

# GenThes: A General Thesaurus Browser for Web-based Catalogue Systems

- A Step towards Component based Catalogue Systems -

Ralf Nikolai, Ralf Kramer, Marc Steinhaus

[Forschungszentrum Informatik \(FZI\)](#)

Haid-und-Neu-Strasse 10 - 14

D-76131 Karlsruhe, Germany

phone: +49 721 9654-744

fax: +49 721 9654-709

[{nikolai,kramer,steinhaus}@fzi.de](mailto:{nikolai,kramer,steinhaus}@fzi.de)

Paolo Plini, Bruno Felluga

[CNR - Consiglio Nazionale delle Ricerche](#)

Istituto Tecnologie Biomediche

Reparto Ricerca e Documentazione Ambientale

Via G.B. Morgagni 30/E

I-00161 Rome, Italy

phone: +39 06 4403728

fax: +39 06 44230229

[{felluga,plini}@itbm.rm.cnr.it](mailto:{felluga,plini}@itbm.rm.cnr.it)

## ABSTRACT

Thesauri have been proven means to identify documents in libraries for centuries. In this paper, we show how this approach can be combined with most recent Internet technologies. The Java-based general thesaurus browser GenThes is able to handle several heterogeneous, multilingual thesauri. With the General European Multilingual Environmental Thesaurus (GEMET), GenThes is currently being used with several environmental catalogue systems. Feedback from users reveal that this approach greatly facilitates search and retrieval as compared to free-text only search. Performance which is crucial in Web applications as well has been improved by reducing the transfer volume of data and code. The software architecture of GenThes supports easy configuration and adaption to the individual needs of different systems that it is connected to. Ongoing work as well addresses to use GenThes as a query expansion module for multilingual search in distributed document collections.

**Keywords:** thesaurus browser, multi-thesaurus systems, heterogeneous resource indexing and retrieval, environmental thesauri, environmental catalogue systems

## 1.0 INTRODUCTION

As a result of the distribution of interrelated information over several different information systems, the interconnection of information systems has increased in recent years. However, a purely technical interconnection is insufficient for the users who need to find their way to the information they are looking for. In the ever growing amount of distributed data and other information that has been collected for different purposes, metadata which is often managed in so-called meta information or catalogue systems (such as the Catalogue of Data Sources (CDS) [1] of the European Environment Agency (EEA) [2]) is the most important means to facilitate search and retrieval of distributed, heterogeneous data.

Thesauri have been proven means to identify documents, e.g., books of interest in a library. They are used in catalogue systems as well, e.g., for indexing and retrieval of information bearing objects (IBOs) [3], [4], [5], [6]. The purpose of thesauri in information retrieval is to provide a uniform and consistent vocabulary for indexing (meta-) data objects in information or catalogue systems and to supply users with a suitable vocabulary for the retrieval of objects in such systems. Additionally, multilingual thesauri such as the General European Multilingual Environment Thesaurus (GEMET) [7] are an important tool to support cross-lingual information retrieval [8], [9]. Users can formulate queries in their preferred languages and retrieve all relevant (meta-) data objects in whatever language the objects are stored.

Most recent advances in open Internet technologies (e.g., World-Wide Web, Java) offer the potential of completely new technical solutions for employing thesauri in information systems. Furthermore, thesauri are currently used in many catalogue systems (e.g., CDS of the EEA [10], [11], Norwegian CDS, Austrian/German UDK (Umweltdatenkatalog, environmental data catalogue) [12], [13]). Thus, the idea came up to develop a widely usable user-friendly thesaurus browser that could be used with different thesauri and different Web-based catalogue systems: GenThes.

The remainder of this paper is organized as follows: In the following [Section 2](#) the approach of GenThes is compared with other approaches. Some details about GEMET, the General European Multilingual Environment Thesaurus, which is the most important thesaurus GenThes works together with, are given in [Section 3](#). In [Section 4](#) the requirements for a general thesaurus browser are described. Furthermore, the multi-layer, component-based architecture of GenThes is presented. In [Section 5](#) the current uses of GenThes in three different catalogue systems, a multilingual search service and a facility for getting statistic information about the use of the thesauri are presented. Our experience when using GenThes as a thesaurus browser for indexing and retrieval with these systems over the past two years are described in [Section 6](#). The paper concludes with a brief summary and an outlook on future works in [Section 7](#).

## 2.0 Related Work

In this section, the approach of GenThes is compared with other approaches and related systems. We focus on those approaches and systems that we think are most closely related to our work.

First, we want to have a closer look at catalogue systems. Catalogue systems are used in many areas, such as e-commerce, environmental information systems, systems for earth observation, electronic yellow pages, and many more. Concerning the "vocabulary dimension" they can be organized according to the level they use a controlled and structured vocabulary for search and retrieval. Controlled vocabulary means that only a distinct set of words may be used. Structured vocabulary means that there are certain relationships between the words.

