

Špela Vintar*, Uroš Stepišnik**



TERMFRAME: A SYSTEMATIC APPROACH TO KARST TERMINOLOGY

Original scientific article
COBISS 1.01
DOI: 10.4321/dela.54.149-167

Abstract

We describe a systematic and data-driven approach to karst terminology where knowledge from different textual sources is structured into a comprehensive multi-lingual knowledge representation. The approach is based on a domain model which is constructed in line with the frame-based approach to terminology and the analytical geomorphological method of describing karst phenomena. The domain model serves as a basis for annotating definitions and aggregating the information obtained from different definitions into a knowledge network. We provide examples of visual knowledge representations and demonstrate the advantages of a systematic and interdisciplinary approach to domain knowledge.

Keywords: definitions, frame-based terminology, definition types, karstology, karst

* Department of Translation, Faculty of Arts, University of Ljubljana, Aškerčeva 2, SI-1000 Ljubljana

** Department of Geography, Faculty of Arts, University of Ljubljana, Aškerčeva 2, SI-1000 Ljubljana

e-mail: spela.vintar@ff.uni-lj.si, uros.stepisnik@ff.uni-lj.si

TERMFRAME: SISTEMATIČEN PRISTOP H KRAŠKI TERMINOLOGIJI

Izvleček

Prispevek opisuje sistematični in na podatkih utemeljeni pristop h kraški terminologiji, pri katerem skušamo izluščiti znanje iz različnih besedilnih virov in ga strukturirati v celovito večjezično reprezentacijo znanja. Naš pristop izhaja iz modela specializirane domene, ki smo ga zgradili v skladu z načeli terminologije shem in analitske geomorfološke metode opisovanja kraških pojavov in procesov. Model specializirane domene predstavlja ogrodje za označevanje definicij, nato pa se podatki iz različnih virov združujejo v mrežo znanja. V prispevku predstavimo nekaj primerov vizualizacij znanja, s katerimi ponazarjamo prednosti sistematičnega in interdisciplinarnega pristopa k urejanju specializiranega znanja.

Ključne besede: definicije, terminologija shem, tipi definicij, krasoslovje, kras

.....

1 INTRODUCTION

Karst is a type of Earth's surface that got its name after the Karst region in the hinterlands of the Gulf of Trieste in present-day Slovenia and Italy. The science of studying karst is called karstology. Its development was vastly expedited by research of the northern part of the Dinaric Karst, which was the site of the first explorations of this kind of terrain and was hence designated as the *Classical Karst* (Mihevc, 2010). There are several reasons why Slovenian Karst was the one to become the synonym for the scientific term and not some other karstic area in Europe. The most important factor is its geographic location and its geopolitical position in the period when karstology was developing between the 16th and 19th century, as the southern part of the Balkan Peninsula had been a part of the Ottoman Empire. At the time, Istria and a part of Karst were part of the Habsburg Monarchy and Trieste had become an important commercial hub. In light of all this, the region in the hinterlands of Trieste managed to impress the travellers of that day, thus becoming a synonym for a barren, rocky surface (Kranjc, 1994).

Since the beginning of the scientific study of karst in the middle of the 19th century, in addition to the general term *karst*, many other karst terms were derived from South Slavic languages or local dialects within the area of the Classical Karst. They are still used today in international karstology describing mostly basic surface karst features such as *dolina*, *uvala*, *polje*, *hum*, *ponor*, etc. (Kranjc, 2008).

Because of the strong interactions between the international karst nomenclature and South Slavic languages covering prominent karst regions, within *TermFrame*:

*Terminology and Knowledge Frames Across Languages*³ we explore, model and systematically represent karst terminology and knowledge in three languages: English, Slovene and Croatian. In line with the state-of-the-art frame-based approach to terminology (Faber et al., 2012), the TermFrame project aims to propose a systematic domain model of karstology comprising concept categories, relations and definition frames. Such a domain model allows us to build a knowledge base for karst using a comprehensive collection of relevant texts as the primary source, and employing advanced methods of text mining and natural language processing to extract the information we do not find in existing reference works for karst terminology.

The aim of this paper is to present the advantages of our frame-based and data-driven approach to describing and representing karstology, especially if compared to existing karst terminologies. In Section 2 we thus first describe past attempts to collect and describe karst terminology, then proceed with a more detailed description of the data sources and methods used in the TermFrame project in Section 3. Section 4 presents the main outputs, namely the annotated collection of definitions which serves as the basis for the structured knowledge base, and its potential uses by experts, researchers, students and other karst enthusiasts. We conclude with a brief discussion and plans for future work.

2 KARST TERMINOLOGY – OVERVIEW OF EXISTING WORKS

Several attempts to organise international karst terminology have been made in the past. Among the first such attempts was the Glossary of Karst Terminology by Watson H. Monroe (Monroe, 1970). According to its preface, this glossary includes mostly terms used in describing karst geomorphologic features and processes as used in the literature of English-speaking countries, but a few of the more common terms in French, German, and Spanish are included, with references to the corresponding English terms where they are available. The glossary also includes simple definitions of the more common rocks and minerals found in karst terrain, common terms of hydrology, and a number of the descriptive terms used by speleologists. The glossary contains around 450 terms.

