Tempus

TERMINOLOGICAL AND LEXICAL RESOURCES USED TO PROVIDE OPEN MULTILINGUAL EDUCATIONAL RESOURCES

BILJANA LAZIĆ,[1] **DANICA SENIČIĆ,**[2]
ALEKSANDRA TOMAŠEVIĆ,[1] **BOJAN ZLATIĆ**[1]

[1]*University of Belgrade, Faculty of Mining and Geology*
[2]*Université catholique de Louvain, Faculté de philosophie, arts et lettres*

BAEKTEL

Content

Natural Language Processing (NLP) and Education

Terminology

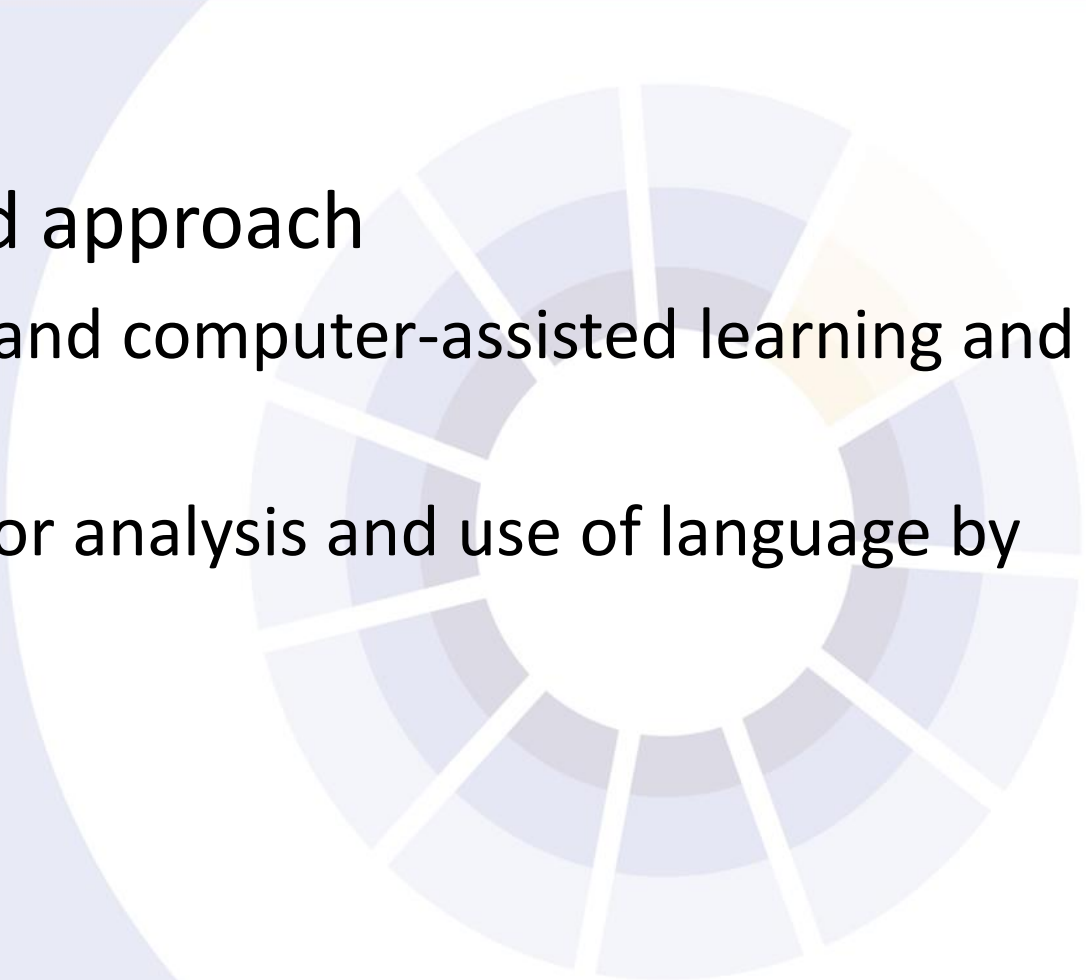BAEKTEL and multilinguality

Lexical resources

Terminology extraction

Termi

Natural Language Processing (NLP) and Education

- two-faceted approach
  - E-learning and computer-assisted learning and instruction
  - NLP tools for analysis and use of language by machines

