XABIER ARREGI
ANA ARRUARTE
XABIER ARTOLA
MIKEL LERSUNDI
IGONE ZABALA
University of the Basque Country, UPV/EHU
Basque Country
{xabier.arregi | a.arruarte | xabier.artola | mikel.lersundi | igone.zabala}@ehu.es

TZOS: An On-Line System for Terminology Service

1. Introduction

Academics use terms in their everyday activity. Terms –specialized lexical units that must be shared by experts of a discipline– are needed to develop and transfer specialized knowledge. But, how can we build, manage, and share term collections –glossaries– efficiently? How can we learn and discuss about terms used or proposed by other colleagues? What about specialized computer applications and support for this kind of activities?

These questions led us to develop TZOS (*Terminologia Zerbitzurako On-line Sistema*, On-line System for Terminology Service), the on-line system for terminology service presented in this paper. Convinced of the fact that terminology must be a collective activity, TZOS has been designed as a tool which grants a more active role to the users of terminology which are also engaged in terminology creation. In some sense, it represents an attempt towards what is known as *collaborative terminology*.

Plenty of organizations and initiatives geared to promote, develop, standardize and disseminate terminological resources emerged in the last years. The most important goals of these organizations and initiatives are to set up methodological criteria and to develop standard and interchangeable resources, in order to support collaborative work on the field. Along with those organizations, several networks and communities were created, since terminology work can not be carried out in an isolated way. With this goal in mind were born the *European Association for Terminology* (EAFT)[1], the *International Information Centre for Terminology* (*Infoterm*)[2], the *International Network for Terminology* (*TermNet*)[3], and some other forums.

Thanks to the initiative of these organizations and communities, and to the support of many professionals of different fields, nowadays a profusion of diverse terminological databases can be consulted in the internet. In this context, it is worth mentioning *Inter-Active Terminology for Europe* (IATE)[4]. IATE is the terminological databank of reference in the European Union.

Closer to the Basque sociolinguistic reality, there are some other initiatives. Quebec constitutes an important reference. *Le grand dictionnaire terminologique*[5] is a well-known resource developed by the *Office Québécois de la Langue Française*[6]. *Inventerm*[7] is another interesting tool developed by the same charter.

The activity of terminology standardization has great significance in the context of minority languages. This was made clear at the Special Seminar on *Minority Languages and Terminology Policies*[8], organized the EAFT in 2007. It is interesting in this context the Catalonian reference. The Centre for Terminology TERMCAT has been working there since 1985, standardizing terminology and developing dissemination tools, both general and domain-based. *Cercaterm*[9] is their best-known tool.

There exist as well terminological resources similar to TZOS that have been developed at universities. For instance, in Catalonia, the terminological database UBTERM[10], built at the University of Barcelona, or the multilingual technical terminology UPCTERM[11], developed at the Polytechnic University of Catalonia. Both tools provide users with a simple query interface, allowing them to specify the search term along with the knowledge field when consulting these resources.

In the Basque Country, the best-known terminology databank is *EuskalTerm*[12]. The Elhuyar Foundation has recently published the science and technology dictionary ZTH (*Zientzia eta Teknologiaren Hiztegi Entziklopedikoa*)[13]. There is also a project named *WNTerm* (Pociello *et al.,* 2008), which aims to integrate ZTH with *WordNet*[14]. The main goal of *WNTerm* is to develop a multilingual ontology of science and technology.

---

[1] www.eaft-aet.net/en/index/
[2] www.infoterm.info/
[3] linux.termnet.org/
[4] iate.europa.eu/
[5] www.granddictionnaire.com
[6] The Charter of the French Language: www.oqlf.gouv.qc.ca/.
[7] www.inventerm.com/
[8] www.eaft-aet.net/en/activities/special-eaft-seminar-minority-languages-and-terminology-policies/
[9] www.termcat.cat/ca/Cercaterm/Fitxes/
[10] www.ub.edu/slc/ubterm/td_Arrencada.html
[11] www.upc.edu/slt/upcterm/
[12] www1.euskadi.net/euskalterm/indice_i.htm
[13] Encyclopedic Dictionary of Science and Technology: zthiztegia.elhuyar.org/
[14] wordnet.princeton.edu/

Although we have not mentioned so far but terminology query systems, in what regards the functionality of TZOS we should also have a look on terminology management systems and tools.

Some terminology management systems worth mentioning are *evoTerm*[15], *heartsome*[16], *Terminology Management Software*[17] and *Enterprise Terminology Management*[18]. All they are concept-based tools, i.e., systems in which terms in several languages are attached to the same underlying concept.

