Proceedings of the 6th International Workshop on Computational Terminology (COMPUTERM 2020), pages 55–61
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

The NetViz terminology visualization tool and the use cases in karstology domain modeling

Senja Pollak[1], Vid Podpečan[1], Dragana Miljkovic[1], Uroš Stepišnik[2] and Špela Vintar[2]
[1]Jožef Stefan Institute, Ljubljana, Slovenia
{senja.pollak,vid.podpecan,dragana.miljkovic}@ijs.si
[2]Faculty of Arts, University of Ljubljana, Slovenia
{uros.stepisnik,spela.vintar}@ff.uni-lj.si

Abstract

We present the NetViz terminology visualization tool and apply it to the domain modeling of karstology, a subfield of geography studying karst phenomena. The developed tool allows for high-performance online network visualization where the user can upload the terminological data in a simple CSV format, define the nodes (terms, categories), edges (relations) and their properties (by assigning different node colors), and then edit and interactively explore domain knowledge in the form of a network. We showcase the usefulness of the tool on examples from the karstology domain, where in the first use case we visualize the domain knowledge as represented in a manually annotated corpus of domain definitions, while in the second use case we show the power of visualization for domain understanding by visualizing automatically extracted knowledge in the form of triplets extracted from the karstology domain corpus. The application is entirely web-based without any need for downloading or special configuration. The source code of the web application is also available under the permissive MIT licence, allowing future extensions for developing new terminological applications.

Keywords: Terminology visualization, Karstology, Domain modeling, Networks

1. Introduction

Visual representations of specialized domains are becoming mainstream for several reasons, but firstly as a natural response to the fact that "concepts do not exist as isolated units of knowledge but always in relation to each other" (ISO 704, 2009). In recent terminological projects, visualization has been considered an important asset (Faber et al., 2016; Carvalho et al., 2017; Roche et al., 2019). We believe that the visualization of terminological knowledge is especially well-suited to the needs of frame-based terminology, aiming at facilitating user knowledge acquisition through different types of multimodal and contextualized information, in order to respond to cognitive, communicative, and linguistic needs (Gil-Berrozpe et al., 2017). Moreover, it has been shown that domain experts are often able to interpret information faster when viewing graphs as opposed to tables (Brewer et al., 2012). More generally, as has become evident in the rising field of digital humanities, digital content, tools, and methods are transforming the entire field of humanities, changing the paradigms of understanding, asking new research questions and creating new knowledge (Hughes et al., 2015; Hughes, 2012).

As this workshop demonstrates, terminological work has undergone a significant change with the emergence of computational approaches to extracting various types of terminological knowledge (e.g., term extraction, definition extraction, semantic relation extraction), which enhances the potential of visualization not only to represent manually annotated data, but also for automatically and semi-automatically extracted knowledge, which we also show in our use cases.

We focus on the field of karstology, the study of specific relief which develops on soluble rocks such as limestone and is characterized by caves, typical depressions, karst springs, ponors and similar. It is an interdisciplinary subdomain of geography bordering on geomorphology, geology, hydrology and chemistry. In karstology, the main objects of interest are its typical landforms usually described through their form, size, location and function, and the environmental and chemical processes affecting their development such as dissolution and weathering.

The proposed semantic network visualization tool NetViz[1] used in the presented karstology domain modeling experiments, complement our previous research in the TermFrame project including work of Vintar et al. (2019) where frame-based annotation of karst definitions is presented, Pollak et al. (2019) presenting results of term and definition extraction from karst literature, Miljkovic et al. (2019) with term co-occurrence network extraction and Grčić-Simeunović and De Santiago (2016) where semantic properties of karst phraseology are explored.

