THE DEVELOPMENT OF THE GEOLISSTERM TERMINOLOGICAL DICTIONARY

Ranka Stanković[1], Branislav Trivić[2], Olivera Kitanović[3], Branislav Blagojević[4], Velizar Nikolić[5]

[1, 2, 3] Faculty of Geology, University of Belgrade, [4] SEE, Belgrade
[5] Ministry of the Environment, Mining and Spatial Planning of the Republic of Serbia

Translated from Serbian by Jelena Bajić

Abstract: This paper presents the structure and implementation of the electronic dictionary of geologic terms (GeolISSTerm) as a special-purpose taxonomy of basic geologic concepts and terms. GeolISSTerm can be an elementary electronic resource in the process of domain formation in the Geologic Information System of Serbia (GeolISS). As such, it is the core of GeolISS through which validation, classification and specification of attributes of the observed and the interpreted takes place.

GeolISSTerm is a result of the work on the project *The Development of Geologic Terminology and Nomenclature for the Geologic Database of Serbia* carried out by the Faculty of Mining and Geology, University of Belgrade, funded by the Ministry of the Environment and Spatial Planning of the Republic of Serbia. A short history of GeolISSTerm and an overview of the structure and the way of arranging terms, in view of the terminological resources and the position of the GeolISSTerm dictionary in the semantic spectrum is followed by a description of the approach to the development of the geologic terminology and the structure of the vocabulary using the UML (*Unified Modeling Language*) model. The concept of domain control in the Geologic Information System of Serbia has been explained in detail. Software solutions for browsing and searching the dictionary online and the solution for domain control, namely, integration into GeolISS are illustrated by screenshots of the interface. The possibilities of exporting the vocabulary to standard ontology languages, such as the OWL, modelled on the GeoSciML initiative are planned to be provided. The electronic edition of the dictionary is complemented by the printed version.

Keywords. Terminological resources, geology, GIS, Geologic Information System, geologic vocabulary, electronic dictionary.

[1]ranka@grf.bg.ac.rs, [2]btrivic@rgf.bg.ac.rs,[3]olja@rgf.bg.ac.rs,
[4]bsblagojevic@yahoo.com, [5]velizar.nikolic@ekoplan.gov.rs

1. Introduction

The physical implementation of the Geologic Information System of Serbia (GeolISS) in mid-2006 marked the beginning of development of the geologic terminology and nomenclature. The main aim of the development of this electronic resource was the creation of a standard dictionary of geologic terms for the purpose of providing logically consistent description, interpretation and classification of geological units, geologic structure, mineral and hydrological resources and other geologic features in the domain of applied disciplines by using GeolISS.

The development of the electronic dictionary of geologic terms intensified in 2008 as part of a separate GeolISS project named *The Development of Geologic Terminology and Nomenclature for the Geologic Database of Serbia*. The goal of this project was to add information on the existing geologic concepts and terms and expand the dictionary by adding new ones, primarily those from specialized fields of geology that were not present in the initial version to a satisfactory degree. The work on the dictionary continued after the project ended and the current version containing around 3800 concepts in Serbian and English is a functional geologic terminology resource that will be capable of meeting the needs of the Geologic Information System of Serbia – GeolISS (Blagojević, 2007). Experts in a variety of geologic disciplines have contributed to the work on terminology[1]. Although GeolISSTerm has been developed for the purpose of data processing, the electronic dictionary is complemented by a print edition (Trivić, 2011).

Special attention has been paid to the structure of concepts and the manner of their organization in the terminological resources and the GeolISSTerm dictionary, as well as to their position on the semantic spectrum (Section 2). In addition to a detailed description of the approach to the geologic terminology and the UML model of vocabulary structure, Section 3 presents the concept of controlling the allowed values of domains in the geologic information system. The developed software solutions are shown in Section 4. It features a description of not only the tools used to develop the dictionary and the application for browsing and searching its content online, but also the component for domain control, namely, integration into GeolISS. The application of GeolISSTerm as a bilingual resource is illustrated by automatic map symbolization in Serbian and English. The directions for further development of this important terminological resource are proposed in the closing section.

[1] http://geoliss.ekoplan.gov.rs/term/OProjektu.aspx

2. Terminological resources

The resources intended for data processing can be organized in a variety of ways and serve different purposes. The primary, although not the only function of GeolISSTerm is controlling the allowed values of domains. A controlled vocabulary is a carefully selected collection of terms (words and phrases) chosen to suit a special purpose determined by its author. These vocabularies are used for tagging information in databases and documents, making it easier to locate them when performing searches.

Depending on their content, structure and organization, vocabularies, as terminological resources can, in a general sense, be viewed as indexes, glossaries, taxonomies, thesauruses and ontologies. Although the boundaries between these types are not always clear-cut, each is an extension of the one that precedes it. The semantic scale (Figure 1) is based on the ontological spectrum from Daconta M.C. (2003) and Deborah L. McGuinness (2003). The starting point in the lower left corner is an **index**, as the simplest list of terms (usually arranged in the alphabetical order), followed by a glossary, which, like an index contains definitions of terms.

Glossaries are lists of terms with definitions that can be monolingual, bilingual or multilingual. In the case of a multilingual linking of terms, parallel multilingual or bilingual lists, namely, lists of paired terms are often made. These lists usually do not contain definitions of terms.