TermWiki[19] is also a similar tool. However, its conception goes beyond the traditional approach in what concerns the management of terminology, since it provides the users with an environment for collaborative work.

Moving from the commercial grounds towards academia, we would mention *Terminus*[20]. This tool, developed in Catalonia at the *Institut Universitari de Lingüística Aplicada* of the Universitat Pompeu Fabra, constitutes a real workstation for terminologists.

We have shown that there are plenty of terminology resources and tools here and there. Most of them are robust tools, full-featured, and, so, they are good references for the work presented here. However, it must be said that all of them are proprietary software, and only a few are aimed at collaborative and open work.

In our opinion, one of the most interesting contributions from the point of view of collaborative work is the one described by Désilets *et al.* (2009). In this work the authors give the idea of *Collaborative Multilingual Terminology* (CMT) as follows:

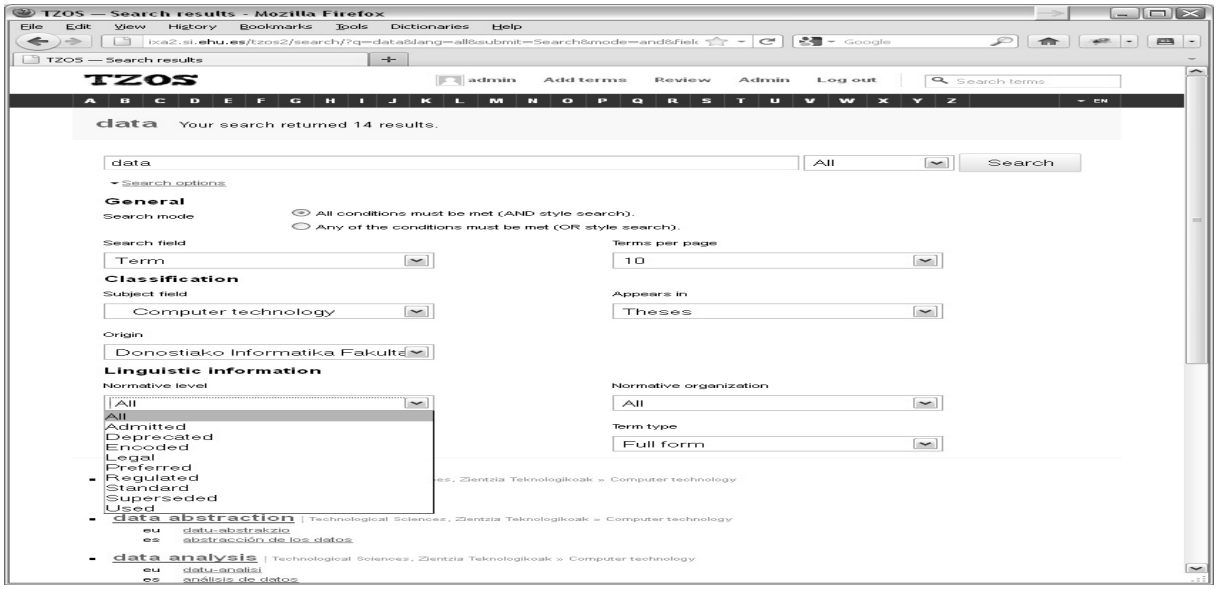
> ... a massive collaboration process like Wikipedia could have beneficial effects on TDBs (Terminology Databases), by spreading their creation and maintenance costs across a large number of individuals, and by fostering collaboration between terminologists, translators, domain experts, and even members of the general public.

It is precisely this approach which lies under the conception of TZOS. Academia offers very good opportunities towards this goal.

Once having shown an overview of applications and tools related to our work, we will give a description of TZOS in section 2. In section 3, the representation schema of terminological information used in TZOS will be presented. Finally, in section 4, we will outline some conclusions and future work.

2. TZOS: users and functionality

TZOS is an on-line platform to work on terminology collaboratively. It is addressed to five types of users: visitors, terminologists, correctors, supervisors, and administrators. Visitors are the only group who can use the system without previous registration but their actions are exclusively restricted to search operations. It is responsibility of terminologists to propose new terms. Correctors verify whether the proposed terms are correct regarding orthography and orthotypography, according to the standard. Supervisors are responsible for ensuring the suitability of the proposed terms. Finally, administrators are in charge of managing users and the system itself: create and remove users, make changes to the system (subject field changes, additions…), forum administration, backups, and so on. Next, the main functionality of TZOS is presented.