Unesco's Glossary and Multilingual Equivalents of Karst Terms (1972) was launched by the General Conference of Unesco in order to promote cooperation in scientific hydrology research around the world. The glossary includes 227 terms with definitions in English, and translation equivalents in eight languages (French, German, Greek, Italian, Spanish, Turkish, Russian and Yugoslav⁴). In addition, it incorporates a classification of karst terms in a separate chapter.

3 Basic research project funded by the Slovenian Research Agency under grant J6-9372, 2018-2021.

4 The authors referred to two of former Yugoslavia's official languages, Serbo-Croatian and Slovenian; Yugoslav as a language does not exist.

In 1990s, *Cave and Karst Terminology* (Jennings, 1997) was published as a result of the efforts of the Australian Speleological Federation (Matthews, Matthews, 1968; Jennings, 1979; Australian karst index, 1985). The glossary is a highly selective list of terms recommended for use within the borders of Australian karst research and does not aim to be a comprehensive collection of global karst terminology.

About the same time, the British Cave Research Association (B.C.R.A.) published and updated a dictionary that covers the general area of karst and caves, namely the *Dictionary of Karst and Caves: A Brief Guide to the Terminology and Concepts of Cave and Karst Science* (Lowe, Waltham, 2002).

Since then, many new terms related to karst in general have come into use throughout the world mostly related to the upsurge in environmentalism. A *Lexicon of Cave and Karst Terminology with Special Reference to Environmental Karst Hydrology* (Field, 2002) was published by the U.S. Environmental Protection Agency with the aim to unify karst terminology and serve as a technical guide for karst researchers. It includes karst-specific terms and terms related to the field of environmental karst.

Since Slovenian karst terminology is an important part of international terminology and karstology is among the few scientific disciplines that originate from Slovenian territory, we would expect important works by Slovenian authors in this field. Nevertheless, only three basic works covering the field of karstology have been published so far (Gams et al., 1962; Gams, Kunaver, Radinja, 1973; Šušteršič, Knez, 1995).

The first attempt to collect and systemize karst terminology in Slovenian was made in the 1960s in the form of scientific article. It was presented as a report at the symposium organised by the Association of Slovenian Geographers and the Slovenian Geological Society held in 1962 on the topic of karst terminology. The article was written by Gams, Kunaver, Novak, Jenko and Savnik, and published in the *Geographical Bulletin*, the Association of Slovenian Geographers' official publication, under the simple title *Karst Terminology* (Gams et al., 1962). The contents of the article are divided into sections based on the short papers presented at the symposium, each paper addressing a subcategory within the karst domain, e.g. larger karst landforms, karst hydrology, karst caves etc. The terminology is thoroughly described and discussed in the Slovenian language, and approved by the symposium's programme committee.

This contribution was followed by the publication of *Slovenian Karst Terminology* (Gams, Kunaver, Radinja, 1973) a decade later. The all-encompassing collection of karst terminology in Slovenian was published under the auspices of the Department of Geography (Faculty of Arts, University of Ljubljana). Approximately 200 core dictionary entries, consisting of karst terms and their descriptions are often further elaborated and expanded by related karst expressions and definitions. The dictionary entries and all accompanying parts of the dictionary are presented in Slovenian. The majority of entries, however, include English, French and German translation equivalents as well. The dictionary remains the most important reference in terms of karst terminology in Slovenia to this day since it has never been fully revised yet.

The third important work covering karst terminology in Slovenian, A Contribution to the Slovenian Speleological Glossary, was published as a scientific article in the Bulletin of the Speleological Association of Slovenia by Šušteršič, Knez (1995). The collection includes the explanation of 88 terms with references to Slovenian Karst Terminology, Slovenian Technical Vocabulary and to the Dictionary of the Slovenian Language, focusing on the latest developments in the field of speleology and related scientific fields. The presented entries do not include translation equivalents.

In both international (English) and Slovenian karst terminologies, the authors attempted to be as inclusive as possible in that their glossaries incorporated terms related to karst geomorphology, speleology, hydrology, and karst rock geology. The glossaries usually include a sufficient number of terms describing karst and do not differ significantly from each other in terms of coverage. However, there are major inconsistencies in the description of terms, reflecting the author's expertise and focus which may be either geological, hydrological or geomorphological, but also sometimes resulting from hereditary citing of definitions from older sources (e.g.: *sifon je odsek rova, kjer sega skalni strop do vode / a syphon is a section of a passage where the cave ceiling is reaching water* (Gams, Kunaver, Radinja, 1973); *sifon je kolenasta poglobitev jamskega dna, kjer naj bi na krajšo razdaljo podzemna reka tekla ob pritisnjeni gladini / a syphon is a knee-shaped lowering of cave floor where a short section of the subsurface stream flows along a lowered watertable level* (Šušteršič, Knez, 1995). Furthermore, traditional definitions typically focus on one or two selected attributes of a term rather than presenting a comprehensive overview of all known attributes.

