Management and exploitation of conceptual data and information in technical termbases: the electrotechnical vocabulary

Laura Giacomini

Department of Translation and Interpreting, University of Heidelberg, Plöck 57a, 69117 Heidelberg (Germany)
E-mail: laura.giacomini@iued.uni-heidelberg.de

Abstract

This paper addresses the lexicographic challenges related to the management and exploitation of conceptual data and information by examining the example of electrotechnical vocabulary. Four online tools with different source, typology and reference language will be presented and compared from the point of view of the user's needs. By focusing first on the conceptualization level of the underlying database and then taking into account how this interfaces with the terminological component, the paper will progressively provide specific insights into data availability, ease of access and consistency, and will hint at possible ways to improve conceptual representation in LSP e-lexicography.

Keywords: e-lexicography; term; concept; termbase; technical domain

1. Introduction

In order to evaluate the potential of e-lexicographic tools concerning the quality of data representation for the end user, usability tests are required to highlight the level of effectiveness, efficiency and user satisfaction that a specific tool can achieve (Heid, 2012; Giacomini, 2014). In particular, a satisfactory level of effectiveness, i.e. degree of task completion, and efficiency, i.e. the amount of time needed to perform a task, largely rely upon formal and content-related coherence of the underlying termbases.

With reference to current database and knowledge management theories (Alwert & Hoffmann, 2003; Halpin & Morgan, 2010; Pratt & Adamski, 2011), data are defined as raw lexical and conceptual items, which can be classified, condensed and contextualized to obtain conceptual and terminological information. This implies that two dichotomies need to be taken into account at the same time in this study: on the one hand, the dichotomy between conceptual and terminological items, and on the other hand, that between cognitively unprocessed data and information conveyed by data during consultation.

Starting from a quite comprehensive definition of e-lexicographic tools as information tools of a lexicographic kind (Leroyer, 2012), which can be referred to as, for instance,

dictionaries, glossaries or wiki tools, this paper addresses the challenges related to the treatment of conceptual data and information in terminology databases that serve as a lexicographic basis. The final goal is to explore the extent to which structured and consistent management of conceptual items goes hand in hand with their direct exploitation by dictionary users, which results in increased effectiveness and efficiency of the tool. The paper aims to illustrate this topic through an examination of electrotechnical vocabulary. Section 2 describes the set of example resources that have been taken into consideration in this study, the ideal user that is addressed in the analysis and the method employed in the study. A comparative analysis of the different resources and its results are presented in Section 3, while Section 4 contains some final observations.

2. Representative tools, the applied method and the addressed user

Online resources with different distributions of source and reference language have been selected to make a comparison from the point of view of a user's needs (Tarp, 2008; Koplenig, 2011). This selection is not intended to be exhaustive and should be seen as a way of exemplifying the procedure and drawing first conclusions on the correlation between management and exploitation of conceptual items from a lexicographic point of view. The representativeness of these tools lies in the fact that they exhibit some of the most widespread LSP e-lexicographic structures and cover the prevalent types of sources consulted by translators as the user group addressed in this study. By considering the Function Theory of lexicography (cf. Tarp, 2008) as the theoretical basis of this analysis, the ideal target user group of these lexicographic resources has been, in fact, identified as professional translators performing a passive translation task or producing a specialised text in their native language (Mayer, 1998). The concrete usage situations are primarily of a communicative kind, but consultation for cognitive purposes, i.e. for knowledge acquisition (Tarp, 2008), is also contemplated, especially in the case of monolingual tools. Table 1 illustrates the combination of source and languages in each of the tools. Specific content-related and formal features on the macrostructural and microstructural level will be introduced and discussed in the next section.

TOOL	SOURCE	LANGUAGE(S)
International Electrotechnical Vocabulary (IEV, or Electropedia) and IEC Glossary	standardization organization	multilingual
IATE database (Electronics and electrical engineering section)	institution: EU	multilingual
Open Energy Information Glossary (OpenEI)	open source wiki	English
Electrical Glossary (Fluke Electronics)	company-internal terminology	English

Table 1: The selected tools

This choice allows for a broad assessment of knowledge representation and of the extent to which its formalisms affect the consultation performance in terms of task completion and time investment (cf. Schärfe et al., 2006).

