Developing Termbases for Expert Terminology under the TBX Standard

Ranka Stanković, Ivan Obradović, Miloš Utvić

Developing Termbases for Expert Terminology under the TBX Standard

Ranka Stanković[1], Ivan Obradović[1], and Miloš Utvić[2]

[1] University of Belgrade, Faculty of Mining and Geology,
Đušina 7, 11000 Belgrade, Serbia
{ranka.stankovic,ivan.obradovic}@rgf.bg.ac.rs,
[2] University of Belgrade, Faculty of Philology,
Studentski trg 3, 11000 Belgrade, Serbia
misko@matf.bg.ac.rs

Abstract. Termbases have played a crucial role in translation and localization for many years now. A team at the University of Belgrade Faculty of Mining and Geology (FMG) has initiated several years ago the development of terminological resources in the area of mining amd geology, both monolingual for Serbian and multilingual with Serbian as one of the languages. In this paper we describe the approach to development of these termbases, as well as the role of the TermBase eXchange (TBX) standard in this approach. Namely, paying special attention to portability, simple and speedy transformation of subsets of concepts from termbases as central resources in a custom in-house scheme to standard formats such as TBX has been provided, by a wizard integrated in the terminological information system supporting the termbases.

Keywords: Termbases, TBX standard, Language Resources, Terminology Integration and Portability

1 Introduction

Translation memory (TM) systems have been the major language technology to support the translation and localization industries for the last two decades [11]. Their essential components are termbases, which can broadly be defined as databases containing structured concept-oriented terminological data, that is, domain-specific concepts and terms that designate them. Termbases are monolingual, bilingual, or multilingual language resources related to specific domains of knowledge. TM technology is nowadays increasingly challenged by machine translation (MT), especially statistical machine translation (SMT), an approach developed at IBM in the late 1980s, now the state-of-the art paradigm in MT. The exponential growth of aligned multilingual corpora greatly improved the efficiency and accuracy of SMT in general, and many tools based on this approach, such as Google Translate, are thus being more and more widely used. This has led to a debate whether the development of termbases is still worth the effort. However, the use of SMT tools in translation of documents related to specific expert domains often produces moderate to poor results. Thus, termbases

are still bound to maintain their importance in the case of expert terminology in domains where aligned corpora are sparse [10], such as, for example mining engineering or geology.

In order to secure terminological consistency in one or more termbases, and to avoid locking of termbases into specific TM software, an international standard, TermBase eXchange (TBX), has been defined by ISO and the Localization Industry Standards Association (LISA). TBX defines an XML format for the exchange of terminology data, where a terminology database that is to be represented in TBX must conform to the Terminological Markup Framework (TMF), an abstract data model also defined by ISO.

In this paper we describe an approach to development of termbases for the field of mining engineering and geology, as well as the role of the TBX standard in this approach. In the next section we give an outline of three terminological resources developed at the University of Belgrade, Faculty of Mining and Geology (FMG). The third section is dedicated to a general discussion on TBX, whereas the fourth describes its use in securing the integration and portability of our termbases. The paper ends with a section with conclusions.

2 Terminological Resources Developed @ FMG

Recognizing the importance of terminology in education of future mining engineers a team at FMG has initiated several years ago the development of terminological resources in the area of mining engineering and geology, both monolingual for Serbian and multilingual with Serbian as one of the languages.

The resources have been developed within the scope of various projects, but using the same platform, namely RDBMS SQL Server, with MS Visual Studio .NET and C# programming language for application development. Besides the part for standard manipulation of terminological records these applications offer possibilities for export of termbases or their parts to other formats, such as TBX or Open Lexicon Interchange Format (OLIF), used in a variety of natural language processing applications and general language technology environments (e.g., TM systems) [9].

Termbases developed at FMG followed the principle that terminology, as every other theory, should have an applied side from which applications can be generated to solve problems. To that end they have to describe real data and must be internally consistent [2]. Besides being used for translation purposes, termbases developed at FMG are thus used for control of domain values, classification and search within mining and geological software systems.

2.1 GeolISSTerm

A publicly available bilingual terminological resource, GeolISSTerm (`http://geoliss.mprrpp.gov.rs/term/`), was developed at FMG for the Ministry of the Environment, Mining and Spatial Planning of the Republic of Serbia, in the form of a thesaurus for geological terms. GeolISSTerm now contains more than

3000 terms and their English equivalents [15], divided in several subdomains: petrology, mineralogy, hydrogeology, geophysics, structural geology etc.