Our approach aims to overcome these drawbacks. Firstly, we rely on data-driven methods to determine the relevance of terms. This means that we first compiled a balanced and representative corpus of texts including the above-mentioned glossaries which we use to extract terms and definitions (see Section 3.2). Thus, our coverage is more comprehensive and less subjective. Secondly, the frame-based approach defines a definition template, a so-called "ideal definition" for each concept category in our domain model. This allows us to generate term descriptions which contain all known attributes of a term, even if these attributes are not explicitly mentioned in any of the definitions. Finally, our approach is not aimed towards building a glossary but a knowledge base, the main difference being that all karst concepts are parts of a large knowledge network where the underlying structures reveal true facts about the domain.

3 METHODS AND RESOURCES

3.1 Building the domain model

A systematic description of individual shapes, processes and materials is possible by combining existing geomorphological methods with a systematic and comprehensive approach. Amongst different approaches, we believe that the analytical geomorphological method (Pavlopoulos, Evelpidou, Vassilopoulos, 2009) is the most appropriate and the most systematic for the description of geomorphologic features and processes.

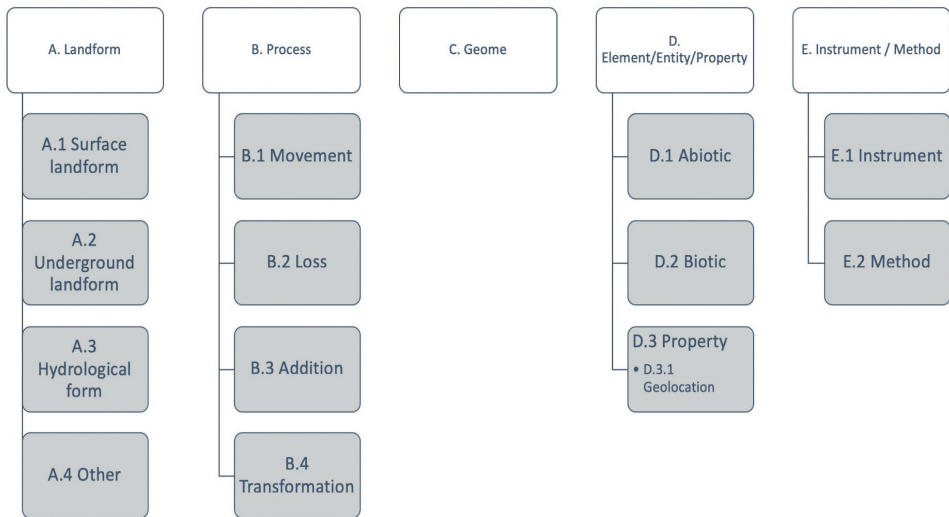
The analytical geomorphological method (Pavlopoulos, Evelpidou, Vassilopoulos, 2009) includes five basic aspects of analysis, namely morphographic or morphological, morphometric, morphogenetic, morphochronological, and morphodynamic. To this set of methods we added the morphostructural analysis (Gerasimov, 1946) which is not included in the classical analytical geomorphological approach, but is also crucial from the point of view of an integrated geomorphological approach.

The morphographic (or morphological) analysis contains the identification and qualitative description (documentation) of geomorphic forms and their distribution in the studied area or characteristic environment (geome) of occurrence. The morphometric analysis refers to the quantitative description of geomorphic aspects. The morphostructural analysis is a set of methodological approaches aimed at explaining the direct or indirect connections between today's relief and the structure of the Earth's interior, or to determine important elements of geological structures in the study area (Gerasimov, 1946). The morphogenetic analysis is a detailed description of the formation of geomorphic forms and includes processes, morphogenetic systems and mathematical simulations of relief design. The morphochronological analysis is the determination of the age of an individual geomorphic form on the basis of absolute and relative dates, correlations of sediments and geomorphic forms on the basis of their age and position. The morphodynamic analysis includes all the dynamic processes on Earth that form a relief. It is a study of geomorphic processes operating today and those processes that will be active in the future.

Top-level categories (Figure 1) indicate the type of individual elements in terms of geomorphological form (A. Landform) or process (B. Process). Since typical geomorphological or hydrological environments also appear in definitions, we defined them as geomes (C. Geome). In addition, we also encounter landforms, materials and their characteristics that are not directly related to karst geomorphology or hydrology but still contribute to domain knowledge (D. Element / Entity / Property), as well as methods of study (E. Instrument / Method). All elements are divided into subcategories according to their spatial distribution (A.1 Surface landform, A.2 Underground landform) and according to the predominant hydrological function (A.3 Hydrologic landform). Forms that are directly related to karst and could not be classified in any

of the above subcategories were labelled as such (A.4 Other). We also divided the processes according to their mode of operation into transport (B.1 Movement), erosion or denudation (B.2 Loss), accumulation and aggradation (B.3 Addition) and transformation (B.4 Transformation). Abiotic (D.1 Abiotic) and biotic (D.2 Biotic) forms and processes and their characteristics (D.3 Property) were classified in category D. Under this category, we also include geolocation (D.3.1 Geolocation), which is of special importance in understanding karst geomorphology and hydrology. In the last category (E.) we used two subcategories that define the methods of study to instruments (E.1 Instrument) and methods (E.2 Methods).