In order to reach this goal a three-step procedure has been applied, which will now be introduced. By focusing first on the conceptualization level and then taking into account how this interfaces with the terminological component, the paper will progressively provide specific insights into

a) conceptual structure availability (presence and depth/granularity of conceptual networks, e.g. the one including ELECTRIC CURRENT, ALTERNATING CURRENT, DIRECT CURRENT, etc.)[1],
b) ease of access (degree of transparency of term-related concepts, e.g. to what extent the user can retrieve, view and consult the conceptual network of ELECTRIC CURRENT by directly accessing the conceptual layer of the database or while performing a term search) and
c) consistency (regular and logical correspondence between concepts and terminological designations, e.g. between ELECTRIC CURRENT and related simple terms, multiword terms, abbreviations, acronyms and their variants: *ampere*, *ampère*, *amp*, *A*, *ampere-hour*, *ampere-hour meter*, etc.).

3. Analysing the management and exploitation of conceptual data in the selected resources

3.1 Conceptual structure availability and properties

The availability of conceptual structures has been assessed on the grounds of the macrostructural properties of the selected tools and is summarised in Table 2, with special focus on the example of the concept alternating current. Not all principles of terminology management proposed by the German terminology association DTT (2014) can be properly evaluated by taking into consideration the only surface features of the lexicographic resources. This paper concentrates on the depth of the conceptual structures, the available relations involving the lower conceptual level (the bottom level, which can be taken into account independently of the typology of the superordinate structures) and the presence of conceptual networks as the criteria that apply particularly to the treatment of the conceptual layer. An important, initial assumption is that correlations between concepts and terms can be multivocal in both directions: a concept may be verbalized by means of more than one term, and a term may designate more than one concept.

[1] Concepts are written in small caps, terms in italics.

TOOL	DEPTH OF CONCEPTUAL STRUCTURES	CONCEPTUAL RELATIONS AT THE LOWER LEVEL
IEV Online	subject area > section > specific concept / term: e.g. CIRCUIT THEORY > GENERAL > ALTERNATING CURRENT / *alternating current*	a) multivocal lower – higher level: e.g. ALTERNATING CURRENT < CIRCUIT THEORY / ROTATING MACHINERY / INDUSTRIAL ELECTROHEAT / … b) no relation lower – lower level: e.g. ALTERNATING CURRENT ? DIRECT CURRENT
IATE	domain > specific concept / term: e.g. ELECTRONICS AND ELECTRICAL ENGINEERING > ALTERNATING CURRENT / *alternating current*	a) multivocal lower – higher level: e.g. ALTERNATING CURRENT < ELECTRONICS AND ELECTRICAL ENGINEERING / ELECTRICAL INDUSTRY / TOWN PLANNING b) multivocal lower – lower level: e.g. ALTERNATING CURRENT – DIRECT CURRENT (antonym) / PULSATING CURRENT (related)
OpenEI / Fluke Glossary	no conceptual structure: e.g. Ø > ALTERNATING CURRENT / *alternating current*	a) no relation lower – higher level b) univocal lower – lower level: e.g. ACTIVE POWER – AMPÈRE-HOUR (Fluke Glossary)

Table 2: Conceptual structures availability and properties

IEV Online, or Electropedia, is an electrical and electronic terminology database comprising around 20,000 terms. Created by the standardization organization IEC (International Electrotechnical Commission), it has definitions provided in English and French, and equivalents in several other languages. From a macrostructural viewpoint, IEV Online can be classified as a resource with a complex, not fully developed organizational system and a primarily systematic arrangement. The prevalent conceptual criterion in IEV Online is a classification in which the main subject areas of the electrotechnical field are recorded and further subdivided into more specific sections, eventually leading to final-level concepts and terms. The tool displays a multivocal directionality of conceptual relations: a final-level concept may be attributed to more than one superordinate section or subject area, e.g. ALTERNATING CURRENT can refer to six different subject areas, taking the form of 11 different terminological entries, among which *alternating current machine* or *capacitor fed alternating current track circuit* (the same term is never recorded under more than one category). However, despite the clear hierarchical categorization, no definite relations (e.g. co-hyponymy) can be identified among the large number of final-level items. For instance, in the structure Area: CIRCUIT THEORY > Section: GENERAL no relation can be established between CIRCUIT : ELECTRIC CIRCUIT : MAGNETIC CIRCUIT : … CIRCUIT or between DIRECT CURRENT : ALTERNATING CURRENT : ACTIVE

