

## **24. Developments in computational terminology management and its influence on terminology science and terminology work**

**Klaus-Dirk Schmitz**

### **SUMMARY**

Using computers for terminology management has an almost 50 years old history. At the beginning of the sixties of the last century main frame computers were programmed to elaborate, maintain and retrieve terminological data collections. Later mini computers and (networking) personal computers were used to manage terminology for certain subject fields and the way of supporting terminological activities changed. Today the internet provides not only a medium for terminological research but also an online platform for world-wide cooperative terminology work.

The paper describes the different stages of computational terminology management with special attention to the possibilities how the software solutions can support (and impede) the terminologist. In particular it is explained in detail how specific principles and methods of terminology theory were influenced by the technology and how terminological working methods have changed over the time. Special attention is drawn to ISO standards for terminological data modelling and the possibilities the world wide web offers to terminologists.

### **Εξελίξεις της υπολογιστικής διαχείρισης ορολογίας και επίδρασή της στην επιστήμη Ορολογίας και στην ορολογική εργασία**

**Klaus-Dirk Schmitz**

### **ΠΕΡΙΛΗΨΗ**

Η χρήση των υπολογιστών με σκοπό τη διαχείριση της ορολογίας έχει ιστορία πενήντα χρόνων περίπου. Στην αρχή της δεκαετίας του 60 του περασμένου αιώνα, οι κεντρικοί υπολογιστές είχαν προγραμματιστεί για να εκπονούν, να διατηρούν και να αντλούν συλλογές ορολογικών δεδομένων. Αργότερα, οι μικρο-υπολογιστές και οι προσωπικοί υπολογιστές χρησιμοποιήθηκαν για την ορολογική διαχείριση μερικών θεματικών πεδίων και ο τρόπος υποστήριξης των ορολογικών δραστηριοτήτων άλλαξε. Σήμερα, το Ίντερνετ αποτελεί όχι μόνο μέσο για την ορολογική έρευνα αλλά και μια επιγραμματική πλατφόρμα για παγκόσμια συνεργατική ορολογική εργασία.

Η ανακοίνωση περιγράφει τα διαφορετικά στάδια της διαχείρισης της υπολογιστικής ορολογίας εστιάζοντας στον τρόπο που οι λύσεις λογισμικού μπορούν να υποστηρίξουν (ή να εμποδίσουν) τον ορολόγο. Πιο συγκεκριμένα, επεξηγεί με λεπτομέρεια πώς συγκεκριμένες αρχές και μέθοδοι της θεωρίας της ορολογίας δέχθηκαν επιδράσεις από την τεχνολογία και πώς οι ορολογικές μέθοδοι εργασίας άλλαξαν με την πάροδο του χρόνου. Τέλος, δίνει ιδιαίτερη προσοχή στα πρότυπα ISO που αφορούν την μοντελοποίηση ορολογικών δεδομένων καθώς και στις δυνατότητες που προσφέρει ο παγκόσμιος ιστός στους ορολόγους.

## 0 INTRODUCTION

Based on the fundamental research and publications of Eugen Wüster (see e.g. Wüster 1991, Picht/Schmitz 2001), several definitive works have been published in the eighties of the last century. These (text) books deal with the theoretical foundations of terminology science as a scientific discipline as well as the methods and principles of terminology work (Felber 1984, Dubuc 1985, Picht/Draskau 1985, Arntz/Picht 1989, Felber/Budin 1989, Sager 1990). Besides the field of terminology standardisation, the practical transfer and realisation of most of the described principles of terminology work have been discussed and judged very critically. One of the reasons for this is that traditional tools for terminology work such as file cards, glossaries, term lists and special dictionaries in most cases do not allow the terminologist to follow the terminology principles described. Since the middle of the nineteen hundred sixties computer technology has provided adequate hardware for terminology management. The first main frame databases and PC-based terminology management programs only allowed with major difficulties to develop terminology by a concept-oriented approach. But today we find a number of reasonable computer-based tools supporting the defined principles of terminology science and terminology work. Using such modern terminology management systems have not only changed the working methods of terminologists but also have had repercussions to the theory of terminology science.

## 1 Historical evolution

The historical evolution of terminology theory, working methods and computational terminology management will be shown by the following chronologically ordered tables.