2. Related Work

There are several projects which consider *terminology visualization* as an important asset of specialized knowledge representation. One such project is the EndoTerm, a knowledge-based terminological resource focusing on endometriosis (Carvalho et al. 2016, Roche et al. 2019). EndoTerm includes a visual concept representation developed via CMap Tools and organizes knowledge into semantic categories linked with different types and levels of relations, while ensuring compatibility with existing medical terminology systems such as SNOMED. The most closely related project to ours using a visual representation of specialized knowledge is the EcoLexicon (Faber et al., 2016), where terms are displayed in a semantic network linking the central query term to related terms and its translation equivalents in up to 5 other languages. The edges of the network represent three types of relations, namely the generic-

[1]`https://biomine.ijs.si/netviz/`

specific (is_a) relation, the part-whole relation and a set of non-hierachical relations (made_of, located_at, affects etc.). While the EcoLexicon remains impressive with the abundance and complexity of data it offers, our own approach differs mainly in that we use natural language processing techniques to infer data, and that we envisage different types of visual representation depending on the task or end-user.

In terms of *domain modeling of terminological knowledge*, we can first mention the field of terminology extraction. In automatic terminology first the distinction was between linguistic and statistical approaches, but most state-of-the-art systems are hybrid. Many terminology extraction algorithms are based on the concepts of termhood and unithood (Kageura and Umino, 1996), where termhood-based approaches include work by Ahmad et al. (2000) and Vintar (2010), while Daille et al. (1994) and Wermter and Hahn (2005) use unithood-based measures, such as mutual information and t-test, respectively. More recently, deep learning and word embeddings (Mikolov et al., 2013) have become very popular in natural language processing, and several attempts have already been made to utilize these techniques also for terminology extraction (Amjadian et al., 2016; Zhang et al., 2017; Wang et al., 2016) and terminology expansion (Pollak et al., 2019). Next, for defining relations between terms, there are several relation extraction methods, which can roughly be divided into categories: co-occurrence-based, pattern-based, rule-based and machine-learning approaches (Bui, 2012; Sousa et al., 2019). Co-occurrence is the simplest approach which is based on the assumption that if two entities are frequently mentioned together in the same sentence, paragraph or document, it is probable that they are related (Song et al., 2011). The pattern- and the rule-based differ in that the former use template rules, whereas the latter might additionally implement more complex constraints, such as checking negation, determining the direction of the relation or expressing rules as a set of procedures or heuristic algorithms (Kim et al., 2007; Fundel-Clemens et al., 2007). Machine-learning approaches usually set the relations extraction tasks as classification problems (Erkan et al., 2007). Recently, the proposed approaches often use the power of neural networks as in Lin et al. (2016), Sousa et al. (2019), Luo et al. (2020).

The focus of this paper is the visualization tool and its use in karstology domain modeling. For data extraction, we employ several techniques mentioned above. Pattern-based methods (Pollak et al., 2012) are used for definition extraction in the first use case (Section 4.3.) providing definition candidates for further manual annotation of domain knowledge, while in the second use case (Section 4.4.) we use statistical term extraction techniques (Vintar, 2010; Pollak et al., 2012) coupled with co-occurrence analysis and relation extraction using Reverb (Fader et al., 2011).

3. NetViz

Network visualization is of key importance in domains where an optimized graphical representation of linked data is crucial in revealing and understanding the structure and interpreting the data with the aim to obtain novel insights and form hypotheses. There is a plethora of software which deals with network analysis and visualization. For example, Gephi (Bastian et al., 2009), Pajek (Batagelj and Mrvar, 2002) and Graphviz (Ellson et al., 2001) are among the most popular classic software tools for these tasks and have been used in very diverse domains. However, every domain and every task poses specific requirements and using tools which are too general is often a poor choice which has adverse effects on usability. Therefore, our aim was to provide a minimal environment which enables zero effort network visualization for specific tasks such as terminology. We developed NetViz (`https://biomine.ijs.si/netviz/`), a web application which enables interactive visualization of networks. NetViz builds upon our previous work on visualization and exploration of heterogeneous biological networks (Podpečan et al., 2019). where several large public databases are merged into a network which can then be explored, analyzed and visualized. We applied the same principles and created a domain independent network visualization tool which was then applied to karstology domain modeling and exploration.