Further arrangement of the resources on the semantic scale (from weakly to strongly represented semantic descriptions) is aimed at the introduction of relationships among terms, namely, concepts represented by concrete terms. The primary and basic relationships among concepts represented by the terms are: hypernym and hyponym that can also be viewed as the broader and the narrower term respectively. The terms in the hierarchy are linked via the *is a* relation which can be:

• The hyponym of the same type as the hypernym
• The hyponym is part of the hypernym
• The hyponym is an instance of the hypernym

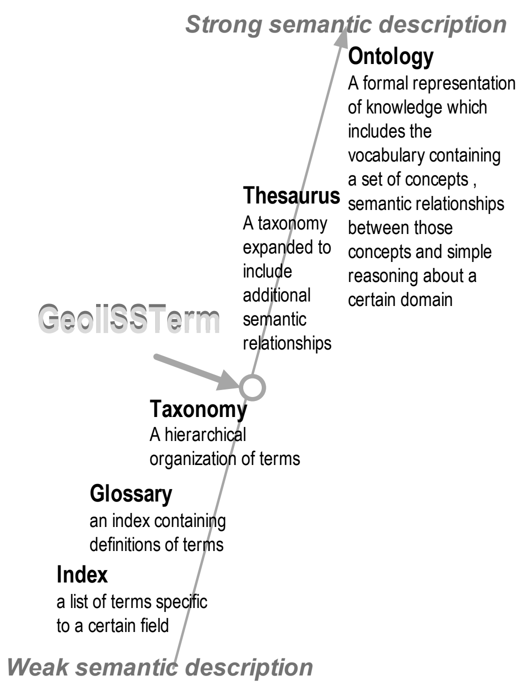


Figure 1. Semantic scale of terminological resources

A **taxonomy** is a hierarchical classification of concepts, things, objects, beings, events or principles that are classified. A trunk or tree is a structure that makes it possible to define a hierarchical relation between concepts. The parent-child relations within a tree represent the hyponym - hypernym relationship between the concepts related to terms. More elaborate relationships between concepts that include meronyms, antonyms, derived terms, etc. cannot be represented by using a tree structure, which is why graph structures are applied to describe semantic relations. Such structures are encountered in many thesauruses, semantic networks and ontologies.

Taxonomies in the broadest sense of the term involve the study of the general principles of scientific or systematic classification. The notion of a taxonomy or systematic classification usually implies an ordered classification of plant and animal world, following the supposed natural relations existing in it. In information science, taxonomies are products of engineering rather than just an abstract result of human thought. The understanding of taxonomies in information science is therefore different and they are viewed as classifications of entities in the form of hierarchies in accordance with the assumed relationship between real world objects represented by these information objects.

Thesauruses are above taxonomies in the semantic spectrum. They can be defined as vocabularies structured in such a way so as to display equivalence, homography, hierarchy and association relations among terms in a clear manner and allow their easy recognition through standard indicators. The primary role of a thesaurus is to facilitate finding documents and achieve consistency in the indexing of stored documents or records in a database. In order to meet this requirement, thesauruses usually link the approximate meaning of one term to that of another. The terms are most often confined to a single domain, for example military equipment. The controlled vocabulary has been conceived to support finding information, while at the same time providing assistance to the individual who associates the terms, namely indexes a document (or a file) in the database or an individual who wants to find or search information with the help of the same terms used for indexing. All this leads to the conclusion that thesauruses are primarily intended to be used by humans, although they are used in the IT environment.

The definitions of the boundary between a thesaurus and an ontology in the literature differ, ranging from those stating that ontologies and thesauruses are one and the same, the difference being only in their respective purposes, to the ones that regard ontologies as only those resources that allow derivation of new knowledge from the existing pieces of knowledge.

It should be kept in mind that both in the literature and in practice, the notion of ontology often includes structures ranging from simple taxonomies with minimal hypernym or hyponym structure, taxonomies containing words and synonyms, conceptual models featuring complex knowledge all the way through to theories of logic with rich, complex, consistent and meaningful knowledge.

Based on all of the above, it could be said that an **ontology** defines the words and concepts used for describing a field of knowledge, thereby standardizing meanings. Ontologies are used by humans, databases and applications sharing information from the same domain. They code the knowledge from a single domain, as well as the knowledge that covers several domains. In this way, ontologies make it possible to reuse coded knowledge. An ontology consists of:

– Classes (or general templates of entities) belonging to different domains of interest

– Instances (individual entities)

– Relations between these entities

– Properties and the values of properties of these entities

– Functions and processes that these entities are included in

– Constraints and rules that refer to these entities

Therefore, an ontology can be viewed formal representation of knowledge which includes the vocabulary containing a set of concepts, semantic relationships between those concepts and simple reasoning about a specific domain. It can also be regarded as a data model describing a certain field, namely, a domain and is used for drawing inferences on the basis of the information stored in the objects within that domain and the interconnections existing between them. Ontologies are used in important fields of computer and information science, such as artificial intelligence, semantic web, information systems, etc., as a form of representation of knowledge about the world or some part of it. Generally speaking, ontologies describe: objects at the basic level, classes as sets, collections or types of objects, as well as attributes as properties, the characteristics or parameters that can be shared by objects and the relations interconnecting objects.

The electronic lexical database SWN (Serbian WordNet), as a semantic network of words of Serbian (Krstev et al., 2008) is indispensable in terms of the development of terminological resources with a rich semantic description. The first network of this kind was built for the English language at Princeton and provided the basis for the creation of WordNet for many other languages. The versions of WordNet built as part of European projects EuroWordNet and BalkaNet, among which the Serbian variety have been aligned with the Princeton WordNet, which allows their use in many multilingual applications (Obradović et al., 2008).

WordNet has become a *de facto* standard for semantic networks based on the premise that in the human mind, words, as the fundamental elements of language, group around concepts, abstract ideas, namely, mental symbols. Concepts include objects belonging to a certain category, domain or a class of entities, interactions or phenomena. One of the domains included in the English WordNet is geology, represented by 625 concepts. However, the selection and range of geology-related concepts and the relationships between them did not meet the specific requirements of GeolISS.

On the semantic scale populated by different terminological resources and properties, the current implementation of GeolISSTerm is positioned between a taxonomy and a thesaurus, although the data model allows storing richer semantic relations. Further development will be focused on enriching the semantic relations existing between the concepts already included in the vocabulary, such as meronymy, antonymy, etc., as well as "is_used for", "is derived_from" and those specific to geology, which will improve the position of GeolISSTerm on the semantic scale, promoting it to the status of an ontology.