Figure 1: Structure of concept categories in the TermFrame domain model.



The second step involved determining the semantic relations governing knowledge structures in karst. The relations were partly taken from the EcoLexicon⁵ but adapted to karstology upon examination of corpus evidence. Our final version of the domain model defines the following 15 relations, of which some occur with a very low frequency: HAS_FORM, HAS_SIZE, COMPOSITION_MEDIUM, HAS_CAUSE, HAS_TIME_PATTERN, HAS_FUNCTION, HAS_LOCATION, HAS_POSITION, AFFECTS, HAS_RESULT, CONTAINS, MEASURES, STUDIES, DEFINED_AS, HAS_ATTRIBUTE.

Relations are more or less closely tied to a more detailed interpretation of individual categories. The category Landform (A.) invokes relations linked to the geomorphological analytical method (Pavlopoulos, Evelpidou, Vassilopoulos, 2009) and

5 https://ecolexicon.ugr.es/visual/index_en.html

defines morphographic (HAS_FORM), morphometric (HAS_SIZE), morphostructural (COMPOSITION_MEDIUM), morphogenetic (HAS_CAUSE) and morphochronologic (HAS_TIME_PATTERN) attributes of surface, subsurface and hydrological karst features. In addition to the relations that are closely associated to the geomorphological analytical method, we also use relations which spatially associate the categories with geomes (HAS_LOCATION) and geolocations (HAS_POSITION). The category of karst processes (B.) invokes semantic relations connected to the effects and results of these processes (AFFECTS and HAS_RESULT). The category of geomes (C.) is usually tied to the characteristic landforms, materials or groups of processes that shape them, so in addition to other semantic relations, their definition frequently lists typical karst elements they encompass (CONTAINS). The category defining the activities related to karst studies (E.) invokes the relations defining those activities (MEASURES and STUDIES). The category defining forms, processes and characteristics that are not directly related to karst geomorphology or hydrology (D.) may invoke all the listed semantic relations. In the event that semantic relations denote any other property of categories, we have defined them generally (DEFINED_AS, HAS_ATTRIBUTE).

The typical and expected combinations of categories and relations explained above constitute *frames*; cognitive templates which represent fragments of specialized knowledge about the domain (Faber et al., 2012).

3.2 Resources

Within the project we built English, Slovene and Croatian specialised corpora⁶. All three corpora are comprised of relevant contemporary works on karstology which were carefully selected. The corpora include specialised texts (books, articles, doctoral and master's theses, glossaries and dictionaries) from the field of karstology, whereby individual works partly overlap with one or several related fields such as geomorphology, geology, hydrology, speleology, biology etc.

	English	Slovene	Croatian
Tokens	2,386,075	1,208,240	1,229,368
Words	1,968,509	987,801	969,735
Sentences	87,713	51,990	53,017
Documents	54	60	43

Since the exploration of differences between the international karst terminology in English and local Croatian and Slovene terminologies lies at the core of our project, we took great care to include all major reference works in English, e.g. Karst

6 A corpus in linguistics is a digital collection of texts selected according to specific criteria in order to represent a language or language variety.

Hydrology and Geomorphology (Ford, Williams, 2007), Karst Hydrology and Physical Speleology (Bögli, 1980), Encyclopedia of Caves and Karst Science (Gunn, 2004) as well as other relevant works published in the past four decades of karst research (see Section 2). For Croatian and Slovene karstology, fewer comprehensive books had been published, we therefore included more PhD theses and scientific articles.

For definition extraction we used the Clowflows definition extractor (Pollak et al., 2012). The tool tries to identify sentences which could be definitions on the basis of various language-specific patterns, e.g. *X is a subtype of Y which [...]*. The definition candidates were later manually validated and only examples with valuable explanatory information about karst concepts were retained (yield ~ 20%). All definitions types (intensional, extensional, functional, paraphrase etc.) were considered, therefore not all obtained definitions have the traditional structure: the definiendum may appear in different positions in the sentence, the genus may or may not be present, the term may be defined only through its hyponyms etc. After validation the yield was 215 and 259 definitions for English and Slovene respectively.

3.3 From definitions to structured knowledge

As pointed out in Section 3.1, a systematic approach to describing karst phenomena would propose for each category of concept (e.g. Surface landform, Underground landform, Process etc.) a set of attributes which need to be specified in order to make the description complete. Such attributes include SIZE, FORM, CAUSE, COMPOSITION, FUNCTION, LOCATION or RESULT, but they vary depending on the type of concept we are describing. Thus, a surface landform should ideally be described through its FORM, SIZE, CAUSE, LOCATION and COMPOSITION or MEDIUM in terms of typical geological and/or geographical environment, but it will almost never be described through its FUNCTION or RESULT, as these can be expected in more dynamic karst entities such as hydrological forms and processes.