Natural Language Processing (NLP) and Education

- Ellis Batten Page - he is considered to be the father of automated essay scoring

- Technology Enabled Assessment (TEA)

- Intelligent Tutoring Systems (ITS)

- Massive Open Online Courses – MOOCs - changing the face of distant learning and education altogether by erasing geographical and spatial constraints, leaving the traditional education model behind

NLP tools

- Interactive forums and teaching assistants rely greatly on various NLP tools to help them cater to a large number of students from all over the world
- These tools may include:
  - assessment of text and speech,
  - writing assistants,
  - automatic generation of exercises,
  - wrap up questions and
  - online instructional environments

Natural Language Processing (NLP) and Education

- The main goal of NLP tools in education is to automate time-consuming, laborious teachers' tasks such as curriculum creation or assignment assessment and do so in a timely manner.

- It is important to note that NLP requires multidisciplinary collaboration in all domains of its application.

- For example, NLP tools for language learning must connect to Second Language Acquisition (SLA) and Foreign Language and Teaching (FLTL) research with insights from Cognitive Psychology and Empirical Educational Science.
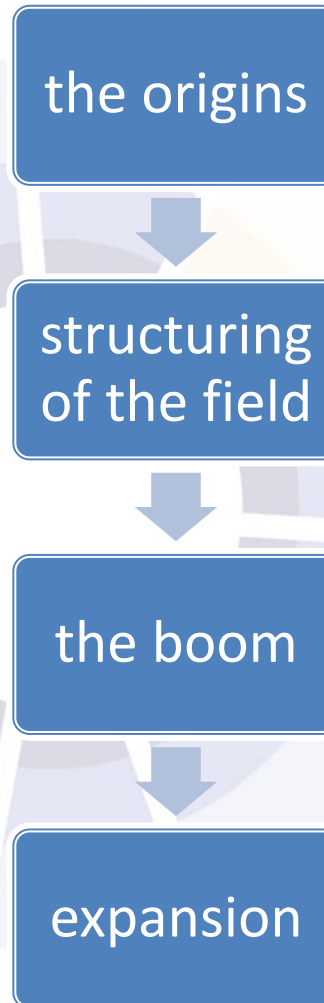
Terminology

- a young interdisciplinary scientific field

- The interest in it arose in 1930s when an electrical engineer, **Eugen Wüster**, became engaged in publishing papers concerning terminology as an individual discipline.

linguistics (lexicography)

? (mining, mathematics etc.)