The core of UML model of this resource is presented in Figure 1. The class *GeološkiRečnik* (Geologic Vocabulary) in the model is a lexicographic superclass whose instances are inherited by *Koncept* (Concept). It enables the entering of general geologic concepts and terms common to all geologic disciplines and centralizes individual classifications: *PetrološkaKlasifikacija*, *MineraloškaKlasifikacija*, *StratigrafskiLeksikon*, *HronostratigrafskaSkala* (petrologic, mineralogic, stratigraphic, chronostratigraphic). The hierarchical structure of the vocabulary is implemented through involution, i.e. a recursive relation modeling the hypernym/hyponym relation in such a way that any (hyponymous) term in the vocabulary hierarchy can appear only once and have just one hypernym. Every term can have an equivalent in one or more foreign languages via the *MultijezičkiLeks* (Multilingual Lex) class. The relations between different terms (e.g. derived from, having broader meaning than, lexical variant, etc.) can be recorded in the class *RelacijeTermina* (Term Relations). 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.

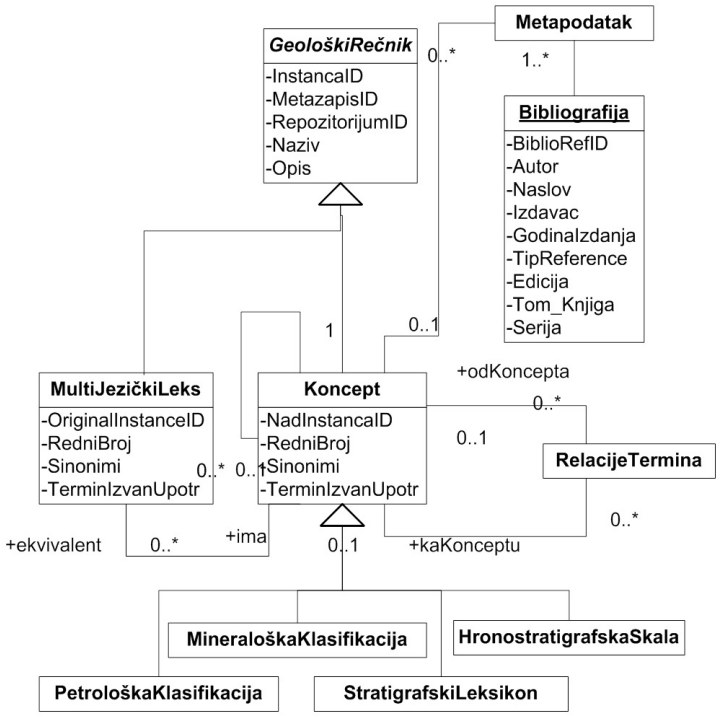

Fig. 1: UML model of the GeolISSTerm structure

An example of a geological term follows. Only the English part will be given, noting that a Serbian equivalent also exists.

Name: Deposits of mineral resources
Definition: Mineral deposits are the geological bodies limited by natural or artificial (industrial) borders. They are formed in the nature, through complex geological and chemical processes. As an integral part of the Earth's crust, mineral deposits are built out of different minerals, elements and compounds, which are suitable for industrial use in natural or refined form. Mineral deposits represent a geo-economic category, which means that under favorable circumstances, they can be considered to have economic potential.
Hyperonyms: Base economic geology terminology and classifications
Hyponyms: Occurrences of mineral resources, Ore body, Ore, Division of mineral deposits
Reference: Ležišta metaličnih mineralnih sirovina. Jelenković, R., 1999. Rudarsko-geološki fakultet Univerziteta u Beogradu, Beograd.

2.2 RudOnto

Another important terminological resource developed at FMG is RudOnto, an in-house project, aimed at covering the larger area of mining engineering and geology and becoming the reference resource for mining terminology in Serbian. RudOnto is managed by a terminological information system, and one of its intended uses is the production of derived terminological resources in sub-fields of mining engineering, such as planning and management of exploitation, mine safety or mining equipment management.