3. The GeolISSTerm data model

3.1. The approach to the development of geologic terminology

The initial ideas about the development and role of terminology in the geologic information system revolved around a variety of views, from the opinion that lists of geology terms for each domain should be made or favouring the idea of simply taking the terms from *Geologic Terminology and Nomenclature* edited by Kosta Petković, through to the belief that taking the glossary from *Glossary of Geology* (Bates, and Jackson, Ed., 1995) would be the best course of action. These ideas ultimately resulted in a partial analysis of the book *Geologic Terminology and Nomenclature* (Petković, 1975) and an analysis of the terms used when making certain sheets of the General Geologic Map (GGM) (fieldwork logs, map legends and the descriptions found in them) and an analysis of the available terminology compiled for geologic information systems (U.S. Geological Survey - USGS, California Geological Survey - CGS, British Geological Survey - BGS, International Union of Geological Sciences - IUGS).

The results of these analyses showed *Geologic Terminology and Nomenclature* (Petković, 1975) to be undoubtedly the most comprehensive local geologic publication, but it also demonstrated that it abounded in synonyms, terms with different, often contradictory meaning, homonyms of the same meaning, archaisms whose etymology is not clear enough or is even unknown, imprecise definitions of certain terms, etc. The terminology used in the GGM is very abundant, but the descriptions made at the geologic observation points are sometimes too elaborate or too short and cryptic, the explanations contain a rather large number of commonplaces, without references to concrete data.

As far as the terminologies developed for geologic information systems are concerned, the American standard of scientific vocabulary for the digital geologic map database (Soller, D.R., 2004) is the most thorough and in many ways it is a terminological system. It focuses not only on the classification of the geologic material, but also on very precisely defined terms that are used to describe the properties of that material, including its genesis. The terminology of the Geological Survey of Canada, originally developed independently of others (Struik, et. al., 2002) features a strict petrologic classification of rock material together with the properties collected from different, primarily cartographic documents. This is how a simple search, comparison and grouping of rocks or the units where they are present is made possible, in the database models that support ontologies, on the bases of a single key word or a number of properties. This approach was later incorporated into the terminological system in the United States too.

The classification of rocks of the British Geological Survey (Gillespie, and Styles, 1999; Hallsworth, and Knox, 1999; Mcmillan, and Powell, 1999) is outstanding from the petrologic and practical point of view of geology, since it features almost all the criteria for the determination and multihierarchical classification of rocks (genesis, composition, structure, texture).

IUGS (International Union of Geological Sciences) publications related to the classification of volcanic and plutonic rocks (Le Maitre, et. al., 2002), the stratigraphic guide and stratigraphic terminology (Salvador, 1994), as well as the chronostratigraphic division (Gradstein, et. al., 2004), have been widely accepted by national geologic communities and academic institutions. As such, they provide both good resources of terms and guidelines for the development of geologic terminology.

The above-mentioned publications made it clear that:

• the terminology, in the geologic information system, cannot consist of a list of geologic terms only, but that every term or concept must be clearly and unequivocally determined, namely, defined and that only then it can have a communications role;

• analytical requirements of comparison, correlation, grouping and searching geologic similarities and differences, details and unique features of rock material, geological units and structures call for assigning the central role to terminology in the information system;

• older, inherited terminology must be included in the GeolISS vocabulary, because that is the only way for the original concepts from the historical geologic documents to be entered into the digital database, to track their evolution, i.e. possible changes of meaning over time;

• the progressive and recursive nature of the observation and interpretation process has a hierarchical, namely, family tree structure; from the initial generalization to gradual distinguishing between specific categories that provide more detailed information about the geologic material (Soller, 2004).

3.2. GeolISSTerm as a controlled vocabulary

In view of the function that the geologic vocabulary has in GeolISS and the international practice, GeolISSTerm is organized as a taxonomy with a definition for each term, synonyms and bibliographical references of the sources from where the terms were taken, as well as the equivalent terms in other languages (currently, only English equivalents are present, see Example 1). The size of the vocabulary, namely, the number of concepts is not finite, but the current version features around 3800 basic, universally accepted geologic terms with the highest frequency of use in practice, with almost as many English equivalents.

The relationships between the concepts represented by the terms are: hyponymy and hypernymy that can be regarded as referring to a broader and a narrower term respectively. The relations can be further expanded by including synonymy, where the basic term becomes associated with the terms of the same meaning. The relationships include establishing a link to the terms in other languages too.

The validation, classification and specification of the value of attributes, namely the fields in the database tables are done via controlled vocabularies derived from GeolISSTerm. The examples of the vocabularies derived from GeolISSTerm are petrologic classification, mineralogic classification, geochronological scale and a lexicon of geological units. These are followed by the vocabularies, extracted from the terminology, belonging to the fields of structural, economic, engineering geology and hydrology that can be browsed at: http://geoliss.ekoplan.gov.rs/term/recpj.aspx.

3.3. GeolISS vocabulary structure

In order for the vocabulary to be able to meet all previously derived requirements and be functional within GeolISS, the UML model with a special structure was developed (Figure 2). The class *Rečnik* (*Vocabulary)* in the model is a lexicographic superclass whose instances are inherited. *GeološkiRečnik* (*GeologicVocabulary)* has been implemented as an abstract class, since the class *Koncept* (*Concept)*, above all, allows entering general geologic concepts and terms common to all geologic disciplines and centralizes individual classifications (petrologic, mineralogic, stratigraphic, chronostratigraphic). The term "concept" itself (lat. *conceptus* – notion) naming the central class has been used in its original meaning to refer to an abstract or a general idea of an assumed or concrete instance (Angeles, 1981).