We can see from the example below that definitions in existing reference works focus on different aspects of the definiendum, but rarely list all of them. In a), *bedding-plane cave* is defined through its SIZE (*has not enlarged by growth into a major tube or canyon*) and LOCATION (*remained almost entirely on the bedding plane*). In b), we have LOCATION and CAUSE (*difference in susceptibility to corrosion in the two beds*), and in c), we have LOCATION and FORM (*elongate in cross-section*).

- a) The term bedding-plane cave is strictly applied to a passage that has not enlarged by growth into a major tube or canyon, but has remained almost entirely on the bedding plane.
- b) bedding-plane cave: A passage formed along a bedding plane, especially when there is a difference in susceptibility to corrosion in the two beds.
- c) bedding-plane cave: A cavity developed along a bedding-plane and elongate in cross-section as a result.

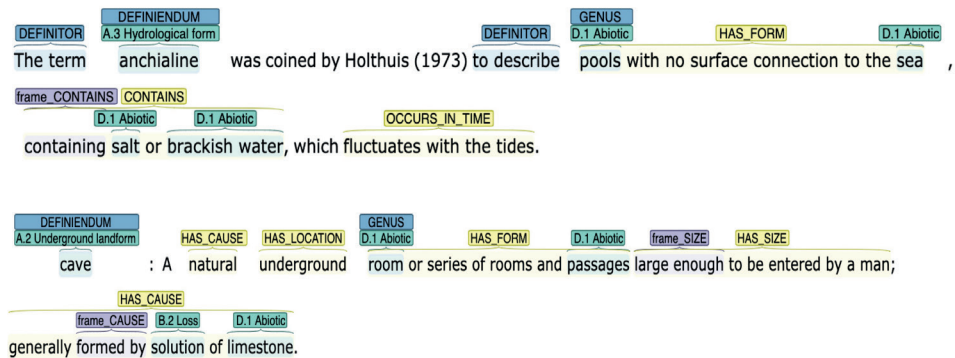
Our aim is to overcome such limitations of “natural” definitions and aggregate knowledge from different sources in order to create the most comprehensive concept description possible. The definitions we collected from different sources were loaded into the WebAnno annotation environment (Castilho et al., 2014) and manually annotated on several levels. For each definition we mark:

- the definition elements: DEFINIENDUM, GENUS, DEFINITOR
- concept categories: e.g. Surface landform, Underground landform (see Figure 1)
- relations describing the concept, e.g. FORM, SIZE, CAUSES, LOCATION (see Section 3.1)

Each definition was annotated by two persons and any discrepancies between the two annotators were later resolved by a domain expert. In addition to this, regular meetings of annotators and domain experts took place in order to discuss borderline cases and ensure the consistency of annotations.

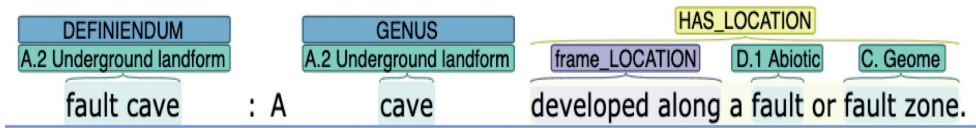
The two examples below illustrate the result of multi-level annotation, where the term *anchialine* is defined through its form (*pools with no surface connection to the sea*), its contents (*salt or brackish water*) and its time pattern (*fluctuates with the tides*), and *cave* is defined through its origin or cause (*natural; formed by solution of limestone*), location (*underground*), form (*room or series of rooms and passages*) and size (*large enough to be entered by man*).

Figure 2: Examples of annotated definitions for *anchialine* and *cave*.



Despite the care taken to produce consistent and logical annotations, many contexts may have multiple meanings or could be assigned different relations. In the example below, the *fault cave* is defined through its location (*developed along a fault or fault zone*), but this also indicates the cause of its formation. In such cases the decision was to retain the most overt meaning and not to assign double or triple relations to the same part of a sentence.