3.1. Features

- **Open source**. Netviz is available under the liberal MIT license on the open source portal GitHub[2].
- **Single page, client-only web application**. NetViz is implemented as a client-only web application. As a result, NetViz requires no hosting and server configuration and can be also run locally simply by downloading and opening its html page in a web browser.
- **High performance network visualization**. NetViz implements a user interface around the `vis-network` module of the `vis.js` visualization library. `vis-network` is a fast, highly configurable library for network visualization in the browser and NetViz builds upon its visualization engine.
- **Visualization and editing features**. A set of fundamental network editing and visualizaton features are implemented. The network can be modified after visualization by adding or removing nodes and edges. Several settings controling the physics simulation which does the layouting can be adjusted before, during or after the visualization. Context menus which are available on all elements (node, edges and the canvas itself) provide a few basic options which can be extended according to the requirements of the specific domain.
- **CSV data format**. In order to make the use of NetViz as simple as possible its data input format is a comma separated file (CSV) with header. Two files are used: the first one which is mandatory defines edge properties while the optional second file defines node properties. The header for edge definition file supports the following columns: `node1`, `node2`, `arrow`, `label`, `text`, `color`, and `width` where `node1`, `node2`, and `arrow` are mandatory and the rest is optional. The header for node definition file supports the following columns:

[2] `https://github.com/vpodpecan/netviz`

node, text, color, and shape. We expect that the list of supported columns (features) will grow and adapt to specific domains where NetViz will be used. We will also add the option to export the current network so that the user modifications of the network will not be lost upon closing the application.

The intended users are domain experts in the process of construction of a domain ontology, terminologists, as well as students and teachers. It also has potential for being used by larger public with some modifications and a fixed domain knowledge base.

4. Karstology Domain Modeling

4.1. The TermFrame Project

The context for this research is the TermFrame project which employs the frame-based approach to build a visual knowledge base for karstology in three languages, English, Slovene and Croatian. The main research focus of the project is to explore new methods of knowledge extraction from specialized text and propose novel approaches to knowledge representation and visualization (see previous work in the project described in Vintar et al. (2019), Pollak et al. (2019), Miljkovic et al. (2019)).

The frame-based approach in terminology (Faber, 2012; Faber, 2015) models specialized knowledge through conceptual frames which simulate the cognitive patterns in our minds. According to this view, a frame is a mental structure consisting of concept categories and relations between them. Unlike hand-crafted ontologies, frame-based terminology uses specialized corpora to induce frames or event templates, thus consolidating the conceptual and the textual level of a specialized domain.

Such an approach to knowledge and terminology modeling has a lot to gain from graph-like representations, because its building blocks are concept categories, concepts and terms as nodes, and various types of hierarchical and non-hierarchical relations as edges. By selecting different layers of representation it is thus possible to visualize the dynamic and multidimensional nature of specialized knowledge.

In the TermFrame project we combine manual and computational methods to extract domain knowledge. However, in an ideal scenario, as many steps as possible would be automated requiring only minimal manual validation. The main steps of our proposed domain modeling workflow can be summarized as follows:

- Convert documents to plain text format.
- Identify domain terms.
- Identify domain definitions.
- Identify semantic categories.
- Identify semantic relations.
- Select information for network visualization.
- Visualize the network.
- Interactively explore and modify the terminological resource.

Details on automated knowledge extraction for several of these steps are provided in Pollak et al. (2019). In the following subsections, we present the corpus, as well as two experiments on karstology domain modeling, where a subset of steps above are performed manually or automatically, before the final steps of visualization and interactive exploration using NetViz, which is the focus of this paper and common to both experiments.

4.2. Corpus

The English part of the TermFrame corpus, which was used in these experiments, contains 56 documents of different length, all pertaining to karstology. It includes books, research articles, theses and textbooks (for more details see Vintar et al. (2019)). We used Google Documents feature for conversion of documents from pdf to text format. Frequently such conversion introduced errors into the document such as additional line breaks or orphaned figure captions in the middle of paragraphs. Such errors were corrected in the post-processing phase either manually or using simple scripts.

4.3. Visualizing Manually Annotated data

In this experiment we use manual annotations of domain definitions. Specialized definitions were first either identified in dictionaries and glossaries or using definition extractor from domain texts (Pollak et al., 2012)[3], and next annotated with a hierarchy of semantic categories and a set of relations which allow to describe karst events. For an example of annotated definition see Figure 1. The annotation process—performed by linguists and domain experts—is described in detail in Vintar et al. (2019) and briefly summarized below.