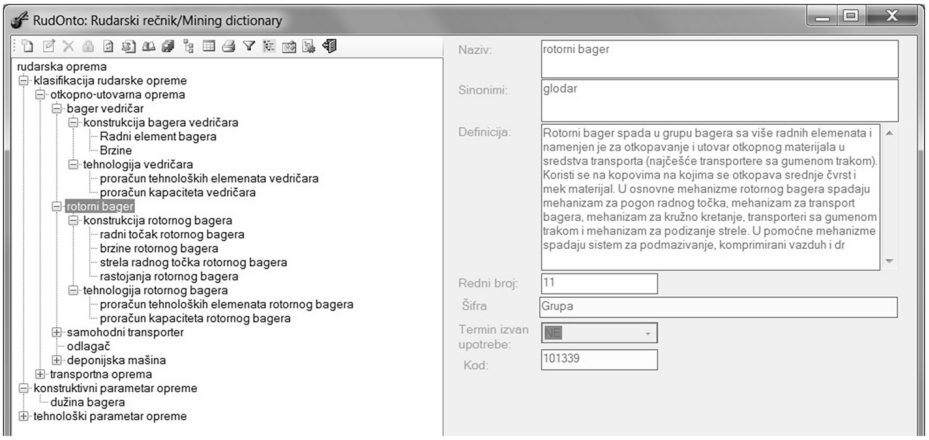

Fig. 2: Management of RudOnto hierarchy of concepts in Serbian

Figure 2 features panels from the terminological information system module that manages RudOnto; the left hand side of the larger panel shows the hyper-

nym/hyponym hierarchy of concepts, while the right hand side offers the full entry for the selected concept in the hierarchy. The entry consists of the basic term used for this concept, its synonyms (none in this case), and its definition.

Figure 3 offers an example of another panel from the same information system, showing all available translational equivalents of a term in other languages on the left hand side and details of the Russian translation on the right.

Both GeolISSTerm and RudOnto are available for browsing and searching of the available terminology to the general public (without log-in). Figure 4 depicts the web interface of RudOnto in Serbian: the right-hand side features a tree structure providing a hierarchical view of entries, with more detailed information on a selected term in Serbian to be found on the left-hand side. In this detailed view, the term itself is displayed as a subtitle, on top of a tree structure showing its place within the taxonomy of its hypernyms. The definition, parameters, hyponyms and bibliographic reference of the term are also given, and for some terms, an appropriate illustration is available. The user can easily switch from one language to another by choosing the appropriate tab on the right-hand side of the page and thus obtaining the hierarchical tree structure in the appropriate language.

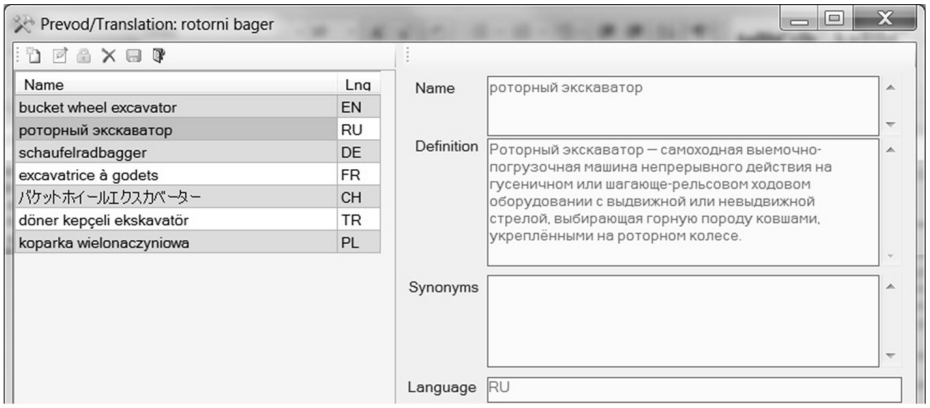

Fig. 3: Available translations of a selected term and details in Russian

2.3 Morphological E-Dictionary of Mining and Geology

The main shortcoming of GeolISSTerm and RudOnto is their lack of lexical information, such as part of speech, grammatical gender, inflection class or word forms. This information is essential for proper processing of all texts, such as lematization, morphological analysis, named entity recognition and the like. This is especially important in the case of domain specific texts as in the fields of geology or mining. Thus, appropriate electronic morphological dictionaries are

Fig. 4: Web application for RudOnto browse and search

needed [18]. The system of morphological dictionaries for Serbian, both of simple and compound words has been developed over a very long period, follows the so-called DELA format [8]. The DELAS dictionaries of simple words have reached a very high level of coverage of Serbian, while the DELAC dictionaries of compounds are still being intensively developed. Serbian e-dictionaries are being widely used for various language technology tasks, including termbase applications. However, the system of dictionaries still lacks domain specific terms for some areas, such as mining and geology, which motivated the language technology team at FMG to build such a dictionary for these areas. For its production and management we used an integrated and easily adjustable tool for language resources, LeXimir, also developed at FMG within the Human Language Technology group at the University of Belgrade [8]. This tool can handle several language resources simultaneously, thus enhancing the potential of each particular resource in realizing a task, in this case production of morphological dictionaries [13].