1. Advanced search query

[15] www.evoterm.net/
[16] www.heartsome.net
[17] www.termbases.eu/
[18] www.acrolinx.com/terminologymanagement_en.html
[19] www.csoftintl.com/termwiki.php
[20] terminus.iula.upf.edu

- *Search query: simple and advanced*

When a user makes a search query, the query term must be written, specifying the language in which to search. In addition to this simple query, there is also the possibility of making advanced searches by specifying the subject field, origin, the status of the term, and so on. Figure 1 shows the resulting screen for the term *data* when searching only in PhD *theses* written in *Donostiako Informatika Fakultatea* (Fac. of Computer Science of Donostia) in the *Computer Technology* field.

- *Glossary search*

TZOS allows the user to view all the terms starting with a given letter.

As it was pointed out above any type of user is allowed to make search and glossary queries.

- *Manual term proposals*

When proposing new terms manually, TZOS allows the user to store certain data associated with these terms such as language (Basque, English, Spanish, French, German or Latin), equivalent terms in other languages, subject field, origin, originating person, definition, examples, comments, context, related terms, and so on.

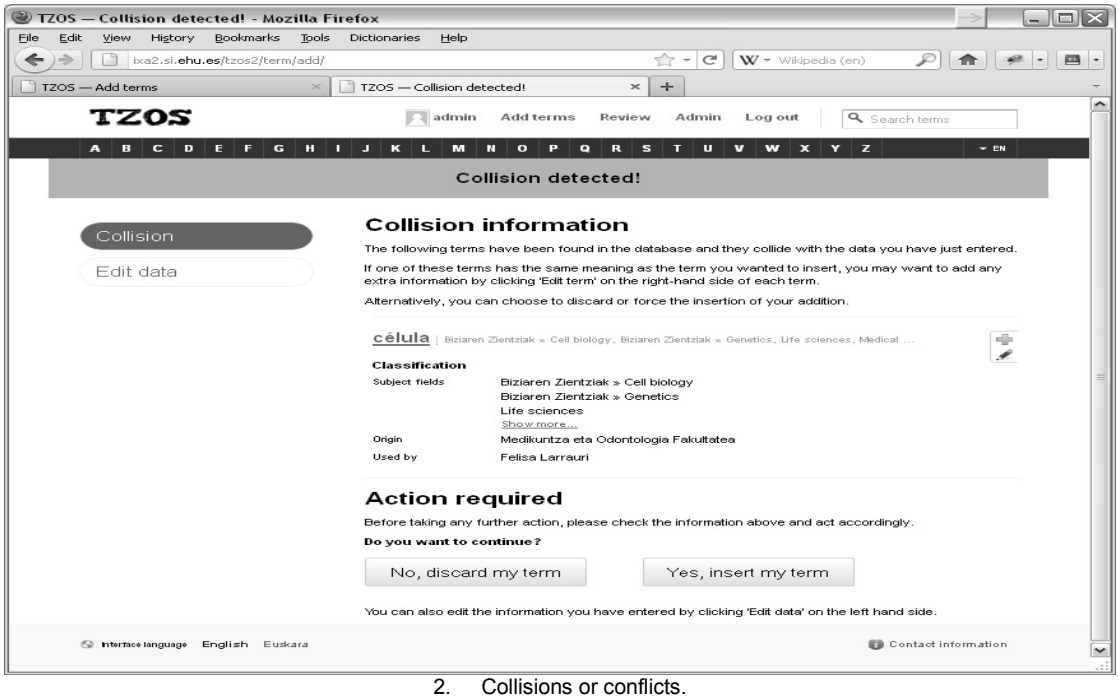
In order to propose a term it is necessary for the user to be registered in the system. Only registered users (called *terminologists* in TZOS) may propose terms.

When adding a new term some fields are compulsory: spelling, language, subject field, and origin. Besides, there are some features related to a term, such as originating person, working status (imported, starter, working, consolidated, archived), administrative status (preferred, admitted, deprecated, superseded) or normative level, which are automatically recorded. Even more, TZOS logs the evolution each term has suffered since it was first proposed in TZOS. For example, the person responsible of modifying the administrative status of a term or the reason to do it can be recorded. By default, terms stored in TZOS –and approved by correctors– are public terms that can be consulted by any user, even by visitors.

