

## **OpenNetTerminologyManager- a Web and Standards based OpenSource Terminology Management Tool**

Dr. Klemens Waldhör  
Friedrichstr. 17  
90574 Roßtal  
Germany

[dr.klemens.waldhoer@waldhor.com](mailto:dr.klemens.waldhoer@waldhor.com)

### **Key words:**

Terminology management, open source, free software, MARTIF, terminology standards

## OpenNetTerminologyManager - a Web and Standards based OpenSource Terminology Management Tool

Dr. Klemens Waldhör

### Summary

OpenNetTerminologyManager is a privately started Open Source project which aims at developing a freely available pure web based concept terminology management system. It runs with any browser supporting JavaScript. The server side requires MySQL, Apache Web Server and Perl. The system is currently available through sourceforge.net at <http://openwebterm.sourceforge.net>. OpenNetTerminologyManager supports different terminological models. A version which is based on MARTIF has been implemented.

### Background

Through the years the world has seen the attempt to establish several different terminology standards starting from MARTIF, Geneter to TBX, XLT (SALT), TMF (ISO 16642) and so on. The author himself was part of one of the older efforts which started 1990 where within the MULTILEX project an attempt was made to create a standard description for areas like mono- and multilingual lexicons, machine translation etc. The basic idea there was to use SGML as the description language. This was followed up in projects like EAGLES, Otelo (OLIF) etc. In parallel other attempts have been made like Geneter. Sometimes one is really puzzled how creative the terminology community is in inventing new ideas and standards. Often it is really hard to follow what is going on. This is the one side of the coin. On the other side the industry uses "**quasi standards**" like the export format from MultiTerm™ from Trados™. Several products of competitors like TermStar™, UniTerm™ provide import and export features to and from the MultiTerm™ format, simply because MultiTerm™ is the market leader in this area. Otherwise getting into this area for new systems is nearly impossible as most customers either use MultiTerm™ or at least provide data in this format.

Interestingly enough Open Source terminology software was never really part of the terminology game, in contrast to other areas like web servers where Apache is the dominating server (60 % of the world web server market). If you search for "open Source terminology management" in yahoo and inspect the results found there are only two other relevant matches, the ForeignDesk match and OpenGALEN. In the last half year Lionbridge has made its software **ForeignDesk** available through open source. Another notable effort is **RosettaWerks** which deals implementing a set of tools for the localisation process.

But what is really missing is a terminology tool which is available on several operating systems (not just Windows) and can be used through the web and itself is built on free available software. This is not the place to discuss the advantages of an open source model. A lot of discussion is going on this area, but I just want to add that one clearly has to distinguish the open source model from models which are offered by software suppliers where one can get the executable software for free, but has no access to the source code. Several providers of terminology software supply down-graded or full versions of their tools – mainly viewers - e.g. UniLex from Acolada GmbH, but this does not bring any advantage to the user as he still relies on the provider to fix bugs etc. In addition it is hard to check if there are any hidden traps in the software. As professional terminology management will contain company or customer information security aspects and the ability to check this will be an important aspect of choosing a system in the future. Based on this observations – and being also a fan of the open source community - I started developing a terminology management software which should fill this gap. Thanks have to given at this place to SourceForge which provides an excellent – and free – way to make open source projects available through the web.

### OpenNetTerminologyManager Terminology Model

One idea of the system is to be able to support different terminology models. The user should have the ability either to create his own model or to adapt an existing model. It should also be possible to keep track with on-going changes in the standardisation community. This has been realised in the system by not directly mapping the attributes (elements) of the terminology model to database tables, but keeping this information in a specific field where the structure can be freely defined. The actual mapping of this field to the attributes is defined in model files. The advantage of this approach is a)

that it keeps the number of databases table to a minimum and b) as a result is quite fast in searching and reading entries.

The kernel OpenNetTerminologyManager approach is concept oriented as it used in most modern terminology systems. In this approach a concept corresponds to one meaning of a word. The language specific parts of a concept are called "**language term**" or simply "**term**". Each concept is tagged with a unique identifier, while each term related to the concept uses the concept identifier plus a language identifier and an internal term counter as identifier.

**Example:** The German term "Birne" (three meanings: Glühbirne, Frucht, Kopf = bulb, pear, nut) will be represented with three concepts:

- a) One concept with the meaning of "Frucht = fruit" and
- b) One with the meaning of "Glühbirne = bulb" and
- b) One with the meaning of "Kopf = head".

