WG 2

UniDive

2nd General Meeting

University of Naples "L'Orientale

Naples, Italy, 8-9 February 2024

https://unidive.lisn.upsaclay.fr/

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

Bridging the Geological Lexicon and Corpus with Focus on MWEs Extraction

Biljana Rujević[1], Cvetana Krstev[2], Mihailo Škorić[3]

[1,3]University of Belgrade, Faculty of Mining and Geology

[2]Association for Language Resources and Technologies Belgrade, Serbia

Terminology extraction: from corpus to lexical entries

Corpus GeoSrpKor

- Geological domain corpus
- 69 documents in Serbian language
- Available on NoSketch instance: https://noske.jerteh.rs/

Rule-based MWE extraction

- **AXN** – an adjective followed by a noun; the adjective and the noun must agree in gender, number, case, and animateness (1162), ***ugljevita glina 'coal-clay'***
- **N2X** – a noun followed by a word that does not inflect in the MWE (309), ***mineral gline 'clay mineral'***
- **N4X** – a noun followed by two words that do not inflect in the MWE (190), ***izrada geološke karte 'creation of a geological map'***
- **2XN** – a noun preceded by a word that does not inflect in the MWE (70), ***alevrit glina 'clay aleurite'***
- **AXN2X** – a noun preceded by an adjective that agrees with it in four grammatical categories and followed by a word that does not inflect in the MWE (55), ***centralni deo masiva 'central part (of a) massif'***
- **2XAXN** – an adjective followed by a noun that agrees in all four grammatical categories and preceded by a word that does not inflect in the MWE (35), ***jezersko-barski sedimenti 'lake-swamp sediments'***
- **AXAXN** – a noun preceded by two adjectives that agree with it in four grammatical categories (19), ***lecki andezitski masiv 'Lece adensite massif'***
- **NXN** – a noun followed by a noun that agrees with it in number and case, where the separator can be a hyphen (13), ***facija mrtvaja*** '***oxbow facies***'
- **AXN4X** – a noun preceded by an adjective that agrees with it in four grammatical categories and followed by two words that do not inflect in the MWE (12), ***ugljevita glina sa proslojcima 'carbon clay with layers***'
- **N6X** – a noun followed by three words that do not inflect in the MWE (4), ***krečnjak sa proslojcima rožnaca 'limestone with layers of chert'***

leximirka.jerteh.rs/LexicalEntry/Delaf?eid=217258

magmatska stena,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fs1q
magmatska steno,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fs5q
magmatske stene,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fp1q:fp4q:fp5q:fs2q:fw2q:fw4q
magmatskih stena,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fp2q
magmatskim stenama,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fp3q:fp6q:fp7q
magmatskima stenama,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fp3q:fp6q:fp7q
magmatskoj steni,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fs3q:fs7q
magmatskom stenom,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fs6q
magmatsku stenu,magmatska(magmatski.A2:aefs1g) stena(stena.N600:fs1q).N:fs4q

Leximirka Categories Files Entries Corpora Log off

Lexical Entry #217258 Edit

magmatska stena NC_AXN Delac-im-dec14.dic

Check in dictionaries:

- show WordNet

ENG30-14931879-n, BCS=, +0.00000, -0.00000
Stena koja je nastala hlađenjem i stvrdnjavanjem magme.
magmatska stena 1
DOM: geology, SUMO: Mineral +
---> hypernym: ENG30-14696793-n kamen
... <--- hyponym <- hypernym: ENG30-14932303-n vulkanska stena
... <--- hyponym <- hypernym: ENG30-14725364-n andezit
... <--- hyponym <- hypernym: ENG30-14932554-n diorit
... <--- hyponym <- hypernym: ENG30-14932645-n gabro
... <--- hyponym <- hypernym: ENG30-14726484-n aplit
... <--- hyponym <- hypernym: ENG30-14932741-n Pegmatit
... <--- hyponym <- hypernym: ENG30-14933236-n Liparit; riolit
... <--- hyponym <- hypernym: ENG30-14933314-n Efuzivna stena, vulkanska stena; Izlivna stena
... <--- <- holo_portion: ENG30-15044648-n Sodalit

- show Terminološki

Magmatska stena sin: ;Stena obrazovana očvršćavanjem rastopa.
eng: Magmatska stena; syn: Stena obrazovana ocvršcavanjem rastopa. . Domen: Geologija. Izvor: GeolISSTer

Check in external dictionaries: Wiktionary Babelnet Termi Globi

Search corpora: GeoSrpKor Concordances Attestations Form Frequences Lemma Frequences

Senses (1):

1. +Comp+Conc+Stena+DOM=Geol

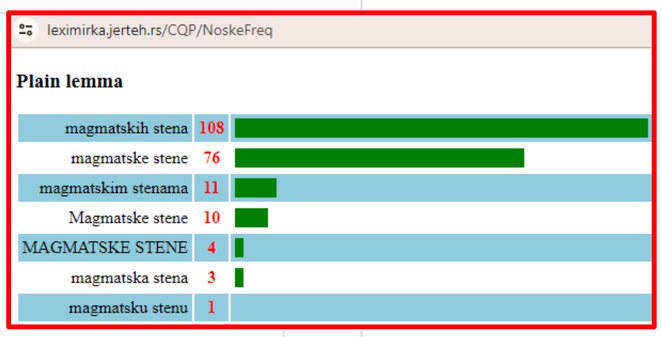
Domains:	geologija
Properties:	polileksemska jedinica, konkretna imenica,
Note:	WN;sep07

