

*Afruz M. Gurbanova*

DOI: 10.25045/jpis.v07.i1.08

Institute of Information Technology of ANAS, Baku, Azerbaijan

[afruz1961@gmail.com](mailto:afruz1961@gmail.com)

## COMPUTATIONAL TERMINOLOGY: STAGES OF FORMATION AND DEVELOPMENT TENDENCIES

*The article explores the history of formation and the current state of computational terminology. The development stages of methods of computational terminology and the process of management of computational terminology are described. The article explains the impact of technology on specific principles and techniques of computation terminology and the evolution of computation terminology techniques. Terminology management systems are investigated and the working principles are illustrated in the article. The conducted research in the field of computational terminology are explored and the development trends are specified as well.*

**Keywords:** *computation terminology, linguistic informatics, natural language processing, ontology, semantic relations, terminology databases.*

### Introduction

The application of computers in terminology management has a 50-year history. Starting from the 60's of the past century, computers have started to be applied in the processing, storage and collection of terminology data sets. After a while, mini and personal computers have started to be used for the management of terminology in specific subject matters and the technique of support of terminology activity has altered [1].

Nowadays, the Internet provides an environment for terminology research as well as an online-platform for international cooperation on terminology activity.

The management of terminology is determined as “the purposeful management of any terminology information” and carried out by the realization of the following actions while updating the published information:

- inclusion of new terms and notions;
- review of existing concepts (their integration or division);
- declaration of the obsolescence of terms.

Depending on how the terminology is published (traditional book, web-data base and etc.), problems on the realization of these works are increasing. The field of computation terminology studies the application issues of computation techniques during the conduction of terminology work. It is to be noted that the computer application and realization of terminology-related issues is relatively weaker, in comparison with other fields, due to different reasons.

The management of computational terminology is a developing field of science. The facilitation of this process is based on the application of computation techniques. The facilitation can be pursued by automatic (or semi-automatic) data processing or software introduction. This will allow the terminologists, researchers, engineers and others to manage the terminology information with more advanced and effective techniques. The computational terminology system is regarded as a new research trend in the crossroads of several scientific-research fields. The roots of this field belong to the terminology management. Several issues pertaining to the computation terminology management are mentioned below:

- terms extraction;
- extraction of notions;
- assistance to terminologist;
- terminology service issues (detection of updated terminology under new conditions in new documents);
- control over the content quality and etc.

The opportunities created by program solutions for a terminologist, especially the impact of technology in specific principles and techniques of the theory of terminology, and the evolution of techniques of terminology activity during this period are clarified in various issues of computation terminology management. The main focus is directed towards the international standards, the terminology data modelling in these standards and the facilities suggested to terminologists by the world wide web. Based on investigations and works by the founder of the theory of terminology E. Wüster [2], several work has been published in the 80's of the past century [3-8]. In these publications, the theoretical foundations, techniques and principles of terminology activity are specified. Moreover, the standardization of terminology, the practical transmission and application of the majority of principles described in terminological activity are critically assessed. This is related to the fact that traditional tools for terminological activity – file cards, terms list and special dictionaries did not allow terminologists to follow the principles of terminology in most cases. In the middle of the 60's of the past century, appropriate apparatus maintenance has been provided for the terminology management by computer technologies. The terminology processing was conducted with substantial efforts with the help of personal computer-based terminology management software.

Nowadays, several smart software tools exist, which supports specific principles of the theory of terminology and terminological activity. The application of such advanced terminology management systems altered the work principles of terminologists, and demonstrated prominent results for terminology theory.

### **The methods of terminological activity and computational terminology management**

The theory of terminology, the methods of terminological activity and the historical evolution of computational terminology management are chronologically described in the following tables (Table 1, Table 2).

