

„SuperThes“: A New Software for Construction, Maintenance and Visualisation of Multilingual Thesauri

Wolf-Dieter Batschi¹, Bruno Felluga², Rudolf Legat³, Paolo Plini²,
Hermann Stallbaumer⁴, Konrad L. Zirm³

Abstract

In the last years, a General Multilingual Environmental Thesaurus (GEMET) in all the languages of the EU-member states has been developed within the working program of the "European Topic Centre for Catalogue of Data Sources & Thesaurus" (ETC/CDS), European Environment Agency, (EEA). GEMET is meant to support indexing of metadata within the CDS system.

At the same time, an environmental Thesaurus based on the Umwelt Thesaurus (UBA - Umweltbundesamt, Berlin) was produced in co-operation between Germany and Austria for their common metainformation system "Environmental Data Catalogue" (Umweltdatenkatalog, UDK). Austrian UDK at <http://udk.ubavie.gv.at>.

To manage and maintain both the CDS-Thesaurus and the UDK-Thesaurus, a THEsaurus MAINTenance (**THESmain**) system as well as a tool for visualising thesauri (**THESshow**) had been constructed by TBHS on behalf of the ETC/CDS and the Germany/Austria UDK-team.

THESmain is fully operational since May 1997 for both applications (CDS and UDK). Both GEMET and UDK-Thesaurus are maintained in Vienna. UDK-Thesaurus Editorial Board at http://www.cedar.at/wgr_home.

Since 1999, CNR is working on the Environmental Applications Reference Thesaurus (EARTH), a new Thesaurus for the Environment, encompassing and extending GEMET. To fulfil the requirements of future Thesaurus work, a Memorandum of Understanding

¹ Umweltbundesamt Berlin, Bismarckplatz 1, D-14193 Berlin,

email: wolf-dieter.batschi@uba.de, Internet: www.umweltbundesamt.de

² CNR Consiglio Nazionale delle Ricerche, Unità Terminologia Ambientale, IIA - Istituto sull'Inquinamento Atmosferico, Area della Ricerca di Roma-Montelibretti, Via Salaria Km 29,300, I-00016 Monterotondo Stazione, Rome - Italy, email: baf@iia.mlib.cnr.it, plini@iia.mlib.cnr.it;

Internet: <http://www.t-reks.cnr.it> - http://www.t-reks.cnr.it/UTA/uta_main.htm

³ Umweltbundesamt Wien, Spittelauer Lände 5, A-1090 Wien

email: legat@ubavie.gv.at, zirm@ubavie.gv.at, Internet: <http://udk.ubavie.gv.at>

⁴ Fa. TBHS, Favoritenstraße 182, A-1100 Wien, email: hermann@tbhs.co.at

(MOU) has been concluded in late 2000 between CNR - Consiglio Nazionale delle Ricerche, Rome, UBA Berlin and UBA Vienna. It has the aim to build a co-operation in order to develop a new software tool for building and maintaining multi-lingual, poly-hierarchical Thesauri, based on the experiences of the existing Thesaurus software tools used within the UDK co-operation.

General information about SuperThes

With the new Thesaurus Software SuperThes, the five years of experience with the existing Thesaurus maintenance software will be implemented.

"SuperThes" is based firstly on the experience gained working with the current software package THESmain/THESshow and secondly on the specifications of the partners of the MOU.

Major design goals for the new software are:

- Compatibility with data maintained in the existing software;
- Enhanced user interface, in order to ease system operability;
- Integration of "state of the art" operating system support;
- Support to create user-defined Thesaurus structures;
- Support to create user-defined export formats;
- Support for interfacing with Microsoft standard software;
- Provision of maximum flexibility for creating reports.

Description of the Thesaurus Maintenance Software SuperThes

SuperThes is used for visualisation and maintenance of thesaurus data according to DIN 1462 /ISO 2788 and DIN /ISO 5964

The application has been developed in Delphi, a Pascal dialect by Borland/Inprise. The programming language offers true object-oriented programming support as well as the stability common to Pascal programs.