The hierarchical structure of the vocabulary (Figures 2 and 3) is made possible through involution i.e. recursive relation modelling the relation hypernym–hyponym in such a way that any (hyponymous) term in the vocabulary hierarchy can appear only once and have just one hypernym. Moreover, every term can have an equivalent in one or more foreign languages via the *MultijezičkiLeks (MultilanguageLex)* class. The class *AdminTerm* is used for the terms that do not belong to the realm of geology, but are necessary for the functioning of GeolISS. The relations between different terms (e.g. derived from, having broader meaning than, lexical variant, etc.) can be recorded in the class *RelacijeTermina (TermRelations)*. Written source/s from which concepts or terms were taken, together with their meaning are entered into the class *Bibliografija (Bibliography)* and the author who added the new vocabulary entry is registered through the *Metapodatak (Metadata)* class.

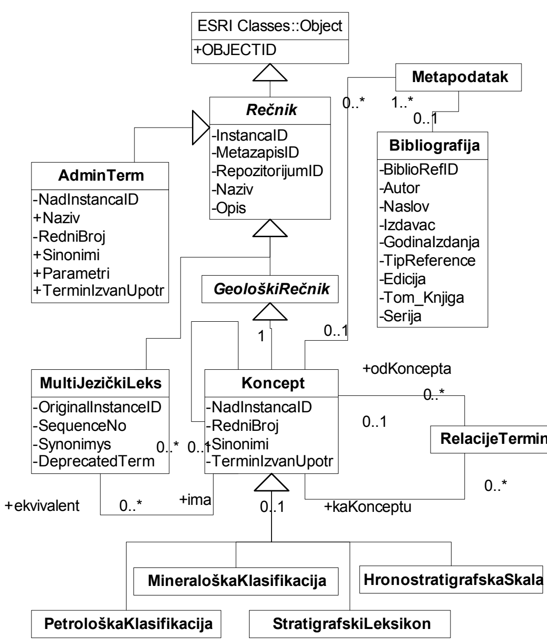


Figure 2. UML model of the structure of the GeolISS vocabulary (ESRI profile)

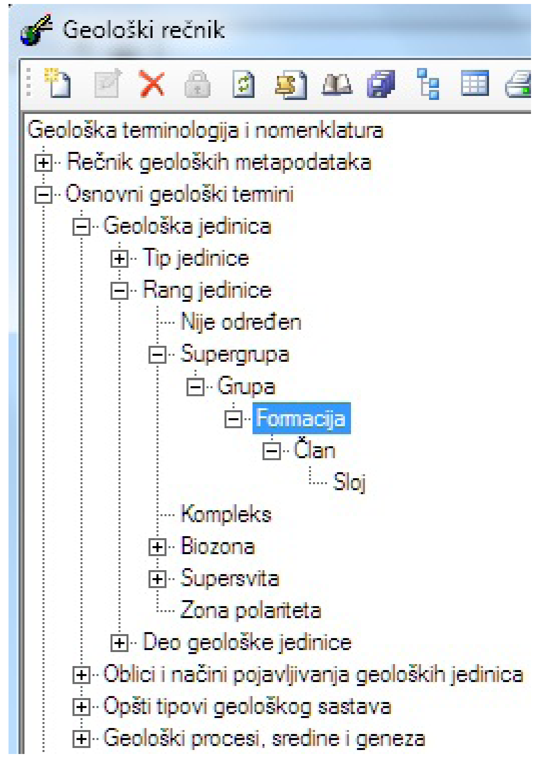


Figure 3. An example of the hierarchical structure of the dictionary of geologic terms

The structure of the individual entry “formacija” shown in Figure 3 is given in Example 1.

Example 1.

Name: Formacija

Definition: Formacija je osnovna jedinica litostratigrafske klasifikacije koja predstavlja stensku masu sa jedinstvenim sedimentološkim i paleontološkim karakteristikama, formiranim u uslovima jedinstvene depozicione sredine nekog vremena.

Direct hypernym in the hierarchy : Grupa

Translation of the term: Formation

Translation of the definition: A persistent body of igneous, sedimentary, or metamorphic rock, having easily recognizable boundaries that can be traced in the field without recourse to detailed paleontologic or petrologic analysis, and large enough to be represented on a geologic map as a practical or convenient unit for mapping and description; the basic cartographic unit in geologic mapping (NACSN, 1983). The formation is the fundamental unit in lithostratigraphic classification. A formation is a body of rock identified by lithic characteristics and stratigraphic position; it is prevailingly, but not necessarily, tabular and is mappable at the Earth's surface or traceable in the subsurface (NACSN, 2005).

Synonyms (English): none

References: North American Stratigraphic Code Grupa autora 2005. AAPG; Internacionalni stratigrafski vodič i rečnik stratigrafskih termina. Jovanović, R., Jovanović, D. 2007 Naftagas, Novi Sad ; Petrology: The Study of Igneous, Sedimentary and Metamorphic Rocks. Raymond, L.A. 2002 McGraw-Hill. 736 str., New York.; Geološko kartiranje. Dimitrijević, D. M. 1978 Izdavačko-informativni centar studenata, 486 str., Beograd.; Sedimentologija Grubić, A., Obradović, J., Vasić, N. 1996 Univerzitet u Beogradu, 436 str., Beograd.; International Stratigraphic Guide. Hedberg, H. D. (Ed.) 1976 J. Wiley & Sons. - Interscience publication, 200 str., New York; NASC (North American startigraphic Code). The North American Commission on Stratigraphic Nomenclature 2005 The American Ass. of Petroleum Geologists. Tulsa-Oklahoma.

The derived relations are descendants and ancestors in Serbian and English.

3.4. Modelling property domains

In order to control property values in GeolISS, it is necessary to ensure extraction of the sets of terms representing the allowed values for individual properties, which is done by creating the appropriate domains. The domains can be lists of terms or a hierarchical tree of terms, but in both cases we are dealing with the terms included in the vocabulary. Figure 4 shows the UML domain management model. The abstract class *MetaAdmin* is inherited by other classes in the model. The *Domain* class is where all the data necessary for a description of some domain are registered: whether it is a list or a hierarchical tree, from which dictionary the data are taken: geologic, administrative, numerical (in which case there is an allowed range of values), which geologic discipline it belongs to and so on. The role of the class *DomainNode* is to determine which terms are incorporated into a domain. This is the class where an individual node and the kind of its integration into a domain are specified. There are four possibilities:

– That node only,

– The node and its children (direct descendants),

– Only the children of the selected node (without the selected node) and

– The node and all its descendants.