- Neither controlled nor structured. Basically, this category represents full text search only; the users can search with any search string they are interested in.
- Controlled, but unstructured vocabulary. There is a controlled vocabulary that serves as an index for search and retrieval (e.g., branches of the economy in electronic yellow pages, such as the activity sectors in <http://www.europages.com/>, kinds of products in e-shops and e-malls, etc.); note that the alphabetic order does not represent any semantic ordering.
- Not controlled, but structured. In this category, there is not a single entity that controls the vocabulary used for indexing on the one hand and for search and retrieval on the other hand. Obviously, this means that users index documents on their own and then in turn use this vocabulary and the structure for search and retrieval. The so-called "secondary catalogues" of former versions (namely versions 1 and 2) of the German/Austrian environmental data catalogue (Umweltdatenkatalog, UDK) are an example [16].
- Controlled and structured. There are different ways in which a semantic structure can be imposed on the vocabulary. A predecessor - successor relationship as, e.g., in the list of valid terms ("Valids") of the Global Change Master Directory (GCMD) is a fairly obvious one [17]. In other systems for earth observation, domain specific thesauri, such as the one from DLR [19], are being used.

To the best of our knowledge, there is no approach in either of these categories in which a general software approach is adopted, that allows to flexibly combine an access module for thesauri with catalogue systems for both search and retrieval and updates.

Furthermore, support for multilinguality in such systems is fairly restricted. If multiple languages are supported, then usually two languages are supplied with English being the 2<sup>nd</sup> language are handled. The GEMET is unique as it covers a domain specific vocabulary fairly broadly in a whole set of different (currently European, but this is not a general restriction) languages. The different lists of parameters in the Global Change Master Directory [18] are restricted to English.

## 3.0 GEMET

GEMET, the General European Multilingual Environment Thesaurus, has been developed for the EEA, as a reference indexing and retrieval tool for the EEA Catalogue of Data Sources and other EEA databases [7].

GEMET is a vocabulary of more than 6.500 controlled terms (keywords), selected as environmentally significant and

not too specific. The terms are arranged in a matrix classification scheme. One axis of the matrix is represented by a facet-oriented systematic and hierarchical arrangement of the terms (with so called super-groups and groups on the top level), according to a broad disciplinary perspective and following the ISO standards. The perpendicular axis presents a complementary arrangement, obtained by clustering the terms in a system of 40 environmentally relevant themes.

The 2.0 version of GEMET will be edited in British and American English and will present the equivalents in 10 European languages. In order to ensure an exact equivalence between the various languages, almost all terms will be endowed with a not too specific definition in English.

To each GEMET term (both descriptor and non-descriptor), a neutral numerical identifier is assigned and never modified. The ID of a term must never be changed, no matter how the latter is presented as a default inside each language and in the different languages. This is an important point since the informatic handling of thesauri lead to the need of a stable reference element to be used as a link for the navigation of information within and outside the thesaurus. The identifier is the stable element for the circulation of information between different applications that use GEMET.

In the environmental information systems which use GEMET as the reference thesaurus, additional terminology maintained in the form of specific lists or thesauri can be used. GEMET and this additional terminology are linked through a common element, the GEMET List of Lists (GLOL). The GLOL is an intelligent repository of the titles of the lists of the information system, accessible through suitably labeled GEMET descriptors and identifiers. In this perspective, it is mandatory that, while indexing, the terminology of the lists is used for a secondary indexing, after the primary indexing performed with the compulsory GEMET terminology.

## **4.0 Requirements and Architecture**

### **4.1 Requirements**

The most important requirements of a general thesaurus browser are the following:

- to handle mono- and multilingual polyhierarchic thesauri according to ISO norms [14], [15], having some additional features stored in different database schemas;
- to handle several thesauri simultaneously;
- to provide three different runtime modes: (1) stand-alone modus for inspecting thesauri without relations to a catalogue system, (2) indexing modus for registering metadata objects in a catalogue system, (3) retrieval modus for searching for metadata objects in catalogue systems;
- to provide different entry points to the thesaurus vocabulary: alphabetic index; hierarchy browser; descriptive search;
- to support navigation within the thesaurus structure;
- to provide statistical data to the thesaurus maintainers about the use of the thesaurus;
- to customize the user interface to the different looks and feels of the catalogue systems;
- to provide a multilingual user interface; and
- to match the specific requirements of Internet applications, e.g., low network bandwidth, heterogeneous user groups.

### **4.2 Architecture**

To fulfil these requirements a three-layer architecture was developed (see [Figure 1](#)). This architecture comprises a database access layer, a thesaurus layer, and a Graphical User Interface (GUI) layer.

