

BG.DBPEDIA.ORG STARTED

VLADIMIR ALEXIEV & BOYAN SIMEONOV

ONTOTEXT CORP



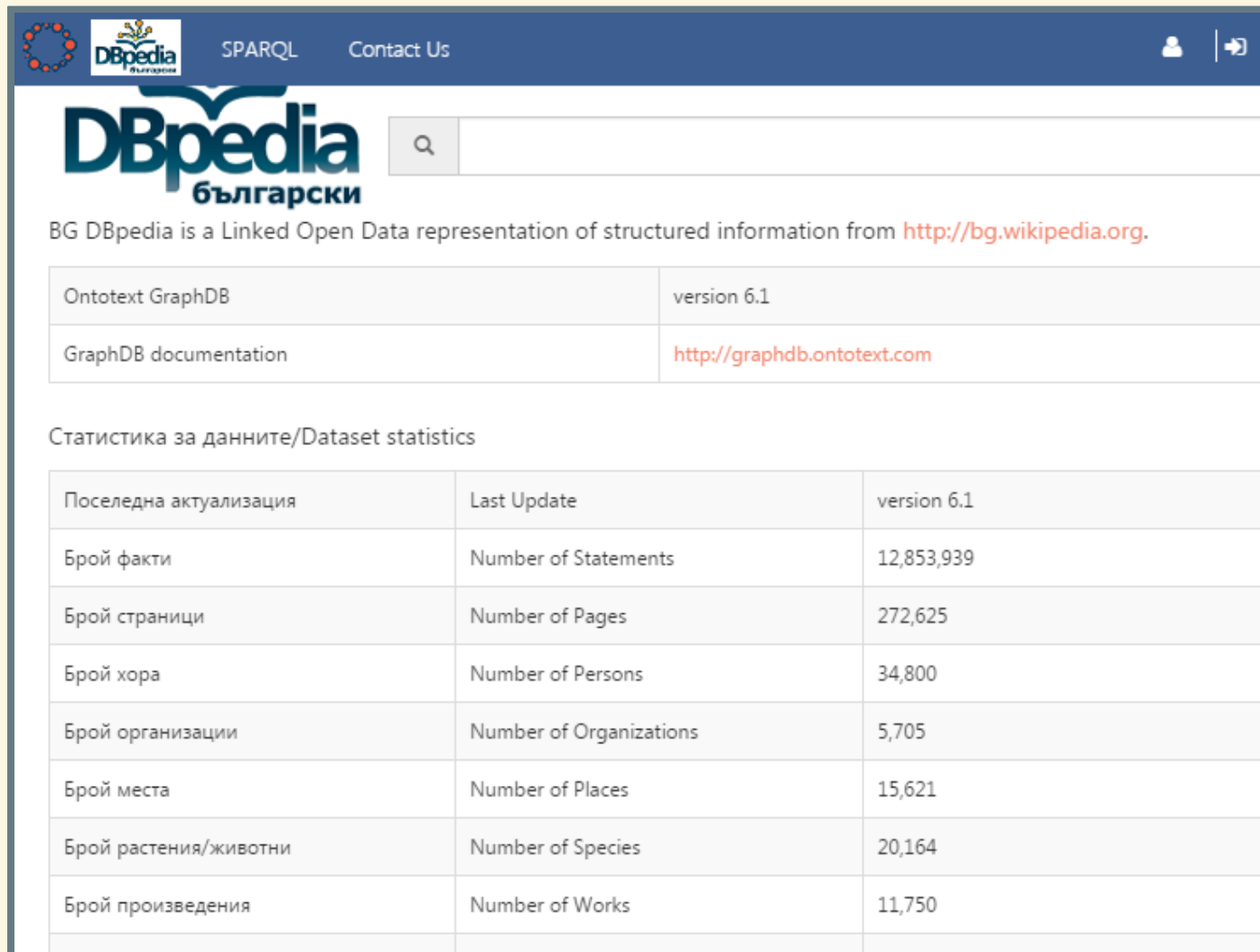
2015-02-08 SOFIA, 2015-02-09 DUBLIN

This is a 2D interactive presentation. Press [O](#) for overview, [H](#) for help.
Proudly made in plain text with [reveal.js](#), [org-reveal](#), [org-mode](#) and [emacs](#).