SuperThes will run by default on all modern 32-Bit Microsoft Operating systems as a client/server application. All persistent data will be kept within the relational database system Interbase.

SuperThes as an application has an English user interface. Language codes follow ISO-standard 639-1. Thesauri created with this application may therefore comprise more than 100 languages. All languages, character sets and glyphs installed on the system may be used. Thus languages like Greek, Russian but also Chinese or Hebrew do not present problems. Input method editors, which are common to Asian languages, are possible. Different fonts are also available.

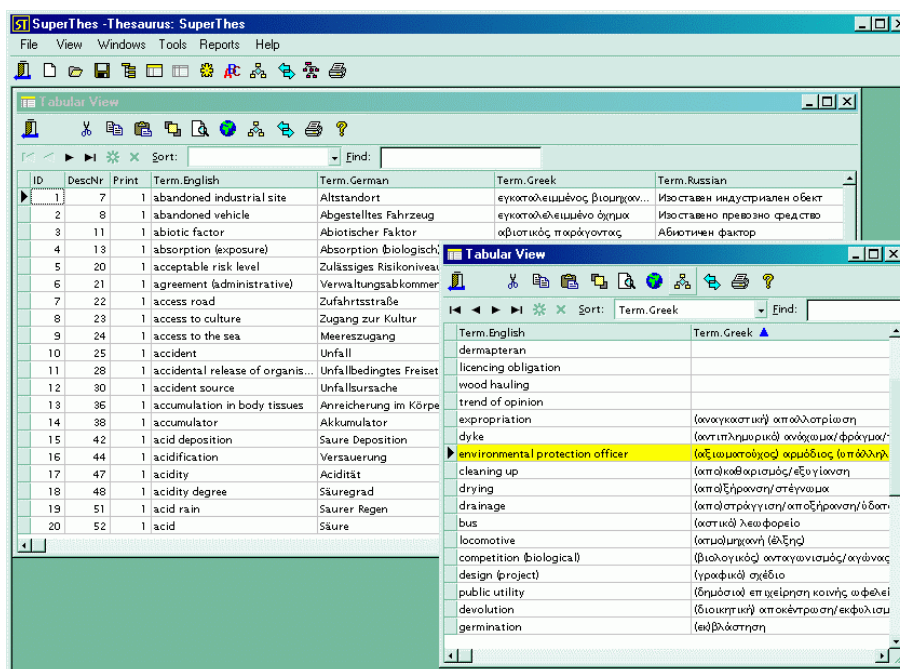


Figure 1: Versatile configurable data windows, in Unicode

Thesaurus Specification

The application provides user interface and functionality to create and maintain monolingual and multilingual thesauri. The structural functionality, as well as the terminology used, conforms to ISO standards 2788 and 5964. The user definable set of hierarchical rules is a superset of ISO standards 2788 and 5964. Other rules, like the minimum requirements for field entry, are also user definable. The adherence to these rules is automatically enforced by the application itself. The main rules are, just to mention the most important ones, the enforcement for uniqueness of concept, the enforcement of reciprocity, the check for arithmetic loops and the check for duplicate entries. Besides the main Thesaurus structure, a SuperThes Thesaurus may contain other user definable tables, which may be used to describe the main Thesaurus in a more detailed way or to keep attached data in them. An example might be the themes-table in GEMET, or a table containing geographic names or other tables.

Thesaurus data are kept in standard database files, which allow the exchange of thesauri, the creation and deletion of thesauri on a simple file base. SuperThes allows to maintain on the same computer system an unlimited number of thesauri and of

course of different Thesaurus systems with different data definitions. More than one instance of the program may be used simultaneously on the same computer, showing different data contents e.g. GEMET together with UDK-Thesaurus. All Thesaurus data may contain linguistic equivalents (“translations”) for up to 30 languages using all character sets and glyphs available in the installation system. Input method editors are also supported.