CURRENT : INDUCTIVE CURRENT : … CURRENT. This aspect can be traced back to insufficient granularity in the lower level of the conceptual structure.

The second resource, the multilingual IATE database, is a comprehensive institutional resource recording terms from a broad range of disciplines, including an electronics and electrical engineering section (domain no. 6826). This domain directly includes a number of concepts/terms with no further groupings (further categories can only be found in the external EuroVoc at http://eurovoc.europa.eu), and, like in the case of IEV Online, final-level items may belong to more than one domain (multivocal relations). Different from the previous resource, the IATE database contains relations between items at the lower conceptual level and labels them accordingly (e.g. antonyms). Although its macrostructure can be defined as fully systematic, IATE's degree of granularity and its conceptual development are clearly unsatisfactory, and no conceptual network is available. The fact that the database covers several different domains may be one of the main causes.

The Open Energy Information Glossary (OpenEI) is an open source monolingual wiki that records data related to the topic of energy. This resource shows a simple and form-determined (i.e. alphabetical) macrostructure and avoids a conceptual organizational system, so that final-level conceptual relations are also not available. If relevant, concepts/terms only seem to be hypertextually linked to each other by means of entry-internal, non-systematic lists labelled "Related Terms" (univocal relations). The same macrostructural type and an analogous kind of approach to conceptually related items can be found in the last resource, the monolingual Electrical Glossary provided by Fluke Electronics, an example of a lexicographic resource reflecting a company-specific view of the domain and its terminology.

3.2 Ease of access to conceptual data

As it has been observed by Giacomini (2015), macrostructural features of LSP e-lexicographic tools, in particular the presence of systematic relations among concepts/terms, may be generally less discernible to the user, since they can often be only partially noticed during consultation. In this section, the actual access to conceptual relations via the microstructure and/or the conceptual structure will be taken into consideration (Table 3).

It has emerged in the previous section that none of the tools contain a structured ontology but, at the most, a conceptual structure based on a closed set of subjects. This structure is available in IEV Online and IATE. The former allows for an external access to its subject areas, which are listed in a hierarchical structure linked to further category and detail pages. The user can choose between performing a term search via a search mask and browsing the available subject areas. In the second case, the search for a specific term may require a longer amount of time, unless the user is already familiar with the recorded subdisciplines and also has an operational knowledge of the

previously mentioned categorization criteria. On the one hand, a term is never recorded under more than one category, whereas combinations of a term may appear under different categorical labels (*alternating current* itself only belongs to the general section of the CIRCUIT THEORY area); on the other hand a concept may be related to different areas (ALTERNATING CURRENT is related to CIRCUIT THEORY / ROTATING MACHINERY / GENERATION, TRANSMISSION AND DISTRIBUTION OF ELECTRICITY / SWITCHING AND SIGNALLING IN TELECOMMUNICATIONS / SIGNALLING AND SECURITY APPARATUS FOR RAILWAYS / INDUSTRIAL ELECTROHEAT, cf. Table 2).

TOOL	ACCESS TO CONCEPTUAL RELATIONS VIA THE CONCEPTUAL STRUCTURE	ACCESS TO CONCEPTUAL RELATIONS VIA THE MICROSTRUCTURE
IEV Online	direct access	direct access; non-specified relations; totally accessible; systematic; hyperlinked
IATE	no access, only filtering function	direct access; specified relations; totally accessible; systematic; not hyperlinked
OpenEI	∅	indirect access via related terms; totally accessible; systematic; hyperlinked
Fluke Glossary	∅	indirect access via related terms; partially accessible; non-systematic; not hyperlinked