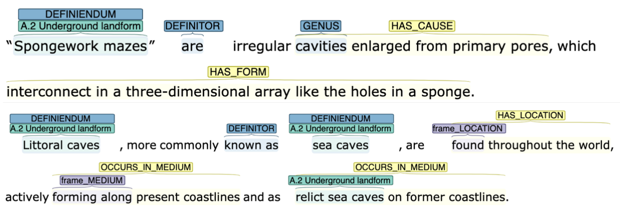

Figure 1: Manual annotation of automatically extracted definitions.

The semantic categories were inspired by the concept hierarchy in the EcoLexicon[4] and adapted to karstology by domain experts. The first three top-level categories, LANDFORMS, PROCESSES and GEOMES, are the most relevant for domain modeling as they contain terms specific to karst, while the rather broad group of ELEMENTS, ENTITIES and PROPERTIES contains broader terms from geography, chemistry, botany and similar. INSTRUMENTS and METHODS are used to categorize karstology-specific

[3]The evaluation of automated definition extraction is described in detail in Pollak et al. (2019). About 30% of extracted definition candidates were judged as karst or neighbouring domain definitions, while about 16% of definition candidates were evaluated as karst definitions used for the fine-grained manual annotation.

[4]https://ecolexicon.ugr.es/en/index.htm

research and/or measurement procedures, but were found to occur rarely in our set of definitions.

The second important level of annotation identifies the semantic relations which describe specific aspects of karst concepts. According to the geomorphologic analytical approach (Pavlopoulos et al., 2009), landforms are typically described through their spatial distribution (HAS_LOCATION; HAS_POSITION), morphography (HAS_FORM; CONTAINS), morphometry (HAS_SIZE), morphostructure (COMPOSITION_MEDIUM), morphogenesis (HAS_CAUSE), morphodynamics (HAS_FUNCTION), and morphochronology (OCCURS_IN_TIME). The ideal definition of a landform would include all of the above aspects, but in reality most definitions extracted from the corpus or domain-specific glossaries specify only two or three. In total, 725 definitions were annotated, 3149 terms were assigned categories.

In this experiment we focus on the visualization of the taxonomy built from manually annotated categories of DEFINIENDUM and their hypernyms, connected by IS_A relation to their subcategories and categories (LANDFORM, PROCESS, GEOME, ELEMENT/ENTITY/PROPERTY, and INSTRUMENTS/METHODS). The top level—taxonomy of categories—can be observed in Figure 2. In Figure 3, we can see lower levels, which correspond to terms from definitions, more specifically terms (definiendums) assigned to specific subcategories of Hydrological forms and Underground landforms. It allows the user to quickly grasp the main conceptual properties of hydrological forms, namely that water in karst continuously submerges underground (*sinking creek, losing streamflow, swallow hole etc.*) and reemerges to the surface (*karst spring, resurgence, vauclusian spring etc.*), depending on the porosity of the underlying bedrock. Amongst underground landforms we can quickly discern various types of caves (*crystal cave, lava cave, active cave, bedding-plane cave, roofless cave*) and typical underground formations found in them (*straw stalactites, flute, capillary stalagmite, column, cave pearl*). The network also shows that certain terms belong to both categories (*blue hole, inflow cave*) as certain forms are both underground and submerged in water or have a hydrological function in karst. In addition, we have noticed that graph-based visualization facilitates the identification and correction of inconsistencies in the manual expert annotation. The final goal is to integrate the visual, graph-based representation into a multimodal knowledge base where frames (Cause, Size, Location, Function etc.) as defined in Vintar et al. (2019) will be presented to the user together with corpus examples, images and geolocations.