TABLE OF CONTENTS

- [bg.dbpedia: Home Page](#)
 - [Search With Autocomplete](#)
 - [Resource Page \(Vassil Levski\)](#)
- [Sample Queries](#)
- [Pages without type](#)
- [Page Stats](#)
- [Settlements](#)
 - [Wrong Data](#)
 - [Place Parents](#)
 - [Places With Biggest Population](#)
- [Musical Artists](#)
 - [Conditional Mapping](#)
 - [Data Improvement](#)
- [Musical Genres](#)
 - [Popular Genres](#)
 - [Making Charts](#)

1 BG.DBPEDIA: HOME PAGE



SPARQL Contact Us

DBpedia
български

BG DBpedia is a Linked Open Data representation of structured information from <http://bg.wikipedia.org>.

Ontotext GraphDB	version 6.1
GraphDB documentation	http://graphdb.ontotext.com

Статистика за данните/Dataset statistics

Последна актуализация	Last Update	version 6.1
Брой факти	Number of Statements	12,853,939
Брой страници	Number of Pages	272,625
Брой хора	Number of Persons	34,800
Брой организации	Number of Organizations	5,705
Брой места	Number of Places	15,621
Брой растения/животни	Number of Species	20,164
Брой произведения	Number of Works	11,750





1.1 SEARCH WITH AUTOCOMPLETE


Search interface showing the query "левс" and the following suggestions:

- Васил Левски**
Васил Левски (Васил Иванов Кунчев) е български националреволюционер, идеолог и организатор на българската национална революция, национален герой. Основател е на Вътрешната революционна организация (ВРО) и на Българския революционен централен комитет (БРЦК). Известен е и като Апостола на свободата заради организирането и разработването на стратегията за освобождаване на България от османско иго. Пътува по страната и създава частни революционни комитети, които да подготвят обща революция.
- ПФК Левски (София)**
- Левски (град)**
- Национален стадион „Васил Левски“**
- Васил Левски (булевард в София)**
- Васил Левски (стадион)**
- Васил Левски (Област Стара Загора)**
- Васил Левски (Област Търговище)**
- Васил Левски (Област Пловдив)**
- Васил Левски (Област Силистра)**

Сuggestion	Value
Левски (град)	versio
Национален стадион „Васил Левски“	12,85
Васил Левски (булевард в София)	272,6
Васил Левски (стадион)	34,80
Васил Левски (Област Стара Загора)	5,705
Васил Левски (Област Търговище)	15,62
Васил Левски (Област Пловдив)	20,16
Васил Левски (Област Силистра)	11,75

1.2 RESOURCE PAGE (VASSIL LEVSKI)

SPARQLContact Us



Download as ▾

Васил Левски (Васил Иванов Кунчев) е български националреволюционер, идеолог и организатор на българската национална революция, национален герой. Основател е на Вътрешната революционна организация (ВРО) и на Българския революционен централен комитет (БРЦК). Известен е и като Апостола на свободата заради организирането и разработването на стратегията за освобождаване на България от османско иго. Пътува по страната и създава частни революционни комитети, които да подготвят обща революция.

Source: http://bg.dbpedia.org/resource/Васил_Левски

Subject (100 of 214) Predicate Object All

Statements in which the resource exists as a subject.

