

## Розета - вишејезички алат за обраду природних језика и Linked Data

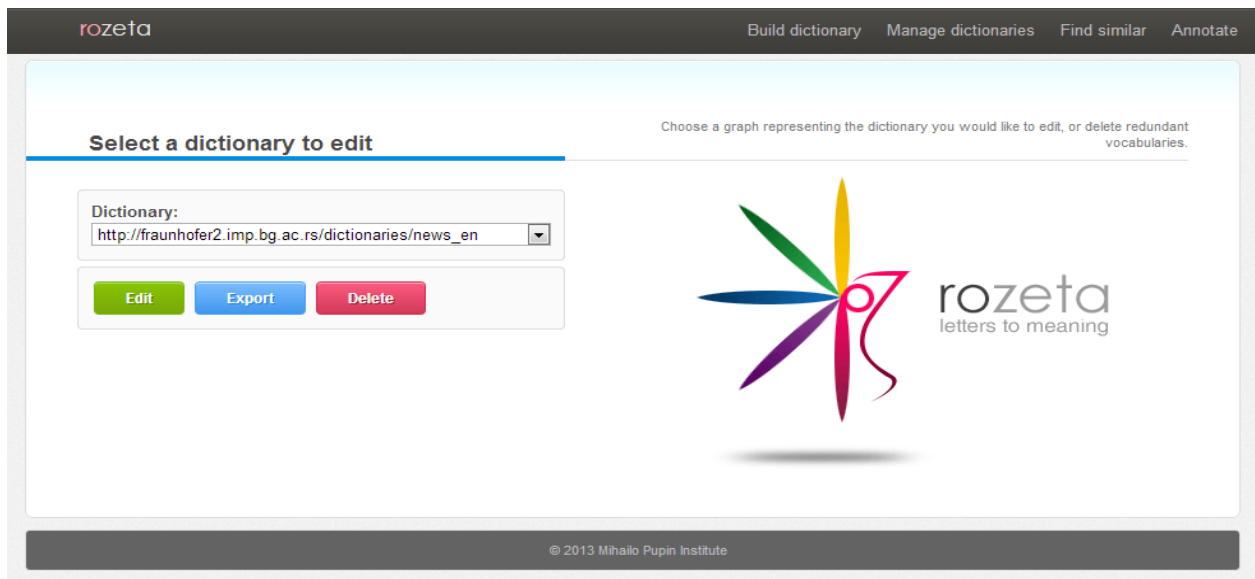
Розета је алат који омогућава претраживање базе текстова на основу неког текста, класификацију текстова и повезивање доменски специфичних термина са појмовима у *Linked Data Cloud*-у. Алат се ослања на STRUTEX технику за представљање докумената писаних у природним језицима у структурираној форми. STRUTEX аутоматски конструише доменске речнике у *Linked Data* формату који заузврат омогућавају лако добављање додатног знања и повезивање појмова са *Linked Data Cloud*-ом.

Алат се састоји из четири модула:

- Аутоматска конструкција *Linked Data* речника
- Претрага докумената по сличности и аутоматска класификација
- Семантичка анотација и обогаћивање докумената
- Ручно уређивање и обогаћивање речника

Велике колекције докумената у природном језику су потенцијалне ризнице знања, али је и извлачење значења из истих, уз минималну интервенцију корисника, велики изазов. Овом проблему Розета прилази кроз семантичке технологије и сервисе за анотацију и откривање нових веза (*DBpedia Spotlight*, *Sindice*, *SPARQL*). За ове потребе, као и за потребе проналажења сличних докумената, приликом спецификације решења, узете су у обзир све препреке и потешкоће приликом рада са скуповима докумената у природном језику, и то у различитим доменима. Резултат је сервис једноставног корисничког интерфејса који самог корисника не излаже комплексним технологијама које се налазе у позадини.

Софтверска компонента је урађена је у оквиру пројекта *LOD2 "Creating Knowledge out of Interlinked Data"*, European Commission, FP7, IP, 2010-2014, GA 257943.