4.4. Visualizing Automatically Extracted Knowledge

In this experiment we used sentences where automatically extracted terms co-occurred, and then identified relations between them. The resulting knowledge is shown in Figure 4. The relation extraction was done using ReVerb (Fader et al., 2011), which is a program that automatically identifies and extracts relationships from English sentences, output the triplets in form `<argument1, relation phrase, argument2>`, usually corresponding to subject-verb-object. It is designed for cases where the target relations cannot be specified in advance, which corresponds to the requirements of this experiment with knowledge discovery in mind. The preprocessing includes tokenization, lemmatization and POS tagging. We used the lemmatized forms. We are interested in triplets that include as arguments only terms from the karst domain. The terms were extracted using (Pollak et al., 2012) and were further validated by domain experts.[5] We also used terms in karstology term list QUIKK [6]. The validated list of domain-specific terms contained 3,149 terms, and triplet arguments extracted with ReVerb were matched against this list. In this way, a huge general triplet network containing less relevant information for domain exploration is reduced and thus made easier for manual inspection. After filtering we retained 302 triplets where arguments exactly match the terms from the list. The most frequent relations include: *be, fill_with, exceed, form_in, associate_with, be_source_of,...*.

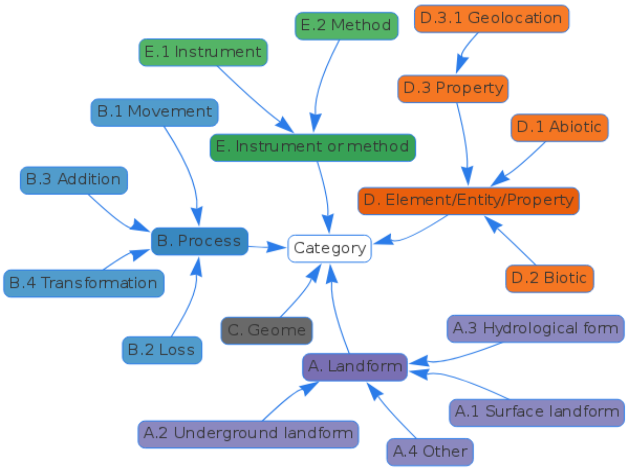

Figure 2: The taxonomy of categories visualized in NetViz.

5. Conclusion and future work

We presented the NetViz terminology visualization tool and two examples of its use for knowledge modeling in the domain of karstology. First, we have demonstrated the visual representation of domain knowledge as extracted from manually annotated definitions. The multi-layer annotations include conceptual categories (Landform, Process, Geome, Element/Entity/Property, Instrument/Method) and their subcategories with which the terms are labelled, and the resulting network can be used by experts, teachers, students or terminologists to explore related groups of concepts, identify knowledge patterns or spot annotation mistakes. Next, we visualized the relations as proposed by the automated term and triplet extraction. This approach is complementary to the manual annotation and may point to previously unknown connections or knowledge structures.

[5] A detailed evaluation of term extraction process is presented in Pollak et al. (2019), ranging from 19.2% for strictly karst terms and 51.6% including broader domain terms and names entities.