Table 3: Access to conceptual data

IATE's users can only employ available domain categories as search filters during term search and cannot directly consult these categories. This results in a necessity for the user to perform rather specific queries and the impossibility to retrieve all terms belonging to the same domain. Moreover, besides the absence of a subdomain categorization (cf. Table 2), only one domain can be selected as a filter at once, which makes it quite difficult for the user to identify terminological cross-references between different disciplines. In comparison to IEV Online, IATE does not clearly highlight the terms containing the search term itself, so the user is often compelled to analyse long lists of search results in order to look for relevant conceptual/terminological combinations. The search for *alternating current*, for instance, produces, among others, results such as *alternating current generation system*, *single-phase alternating current*, *indirect alternating current converter*, *alternating current supply,* etc., which are indistinctly put together.

From the microstructural perspective, the first two resources share another important feature, which is direct access to conceptual relations: the IEV Online microstructure offers non-specified relations, whereas the IATE entries name the multivocal lower–lower level relations already mentioned in Table 2, even though this does not

seem to happen systematically. These kinds of properties are also present in the other two resources, but, as they do not rely on an underlying conceptual structure, they are far less developed. For this reason, the users of OpenEI and the Fluke Glossary can only retrieve indirect information regarding conceptual relations through article-internal cross-references to other terms. Table 3 highlights the following characteristics of the microstructural access to conceptual relations: availability of a direct vs. indirect access (i.e. access to conceptual items vs. access via terminological items), total/partial access, access systematicity, presence of specified relations (i.e. relations which have been attributed a type), access through hyperlinked data. The results show different possible combinations of these characteristics, which can be summarized and evaluated in the following categorisation proposal:

1) access via the conceptual structure (if such structure is available):
1.1) direct access (direct access enables the user to retrieve information concerning the conceptual relations independently of the consultation of the terminological layer, and can thus actively support consultation for cognitive purposes)
1.2) no access
2) access via the microstructure:
2.1) type of access:
2.1.1) no access
2.1.2) direct access (the user can directly access conceptual information while looking up a term. If this condition is given, the type of conceptual relation between the term and other concepts/terms can either be specified or not)
2.1.2.1) specified relations (as a result of this feature, users should be able to look for single types of relations and identify clusters of concepts such as synonyms, hyponyms, troponyms etc.)
2.1.2.2) non-specified relations
2.1.3) indirect access via terminological items
Moreover, 2.1.2 (direct access) and 2.1.3 (indirect access) can be described in terms of:
2.2) total/partial access (the user can access the same conceptual information by looking up any involved term or only some of the involved terms)
2.3) access systematicity (access to conceptual information is coherently implemented for all terms)
2.4) availability of hyperlinked data

3.3 Consistency of concept-term correspondences

This section concentrates on the degree of consistency in the correspondences between concepts and terms in the analysed e-lexicographic resources, an aspect that is closely related to their mediostructural properties. In the ideal case, a resource should contemplate a coherent and recognizable mediostructure, independent of the depth of its conceptual structure and of the access to conceptual relations it provides. This paper leans on a conception of concepts and terms according to ISO 1087-1:2000. This norm, dealing with the vocabulary of terminology, defines a concept as a unit of

knowledge created by a unique combination of characteristics, whereas a term is a verbal designation of a general concept in a specific subject field. Terms can be instances of different kinds, such as simple terms, complex terms (e.g. collocations or compounds), symbols and formulae.

In order to test the consistency of the selected resources despite their heterogeneity, the example of the general concept ELECTRIC CURRENT and of related terms will be taken into consideration. Specific tasks have been accomplished that aim to assess consistency:

- search for the terms related to ELECTRIC CURRENT by accessing the conceptual structure
- search for the terms related to ELECTRIC CURRENT by looking up *electric current*
- is a correlation between ELECTRIC CURRENT and the corresponding terms coherently represented?
- if yes, is it present in both directions, i.e. when moving from the term to the concept and vice versa?