Some examples from the simple word DELAS dictionary of terms related to mining and geology, represented by their lemmas, transducers for their respective inflection classes and semantic markers are:

```
elektrovod,N1+RudOnto+Elektro
aerozagađenje,N300+RudOnto+Ekolog
hidrogeološki,A2+RudOnto+Hydro+PosQ
```

All inflected forms are retrieved using the information about the corresponding inflection class given in the DELAS format (e.g. N1, N300, A2).

An example of a compound term related to mining and geology, which comprises tree simple words, in the more complex DELAC dictionary is:

```
ležište(ležište.N300:ns1q) mineralnih sirovina,NC_N4X+Comp
```

The first simple word *ležište* (deposit) is a noun in nominative case, singular, where N300 is the transducer for its inflection class, while NC_N4X is the transducer for the multiword unit inflection of the entire compound word. This transducer produces 17 compound word forms in DELACF format:

```
ležište mineralnih sirovina,ležište mineralnih sirovina.N:s1qn
ležišta mineralnih sirovina,ležište mineralnih sirovina.N:s2qn
...
ležišta mineralnih sirovina,ležište mineralnih sirovina.N:w4qn
```

Domain specific e-dictionaries are especially important in recognition of compound words in texts featuring expert terminology, as such texts usually abound with compounds having a meaning often very different from the meaning of each of their components. Thus if such a compound is not recognized, but rather treated as a sequence of its components, the text processing results is bound to be much more complicated.

3 TBX

Consistency, portability, and reusability of termbases cannot be achieved without standards. TMF (Terminology Markup Framework), specified by ISO 16642 standard [4], provides a meta-model for the description of terminologies and other onomasiological structures, as well as the guidance on the basic principles for representing data recorded in terminological data collections. This framework offers a meta-model and methods for describing specific terminological markup languages (TMLs) expressed in XML.

TBX, specified by ISO 30042 standard [5], is in itself an application of TMF. The TBX description of the TMLs is based on modular approach, i.e. a particular TML is defined as the combination of two modules expressed in XML. One module is fixed and represents the common core structure of all TBX-defined TMLs. Another module, XCS (eXtensible Constraint Specification), consists of constraints on the core structure, specific for each TML. Both modules are formally defined with the corresponding DTDs (Document Type Definition), i.e. the XML documents representing core-structure and XCS modules of the concrete TML must be valid against those DTDs.

TBX specification of a particular TML describes which varying types of terminological data (data-categories) are allowed and at what levels of a terminological entry they can occur. The default set of TBX data-categories is selected from ISO 12620:1999 (now ISO 12620:2009) [6]. There are four general types of TBX data-categories:

1. A core-structure module data-category is any data-category that is defined in the core-structure module DTD as a XML element.
2. A meta data-category is a general data-category used to group similar data-categories together. It is implemented as a core-structure module data-cate-

gory (XML element) with a `type` attribute. The `type` attribute values are derived from ISO 12620:2009 and listed in a XCS file.
3. A terminological data-category is an instance of the meta data-category with a particular value of the `type` attribute. A value of the `type` attribute represents the name of the corresponding terminological data-category.
4. A simple data-category is one value of a closed set of values, defined in an XCS file, that represents a permissible content of an XML element (meta data-category) having a specific `type` attribute value.

For example, XML element with an open tag `<descrip type="definition">` represents terminological data-category `definition` as an instance of the meta data-category `descrip`, while `<termNote type="grammaticalGender">` corresponds to the terminological data-category `grammaticalGender`, an instance of the meta data-category `termNote`. The former XML element expects free text as its content, while the latter XML element accepts a simple data-category — a value from the list: "masculine", "feminine", "neuter", "otherGender".

A TBX termbase, compliant with specific TML, is a set of TBX XML documents, each representing a record of terminological data and valid against core-structure module DTD and constraints defined in XCS module of that TML. The set of XML elements and attributes used in TBX documents is partly based on TEI (Text Encoding Initiative) P5 Guidelines [1]. Since TEI P5 Guidelines do not provide all the elements and attributes needed to describe terminological data, TBX also defines additional XML elements and attributes. There has been a recent attempt by Romary to customise the TEI P5 Guidelines and enable them to incorporate TBX [12].

TBX XML documents use MARTIF (Machine-Readable Terminology Interchange Format), specified by the ISO 12200:1999 standard [3] (Figure 5).

A whole TBX document, i.e. a root element `<martif>`, represents a terminological data collection, and consists of a `<martifHeader>` element and a `<text>` element.

A `<martifHeader>` element corresponds to the global information section of the TMF meta-model and consists of a description of the whole terminological data collection (`<fileDesc>`), information about the applicable XCS file (`<encodingDesc>`), as well as a history of major revisions to the collection (`<revisionDesc>`).