Predicate	Object
<code>rdf:type</code>	<code>dbo:Person</code>
<code>rdf:type</code>	<code>dbo:Agent</code>
<code>rdf:type</code>	<code>foaf:Person</code>
<code>rdf:type</code>	<code>http://wikidata.dbpedia.org/resource/Q5</code>
<code>rdf:type</code>	<code>owl:Thing</code>

2 SAMPLE QUERIES

Брой факти

```
select (count(*) as ?c) {?s ?p ?o}
```

Брой страници

```
select (count(*) as ?c) {?thing foaf:isPrimaryTopicOf ?page}
```

Брой хора

```
select (count(*) as ?c) {?person a dbo:Person}
```

Брой организации

```
select (count(*) as ?c) {?org a dbo:Organisation}
```

Брой места

```
select (count(*) as ?c) {?place a dbo:Place}
```

Брой растения/животни

```
select (count(*) as ?c) {?place a dbo:Species}
```

Брой произведения

```
select (count(*) as ?c) {?place a dbo:Work}
```

3 PAGES WITHOUT TYPE

Брой страници без тип

```
select (count(*) as ?c) {  
  ?thing foaf:isPrimaryTopicOf ?page.  
  filter not exists{?thing a ?type.  
    filter (?type in (owl:Thing, dbo:Place, dbo:Species))}}}
```

Страници без тип: с техните категории

```
select ?thing (group_concat(?cat; separator="; ") as ?cats) {  
  ?thing foaf:isPrimaryTopicOf ?page.  
  filter not exists{?thing a ?type.  
    filter (?type in (owl:Thing, dbo:Place, dbo:Species))}  
  ?thing dct:subject ?category.  
  bind(strafter(str(?category), "http://bg.dbpedia.org/resource/Категория:") as ?cat)  
} group by ?thing
```

4 PAGE STATS

Най-големи страници

```
select * {  
  ?x dbo:wikiPageLength ?len  
} order by desc(?len) limit 100
```

Страници с най-много линкове

```
select ?x (count(*) as ?links) {  
  ?x dbo:wikiPageWikiLink ?l  
} group by ?x order by desc(?links) limit 100
```


5 SETTLEMENTS

Най-големи нас.места:

```
PREFIX bgdbp:  
PREFIX dbo:  
select ?x ?area {  
  ?x bgdbp:wikiPageUsesTemplate ;  
      dbo:areaLand ?area  
}  
order by desc(?area) limit 100
```

5.1 WRONG DATA

- Най-големите НМ са Мътеница, Дебелец, Дъбовик, Ръжево, Мало Крушево?!?

- [https://bg.wikipedia.org/wiki/Мътеница_\(село\)](https://bg.wikipedia.org/wiki/Мътеница_(село)):

Землище 33 035 km²

- [https://bg.wikipedia.org/w/index.php?title=Мътеница_\(село\)&action=edit](https://bg.wikipedia.org/w/index.php?title=Мътеница_(село)&action=edit)

| площ = 33 035

- Това е 1/3 от площта на България ;-)
 - Ние приемаме, че "," и "." са десетичен знак, ама чак пък интервал...
 - На друго място съм виждал изписано така: 1 nbsp; 234, което се извлича като "1" :-)
- Ако изключим тези грешки, следващото НМ е София
 - После е Батак?
 - <https://bg.wikipedia.org/wiki/Батак>: Землище 461.105 km²: не ми се вярва
 - Спас Колев казва, че е вярно. Това е землището на града (не само градската територия), и Батак включва голяма планинска територия.
 - Надм. височина 1036m? Но Екатте дава интервал Надм. височина, м: 700 - 999 вкл <http://www.ekatte.com/област-пазарджик/община-батак/батак>

5.2 PLACE PARENTS

- Няма структурно поле, което да индицира Settlements само в Bulgaria. В момента информацията за административната структура на местата (и броят население) не може да се извлече
- Причина: в bg.wikipedia, голяма част от информацията е заключена в lookup lists in sub-templates, като "ekatte" се ползва за ключ.
- Напр за Мътеница [https://bg.wikipedia.org/w/index.php?title=Мътеница_\(село\)&action=edit](https://bg.wikipedia.org/w/index.php?title=Мътеница_(село)&action=edit):

```
{{Селище в България
| екатте = 49624
| надм-височина=441 ...
```

- Извън шаблона (в свободен текст) се казва: "*село в Южна България. То се намира в община Хисаря, Област Пловдив*". Тези се извличат като линкове, но не в структурни полета.
- The extraction framework cannot extract data from such lists. See issue [#305](#): resolving lookup-list data out of sub-templates