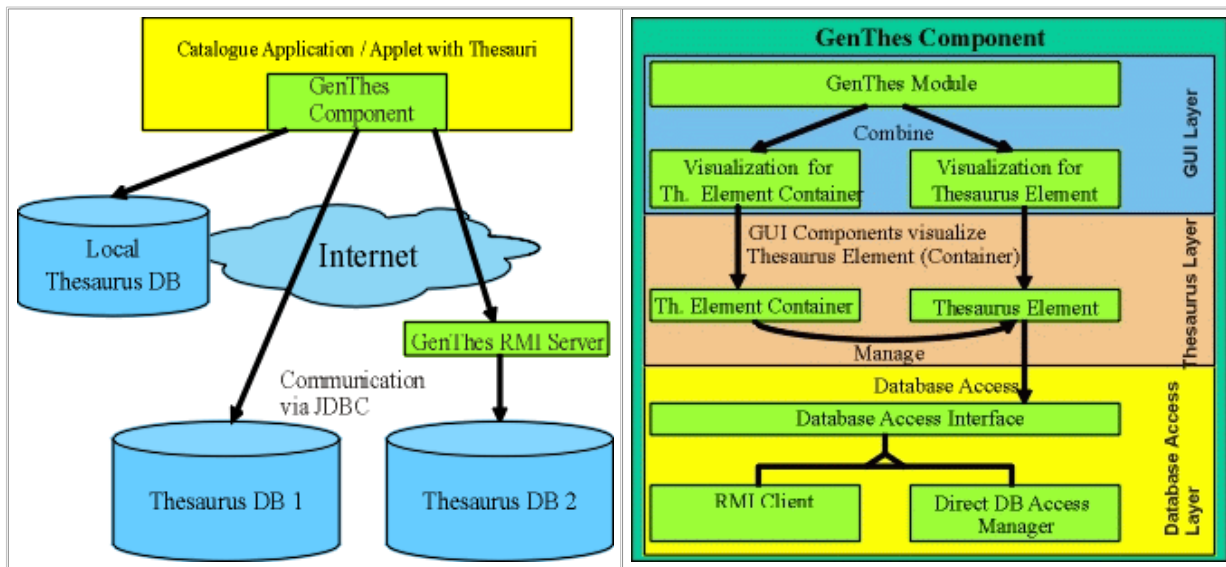


Figure 1: GenThes as embedded component of a catalogue system (left) and the internal GenThes architecture.

The database access layer is the bottom layer. It contains components to control the access to the different databases via Java Database Connectivity (JDBC), which manage the thesauri data. The database access layer allows to access different, distributed databases and – if the catalogue is started as a local application – a local thesaurus database.

The database access layer provides a homogeneous interface called database manager for higher layers. In case the database schema changes or a new schema is added, only the database layer has to be adapted. Therefore, this layer provides wrapper functionality.

The thesaurus layer is the middle layer: it is responsible to model the elements of a thesaurus by mapping the relational view to an application specific object view. Furthermore, methods to manage and sort sets (stored in so-called thesaurus element containers) of these "business objects" are provided. This layer also provides the model for the navigation along the thesaurus structure. It communicates with the database access layer via Java Remote Method Invocation (RMI) and is responsible for caching thesaurus data on client site,

The graphical user interface of GenThes is built by using a collection of GUI components that display objects delivered by the thesaurus layer. These components can be easily customized to the specific needs.

Additionally, the overall application comprises several components, such as the hierarchy browser and the statistical component. This allows further flexibility when configuring GenThes for a specific catalogue system.