[6] `http://islovar.ff.uni-lj.si/karst`

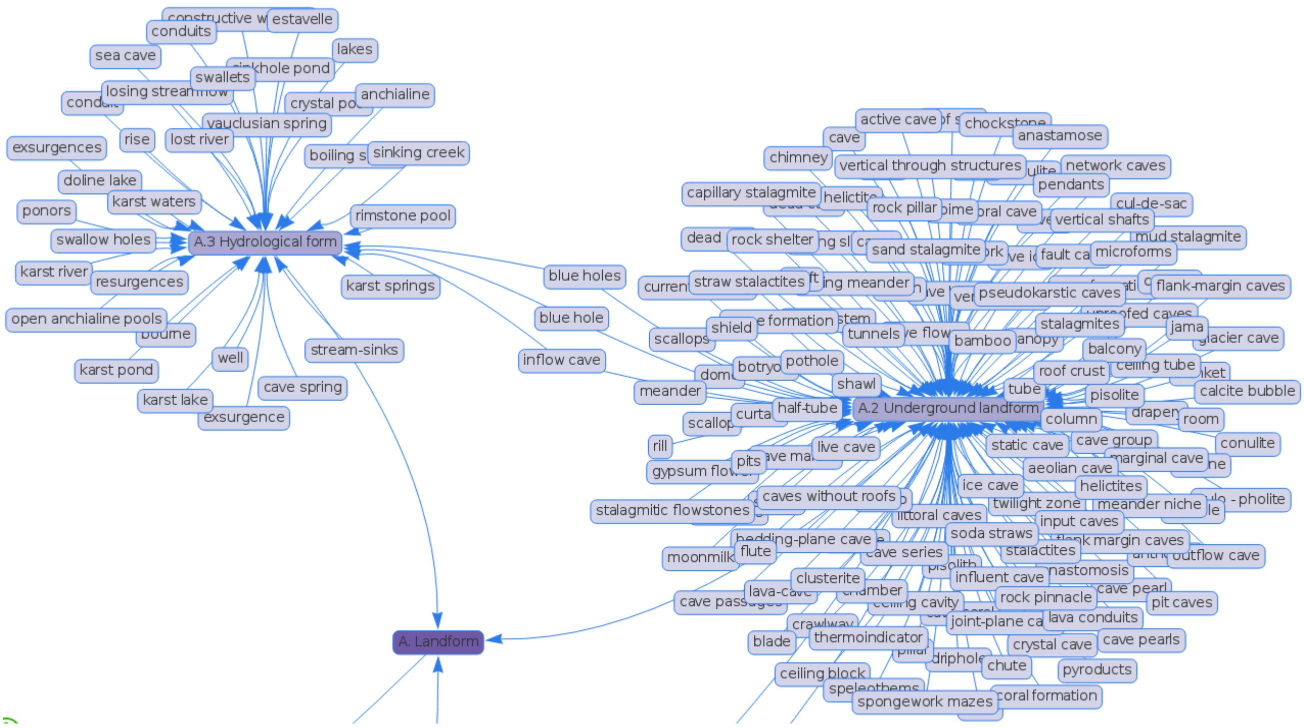

Figure 3: A visualization of a part of categories network which includes hydrological and underground landforms.

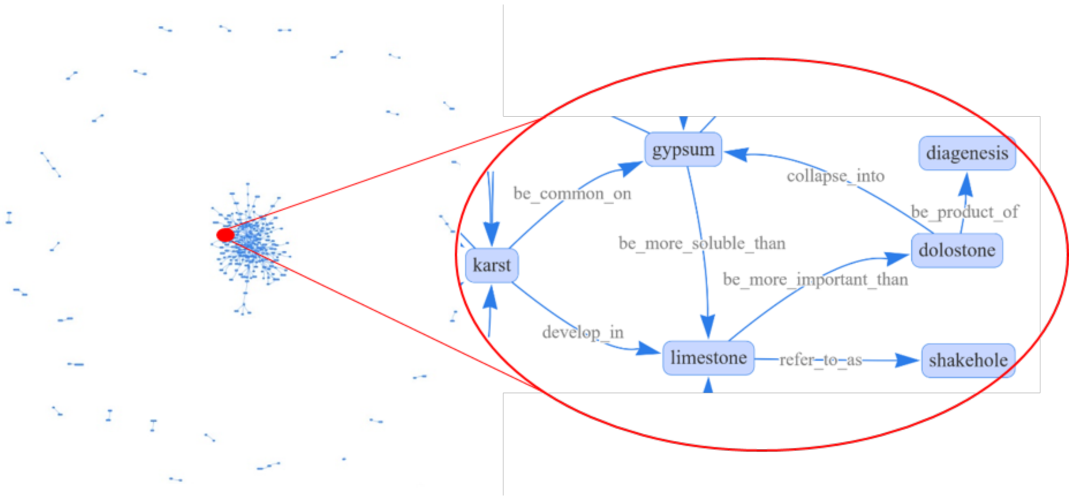

Figure 4: Graph with triplet relations extracted with ReVerb where subject and object match the manually validated list of karst terms.

The simplicity of NetViz allows users to prepare their own input data in the CSV format and create customized visualizations to support their research. For example, in the TermFrame project NetViz is currently used to explore cases where identical or similar concepts have been defined through different hypernyms (e.g. *karst* is a kind of *landscape / terrain / topography / product of processes / phenomenon / area*).