### 1.1 Theory of terminology science with a link to computer applications

before 1965:	theoretical foundation established
1965-1975:	discussion and extension of the theoretical basis; different approaches; no computer-oriented research and development
1975-1985:	first publications about the design of termbanks, used fields, paper templates for recording terminological data and specific aspects of term bases (data retrieval, printing dictionaries, data exchange)

1985-1995:	textbooks contain chapters about termbanks and computer-assisted terminology management; first guidelines and recommendations
1995-today:	(application-oriented) scientific discussion of new methods for computational terminology management; publications and standards

## 1.2 Terminology work and computational terminology management

before 1965:	terminology work only with file cards, glossaries and term lists
1965-1975:	main frame termbanks (Team, Termium, Eurodicautom); first (paper) templates for recording; printed terminology lists; terminology data input and retrieval by data typists and not terminologists
1975-1985:	terminology software on mini computers (e.g. Ericsson Cat); direct access for terminologists; main frame termbanks accessible by in-house terminals; on-line data input and retrieval
1985-1995:	simple PC-based terminology management programs for single users (e.g. Termex-MTX); direct access for terminologist; problems of harmonisation and integration for working groups; modules and procedures for data exchange, data consolidation and reuse
1995-today:	high-quality and powerful terminology management systems (e.g. MultiTerm); cooperative terminology work via networks (LAN and world wide web); client-server-architecture; integration into other tools (e.g. translator workbenches); additional programs supporting terminology work (e.g. term extraction software)

## 2 Change of methods and principles for terminology management

### 2.1 Concept orientation

The design and architecture of the first computer-based tools for terminology management was not really adequate to create concept-oriented terminology collections. Although the

developers of the first main frame termbanks attempted to realize the principle of concept orientation, practice has demonstrated that the database very often contained several terminological entries representing the same concept. The reason was the lack of a homonym controlling mechanism when entering new terms and the lack of a validation routine for so called doublets (double entries). Similar problems came up with the first PC-based terminology programs that were too much designed for managing terms (instead of concepts). Most of them just allowed to store one term in each language per concept / entry, and synonyms had to be recorded in separate entries.

Today, almost all modern terminology management systems are able to create and maintain concept-oriented terminological entries and to store all terminological information belonging to one single concept within one single entry. They also allow to adequately realize hierarchical concept relations (superordinate, subordinate and related concept) by a linking mechanism between different entries.

## **2.2 Treatment and handling of terms**

Old fashioned terminology systems were based on a data model allowing only restricted searching and sorting capabilities. Therefore it was necessary to enter multi-word terms in an inverted order, e.g. "*interface, serial*" instead of "*serial interface*". In some systems an additional field (e.g. keyword) was implemented to support correct sorting and retrieving of entries. Other systems generated auxiliary entries with all possible sequences of word orders, but both solutions led in the long run to inconsistent data collections if for example the term of the main entry was changed or deleted. Modern terminology management systems provide demanding search functions with all possibilities of truncation, wild cards and fuzzy search algorithms allowing to enter a multi-word term in its natural word sequence. This supports to take over terms into other applications (e.g. into a word processor file without changing the term) as well as to automatically detect terms within a text and to align it to the corresponding term in the database.

## **2.3 Term autonomy**

The data model of older termbanks very often supplies only one field or data category per language for the (main) term; sometimes we find one additional field for synonyms. But if there occur several terms or synonyms for the same concept, they are store all together in

one field, separated e.g. by a semicolon. Such an implementation does not only complicate the searching and searching facility of the database software, it also prevents the necessary documentation of each of the terms by term-related data categories such as *grammatical gender*, *part of speech*, *geographical usage* or *context example*. Most of the modern terminology management systems support the principle of term autonomy allowing to repeat the data category *term* with all dependant data categories (*grammatical gender*, *context example* etc.) as much as it is necessary. By this, all terms representing a given concept can be managed, classified and documented in an adequate and reasonable way.

#### **2.4 Terminological data modelling**

Introductory text books, recommendations and other publications dealing with terminological working methods or describing existing terminology databanks do not treat the topic of (computational) terminology management in an adequate and comprehensive way, if they are published several years ago. If at all, only few data categories are mentioned that are important for the management of terminological data collections. Only in rare cases we find additional information about the possible repeatability of data categories and proposals for a (paper-based) input template for terminological data. Newer publications demonstrate different possibilities to define and select data categories for terminological databases as well as proposals for modelling the terminological entry within a computer-based implementation of a terminology management system (see e.g. Schmitz 1996, 1998, 2001a, 2001b and 2004b, Wright/Budin 2001). The works mentioned discuss aspects of terminological data modelling for different user groups and applications scenarios, taking into account principles of granularity, modelling variations, term autonomy and concept orientation. They do not propose a definite fixed data model like older publications, but they support and guide all persons involved in the design, implementation and operation of a computer-based terminology management system to find an adequate solutions for their own specific needs and requirements. The same objective have three ISO standards: ISO 12620 (1999) lists and defines more than 200 data categories that could be serve as a repertoire and proposal for selecting data categories for the own terminology management system design. ISO 12200 (1999) and ISO 16642 (2003) provide not only interchange formats or terminological mark-up frameworks but also a terminological meta model that defines the basic architecture of a terminology management system.