## 5.0 Employment of GenThes

### 5.1 Catalogue Systems

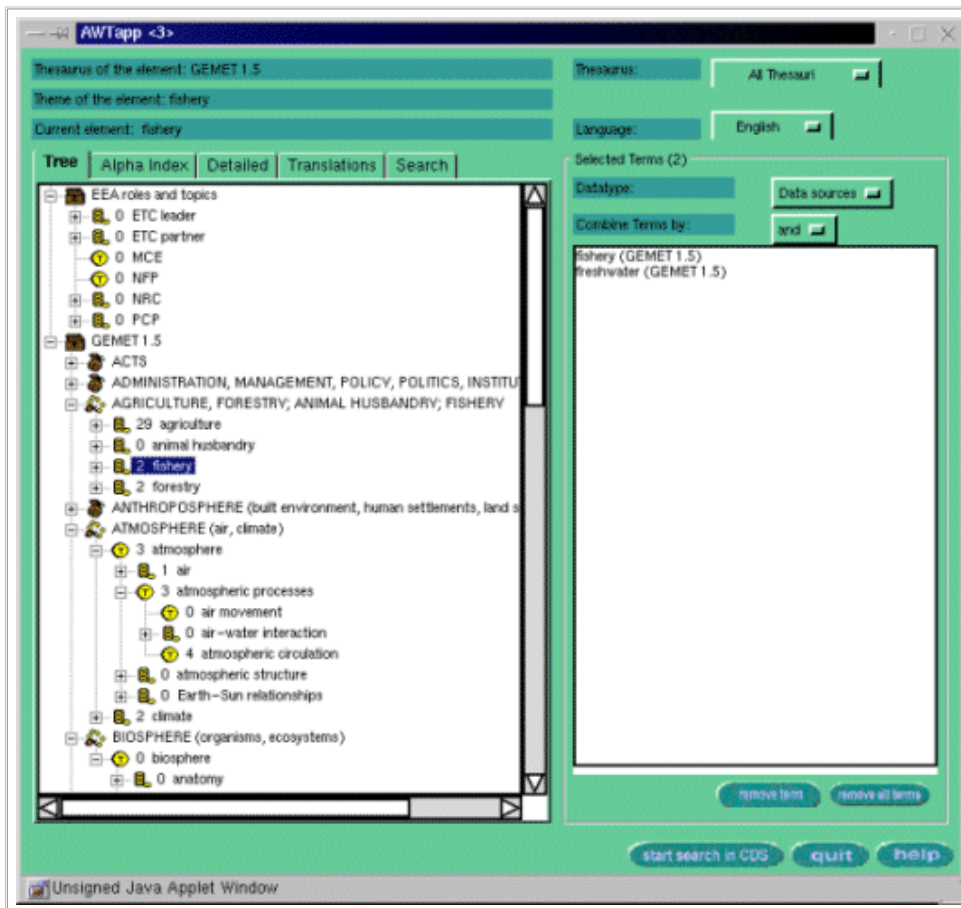


Figure 2: GenThes configured for the Catalogue of Data Sources (CDS). A mini-thesaurus (EEA roles and topics) with hierarchic relations only and GEMET [7] are displayed in the hierarchy browser. They are marked with a treasure case honoring the original meaning of a thesaurus: treasure. In the case of GEMET groups (marked with a moneybag) are displayed as top-level elements. Terms are marked with a coin (no narrower terms) and a stack of coins (narrower terms exist). The number in front of the terms indicates how many objects of the selected data type are indexed with this term. The CDS comprises two different data types: Data sources (meta-data objects) and Addresses.

GenThes is currently used in conjunction with three different catalogue systems and three different thesauri: (1) as indexing tool for the Norwegian Catalogue of Data Sources with the Norwegian Environment Thesaurus available in Norwegian and English, (2) as retrieval tool for the Catalogue of Data Sources (CDS) of the European Environment Agency (EEA) [11] (see Figure 2) with the GEMET available in 12 languages (see Section 3), and (3) as retrieval tool for the Web-based version of the Umweltdatenkatalog (WWW-UDK), with the environment thesaurus of the German Environment Protection Agency (WWW-UDK 3.1 in Austria [20] and WWW-UDK 4.0 currently under development [21]). For an in-depth discussion of the 4<sup>th</sup> Generation of the overall Environmental Data Catalogue (UDK) project for Austrian and German Public Authorities we refer to [22]).

Customizing GenThes was necessary concerning:

- **Functionality:** whereas the main functionality is the same, each of the catalogue systems requires some specific functionality. E.g., for the Norwegian CDS, a mini window allows to enter and check thesaurus descriptors without opening the overall thesaurus browser; the UDK does not need the possibility to combine several thesaurus terms for a query because this is done by the query processing engine. In GenThes, a strict modular approach allows to combine functional components.
- **Database schema:** the Norwegian thesaurus adopts the datamodel of the CDS of the EEA, while the database schema of UDK 3 is different. In GenThes, these differences are hidden by the database layer.
- **User interface:** to support the fairly different ideas about the looks and feels, in GenThes customizable GUI components are provided to build composite GUIs.