cognitive and communication sciences,
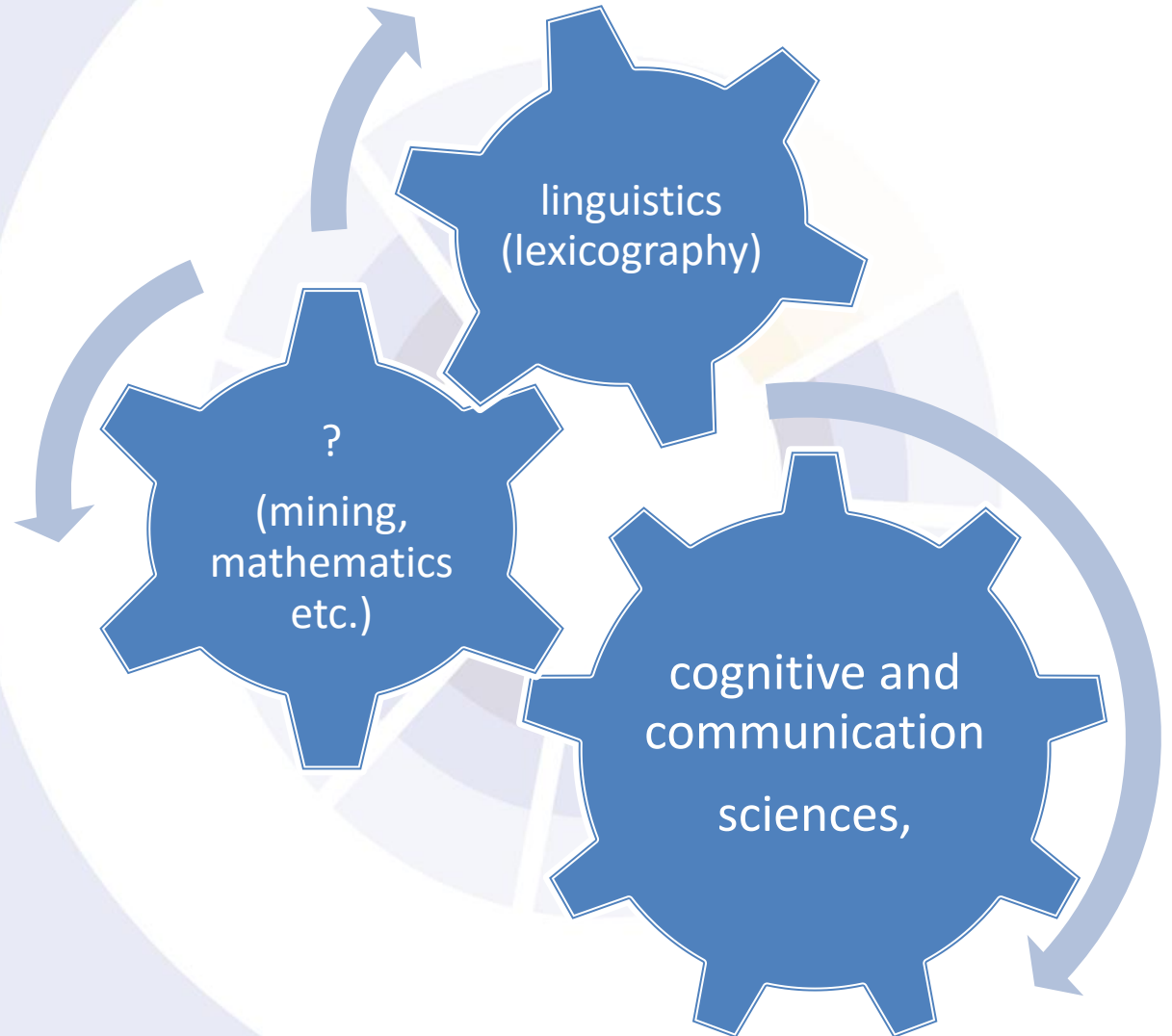
Terminology

- Macmillan Dictionary: "the words and phrases used in a particular business, science, or profession"

- ISO 12620:"The set of designations belonging to the language of a given subject field"