The **kernel** of OpenNetTerminologyManager consists of several tables:

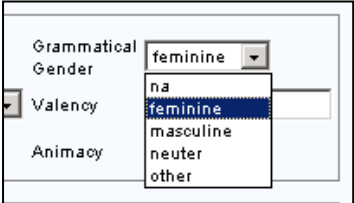
- a) A MONOTERM table which holds all relevant information for a term including term attributes
- b) A MULTITERM table which links MONOTERM terms to a concept and also may also store concept related attributes.
- c) A DETAILS table which contains links from attributes to terms and concepts. This table is only used to optimise the speed when searching with attributes.
- d) A LINK table which establishes links between either concepts or term (e.g. in order to express a relations like "synonym").

Different terminology models are now mapped to the kernel model in a special model file. This model file defines:

- a) The **names** (e.g. "Gender") to be used for the **attributes** of the terminology model to an internal name. This association differentiates between concept related and term related attributes.
- b) The **values** and **forms** to be associated with a such a **name**. As an example associate with the attribute "Gender" three possible values ("male", "female", "neuter") and display them in the browser as a select box.

The following shows a simple section from the MARTIF model:

```
opwdetail140=Grammatical Gender
opwdetail141=Term Type
opwdetail144=Grammatical Number
...
<tr>
  <td>
    <fieldset>
      <legend>Grammar</legend>
      <table>
        <tr>
          <td>@tdopwdetail143=Part Of Speech?10<select!noun|verb|adjective|other
            @tdopwdetail140=Grammatical Gender?10<select!na|feminine|masculine|neuter|other
          </tr>
          <tr>
            <td>@tdopwdetail141=Term Type?10<select!...|variant
              @tdopwdetail146=Valency?10<input
            </tr>
          <tr>
            <td>@tdopwdetail144=Grammatical Number?10<select!na|dual|mass|other|plural|singular
              @tdopwdetail145=Animacy?10<select!animate|inanimate|other
            </tr>
          </table>
        </fieldset>
      </td>
    </tr>
```



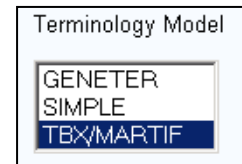
Models can further be differentiated two classes: "**full models**" and "**sub-models**". A sub-model is defined as a subset of attributes of a full model. This is mainly necessary if for a given model (e.g. MARTIF) only specific attributes should be shown or if specific restrictions may apply for attribute values.

The system contains some additional fixed attributes like the owner of the concept, read and write accesses etc.

### OpenNetTerminologyManager Features

The following features are currently available:

- Constraints between attributes can be realised with JavaScript
- New models and sub-models can be created by the user
- Attributes can be defined by the user
- Supports different types for attributes like option fields, text fields, select etc.
- Multiple databases; multi-user read/write support (locking at concept level). Different right combination can be used. Databases are either private (with user and password protection) or public
- Partial UNICODE support. Unicode characters are stored as SGML entities in the database. This will be removed once MySQL supports directly UTF9 or a similar Unicode encoding scheme. Languages like Arabic, Chinese, Japanese etc. can be used through this approach. Once a Unicode implementation of MySQL is available this will be changed to an internal Unicode character set too.



Currently one terminology model based on MARTIF has been (partially) implemented. It normalises the XML definitions into the relational (table based) approach defined above. Others like Geneter are under way.

### The Basic User Interface of OpenNetTerminologyManager



**Login** allows the user define default values like his preferred databases, languages, attributes to be displayed for a model, how search results should be displayed etc.

**Quick Search** offers a simplified search mode which simply looks up the database for a specified term independently of the language.

**Search Concept** supports more detailed search options like languages to be searched and displayed etc. Through this item the user also can edit concepts in the database.

**Add Concept** adds new concepts to the database.

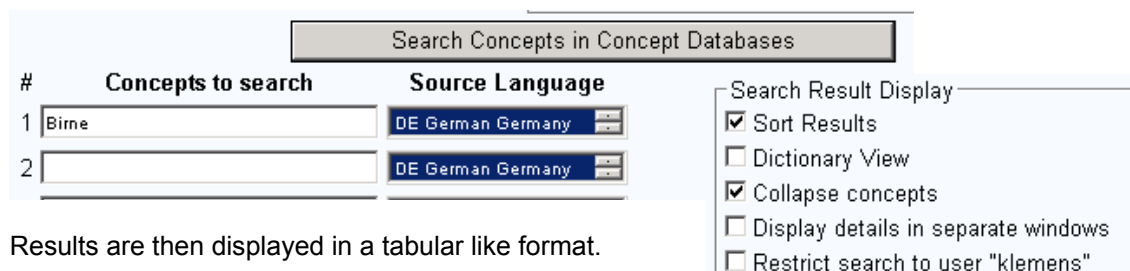
The user can create new databases using **Create Database** where he also specifies the model and access rights for the database.

They can be deleted using **Delete Database**.