## 5.2 Multilingual Search Service for Distributed Environmental Documents

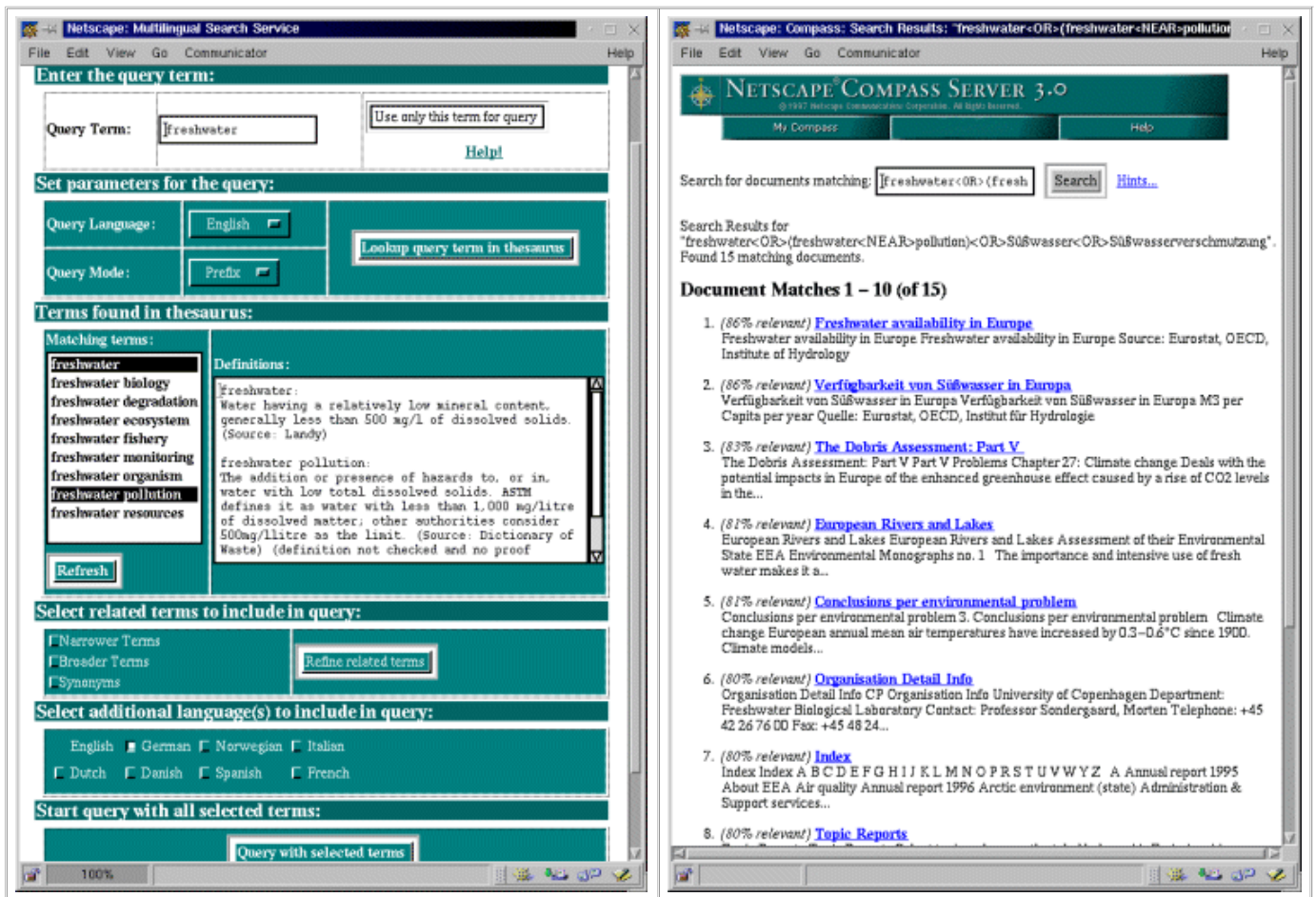


Figure 3: Multilingual Search Service based on Database and Thesaurus Layer of GenThes. Search Results are displayed by a commercial Search Engine.

Furthermore the two bottom layers of GenThes are used for the Multilingual Search Service of the EEA. This service provides cross-lingual access to documents available in the extranet of the EEA by the use of the multilingual list of GEMET [7] and is available as a prototype (see Figure 3). Compared to GenThes it provides a totally different GUI (HTML instead of Java) but the effort to access the thesaurus database is significantly reduced by the use of the thesaurus and of the database layer.

The basic approach to provide the Multilingual Search Service for distributed environmental documents is to combine an off-the-shelf Web Search Engine (comprising – among other things – a Web crawler to index the documents and a Retrieval Engine), in our case Netscape Compass Server, with a concept-aware query expansion component based on the multilingual thesaurus GEMET. Therefore, our approach realizes cross-lingual information retrieval by translating the query terms. In an environment with many different user and document languages (such as the European Union) and a huge amount of documents, this is a more appropriate approach than pre-translating the documents into all different languages. Due to the fact that all documents are covering the theme "environment", the query translation with the domain specific vocabulary of GEMET works well for a high percentage of user queries. The semantic relations between GEMET terms (hierarchy, synonyms, related terms) can be used for further query expansions.