the origins

↓

structuring of the field
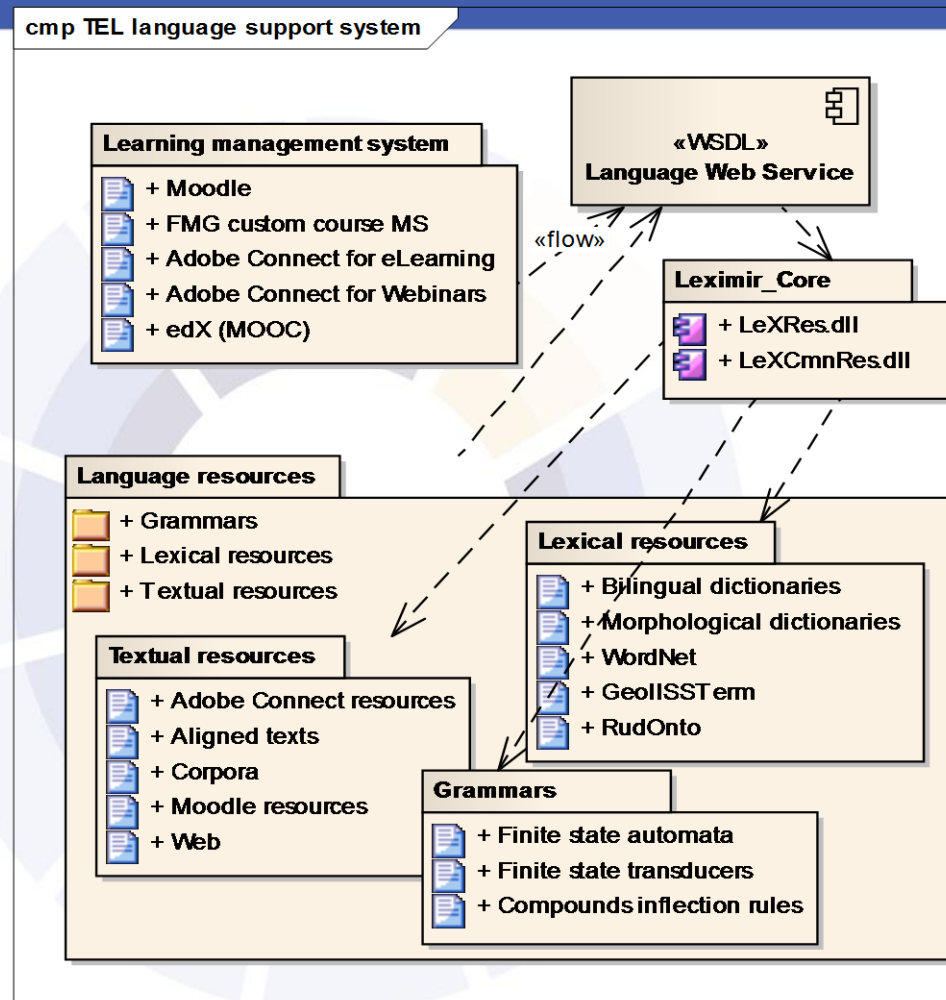
↓

the boom

↓

expansion

BAEKTEL and multilinguality

University of Belgrade (UB) hosts a central repository based on:
• BAEKTEL Metadata Portal (BMP)
• terminological web application for management, browse and search of terminological resources,
• web services for linguistic support (query expansion, information retrieval, OER indexing, etc.),
• annotation of selected resources,
• OER repository on local edX platform.

The BAEKTEL language support system consists of several software components administrating in the same time language resources: grammars, lexical and textual

I. Obradović, R. Stanković, J. Prodanović and O. Kitanović, "A TEL platform blending academic and entrepreneurial knowledge," in Proc. 4th Conference on e- Learning, 2013, pp. 65–70.