The first three kinds are typical of lists of terms, while the fourth is generally used for hierarchical domains. Since several instances of the class *CvorDomena (DomainNode)* can correspond to a single instance of the class *Domen (Domain),* a domain can consist of several branches of the vocabulary. The majority of domains are generated on the basis of the geologic vocabulary (78), but a smaller number is generated from the administrative vocabulary (22) and from meta data about the GeolISS entities (11).

The class *Svojstvo (Property)* models the metadata about the physical and logical attributes in the database, while the class *Entitet (Entity)* comprises instances of all spatial and classes of attributes and also their subclasses, namely subtypes. Among the metadata provided by the relationship class *SvojstvoEntiteta (EntityProperty)* is the domain (Figure4).

The instances of the class *Entitet (Entity)* are for example: MappedUnit, GeologicAge, GeologicMaterial, ExodynamicPhenomenon, etc. Moreover, the instances of the class *Svojstvo* (*Property)* are: Colour, Composition Category, ChemicalComposition, etc., while the relationship class *SvojstvoEntiteta* (*EntityProperty)* connects the instances of entities and properties. For example, the property Composition Category is present in the entities MappedUnit and GeologicMaterial, and takes the values from the domain *dComposition Category* in both cases.

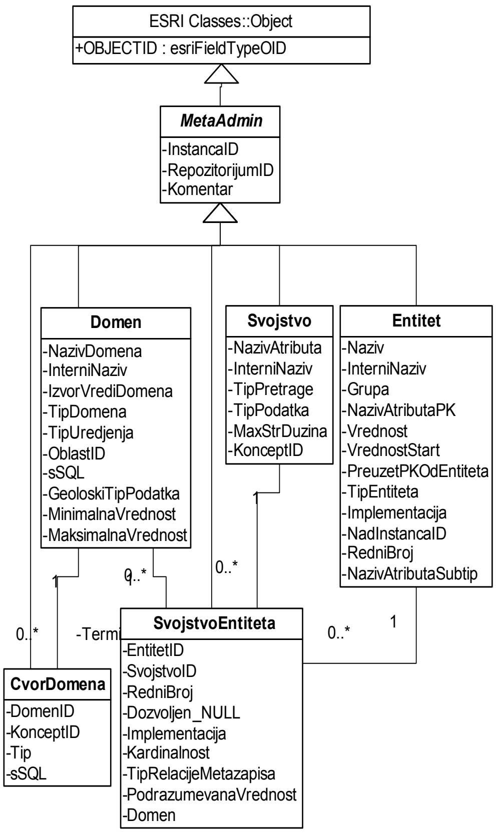


Figure 4. UML model for GeoIISS domain management (ESRI profile)

The generation of instances belonging to the class *Domen (Domain)* is illustrated in Example 2 featuring the basic parameters of the domain *dKategorijaSastava (dWaterPropertyType)* used for all hydrologic phenomena and objects for classification by water type.

Example 2

Domain: dTipPoSvojstvuVode

Domain Value Source: Koncept (geološki rečnik)

DomainType: Integer (prenosi se primarni ključ)

Layout type: Lista

Field: Hidrogeologija

Definisanje sadržaja ovog domena se ostvaruje preko primeraka klase *CvorDomena:*

Klasifikacija podzemnih voda na osnovu porekla (svi čvorovi ispod njega bez njega): Podzemne vode hidrometeorskog porekla, Juvenilna voda, Kosmička voda, Konatna voda, Metamorfna voda;

Voda u nadkritičnom stanju (samo taj čvor);

Voda u vidu vodene pare (samo taj čvor);

Hemijski vezana voda (svi čvorovi ispod njega bez njega): Zeolitna voda, Konstituciona voda, Kristalizaciona voda;

Fizički vezana vode (svi čvorovi ispod njega bez njega): Adheziona voda, Adsorbovana voda, Apsorbovana voda, Higroskopna voda;

Kapilarna voda (samo prvi nivo ispod njega i on): Kapilarno lebdeća voda, Kapilarno podizuća voda

Slobodna voda (samo taj čvor);

Voda u vidu leda (samo taj čvor);

Klasifikacija podzemnih voda na osnovu mineralizacije (svi čvorovi ispod njega bez njega): Malomineralizovana voda, Mineralizovana voda, Rasol;

Mineralna voda (samo prvi nivo ispod njega i on): Termomineralna voda, Lekovita voda, Termalna voda, Industrijska mineralna voda.

The domain management interface shown in Figure 5 has been integrated as part of the GeoIISS software that only the users with administrative privileges can have access to. A user can create a domain in an interactive manner and check its content. The domain browsing tab allows viewing a table listing the terms belonging to a domain (lower left part of Figure 5). However, since there are hierarchical domains too, tree browsing is also possible (lower right part of Figure 5).

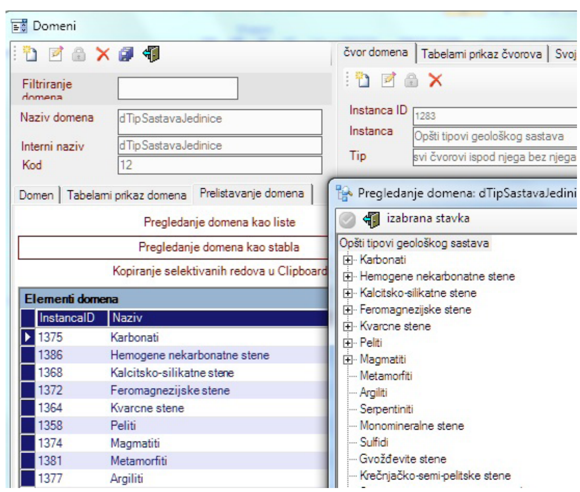


Figure 5. Domain management in GeoIISS

4. Software solution

The application was developed using MS Visual Studio and the programming languages C# and Aspx, while the data were stored in the MS SQL Server 2008 data management system. This section features a brief description of the components of the software solution which consists of several elements:

– Standalone application for entering dictionary entries, which can be used not only in the context of GeolISS, but also as an independently functioning component

– GeolISS component integrated in the Esri® ArcGIS® development environment for the GIS systems (http://www.esri.com/arcgis), as an extension of the ArcMap tool for cartographic content management and

– Web application for browsing and searching the dictionary

The data entry interface (Figure 6) displays the structure i.e. organization of concepts and terms already existing in the dictionary on the left-hand side, while the right-hand side of the interface shows the attributes of the selected concept, namely, dictionary entry. Given the structure of the dictionary, its content is not strictly arranged by theme, but rather following logically related entities. However, the flexibility of structure allows the possibility of rearranging the content, forming new semantically or thematically related groups, as well as sorting and searching the content using different criteria.

Every entry contains a name, i.e. the term itself, its definition, as well as, possibly, synonyms. The number showing the entry's place on the list of terms is provided only for the purpose of correct sorting in the tree structure. The terms that are for some reason regarded as unsuitable for further use are not deleted from the dictionary, but are taken out of use.

The software makes it possible for authorized users to add new entries in the dictionary, where every new entry must contain the appropriate definition of the concept with a mandatory bibliographic reference citing the source from where the definition was taken. The information on the metadata, including source of data, processing method and comments is provided for every individual term entry. A correction of a dictionary entry can be made only by the user who created it, so that the entry authorship requirement is met.

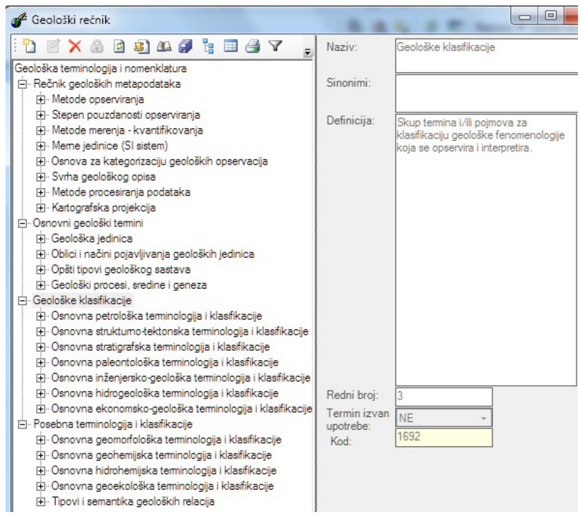
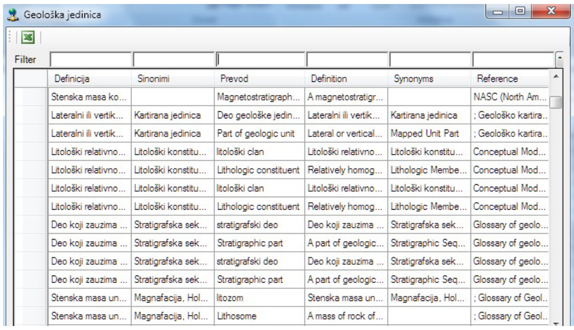


Figure 6. Term entry interface

The application offers possibilities of a tabular view of subordinate terms within the selected node, which is very useful for browsing and updating related terms at the same level of the hierarchy.

Moreover, it allows entering a translation of the selected dictionary entry into one or more foreign languages. At this point, only English is supported, but the structure of the database and the application itself make it possible to add translations to other languages.

This form of a tabular view i.e. the interface for browsing all the descendants (direct and indirect) of the selected node is shown in Figure 7. This example features the term "geological unit“. The users can filter the entry elements they selected applying different criteria, namely, browsing the information sorted in all the columns shown below.



Definicija	Sinonimi	Prevod	Definition	Synonyms	Reference
Stenska masa ko...		Magnetostratigraph...	A magnetostratigr...		NASC (North Am...
Lateralni ili vertik...	Kartirana jedinica	Deo geološke jedin...	Lateralni ili vertik...	Kartirana jedinica	; Geološko kartira...
Lateralni ili vertik...	Kartirana jedinica	Part of geologic unit	Lateral or vertical...	Mapped Unit Part	; Geološko kartira...
Litološki relativno...	Litološki konstitu...	litološki clan	Litološki relativno...	Litološki konstitu...	Conceptual Mod...
Litološki relativno...	Litološki konstitu...	Lithologic constituent	Relatively homog...	Lithologic Membe...	Conceptual Mod...
Litološki relativno...	Litološki konstitu...	litološki clan	Litološki relativno...	Litološki konstitu...	Conceptual Mod...
Litološki relativno...	Litološki konstitu...	Lithologic constituent	Relatively homog...	Lithologic Membe...	Conceptual Mod...
Deo koji zauzima ...	Stratigrafska sek...	stratigrafski deo	Deo koji zauzima ...	Stratigrafska sek...	Glossary of geolo...
Deo koji zauzima ...	Stratigrafska sek...	Stratigraphic part	A part of geologic...	Stratigraphic Seq...	Glossary of geolo...
Deo koji zauzima ...	Stratigrafska sek...	stratigrafski deo	Deo koji zauzima ...	Stratigrafska sek...	Glossary of geolo...
Deo koji zauzima ...	Stratigrafska sek...	Stratigraphic part	A part of geologic...	Stratigraphic Seq...	Glossary of geolo...
Stenska masa un...	Magnafacija, Hol...	litozom	Stenska masa un...	Magnafacija, Hol...	; Glossary of Geol...
Stenska masa un...	Magnafacija, Hol...	Lithosome	A mass of rock of...		; Glossary of Geol...