### 5.3 Providing Information for Thesaurus Maintenance

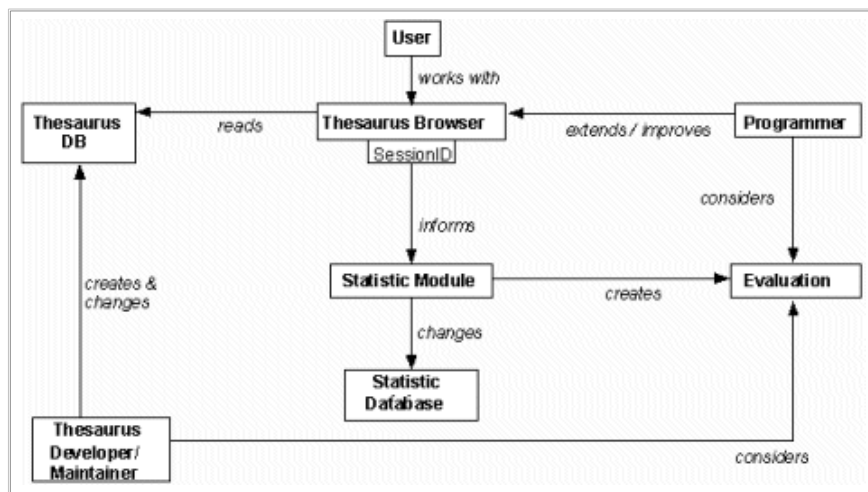


Figure 4: Relationships between the users, the different groups of developers, the GenThes software components and the role of the statistic module.

To optimize the contents of a thesaurus, it is of general importance that the persons in charge of the maintenance of the thesaurus receive information about the use of the thesaurus for indexing and retrieval. For this purpose, the "Statistic Module" of GenThes (see Figure 4) provides qualitative and statistical data about a number of things, such as how often a descriptor has been used for searches, how often a term was focused upon by a user, and which search terms entered by the user could not be found in the thesaurus database. This information is also foreseen to give hints to the software developers for further improvements.

## 6.0 Lessons Learned

The most important lessons we have learned when developing the general thesaurus browser for Web-based catalogue systems in cooperation with four different contracting authorities from three different European countries (Austria, Germany, and Norway) and an European Entity (EEA) are:

- Thesauri are not dead. In spite of some rumors about the ineffectiveness of the development and use of thesauri (too expensive, too complex) in the age of full-text and Web retrieval engines, an outweighing number of users sees the advantages of thesauri for indexing and retrieval in combination with catalogue systems and the heterogeneous IBOs stored in such systems. This especially holds in multilingual environments, where multilingual thesauri support cross-lingual information retrieval. The development of several new environmental thesauri on national (e.g., Norwegian Environment Thesaurus) and international level (e.g., GEMET and the new (3<sup>rd</sup>) edition of AGROVOC [23], the multilingual thesaurus of the Food and Agricultural Organization of the United Nations (FAO) for indexing data in agricultural information systems) and the great interest in these thesauri supports this point of view.
- Powerful tools are needed to make use of thesauri in Web-based catalogue systems. Moreover, the complexity of the thesaurus structure has to be hidden by simple user interfaces. Most of the retrieval users of catalogue systems do not know what and how complex a thesaurus is and (for simple queries) they do not need to know it! Consequently, the user interface of GenThes became simpler with each iteration. The most important role for thesauri in catalogue systems – beside the support of multilingual information retrieval – seems to be to provide a topical guide to the catalogue entries comparable to Yahoo's "Browseable Directory" [24].
- Advanced features such as exploring the structure defined by the different relations of the thesaurus and making consciously use of it for a search in the catalogue system are not used by most of the users. Therefore, these features can be "hidden" from the novice user to reduce the fear of thesaurus contact. For the advanced user, query expansions based on the thesaurus structure (synonyms, broader/narrower/related terms) should be made transparent.
- In general we had a very high reusability rate of our thesaurus browser components. One of the reasons for this is that the general requirements for thesaurus browsers for the different catalogue systems are similar. More important, well-designed components using widely known design patterns [25] and bearing in mind customizability from the initial project phase – and not avoiding the additional effort in this phase – are the basis for the reusability. The Java programming language provides an excellent basis for proper implementations.
- Performance is a crucial issue also for Web-based applications. Users expect almost the same response times from Web applications as from a Windows application. Improvements in performance are directly related to improvement of user acceptance. The techniques we used for performance improvements include:

- The data transfer volume (i.e., the volume of the Java applet itself and the thesaurus data) is reduced to a minimum. Especially the start-up screen of a Java applet has to be displayed as fast as possible.
- Large sets of top terms that have to be transferred at once can be reduced, e.g., by introducing another hierarchy level. The ideal thesaurus for providing a topical guide to the catalogue entries should have less than 50 top-level terms or top-level categories and also less than 20 narrower terms for each term. GEMET comprises a classification scheme, containing 3 super-groups and 37 groups (see [Section 3](#)). The super-groups have been adopted to approach an environmental management perspective and they have also been used to help the hierarchical structuring of GEMET. The groups reflect a category- or discipline-oriented perspective. The groups are displayed as the top level of the hierarchy browser.
- Database operations are speeded up, e.g., by the use of materialized views.