TOOL	CONCEPTUAL RELATIONS ACCORDING TO THE TYPE OF QUERY & DEGREE OF TERMINOLOGICAL COVERAGE		CONSISTENT REFERENCES CONCEPTS-TERMS
IEV Online	search by subject area: leads to different relations (hyponymy, meronymy etc.)	it is not possible to identify ELECTRIC CURRENT if not by browsing the content of all or selected subject areas	no references
	search by term: leads to hyponyms	*electric current* includes 5 hyponymical terms belonging to 3 subject areas	yes; the hypernym is referenced to the hyponyms, but not vice versa
IATE	search by term: leads to hyponyms	*electric current* includes 3 hyponymical terms and 1 synonymous term (*current*)	yes; the hypernym is referenced to the hyponyms, but not vice versa
OpenEI	search by term: leads to a specific term only	*electric current* is related to 6 other terms (non-specified relations)	no: consistency in cross-references is not always given
Fluke Glossary	search by term: leads to a specific term only	*electric current* is not among the glossary terms, but it is referred to in the article of *Ampère*	no consistency in cross-references

Table 4: Consistency of conceptual data

Table 4 summarizes the results of the analysis by presenting information concerning the types of relations the user can find according to the kind of query he/she performs, the corresponding degree of terminological coverage and a general evaluation of the consistency of mediostructural correlations between concepts and terms.

As these results show, taxonomical relations are better captured by a cross-referencing system, and are therefore likely to be rendered in a coherent way. Other types of relations tend to be widely underrepresented even in resources with a well-developed conceptual structure like IEV Online, which hints at the fact that underlying termbases do not reach a sufficiently deep level of ontological coverage when it comes to the identification of all available relations among concepts.

Semantic word-families are another interesting aspect in terminology which seems to be largely neglected in these resources. By observing the term *ampere*, which is connected to the concept ELECTRIC CURRENT through the relation "unit of measurement" (ampere is the unit of electric current according to SI) and its semantic word-family includes both orthographic variants (*Ampère*), abbreviations (*A*, *amp*) and compounds (*ampere-hour*, *ampere-hour meter*), it becomes clear that all of these terms should be systematically cross-referenced to each other. However, none of the selected e-lexicographic tools offers a satisfying and coherent representation of this cluster of terms: *ampere* is always referenced to the concept ELECTRIC CURRENT through the terminological definition, but the other terms a) are only partially available and b) are not coherently cross-referenced to each other (cf. Table 5).

TOOL	CROSS-REFERENCES
IEV Online	ampere > ELECTRIC CURRENT ampere > A, volt-ampere meter, ampere-hour meter, volt-ampere-hour meter ampere <> ampere-turns
IATE	ampere > ELECTRIC CURRENT ampere > A, amp, ampere-turn/ampere turn, ampere-hour capacity, ampere hour/ampere-hour, amperes per metre, kiloVolt Ampere, metre-kilogram-second ampere, volt-ampere-reactive hour meter, volt-ampere, …
OpenEI	ampere > ELECTRIC CURRENT ampere > amp
Fluke Glossary	Ampère > ELECTRIC CURRENT Ampère > A Ampère-Hour > Ampère

Table 5: Treatment of semantic word-families: *ampere*

Table 5 reveals an overall lack of data concerning the relations among these concepts/terms. IEV Online only records the compounds of *ampere* that are relevant from a subject-area perspective. IATE lists a larger number of compounds but without

clustering them into conceptually coherent groups and offering no opportunity to proceed in the opposite direction, i.e. from a compound to *ampere* (IATE always moves from a base to its compounds/collocations and not vice versa), which is made possible by IEV Online (cf. *ampere* <> *ampere-turns*), although not systematically. OpenEI and the Fluke Glossary display the most lacking treatment of this semantic word-family. The former only refers to the abbreviation *amp* and does not cover other related terms belonging to the family: this may be seen as a partial but not incoherent representation, since this resource does not specifically focus on the topic of electrical engineering. The Fluke Glossary, on the contrary, has an approach to the treatment of these terms which is clearly partial and inconsistent.