Lexical resources

- Morphological Dictionaries
- WordNet
- RudOnto
- GeoISSTerm
- Termi
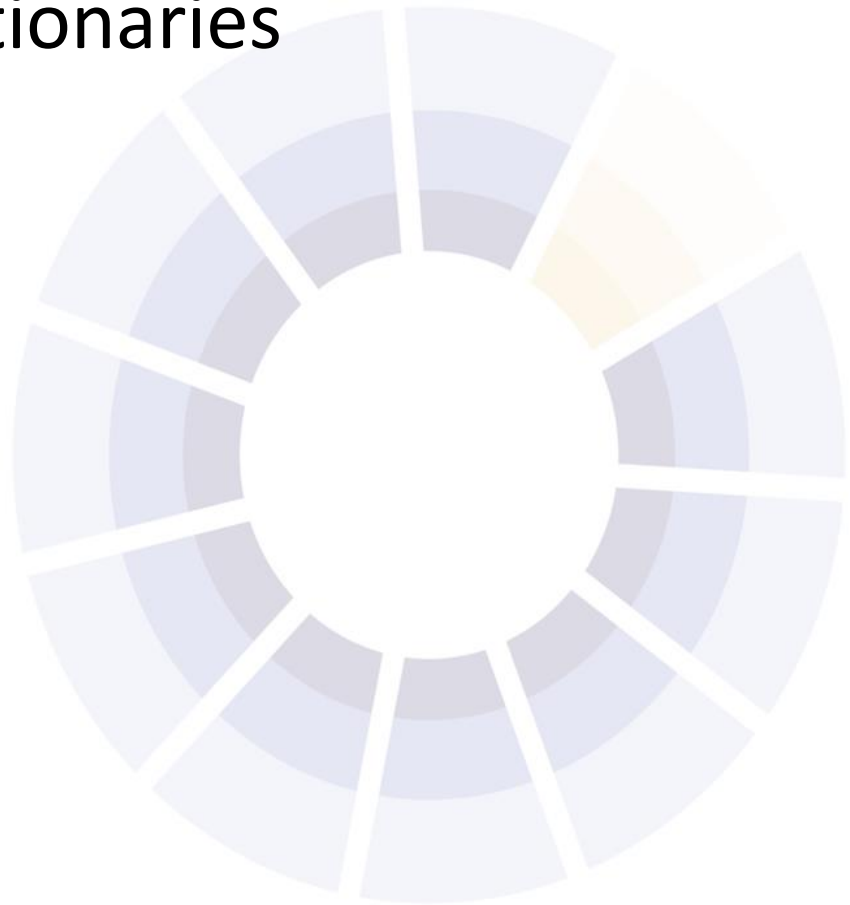
Serbian morphological dictionaries

DELAS

- **učiteljica, N651+Hum+GM:fs4v**

DELAC

- **lekar(lekar. N2:ms1v) akušer (akušer. N2:ms1v),NC_NXN+Comp+Hum**

- **Dictionary of simple words** - 133,361 = 4,581,657word forms.

- **Serbian morphological dictionary of compounds** - 13,717 = 262,686 word forms.

Serbian WordNet

- composed of synsets, or sets of synonymous words representing a concept, with basic semantic relations between them forming a semantic network
- each synset word is denoted by a "literal string" followed by a "sense tag" which represents the specific sense of the literal string in that synset
- Interlingual index (ILI)
- 20 000 synsets from different specific domains (e. g. law, biomedicine, mythology, culinary etc.)