As future work and the end-result, of the TermFrame project we plan to develop an integrated web-based environment for karst exploration which will combine graphs with textual information, images and geolocations. Since a large number of natural monuments worldwide are in fact karst phenomena, we see the potential of such knowledge representations not just for science but also for education, environment and tourism.

6. Acknowledgements

The authors acknowledge the financial support from the Slovenian Research Agency for research core funding for the programme Knowledge Technologies (No. P2-0103) and the project TermFrame - Terminology and Knowledge Frames across Languages (No. J6-9372). This paper is also supported by European Union's Horizon 2020 research and innovation programme under grant agreement No. 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media).

7. Bibliographical References

Ahmad, K., Gillam, L., Tostevin, L., and Group, A. (2000). University of surrey participation in trec 8: Weirdness indexing for logical document extrapolation and retrieval (wilder). 03.

Amjadian, E., Inkpen, D., Paribakht, T., and Faez, F. (2016). Local-global vectors to improve unigram terminology extraction. In *Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016)*, pages 2–11.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

Batagelj, V. and Mrvar, A. (2002). Pajek— analysis and visualization of large networks. In Petra Mutzel, et al., editors, *Graph Drawing*, pages 477–478, Berlin, Heidelberg. Springer Berlin Heidelberg.

Brewer, N., Gilkey, M., Lillie, S., Hesse, B., and Sheridan, S. (2012). Tables or bar graphs? presenting test results in electronic medical records. *Medical decision making : an international journal of the Society for Medical Decision Making*, 32(4):545–553.

Bui, Q.-C. (2012). Relation extraction methods for biomedical literature. *Structure*, 01.

Carvalho, S., Costa, R., and Roche, C. (2017). Ontoterminology meets lexicography: the multimodal online dictionary of endometriosis (mode). In *GLOBALEX 2016: Lexicographic Resources for Human Language Technology Workshop at the 10th International Conference on Language Resources and Evaluation (LREC'16)*, pages 8–15, Portorož, Slovenia.

Daille, B., Gaussier, E., and Langé, J.-M. (1994). Towards automatic extraction of monolingual and bilingual terminology. In *Proceedings of the 15th Conference on Computational Linguistics - Volume 1*, COLING '94, page 515–521, USA. Association for Computational Linguistics.

Ellson, J., Gansner, E., Koutsofios, L., North, S., Woodhull, G., Description, S., and Technologies, L. (2001). Graphviz — open source graph drawing tools. In *Lecture Notes in Computer Science*, pages 483–484. Springer-Verlag.

Erkan, G., Ozgur, A., and Radev, D. (2007). Semi-supervised classification for extracting protein interaction sentences using dependency parsing. In *In Proceedings of the Conference of Empirical Methods in Natural Language Processing (EMNLP '07)*, pages 228–237, 01.

Faber, P., León-Araúz, P., and Reimerink, A. (2016). EcoLexicon: new features and challenges. In *Proceedings of GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference*, pages 73–80.

Faber, P. (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin, Boston: De Gruyter Mouton.

Faber, P. (2015). Frames as a framework for terminology. In Hendrik Kockaert et al., editors, *Handbook of Terminology*, page 14–33. John Benjamins.

Fader, A., Soderland, S., and Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the Conference of Empirical Methods in Natural Language Processing (EMNLP '11)*, Edinburgh, Scotland, UK, July 27-31.

Fundel-Clemens, K., Küffner, R., and Zimmer, R. (2007). Relex - relation extraction using dependency parse trees. *Bioinformatics (Oxford, England)*, 23:365–71, 03.

Gil-Berrozpe, J., León-Araúz, P., and Faber, P. (2017). Specifying hyponymy subtypes and knowledge patterns: A corpus-based study. In *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, pages 63–92.

Grčić-Simeunović, L. and De Santiago, P. (2016). Semantic approach to phraseological patterns in karstology. In T. Margalitadze et al., editors, *Proceedings of the XVII Euralex International Congress*, pages 685–693. Ivane Javakhishvili Tbilisi State University.