Table 1  
The evolution of scientific theory of terminology in software tools

Until 1965	The scientific theory of terminology was established through use of software tools
1965 – 1975	Discussion and expansion of theoretical foundations; various approaches; computer programs oriented to research and development were absent.
1975 – 1985	Early articles regarding the design of terminology data bank (Termbank), the field of application, paper templates for writing the terminology data and the clarification of specific aspects (data extraction, published dictionaries, data exchange) of terminology bank
1985 – 1995	Inclusion of topics about the management of Termbank and terminology with the help of computer programs; the development of early tasks and recommendations
After 1995	Scientific discussion of (application of) new techniques of computational terminology management; publications and standards

Table 2

The evolution of terminological activity and computational terminology management

Till 1965	Realization of terminological activity related to card files, dictionaries and terms list only
1965 – 1975	Development of the main carcass of Termbank (Termium, Eurodicautom); first templates for registration (paper templates); the list of published terminology; the development of inclusion and search of terminology data
1975 – 1985	Terminological software tools in mini-computers (for instance, Ericsson Cat and etc.); direct access to terminology; the access to home terminal of the main carcass of termbank; online data insertion and search
1985 – 1995	Software tools for terminology management for separate users in personal computer base (for example, Termex-MTX); the formation of working group related to the problems of unification and integration of terms; modules and procedures for data exchange; repeated use of data
After 1995	Development of high-quality and powerful terminology management systems (for ex. <i>MultiTerm</i> (translation program)); joint terminological work on the network ( <i>LAN</i> and the world wide web); integration to other tools (for ex. translator device); development of auxiliary programs for the support of terminological work (for example, a software tool for term extraction)

The design and architecture of early program tools for terminology management did not suffice for the establishment of terminology collection. Although the founders of Termbank attempted to realize the principles of the concept in the first place, the practice showed that several terminology articles representing this concept are stored in data bank. The reason was the absence of the mechanism checking the homonyms and the procedure of checking the doublets (two-fold writings) during the insertion of new terms. First programs have been developed for term management in personal computers' base towards the solution of such problems. The majority of these programs allowed to keep one term in each language; synonyms, however, had to be written in a separate writing.

Terminology data bank of the European Commission Eurodicautom was established in 1975 and mainly considered for the use of translators and other employees of the Commission. Eurodicautom, established in six languages, is expanded, and currently operates in seven, nine or eventually, eleven languages.

Three main groups of computation terminology management can be distinguished:

1. terminology creation;
2. terminology support;
3. terminology storage.

The terminology management system is a specially developed software tool and considered for the use by translators, terminologists and other users. Nowadays, almost all modern terminology management systems (*TMS*) are capable to create and support terminological articles and store terminological information. These systems also allow for the realization of conceptual hierarchical links (superior, subordinate and peer notion) from the linking mechanism among various writings:

*Acrolinx* – is a platform of content optimization, oriented towards the linguistics problems and introduces a sentence in an improved form.

*AcrossTerm* (Web) – *Across* Language Network carries out a central function for all terminology companies, and also serves to the translation of those terms.

*Anylexic* – is a new software tool of TMS not related to a particular terminology. It assists the process of terminology translation at each stage – creation, modification, search and the exchange of terminology.

*ApSICXbench* – maintains the quality, and simple and powerful management of terminology in a single package. With this software tool, the download of any file of several formats, supported by *CAT*, can be carried out and the high level of quality is obtained.

*Interplex* – is a glossary for translators; a software tool for *Ipad* and *Iphone*.

*i-Term* – is a tool for management of the state of terminology and knowledge in art, it allows to reserve the structure of notion-related knowledge and search on the Internet.

*qTerm*<sup>TM</sup> – is a software tool for terminology management. It detects and translates critical terminology. Moreover, it is the newest software tool, which explains in detail the use of each term, as well as the language and the history of its use.

*LogiTerm Web* – has an interface providing the access to four terminology databases, *Bitexts*, the whole text and serial number. With the help of this software tool, a terminological writing can be created for the terminology database, modified and reviewed in web-interface or *Microsoft Word*.