Figure 7. Browsing all the descendants (direct and indirect) of the selected node

The web application is available at http://geoliss.ekoplan.gov.rs/term and unlike the standalone component shown above, which is intended for GeolISS users only, it allows free browsing and searching of the dictionary. Thus, only authorized users are permitted to add new entries and update the dictionary, while reading, browsing and searching are accessible to the general public. In Figure 8 the interface of the Serbian and English dictionaries, on the left-hand and right-hand side respectively is shown. The Serbian and English version look similar. On the right-hand side, there is a tree structure providing a hierarchical view of entries, while detailed information about the selected term in the corresponding language is to be found on the left-hand side. In the detailed view, the term itself, displayed as a subtitle (in the interface form in question “Litodemska jedinica”, namely, “Lithodemic unit”), the tree structure showing hypernyms, as well as the definition, synonyms and the bibliographic reference are shown. Switching from one language to another is done by choosing the appropriate tab, namely, the hierarchical tree structure on the right-hand side of the interface.

Figure 8. GeolISSTerm on the web – browsing the dictionary in Serbian and English

The search by key word can be performed in Serbian and English and is available at http://geoliss.ekoplan.gov.rs/term/PretragaRecnika.aspx. The user chooses the key word, for example, “**rang**”, after which the system finds the corresponding terms in Serbian and English. The search is of the “*rang*” type, meaning that the system will find all the terms containing the string of letters “rang“. Figure 9 shows the terms in Serbian: ***Rang*** *jedinice,* ***Rang****iranje ugljonosnih jedinica,* with the terms in English to the left: *Ar****range****d fabric,* ***Range*** *value, Fold* ***range****, Size* ***range***). The user selects one of the proposed terms and then obtains detailed information about it, in Serbian and English, side by side. The definition, synonyms, references, hypernyms and hyponyms, given as hyperlinks leading to more detailed information are displayed in both languages, in addition to the term itself.

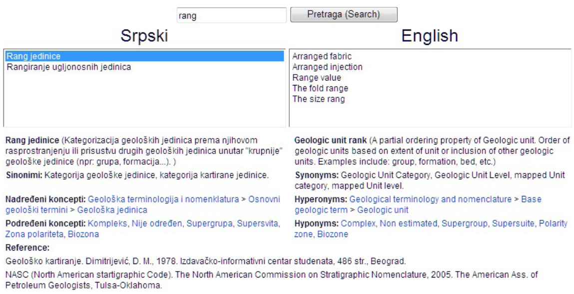
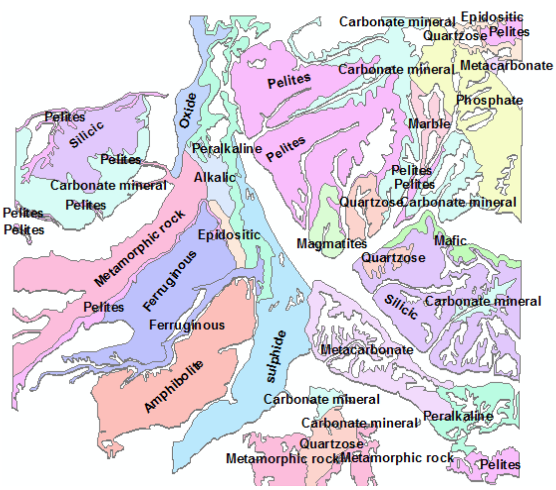


Figure 9. A search by key words

5. The application in cartography

The labelling and annotation of the cartographic material is made possible both in Serbian and English thanks to the relations between Serbian and English equivalent terms, on the one hand and the strong links between the dictionary and the content of the GeolISS database on the other. The steady rise in importance of multilingual annotation reflects the growing need for information exchange with foreign companies and geologic associations.



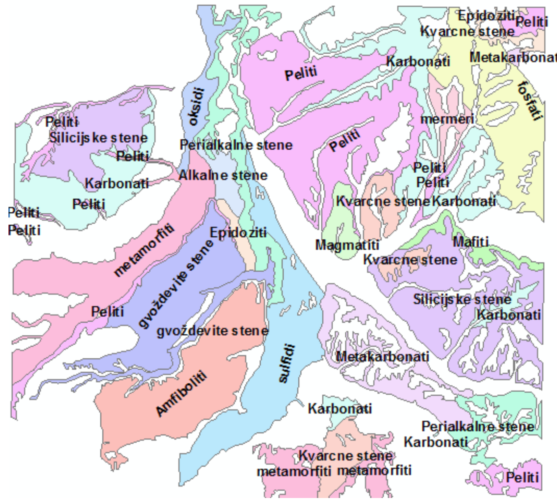


Figure 10. An example of automatic map annotation in English and Serbian

The importance of efficiency and flexibility of database searches and finding information on the web is growing by the day, therefore the use of GeolISSTerm for expanding the queries in the field of geology can play an important role in their improvement (Stanković et al., 2010). In view of the rich morphology of Serbian, one of the ways of improving the searches related to the geology resources on the web would be through the integration of the morphological dictionary of Serbian (Vitas et al. 2003) and GeolISSTerm by adding inflectional class codes to the terms featured in GeolISSTerm.

6. Conclusion

The development and use of the standardized terminology in digital databases is in many respects a novel field that is bound to evolve over time, as more experience is gained in that area. The significant steps and progress that have been made in Serbia in the domain of geology, understandably, suffer from all the shortcomings that accompany such a pioneering project. This is the first time that a consistent glossary that includes the basic geologic classifications in the field of fundamental and applied geology has been integrated into a geologic information system. The absence of unified, unequivocal and universally accepted classifications that can satisfy all the needs of geology as a science must certainly be emphasized. Thus, the expansion of the range of terms included in GeolISS will continue to be a fundamentally and operationally important task, aimed primarily at a clearer determination of the geologic content and improving the quality of geologic data. Immediate and end users of the information system can undoubtedly be instrumental in achieving that goal, above all, geologists, but also all other experts focusing on the geoscience-related research. The possibilities of exporting the vocabulary to standard ontology languages, such as OWL, modelled on the GeoSciML initiative (http://www.geosciml.org/) are planned to be provided.