Hughes, L. M., Constantopoulos, P., and Dallas, C. (2015). Digital methods in the humanities: Understanding and describing their use across the disciplines. In J. Unsworth S. Schreibman, R. Siemens, editor, *A New Companion to Digital Humanities*, pages 150–170). John Wiley & Sons.

Hughes, L. M. (2012). ICT methods and tools in arts and humanities research. In Lorna M. Hughes, editor, *Digital Collections: Use, Value and Impact*, pages 123–134. London, UK: Facet Publishing.

ISO 704. (2009). ISO 704:2009: Terminology work-principles and methods. Standard, ISO, Switzerland.

Kageura, K. and Umino, B. (1996). Methods of automatic term recognition: A review. *Terminology International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(2):259–289.

Kim, J.-H., Mitchell, A., Attwood, T. K., and Hilario, M. (2007). Learning to extract relations for protein annotation. *Bioinformatics*, 23(13):i256–i263, 07.

Lin, Y., Shen, S., Liu, Z., Luan, H., and Sun, M. (2016). Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2124–2133, Berlin, Germany, August. Association for Computational Linguistics.

Luo, L., Yang, Z., Cao, M., Wang, L., Zhang, Y., and Lin, H. (2020). A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature. *Journal of Biomedical Informatics*, 103:103384, 02.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings to The International Conference on Learning Representations 2013*.

Miljkovic, D., Kralj, J., Stepišnik, U., and Pollak, S. (2019). Communities of related terms in Karst terminology co-occurrence network. In *Proceedings of eLex 2019*.

Pavlopoulos, K., Evelpidou, N., and Vassilopoulos, A.

(2009). *Mapping Geomorphological Environments*. Berlin Heidelberg:Springer.

Podpečan, V., Ramšak, v., Gruden, K., Toivonen, H., and Lavrač, N. (2019). Interactive exploration of heterogeneous biological networks with biomine explorer. *Bioinformatics*, 06.

Pollak, S., Vavpetič, A., Kranjc, J., Lavrač, N., and Špela Vintar. (2012). Nlp workflow for on-line definition extraction from English and Slovene text corpora. In Jeremy Jancsary, editor, *Proceedings of KONVENS 2012*, pages 53–60. ÖGAI, September. Main track: oral presentations.

Pollak, S., Repar, A., Martinc, M., and Podpečan, V. (2019). Karst exploration : extracting terms and definitions from Karst domain corpus. In *Proceedings of eLex 2019*.

Roche, C., Costa, R., Carvalho, S., and Almeida, B. (2019). Knowledge-based terminological e-dictionaries The EndoTerm and al-Andalus Pottery projects. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 25(2):259–290.

Song, Q., Watanabe, Y., and Yokota, H. (2011). Relationship extraction methods based on co-occurrence in web pages and files. In *Proceedings of the 13th International Conference on Information Integration and Web-Based Applications and Services*, iiWAS '11, page 82–89, New York, NY, USA. Association for Computing Machinery.

Sousa, D., Lamurias, A., and Couto, F. M. (2019). Using neural networks for relation extraction from biomedical literature.

Vintar, Š., Saksida, A., Stepišnik, U., and Vrtovec, K. (2019). Modelling specialized knowledge with conceptual frames: The TermFrame approach to a structured visual domain representation. In *Proceedings of eLex 2019*, pages 305–318.

Vintar, Š. (2010). Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation. *Terminology*, 16(2):141–158, 12.

Wang, R., Liu, W., and McDonald, C. (2016). Featureless domain-specific term extraction with minimal labelled data. In *Proceedings of the Australasian Language Technology Association Workshop 2016*, pages 103–112.

Wermter, J. and Hahn, U. (2005). Paradigmatic modifiability statistics for the extraction of complex multi-word terms. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 843–850, Vancouver, British Columbia, Canada, October. Association for Computational Linguistics.

Zhang, Z., Gao, J., and Ciravegna, F. (2017). Semrerank: Incorporating semantic relatedness to improve automatic term extraction using personalized pagerank. *arXiv preprint arXiv:1711.03373*.