Слика. Розета алат

rozeta Build dictionary Manage dictionaries Find similar Annotate

URI: [http://fraunhofer2.imp.bg.ac.rs/dictionaries/news\\_en/entry#Balance](http://fraunhofer2.imp.bg.ac.rs/dictionaries/news_en/entry#Balance)

anchorOf: Balance

classAnnotation: Business

file: fraunhofer2.imp.bg.ac.rs.dictionaries.news\_en/Business/Business11.txt  
count: 2  
weight: 8.99962

Wiktionary suggestions

- <http://www.mlode.nlp2rdf.org/semanticquran/Balance>
- <http://www.mlode.nlp2rdf.org/semanticquran/quran42-17-16-en>
- <http://www.mlode.nlp2rdf.org/semanticquran/quran55-7-14-en>
- <http://www.mlode.nlp2rdf.org/semanticquran/quran57-25-18-en>
- <http://wiktionary.dbpedia.org/resource/Balance>
- <http://wiktionary.dbpedia.org/resource/Balance-French-ProperNoun-1en>
- <http://wiktionary.dbpedia.org/resource/Balance-French-ProperNoun-2en>
- <http://wiktionary.dbpedia.org/resource/Balance-German-Noun-1en>
- <http://mlode.nlp2rdf.org/semanticquran/Balance>
- <http://mlode.nlp2rdf.org/semanticquran/quran42-17-16-en>
- <http://mlode.nlp2rdf.org/semanticquran/quran55-7-14-en>
- <http://mlode.nlp2rdf.org/semanticquran/quran57-25-18-en>

Link as skos:exactMatch Link as skos:relatedMatch Link via custom property

Delete Custom link More links

- Уређивање *Linked Data* речника

Build dictionary Edit dictionary Similarity search Phrase tagging

## Phrase extraction & linking

Recommendation sources

- Wiktionary
- DBpedia SPARQL

Please use the above widget to manage your linking recommendation sources (SPARQL endpoints). The sources you add will be saved for future sessions.

fraunhofer2.imp.bg.ac.rs.dictionaries.news\_en/World/US02.txt

Senate panel approves Chuck Hagel defence nomination Chuck Hagel was cross-examined during eight hours of testimony at his Senate confirmation hearing Continue reading the main story Related Stories Pentagon nominee in testy hearing Profile: Chuck Hagel Obama team's inbox of problems A US Senate committee has approved the nomination of ex-Republican Senator Chuck Hagel to be

DS2 dictionary entry: [Senate](#)

DBpedia Spotlight: [http://dbpedia.org/resource/United\\_States\\_Senate](http://dbpedia.org/resource/United_States_Senate) (Similarity: 0.2)

Wiktionary: <http://wiktionary.dbpedia.org/resource/Senate>

DBpedia SPARQL: <http://dbpedia.org/resource/Senate>

Link as skos:exactMatch Link as skos:relatedMatch

- Семантичка анотација и обогаћивање докумената

Class	Filename	Similarity	
World	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/World/US01.txt	0.405386	<a href="#">See document</a>
World	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/World/US05.txt	0.093985	<a href="#">See document</a>
World	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/World/US02.txt	0.0786986	<a href="#">See document</a>
World	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/World/US09.txt	0.072107	<a href="#">See document</a>
World	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/World/US08.txt	0.0599268	<a href="#">See document</a>
Business	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/Business/Business25.txt	0.0588888	<a href="#">See document</a>
World	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/World/LatAm02.txt	0.0388143	<a href="#">See document</a>
Business	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/Business/Business30.txt	0.021498	<a href="#">See document</a>
Business	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/Business/Business27.txt	0.0200877	<a href="#">See document</a>
Health	fraunhofer2.imp.bg.ac.rs.dictionaries.news_en/Health/Health29.txt	0.0199616	<a href="#">See document</a>

[+ Add document](#)

Enrich the vocabulary by adding the uploaded document to the 'World' class (avg. similarity: 0.12482)?

- Претрага документа по сличности