*Multiterm* – is a program facilitating the creation of terminological bank for SDL Trados Studio. In this program, the terminological bases are based not on words, but on notions.

Alongside with the above mentioned tools, *Multitrans Prism*, *TermStar*, *Termbases*, *TermWeb*, *TermWikiPro*, *Taas - Cloud Services for Terminology Work*, *TippyTerm*, *TManager* and etc. terminology management systems have been developed [9].

Three international standards are prepared regarding the development, application and operation of terminology management systems. *ISO 12620* (1999) [10], *ISO 12200* (1999) [11] and *ISO 16642* (2003) [12] standards, developed by *ISO 37 TC* technical committee of the International Standardization Committee, specify more than 200 categories of standard data. These standards serve to the preparation of proposals with the purpose of selection of data categories for specific construction of terminology management systems, and suggest a terminological metamodel specifying the base architecture of a terminology management system.

Hence, computation terminology bears a great importance in text mining, information search and acquisition, request systems, the ontology construction and other fields. Text mining and the extraction of new knowledge are best reflected in the terms.

### **Development trends of computational terminology**

The results of scientific investigations, carried out in various spheres of computational terminology, are discussed in international conferences, symposiums and scientific workshops held in this field.

Starting from 1965, “*Coling*” (*Computational linguistics*) international conference on computational linguistics takes place every two years. First conference was organized in New-York (USA). The 25<sup>th</sup> “*Coling*” International conference was held in Dublin in 2014 [13].

Since 2000, “*CICLing*” (*Conference on Intelligent text processing and Computational Linguistics*) international conferences takes place on text mining and computational linguistics. The 15<sup>th</sup> international conference was held in Kathmandu (Nepal) in 2014.

“*Computational terminology (CompuTerm)*” international scientific workshops are organized regarding the various problems and issues of computational terminology.

First International Workshop “*CompuTerm*” was held in Montreal University of Canada in 1998 [15], second International Workshop was held in University of Pennsylvania (USA) in 2002 [16], third International Workshop in Geneva (Switzerland) in 2004 [17], and fourth International workshop in Dublin (Ireland) in 2014 [18].

By gathering together in international workshops on computational terminology, the experts of this field have discussed the current state of computational terminology and its impact on

applications of natural language processing (NLP); the following development trends have been specified:

- automatic term extraction (*ATE*), recognition and filtration constituting the basis of terminological activity. Automatic term extraction is carried out by using computation techniques. The purpose is the acquisition of potential knowledge for future processing. The techniques developed for automatic term extraction must be realized in terminological activity and in systems assisting the terminology management;

- recognition of the events expanding the notion of terminological integrity from terms bearing its static single meaning to the terms bearing the meaning of a procedure and dynamic process;

- specification of semantic relations among terms – this is applied in updating the existing knowledge base;

- management of variety of terms – this facilitates the work with dynamic nature of the term, its acquisition from various sources, integration, standardization and the introduction in wide-spectrum applications and resources; thus, this problem is encountered while working with various dictionaries, thesauruses, ontologies and text information;

- creation of description encompassing important investigations and oriented to the provision of accurate and univocal definition of terminology object; this definition includes important elements necessary for the formal description of terms and notions in ontology;

- discussion of user expertise which is a new problem in terminological activity. Specific fields here encompass the notions and terms usually not familiar for non-professionals (for example, a patient or bank customer).

- systematic management of terminology and the update of thesauruses and dictionaries of specific fields, which is an important aspect for supporting the existing terminological resources. The number of existing terminological resources are constantly increasing; their continuous update is necessary for the effective use.

- creation of monolingual and multilingual resources – it facilitates the inter-lingual and multilingual applications, and requires specific methods and tools, design and assessment, which is a complex issue;

- reliability and portability of techniques – this allows to apply the techniques used in a specific context to another context and the division of research expertise among those;

- processing of social networks and contemporary media;

- use of terminology in various NLP applications;

- Experiments on the assessment of terminological techniques and tools and etc.

