eds.) *Proceedings of the 19th International Online Information Meeting*, 5-7 December 1995, London, Learned Information Ltd., Oxford, pp. 455–466.

[6] Voigt, K. and R. Brüggemann (1994) Survey of environmental databases, comparison of search results, in D. Raitt and B. Jeapes (Eds.) *Proceedings 18th International Online Information Meeting*, 6–8 December 1994, London, Learned Information Ltd, Oxford, pp. 159–172.

[7] Voigt, K. and R. Brüggemann (1994), Evaluierung von Umwelt-relevanten CD-ROMs, in L.M. Hilty, A. Jaeschke, B. Page and A. Schwabl (Eds.) *Informatik für den Umweltschutz, 8*, Symposium Hamburg 1994, Metropolis-Verlag, Marburg, pp. 187–194.

[8] Voigt, K. and R. Brüggemann (1994), Ansatz eines Scoring-Systems für chemische/umweltrelevante Online Datenbanken, in C. Jochum *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-Jugenheim/Darmstadt, 17–19 November 1993, Gesellschaft Deutscher Chemiker (GDCh) , Frankfurt/Main, pp. 1–15.

[9] Voigt, K. and R. Brüggemann (1994) Scoring approach for environmental databases, in *Proceedings of the National Online Meeting*, New York, 10-12 May 1994, Learned Information Inc., Medford, NJ, pp. 427–439.

[10] Belke, W., D. Graichen and M. Starruss (1979) *Nichtmetrische Klassifizierung von Informationen, Theorie und Anwendung*, Akademie-Verlag, Berlin.

[11] Brüggemann, R., K. Voigt and E. Halfon (1996) Partially ordered sets — a computerized tool to compare environmental databases, in R. Denzer, G. Schimak and D. Russell (Eds.) *Environmental Software Systems, Proceedings of the International Symposium on Environmental Software Systems 1995*, Chapman & Hall, London, pp. 185–195.

[12] Davey, B.A. and H.A. Priestley (1990) *Introduction to Lattice Theory and Order*, Cambridge University Press, Cambridge.

[13] Halfon, E. and M.G. Reggiani (1986) On ranking chemicals for environmental hazard, *Environ.ScienceTech.*, **23**, 1173–11679.

[14] Brüggemann, R. and E. Halfon (1995) Theoretical base of the program 'Hasse', GSF-Bericht 20/95, GSF-Forschungszentrum für Umwelt und Gesundheit, Neuherberg.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").