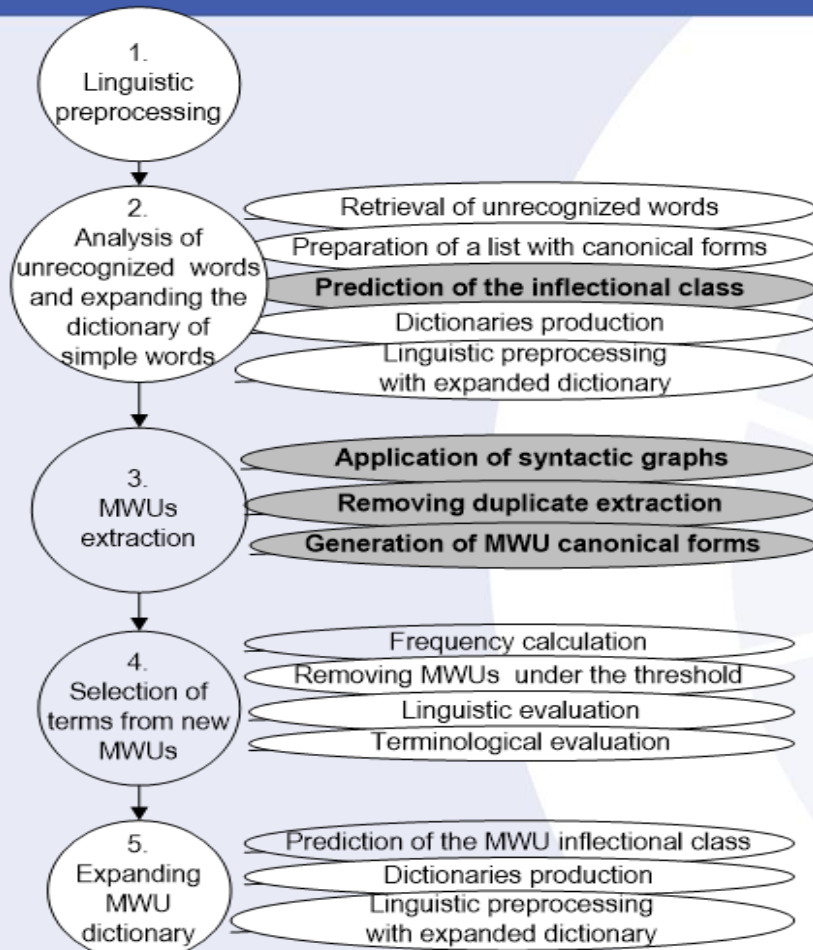# RudOnto and GeoISSTerm

RudOnto

- terminological resource that covers domain of mining
- taxonomy of terms
- definition, synonyms, and bibliographical reference to their source and equivalent terms in other languages

GeoISSTerm

- a thesaurus of geological terms with entries in Serbian and English
- developed as a part of the GeoISS project.
- It contains more than 3000 dictionary entries.

Terminology extraction

1. Linguistic preprocessing

2. Analysis of unrecognized words and expanding the dictionary of simple words
- Retrieval of unrecognized words
- Preparation of a list with canonical forms
- **Prediction of the inflectional class**
- Dictionaries production
- Linguistic preprocessing with expanded dictionary

3. MWUs extraction
- **Application of syntactic graphs**
- **Removing duplicate extraction**
- **Generation of MWU canonical forms**

4. Selection of terms from new MWUs
- Frequency calculation
- Removing MWUs under the threshold
- Linguistic evaluation
- Terminological evaluation

5. Expanding MWU dictionary
- Prediction of the MWU inflectional class
- Dictionaries production
- Linguistic preprocessing with expanded dictionary

C. Krstev, R. Stanković, I. Obradović and B. Lazić, "Terminology Acquisition and Description Using Lexical Resources and Local Grammars," in Proc. 11th Conference on Terminology and Artificial Intelligence, 2015, pp. 81– 89.

- **AXN** – informacioni system (information system)
- **2XN** – audio-zapis (audio record)
- **N2X** – baza podataka (database)
- **N4X** – obrada prirodnih jezika (natural language processing)
- **AXN2X** – optičko prepoznavanje karaktera (optical character recognition)
- **NXN** – bager kašikar (wheel excavator)
- **AXAXN** – morfološki elektronski rečnik (electronic morphological dictionary)
- **N6X** – grupa za jezičke tehnologije (language technology group)
- **AXN4X** – površinska eksploatacija mineralnih sirovina (surface exploitation of mineral resources)
- **2XAXN** – informaciono-komunikacione tehnologije (information and communications technology)

TERMI – an application for terminology management

- http://termi.rgf.bg.ac.rs/
- The application consists of three basic web pages which manage terminology: browse, search and update.
- Each term comes with the **name**, **definition**, **synonyms**, **abbreviations** and **a bibliographic source**. Each term, except the top term in dictionary tree, has a hyperonym term, while each term can have an arbitrary number of hyponym terms.

Conclusion

- Lexical and terminological resources offer priceless aid for better understanding of the available OER contents. Presented resources are also helpful in a sense of appropriate translation option.

- The potential expansion of such resources would inevitably lead to a more fruitful information retrieval and extraction, providing an invaluable education resource, applicable in all of its domains.

- In the further work bilingual terminology extraction will be considered.

Hvala za vašo pozornost

Thank you for your attention

Grazie per la vostra attenzione

Vă mulțumesc pentru atenție

Хвала на пажњи

Hvala na pažnji