Figure 3: Example of annotated definition for fault cave

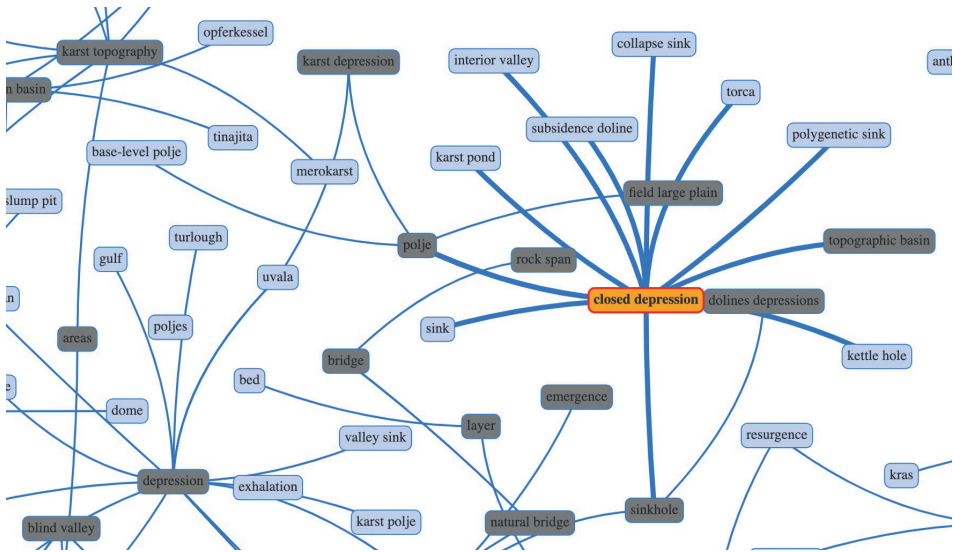


4 RESULTS

At the time of writing this article, the annotation of English and Slovene definitions is complete and for Croatian still in progress. The English data set contains 844 defined terms and the Slovene one 903. For many karst terms the data set contains several annotated definitions which allows us to combine different attributes and generate a more comprehensive description of the concept.

The multi-layered and multilingual annotated database of definitions allows us to explore patterns of knowledge on a large scale, and to compare conceptualisations across languages. Using the visualization tool NetViz (Pollak et al., 2020) which was developed specifically for the purposes of this project, we can draw graphs of the entire knowledge network or just of selected parts thereof. A visualization of the entire network of terms and their categories (Figure 4) will help the expert identify the most common groups of karst concepts and explore their members. For English, the largest group is centered around the category Underground landforms, followed by Surface landforms, Abiotic and Hydrological forms. Looking at the network for Slovenian (Figure 5), we can see that the category of Geomes is more productive than in English, with 156 members as opposed to 103 in English.

Figure 6: Part of the network around depression and closed depression.



Finally, the structured knowledge base allows us to explore the relevant analytical aspects of selected terms as they are typically expressed in different languages. Thus, the surface landform *uvala* is in English defined mainly through its morphographic characteristics (*with undulating floors, floored by sinkholes, large depression*), and the morphogenetic aspect is also provided (*coalescence of several dolines*). In Slovenian, the focus is also on the morphographic attributes (*v tlorisu nepravilnih oblik, v obliki skledaste vdolbine, dolasta ali vrtačasta*), the morphometric attribute specified in relation to its related form (*manjša od kraškega polja*), while the morphogenetic aspect is not present.

Figure 7: Uvala and its attributes in English.

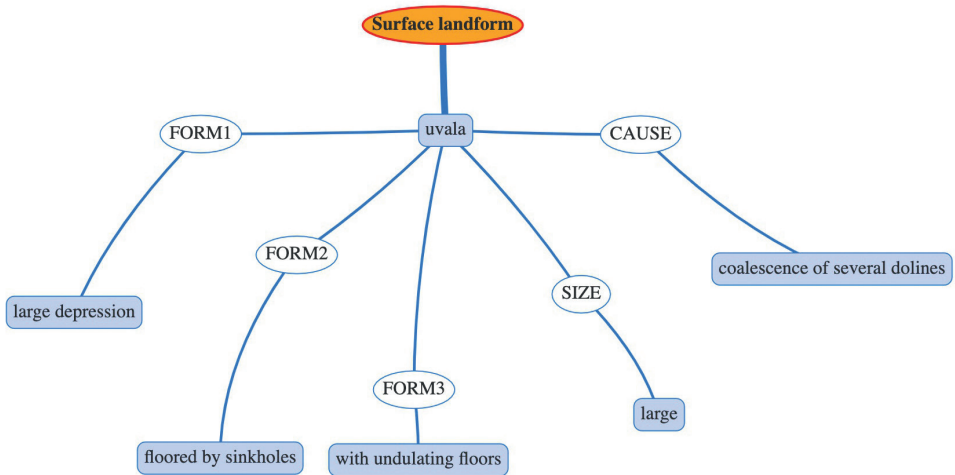
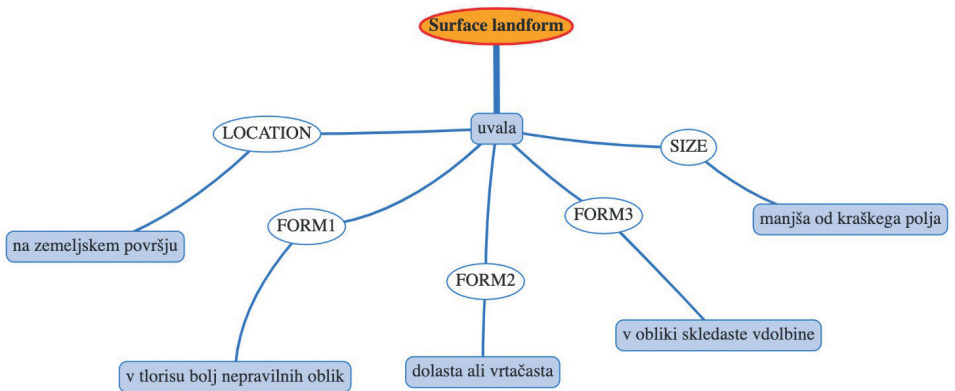


Figure 8: Uvala and its attributes in Slovene.