4. Observations and outlook

Assessment of the management and exploitation of conceptual data in the selected resources has pointed out important differences in their approach. The OpenEI and Fluke glossaries do not develop a conceptual structure, which results in a partial and often incoherent conceptual representation. This can be a great disadvantage to users, particularly non-experts. IEV Online and IATE offer more advanced solutions: they are both based on an underlying conceptual structure and offer a much larger amount of data. Evaluation of conceptual data and information carried out in Section 3 defines the minimum requirements a termbase intended for LSP e-lexicographic resources should comply with:

A) Conceptual structure availability and properties:
- sufficient (multilevel) depth of conceptual structures
- multivocal relations (lower–higher level and lower–lower level)

B) Ease of access to conceptual data:
- direct access via the conceptual structure, with specified relations
- direct access via the microstructure, with specified relations

C) Consistency of concept-term correspondences:
- consistency of cross-references in the search by concept
- consistency of cross-references in the search by term.

IEV Online and IATE do not satisfy all these conditions but combine only some of them. The main drawback of these resources is the absence of a conceptual structure in the form of an ontology. Unfortunately, structures of subject-groups can be systematic and coherent but fail to cover the entire range of semantic relations among the concepts of a discipline. What a subject-group structure does not record is partly compensated for by means of the terminological definition (cf. the example of *ampere* and of its relation to the concept ELECTRIC CURRENT). However, a domain-specific ontology would ensure a definitely higher degree of data accessibility and data coherence. Conceptual structures should account for the existence of different types of relations, such as

1) semantic fields (broadly intended as clusters of concepts displaying, for instance, taxonomic, meronomic, troponymic, or functional relations) and
2) semantic word-families (clusters of concepts/terms with morphological affinity, including abbreviations, orthographic variants and word combinations). The two groups can overlap, but distinctive features should also be taken into consideration to guarantee a possibly comprehensive conceptual representation.

By implementing a method for delivering a detailed description of conceptual data representation in LSP e-lexicographic resources, this study has revealed a series of essential properties and their most effective and efficient combinations. At the same time, new ways to improve terminological representation and exploitation in termbases for lexicographic purposes should be looked for by conducting further investigations on other resources and subdomains, as well as dictionary consultation tests according to specific usage situations (e.g. text production, text reception, active and passive translation).

5. References

Alwert, K. & Hoffmann, I. (2003). "Knowledge Management Tools". In K. Mertins & P. Heisig & H. Vorbeck (eds) *Knowledge Management. Concepts and Best Practices*. Berlin/Heidelberg/New York: pp. 114-150.

Giacomini, L. (2014). "Testing user interaction with LSP e-lexicographic tools: A case study on active translation of environmental terms". In Proceedings of Konvens, Hildesheim 2014, October 8-10.

Giacomini, L. (2015). Macrostructural properties and access structures in LSP e-dictionaries for translation: the technical domain. *Lexicographica* 31.2015 (forthcoming).

Halpin, T. & Morgan, T. (2010). *Information Modeling and Relational Databases*. Burlington, MA.

Heid, U. (2012). "Electronic Dictionaries as Tools: Toward an Assessment of Usability". In *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London: pp. 287-304.

Koplenig, A. (2011). "Understanding How Users Evaluate Innovative features of Online Dictionaries – An Experimental Approach". In Proceedings of eLex 2011: pp. 147-150.

Leroyer, P. (2012). "Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools". In *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. Continuum, London: pp. 121-140.

Mayer, F. (1998): *Eintragsmodelle für terminologische Datenbanken. Ein Beitrag zur übersetzungsorientierten Terminographie*. Tübingen.

Pratt, P. & Adamski, J. (2011). *Concepts of Database Management*. Boston, MA.

Schärfe, H. & Hitzler, P. & Øhrstrøm, P. (eds.) (2006). *Conceptual Structures: Inspiration and Application. 14th International Conference on Conceptual Structures ICCS 2006*. Aalborg.

Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge.* Tübingen.

Websites:

Fluke Electrical Glossary. Accessed at: http://www.fluke.com/fluke/inen/solutions/electrical/electrical%20glossary (15 May 2015)

IATE. Accessed at: http://iate.europa.eu. (15 May 2015)

IEV Online. Accessed at: http://www.electropedia.org. (15 May 2015)

OpenEI. Accessed at: http://en.openei.org/wiki/Glossary. (15 May 2015)