The `<text>` element consists of a `<body>` element and an optional `<back>` element. The `<body>` element contains a list of `<termEntry>` elements. The content of the `<termEntry>` element follows the structure of the TMF meta-model (Figure 6). A `<termEntry>` element is associated with the concept level of the TMF meta-model, i.e. represents a particular concept and contains a list of `<langSet>` elements.

A `<langSet>` element is associated with the language section level of the TMF meta-model and contains a list of `<tig>` and `<ntig>` elements. Each `<tig>` ("term information group") and `<ntig>` ("nested term information group") element represents the complete description of a particular term designating the concept in a given language, and corresponds to the term section level of the

TMF meta-model. Both `<tig>` and `<ntig>` are defined by the TBX standard, but the use of the `<tig>` element is giving way to the `<ntig>` element which is more complex and richer in information. Namely, besides the `<termNote>` element, `<ntig>` also features the `<termNoteGrp>` element, which contains additional information associated with a term, and the element `<termCompList>`, which matches the term component section of the TMF meta-model.

The optional `<back>` element corresponds to the complementary information about terminological data collection.

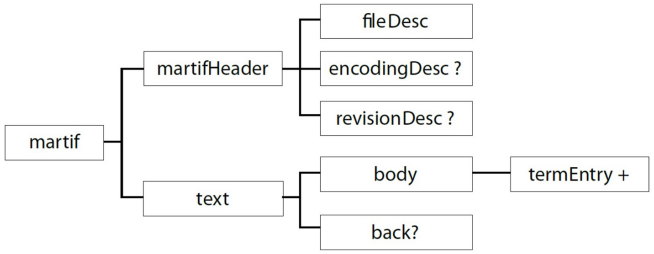

Fig. 5: The MARTIF structure of a TBX document

TBX can be used for interchange, dissemination, analysis and representation of both human-oriented and machine-oriented terminological data within an organization, as well as between an organization and external service providers. In interchange, it can support the flow of terminological data between different technologies and systems, integration of terminological data from multiple sources, and data conversion. Dissemination with TBX can include querying multiple termbases through a single user interface, setting-up data for download, obtaining a response from the user and serving data through web services.

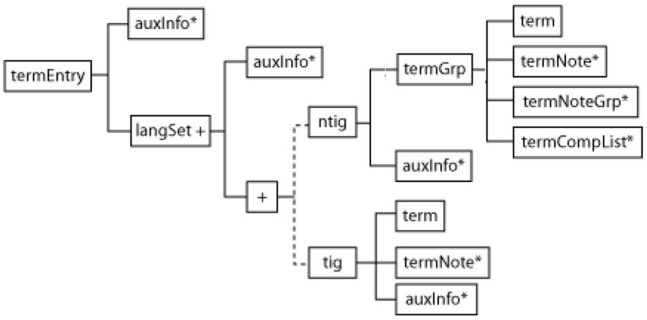

Fig. 6: Levels of a terminological entry

4 Transforming Termbases to TBX for Terminology Integration and Portability

Special attention was given to portability, simple and speedy transformation of subsets of concepts from terminological resources described in Section 2 to TBX, OLIF, OWL, RDF, LMF or MOODLE. This transformation has been secured by a wizard integrated in the terminological information system illustrated by one of its panels in Figure 7.

In the case when hypernymy/hyponymy relations between concepts form a hierarchical tree, and consequently subsets of concepts form sub-trees of this tree, as for example in RudOnto, export of subsets of concepts in the form of sub-trees is possible. In such a case, the user first selects a node (concept) in the hierarchy that represents the root of the sub-tree to be exported. Then, positioned on this node he/she invokes the export wizard and selects the export options [14]. This feature is not so far implemented for GeolISSTerm.