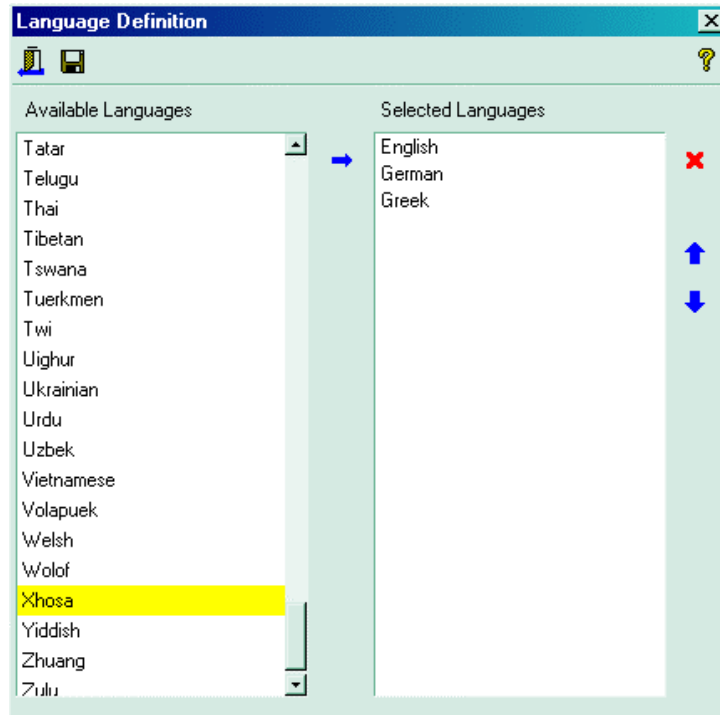


Figure 2: SuperThes, ISO 639-1 compliant, supports many languages

Data Types

In SuperThes, a Thesaurus may be defined according to the required data model. There are only two fields mandatory for any model: a field containing the record neutral identifier and a field for the decimal notation controlling the hierarchies, the content of which is generated dynamically by the application. All further fields are definable.

Besides the standard data types advanced data types like images or formatted text (text forms) are available.

SuperThes will contain predefined templates for GEMET, UDK-Thesaurus and CNR EARTH[®] Thesaurus for the Environment.

Table Definition	
Tablename	Description
THESAURUS	
TestA	

Field properties of selected table				
Fieldname	Datatype	Index	LDF	Description
ID	ID	Unique		
Sort	Numeric		<input type="checkbox"/>	A single field
Term	Date	Unique	<input checked="" type="checkbox"/>	new desc
Percentage	Scientific	Ascend	<input type="checkbox"/>	

Figure 3: Building your own data structure. User definable tables and data fields

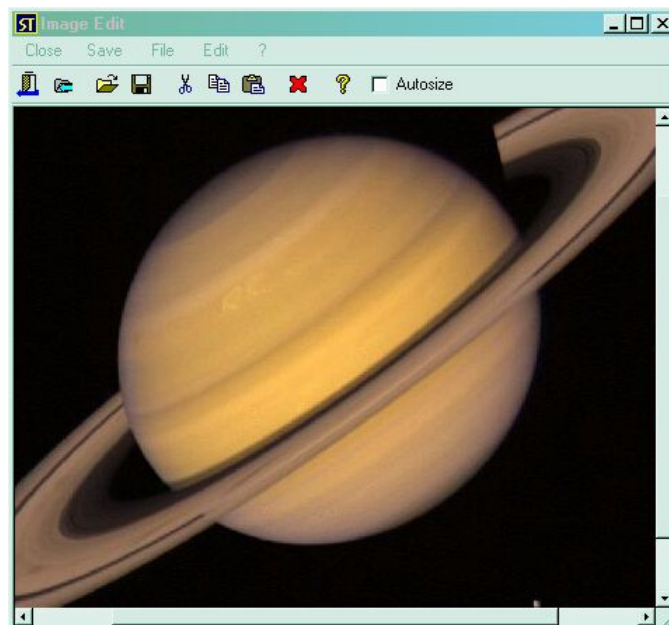


Figure 4: Various editors for text, forms, images etc.