References

Angeles, A. Peter. 1981. *Dictionary of Philosophy*, London: Harper & Row Publisher.

Brodie, K., Fettes, D., Harte, B., and Schmid, R. 2002. *Towards a unified nomenclature of metamorphic petrology: 5. Structural terms including fault rock terms*, A proposal on behalf of the IUGS Subcommission on the Systematics of Metamorphic Rocks, Web version of 31.10.2002: 8 p.

Bates, Robert Latimer & Jackson, A. Julia, Ed. 1995. *Glossary of geology*, Alexandria, Virginia, American Geological Institute.

Blagojević, Branislav, Trivić Branislav, Stanković, Ranka i Banjac, Nenad. 2008. Kratak pregled implementacije Geološkog informacionog sistema Srbije. *Zapisnici Srpskog geološkog društva za 2007.godinu*. Beograd, Srpsko geološko društvo, pp.37-42.

Daconta, Michael. Obrst, Leo & Smith, Kevin 2003. *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*, John Wiley & Sons.

McGuinness, Deborah. 2003. Ontologies Come of Age. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential,* Dieter Fensel, J im Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. MIT Press.

Gillespie, M.R., & Styles, M.T., 1999. *Classification of igneous rocks: Volume 1 of the BGS rock classification scheme*, British Geological Survey Research Report Number RR 99-06, 52 p.

Greidstain, M. Felix. Ogg, James G., Smith. G. Alan at al. 2004. *A Geologic Time Scale*, Cambridge University Press.

Hallsworth, C.R. and Knox, R.W.O'B., 1999, *Classification of sediments and sedimentary rocks: Volume 3 of the BGS rock classification scheme*, British Geological Survey Research Report Number RR 99-03, 44 p.

Krstev, Cvetana. Đorđević, Bojana. Antonić, Sanja. Ivković-Berček, Nevena. Zorica, Zorica. Crnogorac, Vesna and Macura, Ljiljana. 2008. Kooperativan rad na dogradnji srpskog wordneta, *INFOteka: časopis za informatiku i bibliotekarstvo*, vol. IX, No.1-2, pp. 57-75, Beograd.

Le Maitre, R.W. (editor), Streckeisen, A., Zanettin, B., Le Bas, M.J., Bonin, B., Bateman, P., Bellieni, G., Dudek, A., Efremova, S., Keller, J., Lameyre, J., Sabine, P.A., Schmid, R., Sørensen, H., and Woolley, A.R. 2002. *Igneous rocks: A classification and glossary of terms*, Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks: Cambridge, Cambridge University Press, 236 p.

Mcmillan, A.A., and Powell, J.H. 1999. *BGS rock classification scheme: Classification of artificial (man made) and natural superficial deposits*, British Geological Survey Research Report RR 99–04.

Obradović, Ivan i Stanković, Ranka, 2008,"Softverski alati za korišćenje jezičkih resursa za srpski jezik", *INFOteka: časopis za informatiku i bibliotekarstvo*, vol. IX, No.1-2, str. 41-55, Beograd.

Petković, Kosta. 1975. *Geološka terminologija i nomenklatura*, *I-VII*. Beograd: Rudarsko-geološki fakultet.

Projekat "Razvoj geološke terminologije i nomenklature za geološku bazu podataka Srbije", broj 310-02-113/2009-02 od 04. 08. 2009. godine Rudarsko-geološki fakultet, rukovodilac Banjac, Nenad.

Salvador, Amos. ed., 1994. *International Stratigraphic Guide*, Second Edition: Trondheim, Norway, International Union of Geological Sciences, and Boulder, Colorado, Geological Society of America, 214 p.

Soller, R. David, ed., 2004a. *Classification of metamorphic and other composite-genesis rocks, including hydrothermally altered, impact-metamorphic, mylonitic, and cataclastic rocks*, Version 1.0 (12/18/2004), North American Geologic-map Data Model Science Language Technical Team, 56 p.

Soller, R. David, ed., 2004b. *Sedimentary materials: science language for their classification, description, and interpretation in digital geologic-map databases*, Version 1.0 (12/18/2004), North American Geologic-map Data Model Science Language Technical Team, 595 p.

Soller, R. David., ed., 2004c. *Volcanic materials: science language for their naming and characterization in digital geologic-map databases*, Version 1.0 (12/18/2004), North American Geologic-map Data Model Science Language Technical Team, 18 p.

Stanković, Ranka. Obradović, Ivan. Kitanović, Olivera. 2010. GIS Application Improvement with Multilingual Lexical and Terminological Resources, Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2010, Valetta, Malta, pp.2283-2287.

Struik, L.C. Quat, M.B. Davenport, P.H. and Okulitch, A.V. 2002. *Multi-hierarchical rock classification for use with thematic computer-based query systems:, a preliminary scheme.* Geological Survey of Canada, Current Research 2002D, 9 p.

Trivić, Branislav. Blagojević Branislav & Stanković, Ranka. (editori), 2011, *Rečnik geoloških termina i pojmova geološkog informacionog sistema Srbije*, Ministarstvo životne sredine, rudarstva i prostornog planiranja Republike Srbije.

Vitas, Duško. Pavlović-Lažetić, Goradana. Krstev, Cvetana. Popović, Ljuba. Obradović, Ivan. 2003: Processing Serbian Written Texts: An Overview of Resources and Basic Tools, Proceedings of the International Workshop on Balkan Language Resources and Tools, Thessaloniki, Greece, November 2003, S. Piperidis, V. Karakaletsis (eds.), pp. 97-104.

Online resources

GeolISS - Geologic Information System of Serbia, http://geoliss.ekoplan.gov.rs (accessed May, 2011).

Esri® ArcGIS® - The Complete Geographic Information System http://www.esri.com/arcgis (accessed May, 2011).

GeoSciML Resources repository http://www.geosciml.org (accessed May, 2011)

One Geology - Making Geological Map Data for the Earth Accessible http://www.onegeology.org