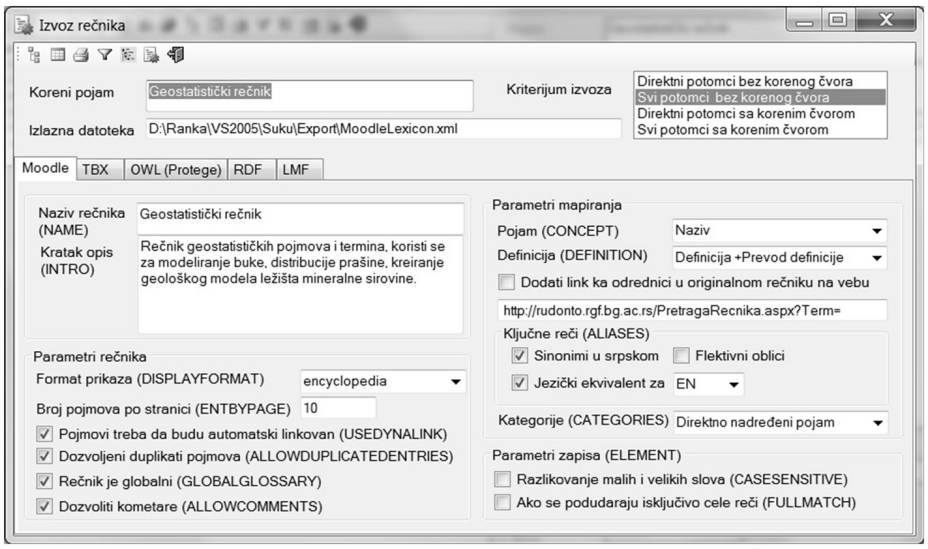

Fig. 7: Module for export from RudOnto into standard formats

Every transformation of termbases must take into account morphological information contained in these dictionaries. Thus in the case of TBX the TBX-default version was used, rather than TBX Basic, as it is more suitable for compounds, with inflectional characteristics more complex than those of single words. Namely, TBX Basic allows a limited set of data categories and cannot accept morphological and some other important information.

All languages used in a TBX document instance must be declared in the elements below element `<languages>` within the XCS file. Serbian is not specified

in the available TBXXCSV02.xcs file, the TBX default XCS file Version 0.5. However, this file can be modified to include additional data categories, where modifications must be specified in a commented-out section in the header of the file. Thus we added a new `<langInfo>` element with a two character language code "sr" for `<langCode>` in compliance with the IETF (Internet Engineering Task Force)[3] language tag and "Serbian" for `<langName>`[4].

Figure 8 represents the header of a TBX file generated by export from our termbases. The header is in compliance with ISO 12200 MARTIF [3] and contains reference to the XCSURI (the URI of the XCS file) constraint specification.

Fig. 8: Header of TBX file with the reference to XCSURI constraint specification

Part of a `<termEntry>` element for the main term *ležište mineralnih sirovina* (deposit of mineral resource) resulting from export is given in Figure 9, showing information that can be associated with the concept level, like `subjectField`, `relatedConcept`, `relatedConceptBroader`, `relatedConceptNarrower` and the like.

A complete description of the term is generated for each language represented in the termbase, e.g. English (`<langSet xml:lang="en">...</langSet>`), Serbian (`<langSet xml:lang="sr">...</langSet>`), etc.

Figure 10 presents an example of the information on the Language level (`<langSet>`) and Term level `<ntig>` for Serbian term *ležište mineralnih sirovina* (deposit of mineral resource) with related broader concept *ekonomska geologija* (economic geology), and related concepts *pojava mineralnih sirovina* (occurence of mineral resource), *rudno telo* (ore body) i *ruda* (ore). Related narrower concepts are also represented: *ležište metaličnih mineralnih sirovina*, *ležište nemetaličnih mineralnih sirovina*, *ležište energetskih mineralnih sirovina* (deposits of metallic minerals, non-metallic mineral deposits, deposits of fossil fuel resources).

Morphological data are supplied within `<termNote>`, with different values for attribute `type`, such as `partOfSpeech` with the value "noun", `termType` with the value "entryTerm" and `grammaticalGender` with the value "neuter". The element

[3] `http://tools.ietf.org`
[4] `http://tools.ietf.org/rfc/bcp/bcp47.txt`

```
<termEntry id="R10001">
  <descrip type="subjectField">Geology</descrip>
  <descrip type="relatedConcept">occurrence of mineral resource</descrip>
  <descrip type="relatedConcept">ore body</descrip>
  <descrip type="relatedConcept">ore</descrip>
  <descrip type="relatedConceptBroader">economic geology</descrip>
  <descrip type="relatedConceptNarrower">deposits of metallic minerals</descrip>
```

Fig. 9: Part of a TBX file with information at the concept level

`<termCompList>` also features different attribute types, such as `lemma` with the value "ležište (ležište.N300:ns1q) mineralnih sirovina", `morphologicalElement` with the value "NC_N4X+Comp", namely the transducer for the multiword unit inflection, as well as `termElement`, a list of components of the compound word. During export, information about the lemma and the inflection class is supplied by a web service developed by University of Belgrade HLT Group based on Serbian electronic morphological dictionaries. The export of information on all inflective forms is currently under development. Finally, hyphenation of each word is specified as type `hyphenation`.

