

03 поговорки #3....

Подготовлено участниками хакатона по
лингвистике-2019

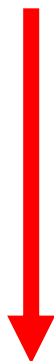
Алекберовым Артёмом, Молодык Петром, Матяш
Дарьей, Орловым Евгением

Ссылка на репозиторий:

https://vk.com/away.php?utf=1&to=https%3A%2F%2Fgithub.com%2FBunnyNoBugs%2FHackaton_2019

ПРЕДИСЛОВИЕ И ЗАДАЧА

Часто, неожиданные сочетания разных частей поговорок (*белыми нитками по воде писано*), вызывают комический эффект.



Мы хотим создать бот, порождающий такие сочетания с определённым временным интервалом.

ДЛЯ ЧЕГО ЭТО НАДО?

Человек сможет расширить или закрепить свои знания русских народных пословиц и поговорок.

КАК?

Бот будет выдавать «склеенное» из двух народных мудростей предложение. Пользователь сможет узнать или вспомнить сразу две поговорки.

ПРОЕКТ В ЦИФРАХ

- 280 пословиц и поговорок
- 9 часов работы
- 4 участника проекта
- более 1100 генерируемых вариантов
«склеенных» грамматически верно поговорок

РЕАЛИЗАЦИЯ. ЧТО БЫЛО СДЕЛАНО И ИСПОЛЬЗОВАНО

- В качестве исходных данных для генерации «пословиц-приколов» были взяты фразы с сайта <http://iamruss.ru/famous-russian-proverb/>
- Пословицы были вручную поделены на две смысловые части в совместной **Гугл-таблице**
- Обработка полученных данных была выполнена на языке программирования **Python** и сервисом **Mystem**
- Программа написана для чат-бота сервиса **Telegram** (@funnyidiom_bot)
- Результат работы загружен в репозиторий веб-сервиса для хостинга IT-проектов **Github**

РЕАЛИЗАЦИЯ. КОДЫ ДЛЯ ГЕНЕРАЦИИ ПОСЛОВИЦ

```
codes = ( 'пов, 2-л',  
          'нп=непрош, ед, изъяв, 3-л',  
          'мн, изъяв, 1-л',  
          'ед, кр, муж',  
          'нп=непрош, ед, изъяв, 3-л, сов',  
          'нп=непрош, ед, изъяв, 3-л, несов',  
          'прош, мн, изъяв, сов, пе',  
          'прош, мн, изъяв, сов, пе',  
          'нп=прош, ед, изъяв, сред, сов',  
          '=V, пе=инф, несов' )
```

РЕАЛИЗАЦИЯ. ПРИМЕРЫ АВТОМАТИЧЕСКОЙ ЛЕММАТИЗАЦИИ МЫСТЕМ

Яблоко{яблоко=S, сред, неод=вин, ед} от{от=PR=} яблони{яблоня=S, жен, неод=вин, мн}
Пришла{приходить=V, нп=прош, ед, изъяв, жен, сов} беда{беда=S, жен, неод=им, ед} –
Что{что=CONJ=} в{в=PR=} лоб{лоб=S, муж, неод=вин, ед}, что{что=CONJ=}
Как{как=CONJ=} корова{корова=S, жен, од=им, ед}
Чем{чем=CONJ=} богаты{богатый=A=мн, кр},
Не{не=PART=} рой{рыть=V, несов, пе=ед, пов, 2-л} яму{яма=S, жен, неод=вин, ед} другому{другой=APRO=дат, ед, муж} –
Волос{волос=S, муж, неод=вин, ед} долог{долгий=A=ед, кр, муж},
Умный{умный=A=вин, ед, полн, муж, неод} бы{бы=PART=} ты{ты=SPRO, ед, 2-л=им} был{быть=V, нп=прош, ед, изъяв, муж, несов} человек{человек=S, муж, од=вин, мн} –
Кому{кто=SPRO, ед, муж, од=дат} война{война=S, жен, неод=им, ед}, а{а=CONJ=}
Без{без=PR=} царя{царь=S, муж, од=вин, ед}
Смелому{смелый=A=дат, ед, полн, муж} и{и=CONJ=}
Укатали{укатывать=V, пе=прош, мн, изъяв, сов} сивку{сивко=S, муж, од=дат, ед}
На{на=PR=} воре{вор=S, муж, од=пр, ед} и{и=CONJ=}
У{у=PR=} страха{страх=S, муж, неод=род, ед}
Хвост{хвост=S, муж, неод=вин, ед} вытацишь{вытаскивать=V, пе=непрош, ед, изъяв, 2-л, сов} –
Тише{тихо=ADV=срав} едешь{ехать=V, несов, нп=непрош, ед, изъяв, 2-л} –
Хорошо{хорошо=ADV, прдк=} там{там=ADVPRO=},
Гусь{гусь=S, муж, од=им, ед} свинье{свинья=S, жен, од=пр, ед}
Сядем{садиться=V, нп=мн, пов, 1-л, сов} рядом{рядом=ADV=} да{да=CONJ=}
Не{не=PART=} в{в=PR=} бровь{бровь=S, жен, неод=вин, ед}, а{а=CONJ=}
Слово{слово=S, сред, неод=вин, ед} – серебро{серебро=S, ед, сред, неод=вин}, а{а=CONJ=} молчание{молчание=S, сред, неод=вин, ед} –

РЕАЛИЗАЦИЯ. ПРИМЕРЫ ГЕНЕРИРУЕМЫХ ПОСЛОВИЦ

Пуганая ворона и вылетит — не воротишь.

Не плюй в колодец — откликнется.

Нет друга, так ищи, а нашёл, так вратъ не мешай.

Любишь гостить — люби и в стойло не лезь.

Баба с возу — ближе к телу.

До свадьбы придется напиться.

РЕАЛИЗАЦИЯ. ПРОБЛЕМЫ И РЕШЕНИЯ

Мало входных данных

Дополнить исходные данные еще бóльшим числом пословиц и поговорок

Расширить специфику бота и генерировать не только народные мудрости, но крылатые фразы и афоризмы

Трудности
морфологического
анализа

Ручная проверка

Усовершенствование сервисов морфологического анализа

ЧТО МОЖНО СДЕЛАТЬ ЕЩЁ

- Сделать базу данных больше
- Увеличить число опций бота и его интерфейс:
 - Выдавать склеенные пословицы, объединённые
 - общей тематикой
 - правилом генерации
 - настроить индивидуальные временные промежутки
- Минимизировать риск повтора одной и той же «склеенной» пословицы