On the other hand, it could happen that when someone proposes a new term, an equivalent term –equal spelling within the same subject field– already exists in TZOS. In that case the system automatically shows a collision message and it is up to the terminologist to decide if it is or not a real collision. By analyzing the information already stored in TZOS, the terminologist will decide: a) to include the proposed term as a new one –in that case both terms will become homographs–; b) to include it as a synonym –a denominative variant– of an already stored term; or c) to discard the proposed term because the denomination has been already stored for the same concept. Figure 2 shows the collision message issued by the system after identifying that the *Spanish* term *célula* specified by the terminologist as equivalent of the *English* term *cell* is already in the system.

- *Term importation*

TZOS allows the terminologist to import a list of terms from an external CSV file that contains, separated by commas, terms in different languages. Each row corresponds to a new term and each column contains the term in a different language. The subject field, origin and originating person must also be specified when importing the terms. Imported terms have the value *imported* as working status, and should be completed once the corrector approves them.



2. Collisions or conflicts.

- *Exportation of query results*

TZOS allows administrators the exportation of the results obtained from a query in an XML file (TBX standard, see section 3). The exportation form requires values for language, subject field and origin of the terms to be included in the results.

- *Discussion forum*

TZOS provides a forum to manage the questions and discussions related to both terms and the system itself.

3. Representation of terminological information in TZOS

3.1. TBX Standard

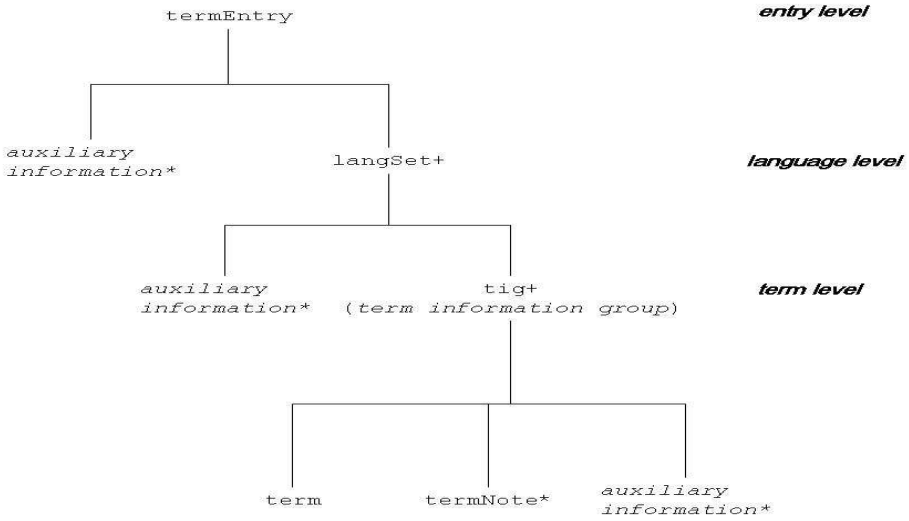
From the beginning, concept-based representation has guided the effort of computationally representing terminology. Starting in the seventies several standards have been defined and used. Currently, those based in mark-up languages are the most widely used. In this new era, TEI (*Text Encoding Initiative*[21]), already in its third guidelines edition (*TEI P3*, 1994), dedicated one whole chapter to terminology, proposing an SGML-based encoding for representing terminological data. Later, TEI proposals moved on ISO (*ISO Technical Committee 37*[22]*)*, and MARTIF (ISO 12200:1999) and the first version of the ISO 12620 data categories were created.

The European SALT project (*Standards-based Access service to Lexicons and Terminologies*[23]) constituted another important milestone in the computational representation of terminology. Together with the *ISO Technical Committee 37*, organizations such as the LISA OSCAR association (*Open Standards for Container/content Allowing Re-use*[24]), the OLIF consortium[25], and TEI itself collaborated to obtain the first draft of TBX (*Term Base Exchange*), which led to the development of TMF (*Terminological Markup Framework*, ISO 16642:2003).

The standard TBX (LISA OSCAR, 2008), therefore, is the result of a long collaboration process between different organizations. LISA OSCAR adopted, in 2002, TBX as an exchange format. Since 2008, when the LISA standard became ISO 30042, it has allowed users to represent specific lexical information.

TBX is an XML-based open standard for representing structured terminological data. It is designed to support various types of processes involving terminological data such as analysis, descriptive representation, dissemination, and interchange in a variety of computer environments. It can be said that the main purpose of TBX is the interchange of terminological data. Taking into account that XML is just a format, it is not adequate for specifying how to visualize terminological information; TBX is only conceived to describe structure and content. Most of the software packages which use terminological databases are equipped with all the necessary to import and export TBX files. By providing a universal mark-up model, enterprise and organizations use TBX not only to manage their internal terminological data but also to share and exchange information with others.

