Planning a domain-specific electronic dictionary for the mathematical field of graph theory

Keywords electronic lexicography; mathematical terminology; patterns; semantic relations

We are planning to create an electronic LSP-dictionary for the mathematical field of graph theory. In mathematics a special purpose language in the sense of (Roelcke, 2010) is used.

A bilingual corpus in German and English is the basis for the dictionary. It is composed of textbooks, scientific papers and lecture notes (DE: 21.000 types, 483.000 tokens, EN: 10.000 types, 379.000 tokens). The selection of texts is based on the bibliography used in the lecture. Additionally, we carried out a survey with 40 mathematics students of graph theory to find out which sources they use during their studies, and included further texts after a qualitative analysis.

The purpose of the dictionary is to help mathematics students to improve their academic writing regarding terminology, because they have to deal with two challenges in their writing process: First, they write their theses in German, but most of the sources are in English. Therefore they have to find German equivalents; this is made even more difficult by the fact that a lot of mathematical terms are also used in the general language (Hischer, 2010), but equivalents are not always straightforward; for example, the German equivalent for *complete* in graph theory is *vollständig* but not *komplett*.

Secondly, the students have to learn the meaning of the terms. We intend to support this by means of an ontology. To examine the effect of the dictionary and ontology use theses will be analyzed with regard to the usage of terminology from a quantitative and a qualitative perspective. So we build two more corpora which contain theses written with and without the dictionary. The dictionary will be used for an intervention study.

Besides the alphabetical access, the dictionary will also provide thematic, onomasiological access; it will contain lemmas in German and English, related terms and equivalence statements. The lexicographic results of Bergenholtz et al. (2008), Bothma et al. (2017), Fuertes-Olivera and Tarp (2014) and Giacomini (2015) are theoretical bases considered in the creation process. In respect of terminology we refer to Faber et al. (2016), Faber and Castro (2014), Faber et al. (2011), Bergenholtz and Kaufmann (1997), Kaufmann and Bergenholtz (1992).

We aim to achieve a highly automated creation process for the dictionary. Therefore we are identifying definition patterns and associating with each pattern a certain (semantic) relation (cf. Table 1 for an example). The relations are based on GermaNet but domain-specifically adjusted (Kunze and Lemnitzer, 2007). In the next step, the terms will be extracted with the aid of the patterns and can be used for directly creating an ontology (cf. Meyer, 2001, Barnbrook, 2002).

	Pattern	**Relation**
(1)	If-clause N1 is called N2	N1 is-hypernym-of N2
(2)	N1 is called N2 If-clause	N1 is-hypernym-of N2
(3)	N is called ADJ	ADJ is-attribute-of N
(4)	N1 is called N2	N1 is-synonym-of N2
(5)	ADJ N1 is called N2	ADJ is-attribute-of N1
(6)	ADJ N1 is called N2	(ADJ N1) is-synonym-of N2
(7)	ADJ N1 is called N2	N1 is-hypernym-of N2
(8)	N1 of N2 is called N3 If-clause	(N1 of N2) is-hypernym-of N3

Table 1: Pattern *is called*

References

Barnbrook, G. (2002). *Defining language. A local grammar of definition sentences.*, volume 11 of *Studies in Corpus Linguistics*. John Benjamins, Amsterdam/Philadelphia.

Bergenholtz, H. and Kaufmann, U. (1997). Terminography and lexicography. a critical survey of dictionaries from a single specialised field. *Hermes*, 18:91–125.

Bergenholtz, H., Tarp, S., and Wiegand, H. E. (2008). Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In Hoffmann, L. et al., editors, *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. Handbücher zur Sprach-*

und Kommunikationswissenschaft, volume 14/2 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, pages 1762–1832. de Gruyter, Berlin/Boston/New York.

Bothma, T. J. D., Prinsloo, D. J., and Heid, U. (2017). A taxonomy of user guidance devices for e-lexicography. *Lexicographica*, 33:391–422.

Faber, P., Araúz, P. L., and Reimerink, A. (2011). Knowledge representation in ecolexicon. In Zanon, N. T., Monje, E. M., and Romero, F. P., editors, *Technical Innovation in the Teaching and Processing of LSPs: proceedings of TISLID' 10*, volume 50308 of *Filologia (UNED)*, pages 367–385, Madrid. Universidad Nacional de Educacion a Distancia.

Faber, P. and Castro, M. B. (2014). Ecolexicon. In Abel, A., Vettori, C., and Ralli, N., editors, *Proceedings of the XVI EURALEX International Congress: The User in Focus.*, pages 601–607, Bolzano. EURAC research.

Faber, P., León-Araúz, P., and Reimerink, A. (2016). Ecolexicon: New features and challenges. In Kernerman, I., Trojina, I. K., Krek, S., and Trap-Jensen, L., editors, *GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference*, pages 73–80, Portorož.

Fuertes-Olivera, P. A. and Tarp, S. (2014). *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminology.* de Gruyter, Berlin/Boston.

Giacomini, L. (2015). Macrostructural properties and access structures of LSP e-dictionaries for translation: the technical domain. *Lexicographica*, 31:90–117.

Hischer, H. (2010). *Was sind und was sollen Medien, Netze und Vernetzungen? Vernetzungen als Medium zur Weltaneignung.* Franzbecker, Hildesheim/Berlin.

Kaufmann, U. and Bergenholtz, H. (1992). *Genteknologisk ordbog. Dansk-engelsk/engelsk-dansk molekylærbiologi og DNA-teknologi.* Gads, København.

Kunze, C. and Lemnitzer, L. (2007). *Computerlexikographie. Eine Einführung.* Gunter Narr, Tübingen.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. In Bourigault, D., Jacquemin, C., and L'Homme, M.-C., editors, *Recent Advances in Computational Terminology*, volume 2 of *Natural Language Processing*, pages 279–302. John Benjamins, Amsterdam/Philadelphia.

Roelcke, T. (2010). *Fachsprachen.* Erich Schmidt, Berlin.