## 7.0 Conclusions and Future Work

GenThes provides a powerful tool for inspecting thesauri and for making use of them for indexing and retrieval. GenThes can be used in combination with different Web-based catalogue and information systems on the Internet as well as on Intra- and Extranets. Currently, GenThes is used together with WebCDS as well as the Austrian WWW-UDK to support retrieval and together with the Norwegian CDS to support indexing of metadata records. In the future, a seamless integration in the advanced Java-based integrated WebCDS Retrieval system as well as the WebCDS Update system is planned. The necessary integration work will include a SWING version of GenThes to make use of the new graphical user interface features of Java. It is also planned to fully integrate GenThes in the new version 4.0 of the German/Austrian WWW-UDK.

Different methods for the 2D-visualization of the graph structure of the thesauri have already been evaluated (e.g., Watanabe's heuristic approach for graph displaying [26] or Eades Spring Embedder approach [27]), to provide another intuitive graphic view to the complex structure of several thousands terms. Visual oriented users seem to be especially attracted from pseudo 3D-visualizations such as the hyperbolic view for hierarchies presented in [28], [29].

Several meta-data standards and catalogue systems have been established in recent years. To provide a single entry point for information seekers the interoperability of catalogue systems (which exist on regional, national and international level) and their respective thesauri is a key issue, see e.g., [30], [31]. In the past, multiple thesauri have been integrated manually into a single "super-thesaurus". This clearly is insufficient for loosely-coupled distributed catalogue and information systems where a loose integration, a so-called federation, of thesauri that allows them to retain their autonomy is more appropriate. In an on-going research project, we are developing concepts for the integration of distributed, autonomous, and heterogeneous thesauri on a technical, a schema and a semantic level [6].

A thesaurus federation offers new ways of indexing and retrieval. The indexing as well as the retrieval process of thesaurus terms benefits from having access to different, general and specialized thesauri which take part in the thesaurus federation. One or more thesauri in the federation provide the entry point for the user (indexer or information seeker). The inter-thesaurus-relationships provide the basis for cross-domain information retrieval.

From a technical point of view, GenThes 1.2 provides all those features needed to be integrated with further catalogue systems including mechanisms for customizing GenThes to different database schemas and different looks and feels. Therefore it provides an important customizable component when building Web-based catalogue systems. We believe that GenThes is an important first step towards a component-based catalogue system. It is our vision that together with further customizable components (e.g., a map-based search tool and a navigator for visualizing the structure of the catalogue entries) future Web-based catalogue systems will be composite applications of ready-to-use, off-the-shelf components. This approach will greatly facilitate the development of catalogue systems.

## 7.0 References

- [1] Catalogue of Data Sources (CDS), URL <http://www.mu.uni-hannover.de/cds/>, December 1998.
- [2] European Environment Agency (EEA), URL <http://www.eea.eu.int/>, January 1999.
- [3] Smith, T. Meta-information in digital libraries. In *Journal On Digital Library, Special Issue on Metadata for Digital Libraries*, September 1997, vol. 1, no. 2, pp. 105-107, URL, <http://link.springer.de/link/service/journals/00799/tocs/t7001002.htm> .
- [4] Yu, C. and Salton, G. The Effectiveness of the Thesaurus Method in Automatic Information Retrieval. Technical Report, Department of Computer Science, Cornell University, No 75-261, Ithaca, New York, 1975.

