

# **Корпус национальных ГИМНОВ**

**Любовь Беликова**

# Этап 1: Идея

# Что?

Корпус гимнов разных стран на русском языке с разметкой, включающей в себя:

- а) данные о стране;
- б) тему(ы) гимна;
- в) текст гимнов с синтаксическим и морфологическим разбором.

# Зачем?

Для создания «патриотического портрета» разных стран и регионов. Примеры исследовательских вопросов:

- Какие темы наиболее часто поднимаются в национальных гимнах?
- Каков образ родины в национальных гимнах?
- Есть ли какие-то тематические особенности у гимнов европейских/азиатских/светских/религиозных государств, стран-бывших колоний и т. д.?
- Какие мотивы наиболее часто встречаются в гимнах – похвала, призыв на бой, обращение к истории, пожелания родине?

## **Этап 2: Сбор и разметка данных**

# Источники и инструменты

Данные	Источник	Инструменты
Тексты гимнов	Сайт <a href="#">Discoveric</a>	Selenium
Данные о странах: площадь, численность и религиозный состав населения, форма государства, зависимый статус в прошлом, континент, название страны на английском и т. п.	Википедия, библиотека ruscountry_convert, Гугл-переводчик	wptools, pycountry_convert
Лемматизированные тексты гимнов	---	Mystem
Тексты гимнов с синтаксическим разбором	---	UDPipe, NLTK
Темы гимнов	---	Scikit-learn
*Стоп-слова для русского языка	НКРЯ, <a href="#">топ-100</a> самых частотных словоформ	---

# Что получилось

JSON-файл с текстами 93 гимнов

Ключ	Значение
name	Название страны
link	Ссылка на текст гимна в источнике
text	Текст гимна
name_en	Название страны на английском языке
state_characteristics	Форма государства - государственное устройство, форма правления
area	Площадь страны в квадратных километрах
population	Население страны
former_colony	Была ли страна зависимой в прошлом
religion	Религиозный состав населения страны
continent	Континент, на котором находится страна
secular	Является ли страна светской
text_lemmatized	Лемматизированный текст гимна
topic_to_similarity	Близость текста каждой из следующих тем: общепатриотическая, война, бог, описание
ud_graphs	Предложения текста гимна с синтаксическим разбором

# Фрагмент корпуса

```
"90": {  
  "name": "ЮАР",  
  "link": "http://discoveric.ru/anthem/yuar",  
  "text": "Господи,\nблагослови Африку\nГосподи, благослови Африку,\nНаставь ее на путь истинный,\nУслышь наши мольбы.\nГосподи, благослови -\n\nГосподи, благослови,\n\nГосподи, благослови Африку,\nНаставь ее на путь истинный,\nУслышь наши мольбы.\nГосподи, благослови нас, детей Твоих,\nСвятой Дух, Святой Дух, снизойди к нам...",  
  "name_en": "South Africa",  
  "state_characteristics": ["unitary state", "dominant-party system", "parliamentary system", "republic"],  
  "area": 1221037,  
  "population": 58775022,  
  "former_colony": 1,
```

# Фрагмент корпуса

"religion": {"christianity": 95, "no": 1},

"continent": "AF",

"secular": 1,

"text\_lemmatized": "господи,\nблагословлять африка\nгосподи, благословлять африка,\nнаставлять она на путь истинный,\nуслышать наш мольба.\nгосподи, благословлять -\n\nгосподи, благословлять,\n\nгосподи, благословлять африка,\nнаставлять она на путь истинный,\nуслышать наш мольба.\nгосподи, благословлять мы, ребенок твой,\nсвятой дух, святой дух, снисходить к мы...",

"topic\_to\_similarity": {"общепатриотическая": 0.029578765834599403,"война": 0.0072533308896319615,"бог": 0.22951459195252744,"описание": 0.0},

"ud\_graphs": [[...],[...],{"address": 3,"word": "благослови","lemma": "благословлять","ctag": "NOUN","tag": "NN","feats": "Animacy=Inan|Case=Dat|Gender=Fem|Number=Sing", "head": 1,"deps": {"punct": [2,6],"appos": [4]},"rel": "conj"}, {...}], ...]