The definition of the term is supplied within the `<descripGrp>` element with type `definition`, followed by a bibliographical reference, described in the `<admin>` element with attribute `type` value `sourceIdentifier`, and described in detail in complementary information. In the `<ntig>` elements that follow, synonymous terms are described, such as for example, *mineralno ležište* (mineral deposit) with `partOfspeach` "noun" and `termType` "synonym".

The element `<refObjectList>` within the `<back>` complementary information with type `bibl` describes the bibliography that has been used (Figure 11), where each bibliographical unit has its identifier given in the attribute ID of the element `<refObject>` (e.g. `jelenkovic00`).

Thus the wizard integrated in the terminological information system secures comprehensive, fully automatic transformation of our termbases from their in-house scheme to standard TBX format, which allows for terminology integration and portability.

5 Conclusion and further development

In this paper we have argued for the necessity of maintaining and developing termbases, especially in the case of text with expert terminology. We have demonstrated how termbases kept in various in-house schemas can automatically be transformed in one of the standard formats, such as TBX. We have also showed how information from termbases can be upgraded during transformation to TMX with morphological information using Serbian electronic morphological dictionaries and a web service developed by HLT Group from University of Belgrade.

There is still much work to be done in this area, in the first place an enhancement of domain specific morphological dictionaries of terminology related

```
<langSet xml:lang="sr">
  <descrip type="relatedConceptBroader">ekonomska geologija</descrip>
  <descrip type="relatedConcept">pojava mineralnih sirovina</descrip>
  <descrip type="relatedConcept">rudno telo</descrip><descrip type="relatedConcept">ruda</descrip>
  <descrip type="relatedConceptNarrower">ležište metaličnih mineralnih sirovina</descrip>
  <descrip type="relatedConceptNarrower">ležište nemetaličnih mineralnih sirovina</descrip>
  <descrip type="relatedConceptNarrower">ležište energetskih mineralnih sirovina</descrip>
  <ntig>
    <termGrp>
      <term>ležište mineralnih sirovina</term><termNote type="partOfSpeech">noun</termNote>
      <termNote type="termType">entryTerm</termNote><termNote type="grammaticalGender">neuter</termNote>
      <termCompList type="lemma"><termComp>ležište(ležište.N300:ns1q) mineralnih sirovina</termComp></termCompList>
      <termCompList type="morphologicalElement"><termComp>NC_N4X+Comp</termComp></termCompList>
      <termCompList type="termElement">
        <termComp>ležište</termComp><termComp>mineralnih</termComp><termComp>sirovina</termComp>
      </termCompList>
      <termCompList type="hyphenation">
        <termComp>leži</termComp><termComp>šte</termComp><termComp>mine</termComp>
        <termComp>ralnih</termComp><termComp>siro</termComp><termComp>vina</termComp>
      </termCompList>
    </termGrp>
    <descripGrp>
      <descrip type="definition">Ležišta mineralnih sirovina su geološka tela prirodnih ili veštačkih (industrijskih) granica.
Nastala su u prirodi u okviru složenih geoloških i fizičko-hemijskih procesa. Predstavljaju sastavni deo Zemljine kore, a izgrađena su od
različitih minerala, elemenata i jedinjenja koja se u industriji koriste u prirodnom ili prerađenom stanju. Ležišta mineralnih sirovina su i
geološko-ekonomska kategorija, što znači da se pri savremenim uslovima eksploatacije mineralnih sirovina i dostignutom
tehničko-tehnološkom nivou pripreme i prerade rude, u privredi koriste dajući pozitivne ekonomske efekte.</descrip>
      <admin type="sourceIdentifier" target="jelenkovic00">jelenkovic00</admin>
    </descripGrp>
  </ntig>
  <ntig>
    <termGrp>
      <term>mineralno ležište</term>
      <termNote type="partOfSpeech">noun</termNote><termNote type="termType">synonym</termNote>
    </termGrp>
  </ntig>
```

Fig. 10: Part of a TBX file with information at the language and term level

to mining and geology. An approach for realizing this task is being developed based on n-grams obtained by processing texts containing expert terminology, such as the textbook "Introduction to Mining". The approach also envisages integration with cascades for named entity recognition such as mining equipment, specific minerals and the like. Building of an aligned Serbian-English corpus of texts in the area of mining and geology from sources like the bilingual journal "Underground Mining" are underway. The possibility of searching such corpora of expert texts would contribute to further development of domain specific terminological resources. Initiating an interactive, web-based Terminology forum would also be beneficial. In mining engineering and geology environments, the volumes of content and subsequently of the required terminology are typically large. Therefore, integrating related terminology into a translation pipeline should be explored [17]. This approach requires a process that is as automated as possible. With term extraction as its cornerstone, it requires a post-processing strategy that repurposes existing lexical resources to maximize efficiency. Terms extracted from corpora and subsequently translated should be channeled into the company termbase, so that they can be leveraged for other purposes.