User Interface

SuperThes is designed to create and edit thesauri using a graphical user interface according to rules defined for Microsoft Windows. The program functions are hierarchically ordered. Each window provides context sensitive help. State of the art features, like drag and drop and right click context menus, will be used where useful. All functions are available *via* mouse control, as well as *via* keyboard control. Operation of similar functions will be similar in all windows.

The following key functions are implemented:

- Management for thesauri, including creation, deletion, and copying thesauri;
- Definition of thesauri due to the required data model, including auxiliary tables;
- Definition of languages;
- Visualisation of all data contents of a Thesaurus, either in tabular or in hierarchical view;
- Editing all Thesaurus data;
- Generation of reports;
- Data exchange with external applications.

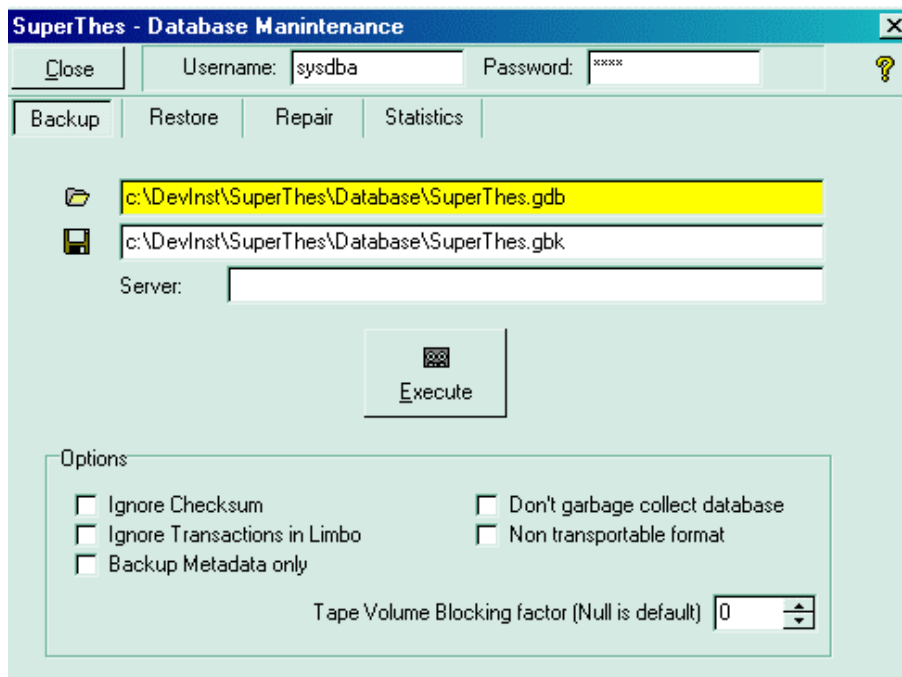


Figure 5: Server management utilities are included

Data Exchange

Data exchange may be performed in several ways:

- **SGML Export and Import**
For the use at the European Environmental Agency, a data exchange format utilising SGML has been designed. This format is already used in a number of related applications. SuperThes will also be able to write and read this data exchange format. Benefits of this method are the general available document type definition and the possibility to read and write every detail of a Thesaurus. The drawback is the amount of work that must be put into a custom reader or writer.
- **Microsoft compatible Export and Import**
Import and Export of "flat" (non-hierarchical) data directly to and from Microsoft applications like MS-Excel or MS-Access. This method was already used with great success in porting GEMET into other languages. It may also be used to import any flat file (list of terms or similar) into the system.
- **Interface to attached applications**
For applications designed as to be used in conjunction with SuperThes, a specialized interface will be available. This feature may be used for THESshow, the standalone visualizer, as well as for the Web visualiser, "THESweb", which is in the design phase.