Is composed of:

Form	Lemma	FST Code	Gram Cat	Separator
magmatska	magmatski	A2	aefs1g	
stena	stena	N600	fs1q	

leximirka.jerteh.rs/CQP/NoskeFreq

Plain lemma



leximirka.jerteh.rs/CQP/Noske

Plain lemma

, magmatske i metamorfne stene) i M. Divljan (	magmatske stene	) ; hemijske analize je uradila S. Crnčević , a
" predstavljaju snažno metamorfisane bazične	magmatske stene	. Teksture su masivne , nekad paralelno
različitih sedimentnih , metamorfnih i	magmatskih stena	. Pored ranijih nalazaka asocijacije
: Q - kvarc-rožnac , F-feldspati - fragmenti	magmatskih stena	, M - liskuni - fragmenti metamorfnih stena ; 1 -
kao i pliokvartar . Neogene je starosti i većina	magmatskih stena	(daciti , andeziti i alkalne gabroidne stene sa
sedimentne stene progresivnog karaktera , a za	magmatske stene	retrogradnog karaktera . Oni su se , obzirom da u
šire okoline Novog Pazara i tercijarne	magmatske stene	Kopaonika i Golije takođe su detaljno obrađeni
je veliki dio terena izgrađen od klastičnih i	magmatskih stena	to ovo područje obiluje sa dosta velikom
su dosta zastupljeni , retki su krečnjaci , a od	magmatskih stena	gabrovi , dijabazi , spiliti a retko i
, dijabazi , spiliti a retko i serpentiniti .	Magmatske stene	, koje se javljaju kako u samoj
, spilita i male mase gabrova . Pošto su ove	magmatske stene	identične analognim stenama koje se javljaju u
su kompetentniji peščari i rožnaci , kao i	magmatske stene	, komadani i pri kretanju manje ili više
u dijabaz - rožnačkoj formaciji su slične .	MAGMATSKE STENE	U DIJABAZ-ROŽNAČKOJ FORMACIJI U
FORMACIJI U dijabaz-rožnačkoj formaciji od	magmatskih stena	javljaju se dijabazi , spiliti , zatim gabrovi i
, ali i blokove . Pritom najveće pojave ovih	magmatskih stena	su u blizini i po obodu ultramafitske mase
ultramafitskog masiva sa sedimentnim i	magmatskim stenama	dijabaz-rož načke formacije , u erozionom
od autometamorfno alterisanih bazičnih	magmatskih stena	. S obzirom da su srednji članovi ovog niza
i vezana je za razlomne zone većeg inteziteta u	magmatskim stenama	. DONjA KREDA Donja kreda predstavljena je
, pojasevi krede , tercijarni sedimenti i	magmatske stene	. Nešto kasnije , A. Viquesnel (1842)
čije je poreklo svakako vezano za kisele	magmatske stene	. Hromita i epidota ima malo , dok su listići

Glosses

ARTEŠKI BUNAR (def. Bunar koji kaptira podzemne vode sa arteškim pritiskom koji se nalazi iznad površine terena.) Sinonimi: Samoizlivni arteški bunar; samoizlivni bunar – Eng. **Artesian well** (def. Well that penetrates groundwater with artesian pressure, which is above the surface.) Synonyms: Flowing artesian well; overflowing well

KVARCNI PESAK (def. Pesak preovlađujuće sastavljen od zrna kvarca.) – Eng. **Quartz sand** (def. Sand predominantly composed of quartz grains.)

MAGMATSKA STENA (def. Stena obrazovana ocvršćavanjem rastopa.) – Eng. ´ **Igneous rock** (def. A rock formed by solidification of a melt.)

NAFTNI PESKOVI (def. Naftni peskovi su značajna mineralna sirovina iz grupe nekonvencionalnih izvora za dobijanje nafte. Poznati su i pod nazivom „tar sands" i predstavljaju nekadašnja (degradirana)
ležišta nafte, osiromašena lakšim ugljovodonicima. Najcešće nastaju erozijom povlatnih sedimenata.) – Eng. **Tar sands** (def. Tar sands are significant mineral resources from the group of unconventional oil
sources. They usually represent the former reservoirs of oil (degraded reservoirs), characterized by reduced amounts of lighter hydrocarbons. The genesis of tar sands is usually related to erosion of overlying sediments.)

PESKOVITI KREČNJAK (def. Krečnjak koji sadrži zrna kvarca kao klasticnu komponentu.) – Eng. **Sandy limestone** (def. Limestone containing quartz grains as its osteoclastic component.)

PODZEMNE VODE (def. Svaka voda ispod površine zemlje (u litosferi), bez obzira na agregatno stanje, vidove, poreklo, fizicke osobine, hemijski, radiološki i mikrobiološki sastav.) – Eng. **Groundwater** (def. Any water that can be found below ground level (in the lithosphere), regardless of their physical state, type, origin, physical properties, chemical, radiological and microbiological
composition.) Synonyms: Subterranean water; underground water; subsurface water.

Conclusion

Bridging Lexicon and Corpus is designed to facilitate dictionary use for the end users. It allows users to check examples in corpora concordances directly from the lexical entry, which saves them time and is suitable for those who may not be familar with corpus queries. Geological domain lexicon and corpus used for MWE extraction are meant for geologists and geology students who may not have expertise in corpus linguistics.

biljana.rujevic@rgf.bg.ac.rs; cvetana@jerteh.rs; mihailo.skoric@rgf.bg.ac.rs