TBX provides two modules for managing data categories in terminological databases (*termbases*), both specified using XML. Meanwhile the first one allows users to define the basic structure, the second one provides a formalism to identify and to establish constraints among data categories (XCS, *eXtensible Constraint Specification*). So, TBX is modular. It must be also remarked that the term TBX is used to refer to the result of the interaction of these two modules.



3. Structure and levels of a terminological entry.

In order to maximize the interoperability among currently existing terminological data, TBX provides a default data category set to be used in terminological databases: the default XCS. The main objective of this default category

[21] www.tei-c.org/
[22] www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=48104
[23] www.ttt.org/salt/
[24] www.lisa.org/OSCAR-LISA-s-Standa.79.0.html
[25] www.olif.net/consortium.htm

is to provide a blind representation mechanism. Its use will enable users to interpret data without consulting providers.

Nevertheless, in some cases applications do not share the same data categories. In these cases, and thanks to the XCS mechanism, TBX is also a valid and flexible option. By using the XCS mechanism each user group can define their own data categories taking into account their requirements. The data categories and the constraints among categories will be stored in an XCS file. So, each user group define its own TML *Terminological Markup Language*), its own terminological mark-up XML language.

3.2. Using TBX in TZOS

Next, we will describe the data model adopted to represent terminology in TZOS, based on the TBX standard.

As mentioned above, TBX provides an XML format for exchanging terminological information. In order to gather some particularities of the terminological system we aimed to build, it was necessary to adequate the default XCS. In this way, TZOS_TML, a TML to be used in TZOS, was defined.

Following TBX, a terminological entry in TZOS will have the structure shown in Figure 3. A terminological collection is composed of such entries, one per concept (`termEntry`). Three levels can be distinguished there, each one of them having associated its own information. At entry level, concept-related information is represented; at language level, information about the concept is expressed in different languages (in the figure, the symbol + indicates that one or more `langSet` can occur, one per language); finally, there is the term level, where the term itself (`term`) and associated information are represented by means of a `tig` element. If several synonyms denoting the concept are to be represented, more `tig` elements will be needed, one per term. As can be seen in the figure, different kinds of auxiliary information can always be represented at any level (the symbol * indicates that it can occur 0 ore more times). Auxiliary information may include elements containing descriptive features of the term (definition, examples, etc.), administrative information (who created the entry, the knowledge field in which the term is used, the evolution of the term since it was entered into the database...), and so on. Moreover, you can find one or more `termNote` elements attached to any term, to indicate, for instance, the status of the term, or the normative authority, in the case the term is a standardized term; `termNote` elements will also be used to indicate the part-of-speech or other features of the term.

4. Conclusion

Terminology creation and spreading is so a hard and laborious task that it is supposed the process would considerably lighter, faster and become more productive when working collaboratively. The aim of TZOS, the system presented in this paper, is to facilitate the collaborative creation and spreading of terms. Although being a technical resource, it has been conceived as a support for working in a particular way. If it really will be fruitful, TZOS requires involvement and participation. We can call *terminology making 2.0* to this way of working, in the sense that it integrates Information and Communication Technologies into classical terminography.

The functional design of the system looks for active users and, in addition to consultation, TZOS provides mechanisms for proposing and updating terms. Even more, the system integrates a forum mechanism to promote debate on terminology. Regarding the representation of terminological information, TZOS is based on the TBX standard, and it is precisely the way proposed by TBX the approach adopted to fulfill our needs and requirements. Using an internationally accepted standard for the management and the exchange of information has many advantages.

The current version of TZOS is being used by the Basque Language Service trying to gather and identify real terminology used in different domains and colleges at the University of the Basque Country, UPV/EHU.

It must be said, however, that TZOS does not integrate corpus creation and management tools, not (semi-) automatic term extraction. TZOS does not, therefore, complete the full chain of terminology activities. In a near future it would be interesting to integrate other tools and resources into TZOS.

5. References

Pociello E., Gurrutxaga A., Agirre E., Aldezabal I. and Rigau G. 2008. “WNTERM: Combining the Basque WordNet and a Terminological Dictionary" . In *Proceedings of the 6th International Conference on Language Resources and Evaluations (LREC).* Marrakech.

Désilets A., Huberdeau L-P., Laporte M., Quirion J. 2009. Building a Collaborative Multilingual Terminology System. In *Proceedings of the Translating and the Computer 31 Conference*. London.

LISA Open Standards for Content/Container and Reuse (OSCAR) 2008. *Systems to manage terminology, knowledge and content - TermBase eXchange* (TBX).