- [5] Kristensen, J. Expanding end-user's query statements for free text searching with a search-aid thesaurus. In *Journal on Information Processing and Management*, vol. 29, no. 6, 1993, pp. 733-744.
- [6] Nikolai, R., Traupe, A., and Kramer, R. Thesaurus Federations: A Framework for the Flexible Integration of Heterogeneous, Autonomous Thesauri. In *Proceedings of Conference on Research and Technology Advances in Digital Libraries (ADL'98)*, April 1998, pp. 46-55.
- [7] CNR, Rome, Umweltbundesamt, Berlin. GEMET - General European Multilingual Environment Thesaurus, Version 1.0. Technical Report, European Environment Agency, September 1997.
- [8] Kramer, R. and Nikolai R. Accessing Multilingual, Heterogeneous Data Sources in Wide Area Networks - Requirements and Approach. In *Proceedings of the 1996 Flexible Query-Answering Systems Workshop (FQAS'96)*, Roskilde, Denmark. Edited by Christiansen, H., Larsen H.L. and Andreasen, T., pp. 255 – 264.
- [9] Oard, D. Alternative Approaches for Cross-Language Text Retrieval, In *Electronic Working Notes of AAAI Spring Symposium on Cross-Language Text and Speech Retrieval* URL <http://www.ee.umd.edu/medlab/filter/sss/papers/>. Edited by Hull D. and Oard D., 1997.
- [10] Kazakos, W., Kramer, R., Nikolai, R., and Rolker, C. WebCDS - A Java-based Catalogue System for European Environment Data. In *Proceedings of International Workshop Issues and Applications of Database Technology (IADT'98)*, pp. 482 – 490, URL <http://www.fzi.de/dbs/publications/kazakos/IADT98.ps>, 1998.
- [11] WebCDS, URL <http://www.mu.niedersachsen.de/system/cds/>, December 1998.
- [12] Kramer, R., Nikolai R., Koschel A., Rolker C., Lockemann P., Keitel A., Legat R., and Zirm, K. WWW-UDK: A Web-based Environmental Metainformation System. In *ACM SIGMOD*, 1997, vol. 26, no. 1, <http://www.cs.umd.edu/areas/db/record/issues/9703/index.html>.
- [13] UDK coordination centre, URL <http://www.mu.niedersachsen.de/udk/>, August 1998.
- [14] ISO 2788 – 1986 Documentation – Guidelines for the Establishment and Development of monolingual thesauri, ISO Norm.
- [15] ISO 5964 – 1985 Documentation - Guidelines for the Establishment and Development of multilingual thesauri, ISO Norm.
- [16] Lessing, H., Swoboda W., and Günther, O. UDK: A European Environmental Data Catalogue. In *Proceedings of Third International Conference/Workshop Integrating GIS and Environmental Modeling (NCGIA)*, URL [http://ncgia.ncgia.ucsb.edu:80/conf/SANTA\\_FE\\_CD-ROM/sf\\_papers/guenther\\_oliver/my\\_paper.html](http://ncgia.ncgia.ucsb.edu:80/conf/SANTA_FE_CD-ROM/sf_papers/guenther_oliver/my_paper.html), 1996.
- [17] Global Change Master Directory (GCMD), URL <http://gcmd.gsfc.nasa.gov/>, January 1999.
- [18] GCMD valids, URL <http://gcmd.gsfc.nasa.gov/cgi-bin/valids>, January 1999.
- [19] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Thesaurus. July 1998, URL <http://isis.dlr.de/thesaurus/>, July 1998.
- [20] WWW-UDK 3.1 Austria, URL <http://udk.bmu.gv.at/>, July 1998.
- [21] WWW-UDK 4.0 Prototype, URL <http://www.fzi.de/dbs/applAreas/udk4.html>, November 1998.
- [22] Swoboda, W., Kruse F., Nikolai R., Kazakos W., Nyhuis D., Rousselle, H. The UDK Approach: The 4th Generation of the Environmental Data Catalogue for Austrian and German Public Authorities. *Third IEEE Meta-Data Conference*, 1999.
- [23] Food and Agricultural Organization of the United Nations (FAO). Introduction to the third edition of AGROVOC. URL <ftp://ftp.iaea.org/dist/agris/outgoing/agrovoc/prefen98.doc> and <http://www.fao.org/LIBRARY/agris/agris-e.htm>, 1998.
- [24] Yahoo! Browsing Yahoo! URL <http://howto.yahoo.com/chapters/8/1.html>, 1998.
- [25] Gamma, E., Helm, R., Johnson, R., and Vlissides, J. *Design Patterns*, Addison Wesley, 1995.

- [26] Watanabe, H. Heuristic Graph Displayer for G-BASE. In International Journal for Man-Machine Studies, vol. 30, pp. 287-302, 1989.
- [27] Eades, P. A Heuristic for Graph Drawing. In Congressus Numerantium, no. 42, pp. 149-160, 1984.
- [28] Lamping, J., Rao, R., and Pirolli, P. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. ACM Conference on Human Factors in Software (CHI '95), Denver, Colorado, ACM, 1995.
- [29] Inxight. Hyperbolic Tree. URL <http://www.inxight.com/Content/243.html>, 1999.
- [30] Nikolai, R., Koschel A., and Kramer, R. Automating metadata updates exemplified by the environmental data catalogue UDK. In Proceedings of 8th International Conference on Management of Data (COMAD'97), Chennai (Madras), India, pp. 263-276, 1997.
- [31] Kramer, R., Nikolai, R., Rolker C., Bjarnason S., Jensen, S. Interoperability Issues of the European Catalogue of Data Sources (CDS). In Proceedings of the 2nd IEEE Metadata Conference, Silver Spring, USA, URL <http://computer.org/conferen/proceed/meta97/papers/rkramer/rkramer.html>, 1997.
- 

## **ACKNOWLEDGEMENTS**

The contributions of Wolf-Dieter Batschi, German Federal Environment Agency (Umweltbundesamt UBA) to the development of GEMET are greatly appreciated.

---

Copyright 1999 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

---