#### 4 Bibliography

[1] Arntz, Reiner; Picht, Heribert (1989): *Einführung in die Terminologearbeit* (1. Auflage). Hildesheim.

[2] Dubuc, Robert (1985): *Manuel pratique de terminologie* (1<sup>er</sup> édition). Quebec.

[3] Felber, Helmut (1984): *Terminology Manual*. Paris.

[4] Felber, Helmut; Budin, Gerhard (1989): *Terminologie in Theorie und Praxis*. Tübingen.

[6] ISO 12200 (1999): *Computer applications in terminology – Machine-readable interchange format (MARTIF) – Negotiated interchange*. Geneva.

[7] ISO 12620 (1999): *Computer applications in terminology – Data categories*. Geneva.

[8] ISO 16642 (2003): *Computer applications in terminology – Terminology markup framework (TMF)*. Geneva.

[9] Picht, Heribert; Draskau, Jennifer (1985): *Terminology - An Introduction*. Guildford.

[10] Picht, Heribert; Schmitz, Klaus-Dirk (eds.) (2001): *Terminologie und Wissensordnung - Ausgewählte Schriften aus dem Gesamtwerk von Eugen Wüster*. Wien.

[11] Sager, Juan C. (1990): *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia.

[12] Schmitz, Klaus-Dirk (1996): „Terminology Management Systems“, in: Owens, Rachel (ed.): *The Translator's Handbook, Third Edition*. London, 221-246.

[13] Schmitz, Klaus-Dirk (1998): „Über wichtige Aspekte bei der Einrichtung einer rechnergestützten Terminologieverwaltung“, in: Lundquist, Lita; Picht, Heribert; Qvistgaard, Jacques (edd.): *LSP Identity and Interface, Research, Knowledge and Society, Proceedings of the 11th European LSP Symposium on Languages for Special Purpose, Copenhagen, August 1997*. Copenhagen, 391-398.

[14] Schmitz, Klaus-Dirk (1999): „Computergestützte Terminographie: Systeme und Anwendungen“, in: Hoffmann, Lothar; Kalverkämper, Hartwig; Wiegand, Herbert Ernst

(Hrsg.): Fachsprachen / Languages for Special Purposes – ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. 2. Halbband. Berlin/New York, 2164-2170.

[15] Schmitz, Klaus-Dirk (2001a): „Systeme zur Terminologieverwaltung“, in: Technische Kommunikation, No. 2/2001, 34-39.

[16] Schmitz, Klaus-Dirk (2001b): “Criteria for evaluating terminology database management programs”, in: Wright/Budin (eds.), 539-551.

[17] Schmitz, Klaus-Dirk (2004a): „Terminologearbeit, Terminologieverwaltung und Terminographie“, in: Knapp, Karlfried et al. (eds.): *Angewandte Linguistik. Ein Lehrbuch*. Tübingen, 435-456.

[18] Schmitz, Klaus-Dirk (2004b): „Die neuen Terminologiedatenbanken: online statt offline“, in: Mayer, Felix; Schmitz, Klaus-Dirk; Zeumer, Jutta (Hrsg.): *Terminologie und Wissensmanagement*. Akten des Symposions, Köln, 26.-27. März 2004. Köln, 180-189.

[19] Wright, Sue Ellen; Budin, Gerhard (eds.) (1997): *Handbook of Terminology Management. Volume 1*. Amsterdam/Philadelphia.

[20] Wright, Sue Ellen; Budin, Gerhard (eds.) (2001): *Handbook of Terminology Management. Volume 2*. Amsterdam/Philadelphia.

[21] Wüster, Eugen (1991): *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie* (3. Auflage). Bonn.

Prof. Dr. Klaus-Dirk Schmitz  
Full professor for Terminology Science  
University of Applied Sciences Cologne  
Mainzer Str. 5, D-50678 Köln, Germany  
Tel.: +49-221-8275-3272, Fax: +49-221-8275-3991  
Mail: [klaus.schmitz@fh-koeln.de](mailto:klaus.schmitz@fh-koeln.de)