**Import Term File** allows existing terminology files to be imported in various formats supporting double detection during import, while **Export Database** exports databases in various formats.

**Management** is mainly intended to give an overview of current system settings and databases. It also supports recreating the database structure.

Concepts can be editing by first searching them with the Search Concept function and using the "Edit Mode" (not choosing "Dictionary View").



Results are then displayed in a tabular like format.

No	ID / Concept	Source Language DE	Translation Term	Language	Database	Operation		
1/1	<a href="#">316071404</a>	<a href="#">Birne</a>	<a href="#">nut</a>	EN	meine	<input type="button" value="Edit"/>	<input type="button" value="Delete"/>	<input type="button" value="Copy Concept"/>
2/2	<a href="#">284021368</a> Glühbirne	<a href="#">Birne</a>	<a href="#">bulb</a>	EN	meine	<input type="button" value="Edit"/>	<input type="button" value="Delete"/>	<input type="button" value="Copy Concept"/>
3/3	<a href="#">312237593</a>	<a href="#">Birne</a>	<a href="#">pear</a>	EN	meine	<input type="button" value="Edit"/>	<input type="button" value="Delete"/>	<input type="button" value="Copy Concept"/>

“Edit” will then display the full entry.

Abstract Concept Name:  Database **meine**

#	Concept to edit	Language	Term Relate
1	<input type="text" value="Birne"/>	DE German Germany	<input type="button" value="Save Details"/>
2	<input type="text" value="bulb"/>	EN English UK	<input type="button" value="Save Details"/>
3	<input type="text"/>	EN English UK	<input type="button" value="Save Details"/>

**Administration Information**

Creation Author  Cre  
 Change Author  Ch.

**Concept Related Details**

**Concept Related Description**

Subject Field  Classification System

Classification Number

**Term Type**

Synonym  Q  
 International Scientific Term  C  
 Full Form  A  
 ot  
 s

Results can also be displayed in a “Dictionary View” mode

(1/1) [Planet](#) [DE] [*EN/meine*] [terrestrial planet](#) ; [giant planet](#) ; [Planet](#)  
 (2/4) [Planet \(innerer/äußerer\)](#) [DE] [*EN/meine*] [planet \(inner or inferior/superior or outer\)](#)  
 (3/5) [Planetarischer Nebel](#) [DE] [*EN/meine*] [planetary nebula](#)  
 (4/6) [Planetarium](#) [DE] [*EN/meine*] [planetarium](#)  
 (5/7) [Planeten](#) [DE] [*EN/meine*] [Planets](#)

Depending on the user settings attributes will be displayed either directly in the window or by a link and display in a separate browser window.

### Software requirements

OpenNetTerminologyManager requires the following software component: Perl > 5.0 (with some additional modules installed), Apache Web server, MySQL, JavaScript enabled Web Browser. Test have been done with Internet Explorer 5.0, 6.0®, Netscape®, Opera® . The system has been tested both on Windows (NT and 2000) and LINUX.

### Next Steps

An important feature which is currently under development is an advanced link concept. This link concept will not only support links in the way as TBX defines them but will allow to create complex typed links between concepts and terms and databases. This will allow the user to search the databases not only as a simple term-lookup tool but to browse through it in a kind of semantic net and find related concepts.

Other developments concern additional import / export formats and simplified form handling for attributes. As there are several opens source project on mapping xml to relation databases on the way (e.g. XML-DBMS) I am currently also looking into replacing the internal structure of the database by a full xml database approach. This will heavily depend on the access speed compared to the current implementation.

### References:

<b>Acolada</b>	<a href="http://www.acolada.de">http://www.acolada.de</a>
<b>ForeignDesk</b>	<a href="http://sourceforge.net/projects/foreigndesk/">http://sourceforge.net/projects/foreigndesk/</a>
<b>Rosettawerks</b>	<a href="http://rosettawerks.sourceforge.net/Default.php">http://rosettawerks.sourceforge.net/Default.php</a>
<b>OpenGALEN</b>	<a href="http://www.opengalen.org/">http://www.opengalen.org/</a>
<b>OpenNetTerminologyManager</b>	<a href="http://openwebterm.sourceforge.net">http://openwebterm.sourceforge.net</a>
<b>Sourceforge</b>	<a href="http://sourceforge.net/">http://sourceforge.net/</a>
<b>Star AG</b>	<a href="http://www.star-ag.ch/eng/software.html">http://www.star-ag.ch/eng/software.html</a>
<b>Trados</b>	<a href="http://www.trados.com">http://www.trados.com</a>
<b>XML-DBMS</b>	<a href="http://www.rpbouret.com/xmldbms/index.htm">http://www.rpbouret.com/xmldbms/index.htm</a>