The development trends of computational terminology encompass important issues and reveal the necessity of conducting serious scientific research in this direction.

## Conclusion

The investigation of the current state, formation and management stages of computational terminology, international standards prepared in this direction, the analysis of scientific-theoretical research carried out in the field of computation terminology shed light on the development perspectives of terminological activity and necessitate the creation and development of terminological knowledge base.

Terminology database – is a database storing key terms as data, which are translated described in a specific definition within the framework of the whole document (project). Each term in terminology database is provided with metadata (additional information about the term). The meaning of the term, equivalent in other languages, short form, synonyms, application fields and the information about the document verifying the term (for example, national and international standards) are provided.

Along with its practical importance, the terminology database is a powerful tool for conducting serious scientific investigations.

*This work was carried out with financial assistance of Science Development Foundation under the President of the Republic of Azerbaijan – Grant № EIF-2014-9(24)-KETPL-14/02/1*

## References

1. Klaus-Dirk Schmitz. Developments in computational terminology management and its influence on terminology science and terminology work. ELETO 5th Conference "Greek Language and Terminology", Nicosia, Cyprus, 13-15 November, 2005, <http://www.eleto.gr>
2. Wüster E., Baum R. Einführung in die allgemeine Terminologielehre und terminologische Lexikographie (the general introduction to terminology and terminological lexicography lesson). 3rd Edition, Romanistischer Verlag, Bonn 1991, XXI, 239 p.
3. Helmut F. Terminology Manual. Paris, 1984, <http://www.unesdoc.unesco.org>
4. Dubuc R. Manuel pratique de terminologie (1er édition). Quebec, 1985, 158 p.
5. Heribert Picht, Jennifer Draskau. Terminology - An Introduction. Guildford, 1985, 265 p.
6. Reiner Arntz, Heribert Picht. Einführung in die Terminologiearbeit (1. Auflage). Hildesheim, 1989, 344 p.
7. Helmut Felber, Gerhard Budin. Terminologie in Theorie und Praxis. Tübingen, 1989, 315 p.
8. Juan C. Sager. A Practical Course in Terminology Processing. Amsterdam/Philadelphia, 1990, 258 p.
9. Terminology Management Systems, <http://www.recremisi.blogspot.com>
10. ISO 12200 (1999): Computer applications in terminology – Machine-readable interchange format (MARTIF) – Negotiated interchange. Geneva: International Standards Organization.
11. ISO 12620 (1999): Computer applications in terminology – Data categories. Geneva: International Standards Organization, <http://www.semanticweb.kaist.ac.kr>
12. ISO 16642 (2003): Computer applications in terminology – Terminology markup framework (TMF). Geneva: International Standards Organization, <http://www.tbxinfo.net>
13. COLING 2014, the 25th International Conference on Computational Linguistics, Helix Conference Centre at Dublin City University (DCU), Ireland, 23-29 August 2014, <http://www.coling-2014.org>
14. 15<sup>th</sup> International Conference on Intelligent Text Processing and Computational Linguistics, April 6–12, 2014, Kathmandu, Nepal, <http://www.cicling.org>
15. First Workshop on Computational Terminology, University of Montreal (Canada), 1998, <http://www.perso.limsi.fr>
16. Second International Workshop on Computational Terminology, Association for Computational Linguistics, Stroudsburg, PA, <http://www.dl.acm.org>
17. CompuTerm 2004 - 3rd International Workshop on Computational Terminology, Proceedings of the Workshop, 29 August 2004, Geneva, Switzerland, <http://www.acl.ldc.upenn.edu>
18. 4th International Workshop on Computational Terminology, COLING 2014 Workshop, 23rd August 2014, Dublin, Ireland, <http://www.perso.limsi.fr>