
russiannames Documentation

Release 1.0.1

Ivan Begtin

Dec 02, 2021

Contents

1	Indices and tables	5
----------	---------------------------	----------

Russian Names

russiannames is a Python 3 library dedicated to parse Russian names, surnames and midnames, identify person gender by fullname and how name is written. It uses MongoDB as backend to speed-up name parsing.

Documentation

Documentation is built automatically and can be found on <https://russiannames.readthedocs.org/en/latest/>

Installation

To install Python library use *pip install russiannames* via pip or *python setup.py install*

To use database you need MongoDB instance. Unpack *db_data_bson.zip* file from https://github.com/datacooon/russiannames/blob/master/data/bson/db_dump_bson.zip

and use *mongorestore* command to restore *names* database with 3 collections: names, surnames and midnames

Features

Database of names used for identification

- 375449 surnames - collection: surnames
- 32134 first names - collection: names
- 48274 midnames - collection: midnames

Detailed database statistics by gender and collection

collection| total | males|females|universal or unidentified |

— | — | — | — | — |

names | 32134 | 19297 | 8278 | 1196 |

midnames | 48274 | 30114 | 16143 | 0 |

surnames | 375274 | 124662 | 111534 | 38827 |

Supports 12 formats of Russian full names writing style

Format | Example | Description |

—— | ———— | ———— |

f | | only first name |

s | | only surname |

Fs | . | first letter of first name and full surname |

sF | . | full surname and first letter of surname |

sf | | full surname and full first name |

fs | | full first name and full surname |

fm | | full first name and full middlename |

SFM | ... | first letters of surname, first name, middlename |

FMs | .. | first letters of first and middle name and full furname |

sFM | .. | full surname and first letters of first and middle names |

sfM | . | full surname, first name and first letter of middle name |

sfm | | full name as surname, first name and middle name |

fms | | full name as first name, middle name and surname |

Supports names with following ethnics identification

9 ethnic types in names, surnames and middle names supported

key | name (en) | name (rus)

— | — | —

arab | Arabic |

arm | Armenian |

geor | Georgian |

germ | German |

greek | Greek |

jew | Jew |

polsk | Polish |

slav | Slavic (Russian) |

tur | Turkic | ()

Limitations

- very rare names, surnames or middlenames could be not parsed
- ethnic identification is still on early stage

Speed optimization

- preconfigured and preindexed MongoDB collections used

Usage and Examples

Parse name and identify gender

Parses names and returns: format, surname, first name, middle name, parsed (True/False) and gender

```
>>> from russiannames.parser import NamesParser
>>> parser = NamesParser()
>>> parser.parse(' ')
{'format': 'sfm', 'sn': '', 'fn': '', 'mn': '', 'gender': 'm', 'text': ' ', 'parsed': True}
>>> parser.parse(' C..')
{'format': 'sFM', 'sn': '', 'fn_s': 'C', 'mn_s': '', 'gender': 'f', 'text': ' C..', 'parsed': True}
```

Gender field could have one of following values:

- m: Male
- f: Female
- u: Unknown / unidentified
- -: Impossible to identify

Ethnic identification (experimental) Parses surname, first name and middle name and tries to identify person ethnic affiliation of the person

```
>>> from russiannames.parser import NamesParser
>>> parser = NamesParser()
>>> parser.classify(' ', ' ', '')
```

(continues on next page)

(continued from previous page)

```
{'ethnics': ['tur'], 'gender': 'm'}  
>>> parser.classify(' ', ' ', ' ')  
{'ethnics': ['slav'], 'gender': 'f'}
```

Supported languages * Russian

Requirements * pymongo * click

Related projects - Slavic names <https://github.com/wb-08/SlavicNames> - same data shipped as SQLite database

Contents:

CHAPTER 1

Indices and tables

- `genindex`
- `modindex`
- `search`