5.3 PLACES WITH BIGGEST POPULATION

Населени места с най-голямо население

```
PREFIX bgdbp:
PREFIX dbo:
select * {
  ?x bgdbp:wikiPageUsesTemplate ;
     dbo:populationTotal ?pop
} order by desc(?pop) limit 100
```

Ура! Тези май са верни :-)

6 MUSICAL ARTISTS

Музикални изпълнители и групи

```
PREFIX dbo:
select * {
  ?x a ?type
  filter (?type in (dbo:MusicalArtist, dbo:Band))
}
```

- 2559. Но само 111 са dbo:MusicalArtist?!? Напр. Лили Иванова и Горан Брегович излизат като dbo:Band (и оттам dbo:Organisation ;-))
- Причината е, че мапингът **Музикален_изпълнител** имаше следният проблем:

```
{{ Condition
| templateProperty = фон
| operator = isSet
| mapping = {{ TemplateMapping | mapToClass = Band }}
}}
```

- Т.е. ако има поле "фон", то приеми че е Група. Напр. Горан Брегович: фон = инструменталист

6.1 CONDITIONAL MAPPING

Сега го оправихме да има следната сложна логика:

- Ако има поле "състав", "бивши_членове" или "създадена" -> Band
 - (Ами "членове"? Не знам)
- Ако "фон" съдържа "група", "квартет", "ансамбъл", "хор" -> Band
- Ако "фон" съдържа "композитор" -> MusicComposer
- Ако "фон" съдържа "диригент" -> MusicDirector
- Ако "фон" съдържа "певица" -> MusicalArtist, gender=dbo:Female
- Ако "фон" съдържа "певец" -> MusicalArtist, gender=dbo:Male
- Ако "фон" съдържа "пианист" -> MusicalArtist, gender=dbo:Male
- Ако има поле "наставка" -> MusicalArtist, gender=dbo:Female
- Иначе -> MusicalArtist, gender=dbo:Male

6.2 DATA IMPROVEMENT

- Вече Лили Иванова не е Band (и Organisation), а MusicalArtist, gender=dbo:Female. Ура!
 - Макар че някои биха оспорили, че де-факто тя е **институция** в българската музика
- По принцип шаблонът https://bg.wikipedia.org/wiki/Шаблон:Музикален_изпълнител казва, че "фон" трябва да е: певец, инструменталист, не_изпълнител, временно.
 - Но има много изключения, напр. "фолк певица" (е може ли без нея)
 - Нашият мапинг следва реалните данни
- Тригонометрична_функция и България_на_Световното_първенство_по_футбол_1994?!?
 - Преди ползваха шаблон "Музикален изпълнител"
 - Но вече не, напр https://bg.wikipedia.org/w/index.php?title=Тригонометрична_функция&diff=6434190&oldid=5927191
 - При следващото извличане ще се актуализират и ще отпаднат от този списък

7 MUSICAL GENRES

Музикални изпълнители за които са посочени жанрове

```
PREFIX dbo:
select ?x (group_concat(?genre;separator="; ") as ?genres) {
  ?x a ?type. filter (?type in (dbo:MusicalArtist, dbo:Band))
  ?x dbo:genre ?g.
  bind (strafter(str(?g), "http://bg.dbpedia.org/resource/") as ?genre)
} group by ?x
```


7.1 POPULAR GENRES

Музикални жанрове по популярност

```
PREFIX dbo:
select ?genre (count(*) as ?c) {
  ?x a ?type. filter (?type in (dbo:MusicalArtist, dbo:Band))
  ?x dbo:genre ?g.
  bind (strafter(str(?g), "http://bg.dbpedia.org/resource/") as ?genre)
} group by ?genre order by desc(?c) limit 20
```

- 520 жанра
- Най-популярните са

Поп_музика	356	(изненада!)
Хард_рок	313	
Хеви_метъл	238	

7.2 MAKING CHARTS

- Charts directly from SPARQL: [VISU](#), [sgvizler](#) (jsfiddle)
- Or Query> Save as> TSV and make it in Excel