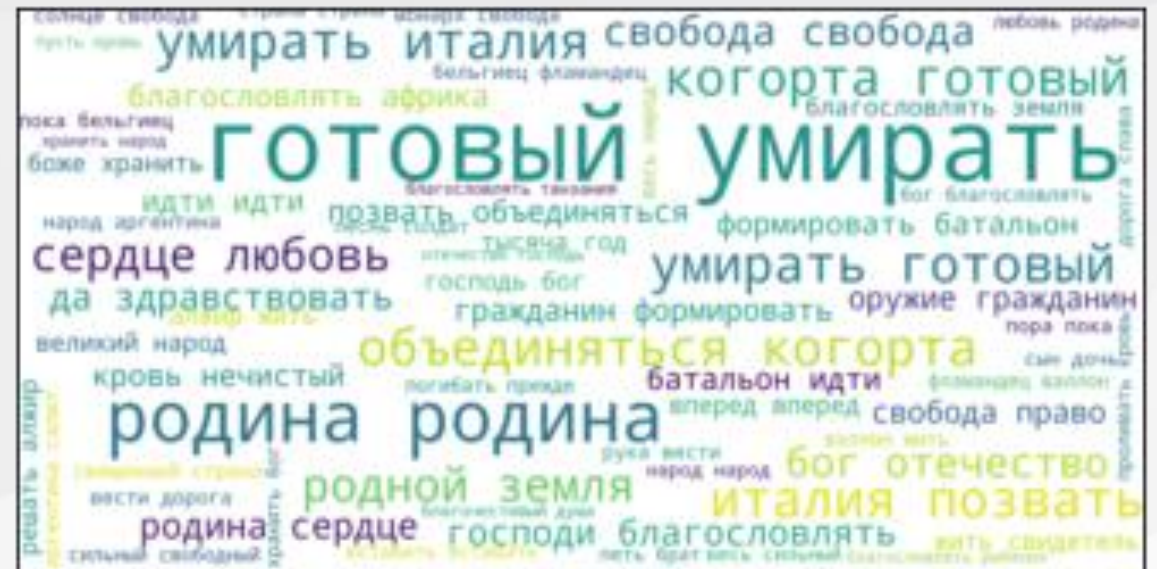
# Этап 3: Анализ данных

# Какие темы характерны для гимнов?

## Частотный анализ



Топ-103 наиболее часто  
встречающихся лемм



Топ-65 наиболее часто  
встречающихся биграмм

# Какие темы характерны для гимнов?

## Тематическое моделирование

Торис 0: земля свобода родина пусть слава  
сердце кровь мир весь честь

Торис 1: благословлять боже хранить земля народ  
дорогой счастье давать защищать дочь

Торис 2: край милый гора труд год день гордый  
родной солнце святой

Торис 3: страна народ отечество земля кровь союз  
всегда год жить родной

Торис 4: вперед вставать единый бой полный  
смерть вместе час новый народ

Торис 5: здравствовать древний гордость гордый  
честь любовь народ свобода слава флаг

Торис 6: свободный умирать великий жить народ  
слава погибать свобода пусть имя

Модель NMF (в основе – TF-IDF,  
min\_df=7, max\_df=0.95)

Общепатриотическая  
Религиозная  
Военная  
Описательная

# Подробнее о тематической разметке

**'общепатриотическая':** 'земля слава сердце  
жить честь труд родной солнце верный  
великий дом гордый гордость  
здоровствовать сын мать отец ребенок  
любить любовь нация флаг'

**'война':** 'вперед вставать бой битва  
смерть война мир час свобода кровь  
борьба герой враг единство единый  
свободный защищать знамя умирать  
погибать победа проливать'

**'бог':** 'бог боже благословлять вера  
господь дух душа святой священный  
хранить защищать'

**'описание':** 'край милый гора солнце  
волна море дорогой скала река сиять  
цвести яркий свет прекрасный'

← косинусное  
→  
расстояние

Текст гимна

# Региональные особенности

- ❑ Самая распространенная тема в гимнах – общепатриотическая, ниже в рейтинге – военная, религиозная и описания.
- ❑ Среди стран, для которых описания в национальном гимне доминируют, в основном страны Европы (Скандинавия и Балканы) и Азии (Япония, Южная Корея и Азербайджан), а также Австралия.
- ❑ Самый «воинственный» континент – Африка. Африканцы поют в гимнах о войне почти так же часто, как об общепатриотических явлениях. На 2-м месте по «воинственности» – Америка. На последнем – Европа.
- ❑ Религиозные страны поют о религии (немного) чаще, чем светские государства.
- ❑ Реже других стран о религии поют Европа и Северная Америка.

# Образ родины в национальных гимнах

Родина – она какая?

{ 'мой': 27,	'благородный': 4,
'наш': 19,	'милый': 4,
'родной': 16,	'священный': 3,
'великий': 7,	'свободный': 3,
'весь': 5,	'любимый': 3,
'дорогой': 5,	'этот': 3,
'прекрасный': 5,	'древний': 3,
'свой': 4,	'родимый': 3}
'независимый': 4,	

Родина – страна чего?

{ 'отцов': 5,
'свобода': 3,
'предок': 2,
'слава': 2,
'круча': 2}

# Проблемы и перспективы

# Что не получилось (но когда-нибудь получится)

- ~~○ Какие темы наиболее часто поднимаются в национальных гимнах?~~
- ~~○ Каков образ родины в национальных гимнах?~~
- ~~○ Есть ли какие-то тематические особенности у гимнов европейских/азиатских/светских/религиозных государств, стран-бывших колоний и т. д.?~~
- Какие мотивы наиболее часто встречаются в гимнах – похвала, призыв на бой, обращение к истории, пожелания родине?



# Спасибо за внимание!

Материалы тут: <https://github.com/meancatt/CL/tree/master/Project1>