By using the systematic domain model which predicts the typical attributes for each category of karst concept, we can generate structured and complete descriptions which inform the user of the most salient properties (Table 1). Such a structured knowledge base also allows us to query according to specific criteria, e.g. surface landforms above a specific size or landforms caused by movement of material.

Table 1: A structured description of škraplja extracted from several definitions.

A.1 Surface landform	škraplja
HAS FORM	<i>razpoka med bloki kamnin</i>
HAS SIZE	<i>od nekaj centimetrov do več decimetrov ali metrov</i>
HAS CAUSE	<i>nastane s korozijo vode</i>
COMPOSITION_MEDIUM	<i>izjedena v trdi kamnini (apnencu ali drugih karbonatnih kamninah)</i>
HAS LOCATION	<i>na površju ali subkutano</i>

For the concepts where the complete set of attributes cannot be retrieved from annotated definitions, several experiments using state-of-the-art text mining and natural language processing techniques are underway in order to extend the manually constructed database and discover new elements of karst knowledge (Miljkovic et al., 2019; Vintar et al., 2020).

5 CONCLUSIONS

We presented the contribution of the TermFrame project towards a comprehensive representation of karst terminology and knowledge. The laborious and complex procedure of compiling the corpora, constructing the domain model, annotating definitions and aggregating knowledge into the final knowledge base required the concerted efforts of an interdisciplinary and multilingual team of experts, including linguists, terminologists, karst researchers, computer scientists and cognitive linguists.

The planned output of the project is a public website delivering the main results of the project through a user-friendly web interface. The basic level of information will provide search and browse functions through the TermFrame Karst Knowledge Base in all three languages. Upon submitting a query, the user will be presented with all the definitions of the query term from different sources, their synonyms and also graphic material. The basic level will be intended primarily for a wider audience and lower grade students interested in karst. Another level of querying the knowledge base will show a visual representation of the relationships between terms (categories) and semantic relations, thus providing the user with a more detailed and comprehensive overview and allowing for comparisons between languages.

For the most salient karst terms (*cave, polje, ponor* etc.), the user will also be offered a map displaying all the toponyms pertaining to the particular landform and their locations. The automatic creation of such maps is made possible through automatic named entity extraction from our corpora and automatic linking with the geolocations from GeoNames.org.

Acknowledgment

This research was funded by the Slovenian Research Agency under grant number J6-9372, 2018-2021, TermFrame: Terminology and Knowledge Frames Across Languages, <https://termframe.ff.uni-lj.si>.

References

- Australian karst index. 1985. Melbourne: Australian Speleological Federation.
- Bögli, A., 1980. Karst hydrology and physical speleology. Berlin, Heidelberg, New York: Springer-Verlag.
- Castilho, R. E. d., Biemann, C., Gurevych, I., Yimam, S., 2014. WebAnno: a flexible, web-based annotation tool for CLARIN. Soesterberg: CLARIN Annual Conference (CAC), pp. 1–6.
- Faber, P., Tercedor, M., Montero Martínez, S., Araúz, P., Prieto Velasco, J. A., Lopez-Rodriguez, C., Reimerink, A., Linares, C., De Quesada, M., Gómez-Moreno, J., San Martín, A., 2012. A cognitive linguistics view of terminology and specialized language. Berlin, Boston: De Gruyter Mouton. DOI: 10.1515/9783110277203.
- Field, M. S., 2002. A lexicon of cave and karst terminology with special reference to environmental karst hydrology. US Environmental Protection Agency.
- Ford, D., Williams, P. D., 2007. Karst hydrogeology and geomorphology. Chichester: Wiley.
- Gams, I., Kunaver, J., Novak, D., Jenko, F., Savnik, R., 1962. Kraška terminologija. Geografski vestnik, 34, pp. 115–137.
- Gams, I., Kunaver, J., Radinja, D., 1973. Slovenska kraška terminologija. Ljubljana: Katedra za fizično geografijo, Univerza v Ljubljani.
- Gerasimov, I., 1946. Opyt geomorfologičeskogo strojenija SSSR. Problemy fizičeskoj geografii, 12, pp. 33–46.
- Glossary and multilingual equivalents of karst terms, 1972. Paris: UNESCO.
- Gunn, J., 2004. Encyclopedia of caves and karst science. New York, London: Fitzroy Dearborn.
- Jennings, J. N., 1979. Cave and karst terminology. AFS Newsletter, 83, pp. 3–14.
- Jennings, J. N., 1997. Cave and karst terminology. Australian Speleological Federation.
- Kranjc, A., 1994. About the name and the history of the region Kras. Acta Carsologica, 8, pp. 82–90.
- Kranjc, A., 2008. Kraška terminologija - pojmi z dinarskega krasa. Geografija v šoli, 17, pp. 3–10.
- Lowe, D., Waltham, T., 2002. Dictionary of karst and caves: A brief guide to the terminology and concepts of cave and karst science. British Cave Research Association.
- Matthews, P., Matthews, P. G., 1968. Speleo handbook. Sydney: Australian Speleological Federation.