Acknowledgments. This research was supported by the Serbian Ministry of Education and Science under the grant #III 47003.

```
<back>
  <refObjectList type="bibl">
    <refObject id="jelenkovic00">
      <itemSet type="author">
        <item type="surname">Jelenković</item>
        <item type="fname">Rade</item>
      </itemSet>
      <itemSet type="book">
        <item type="title">Ležišta metaličnih mineralnih sirovina</item>
        <item type="edition">First</item>
        <item type="isbn">86-7352-056-8</item>
      </itemSet>
      <item type="date">2000</item>
      <itemSet type="pubname">
        <item type="orgName">Rudarsko-geološki fakultet Univerziteta u Beogradu, Beograd</item>
      </itemSet>
    </refObject>
```

Fig. 11: Bibliographical items in the TBX file

References

1. Lou Burnard and Syd Bauman, editors. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium, 2009.
2. M. Teresa Cabré Castellví. Theories of Terminology — Their Description, Prescription and Explanation. *Terminology*, 9(2):163–199, 2003.
3. ISO. Computer Applications in Terminology — Machine-Readable Terminology Interchange Format (MARTIF) – Negotiated Interchange, 1999. Ref. ISO 12200:1999.
4. ISO. Computer Applications in Terminology — Terminological Markup Framework, 2003. Ref. ISO 16642:2003.
5. ISO. Systems to Manage Terminology, Knowledge and Content — TermBase eXchange (TBX), 2008. Ref. ISO 30042:2008.
6. ISO. Terminology and Other Language and Content Resources – Specification of Data Categories and Management of a Data Category Registry for Language Resources, 2009. Ref. ISO 12620:2009.
7. Cvetana Krstev. *Processing of Serbian – Automata, Text and Electronic Dictionaries*. University of Belgrade, Faculty of Philology, Belgrade, 2008.
8. Cvetana Krstev, Ranka Stanković, Duško Vitas, and Ivan Obradović. WS4LR: A Workstation for Lexical Resources. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. European Language Resources Association (ELRA), 2006.
9. Ch. Lieske, S. McCormick, and G. Thurmair. The Open Lexicon Interchange Format (OLIF) Comes of Age. In *Proceedings of the Machine Translation Summit VIII*, 2001.
10. Alan K. Melby. Terminology in the Age of Multilingual Corpora. *The Journal of Specialized Translation*, 18:7–29, 2012.
11. Uwe Reinke. State of the Art in Translation Memory Technology. *Translation: Computation, Corpora, Cognition*, 3(1), 2013.
12. Laurent Romary. TBX Goes TEI - Implementing a TBX Basic Extension for the Text Encoding Initiative Guidelines. *CoRR*, abs/1403.0052, 2014.

13. Ranka Stanković, Ivan Obradović, Cvetana Krstev, and Duško Vitas. Production of Morphological Dictionaries of Multi-Word Units Using a Multipurpose Tool. In K. Jassem, P. W. Fuglewicz, M. Piasecki, and A. Przepi rkowski, editors, *Proceedings of the Computational Linguistics-Applications Conference, October 17-19, 2011. Jachranka, Poland*, pages 77–84. Polish Information Processing Society, 2011.
14. Ranka Stanković, Ivan Obradović, Olivera Kitanović, and Ljiljana Kolonja. Building Terminological Resources in an e-Learning Environment. In D. Milošević, editor, *Proceedings of the Third International Conference on e-Learning, eLearning-2012, September 2012, Belgrade, Serbia*, pages 114–119, 2012.
15. Ranka Stanković, Branislav Trivić, Olivera Kitanović, Branislav Blagojević, and Velizar Nikolić. The Development of the GeolISSTerm Termino logical Dictionary. *INFOtheca*, XII(1):49a–63a, 2011.
16. Duško Vitas, Ljubomir Popović, Cvetana Krstev, Ivan Obradović, Gordana Pavlović-Lažetić, and Mladen Stanojević. *Srpski jezik u digitalnom dobu – The Serbian Language in the Digital Age*. META-NET White Paper Series. Georg Rehm and Hans Uszkoreit (Series Editors). Springer, 2012.
17. Kara Warburton. Processing Terminology for the Translation Pipeline. *Terminology*, 19(1):93–111, 2013.