Bibliography

- Angrik, M., Bös, R., Rüter, M., Bandholtz, T. (2002): "Semantic Network Services (SNS)". In: EnviroInfo 2002, TU Vienna 2002.
- Bandholtz, T., Bös, R., Rüter, M. (2000): "The German Environmental Information Network (GEIN)." In: Cremers, A.; Greve, K. (Eds.): Computer Science for Environmental Protection '00. 2000.
- Batschi, W-D., Legat, R., Stallbaumer, H., (1999): „THESshow und THESmain", Moderne Softwarewerkzeuge zur Erstellung, Wartung und Visualisierung multilingualer Thesauri. In: 51. Jahrestagung der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis e.V. (DGI), Hamburg 21. Bis 23. September 1999, Seite 285ff, Frankfurt am Main: DGI, 1999. Available at: http://www.cedar.at/wgr_home/pub/hamburg.htm.
- EEA, European Environment Agency. Felluga B. & Batschi W.D. (Eds.), (1999) - GEMET, General European Multilingual Environmental Thesaurus. CNR, Rome – UBA, Berlin, Version 2.0. CD-ROM edition. Available at http://www.mu.niedersachsen.de/cds/etc-cds_neu/software.html.

- EEA, European Environment Agency. Felluga, B. & W.-D. Batschi, Eds. GEMET, General Multilingual Environment Thesaurus. Version 2.0, August 1999. Printed edition. pp. xxvi + Volume 1: Systematic List of Descriptors, pp. 44; Volume 2: Thematic List of Descriptors, pp. 78; Volume 3: Alphabetical List of Terms, pp. 550; Volume 4: Concordance List, pp. 127; Volume 5: Multilingual List of Descriptors, pp. 536. CNR, Rome, April 2000.
- EEA, European Environment Agency (1996): Newsletter issue 8, Copenhagen, June 1996
- Felluga, B., Plini, P., Lucke, S., Palmera, M., Eds. EARTH, Environmental Applications Reference Thesaurus[©]. Pp. vi + Volume 1: Elenco Sistemático, pp. 37; Volume 2: Elenco temático, pp. 159; Volume 3: Elenchi Alfabetici brevi, Italiano-Inglese e Inglese-Italiano, pp. 113 + 113; Volume 4: Elenco alfabetico bilingue, pp. 1418; Volume 5: Elenco delle Concordanze, pp. 115; Volume 6: Elenco Termini liberi – assegnati a liste – esclusi/cancellati, pp. 20; totale: vi + 1975. CNR, Roma, Aprile 2002.
- Hashemi-Kepp, H., Legat, R. (2000): "Der Österreichische UmweltDatenKatalog, Erhebung und Strukturierung der Daten" UI 2000 - 14. Symposium Informatik für den Umweltschutz 2000 in Bonn, at: http://www.cedar.at/wgr_home/pub/hashemih.htm
- Kruse, F., Karschnick O., Spöringer A., Eichler M., Behrens S., "The UDK, Present Status and Future Development." In: EnviroInfo 2002, TU Vienna 2002.
- Legat, R., Batschi, W.-D., Hashemi-Kepp, H., Kruse, F., Nikolai, R., Nyhuis, D., Pultz, S., Stallbaumer, H., Swoboda, W., Zirm, K. (1999): "Der Umweltdatenkatalog in Österreich", 5 Jahre Erfahrungen, Workshop Umweltdatenbanken im Web, FZI Karlsruhe. Available at http://www.cedar.at/wgr_home/pub/karlsruhe.htm
- Plini P., Lucke S., Baffioni C. & Felluga B. (2001): T-REKS: a Contribution to the Environmental Information Management through a Computer-Supported Modular Knowledge Organisation System for the Environment. 15th International Symposium Informatics for Environmental Protection, Zurich 2001. "Sustainability in the Information Society". Hilty L.M., Gilgen P. W. (Eds.). October 9-12, 2001. Pp. 691-698.
- Swoboda, W., Kruse, F., Legat, R., Nikolai, R., Behrens, S. (2000): "Harmonisierter Zugang zu Umweltinformationen für Öffentlichkeit, Politik und Planung: Der Umweltdatenkatalog UDK im Einsatz", UI 2000 - 14. Symposium Informatik für den Umweltschutz 2000 in Bonn at: http://www.cedar.at/wgr_home/pub/ui2000.htm.