- Mihevc, A., 2010. Geomorphology. In: Mihevc, A., Prelovšek, M., Zupan Hajna, N. (ed.). Introduction to the Dinaric karst. Postojna: Inštitut za raziskovanje krasa ZRC SAZU, pp. 30–43.
- Miljkovic, D., Kralj, J., Stepišnik, U., Pollak, S., 2019. Communities of related terms in a karst terminology co-occurrence network. *Sintra: eLEX : Electronic lexicography in the 21st century*, pp. 357–373.
- Monroe, W. H., 1970. A glossary of karst terminology. Washington D.C.: U.S. Geological Survey.
- Pavlopoulos, K., Evelpidou, N., Vassilopoulos, A., 2009. Mapping geomorphological environments. Berlin, Heidelberg: Springer.
- Pollak, S., Podpečan, V., Miljkovic, D., Stepišnik, U., Vintar, Š., 2020. The NetViz terminology visualization tool and the use cases in karstology domain modeling. *Marseille: The International Workshop on Computational Terminology COMPUTERM 2020 at LREC 2020*, pp. 55–60.
- Pollak, S., Vavpetič, A., Kranjc, J., Lavrač, N., Vintar, Š., 2012. NLP workflow for online definition extraction from English and Slovene text corpora. *Vienna: KONVENS 2012*, pp. 53–60.
- Šušteršič, F., Knez, M., 1995. Prispevek k slovenskemu speleološkemu pojmovniku. *Naše jame*, 37, pp. 153–170.
- Vintar, Š., Grcic, L., Martinc, M., Pollak, S., Stepišnik, U., 2020. Mining semantic relations from comparable corpora through intersections of word embeddings. *Marseille: Proceedings of the 13th Workshop on Building and Using Comparable Corpora*, pp. 29–34.

TERMFRAME: SISTEMATIČEN PRISTOP H KRAŠKI TERMINOLOGIJI

Povzetek

Prispevek predstavlja pomemben doprinos k urejanju, sistematizaciji in vizualizaciji terminologije na področju krasoslovja, opisani rezultati pa so nastali v okviru raziskovalnega projekta TermFrame: Terminologija in sheme znanja v večjezičnem prostoru. Sodobna terminološka veda se pri opisovanju specializiranega izrazja ne naslanja več zgolj na tradicionalni pojmovni pristop, ampak skuša znanje izbranega strokovnega področja predstaviti v obliki pojmovnih struktur, ki ustrezajo kognitivnim shemam kot podlagam za ekspertno znanje. Opisano teoretično izhodišče, ki je znano kot terminologija shem (frame-based terminology), je uporabljeno na področju krasoslovja za gradnjo obsežne baze znanja, ki poleg specializiranih terminov in njihovih definicij vsebuje tudi pojmovne sheme in iz njih izhajajoči kognitivni model domene.

V prispevku uvodoma opišemo dosedanja terminološka prizadevanja na področju krasoslovja, pri čemer pregledno in jedrnato zajamemo vidnejša slovarska, glosarska in leksikonska dela v angleščini in slovenščini. Nato predstavimo oblikovanje domenskega modela, kar v praksi pomeni oblikovanje hierarhične strukture pojmovnih kategorij in pomenskih relacij, ki jih potrebujemo za opisovanje temeljnih atributov krasoslovnih pojmov. Ob tem se izkaže, da so metode terminologije shem, ko jih implementiramo na krasoslovje, presenetljivo skladne z geomorfološko analitično metodo.

Ker je namen baze znanja, da odraža neidealizirano in avtentično podobo krasoslovja, kot ga opisujejo strokovne in znanstvene objave različnih avtorjev, smo za namene projekta zgradili obsežen in reprezentativen korpus besedil v angleščini, slovenščini in hrvaščini, za luščenje podatkov iz korpusa pa uporabljamo najsodobnejše metode besedilnega rudarjenja in jezikovnih tehnologij. Tako smo iz korpusa za vsak jezik posebej izluščili zbirko terminov in njihovih definicij, v naslednjem koraku pa smo vsako definicijo analizirali in označili s kategorijami in relacijami domenskega modela.

Glavna prednost takšnega pristopa je, da krasoslovnega pojma ne opišemo več le s klasično stavčno definicijo, ampak z vnaprej določenim naborom atributov, ki tipično pripadajo posamezni pomenski kategoriji. Tako denimo za opis površinske kraške oblike pričakujemo navedbo oblike, velikosti, lokacije, nastanka in sestave; ta pričakovani nabor pa predstavlja strukturirano shemo znanja.

Prispevek v zadnjem razdelku podaja primere vizualizacij pojmovnih struktur, pri čemer med jeziki prihaja do različnih odstopanj. Za razliko od formalnih ontologij, ki skušajo znanje in razmerja med pojmi posplošiti do jezikovno, regionalno in kulturno neodvisne reprezentacije, je večjezična krasoslovnna baza znanja TermFrame odraz resničnih in avtentično izpričanih strokovnih razlag in stališč, ki se med avtorji, jeziki in kulturami skorajda nujno razlikujejo, vpogled vanje pa bogati razumevanje področja in olajšuje strokovno komunikacijo.