



Предварительная версия: 23.12.2019



**Проект:  
Контекстуальный  
англо-русский словарь**


Декабрь 2019 г.

# Пример для подражания - 1

 TRANSLATION | DICTIONARY | SPELL CHECK | CONJUGATION | SYNONYMS | GRAMMAR

English Russian 

 specially designed 

Translation of "specially designed" in Russian 

специально разработанные

специально разработанных

специально предназначенные

специально спроектированные

специально разработанное

специально предназначенных

специально предназначены

специально разработанный

специально разработанной

специально разработанного

конкретно предназначенные

специальной конструкции

специально разработанную

All the interviews were recorded using questionnaires that had been specially designed for the purpose.

Students with disabilities had received laptops specially designed to meet their needs.

Lack of specially designed policies on violence against women and girls in Uganda.

Upper secondary education is voluntary and comprises 17 national programmes as well as specially designed and independent study programmes.

Результаты опроса были занесены в специально разработанные для этого вопросники.


Учащиеся с ограниченными возможностями получили компьютеры, специально разработанные с учетом их потребностей.

Отсутствие специально разработанных стратегий, направленных против насилия в отношении женщин и девочек в Уганде.

Старшее среднее образование является добровольным и осуществляется на основе 17 общенациональных программ, а также специально разработанных независимых учебных программ.



# Пример для подражания - 2



## (в меньшей степени)

 русский ↔ английский ▼

🔍




^ Английско-русский словарь  
На данный момент в разработке

specially нар.  — специально нар.  · особенно нар. · [...] 

designed прил./part-II  — созданный прил./part-II  · [...] 

© Словарь Linguee, 2019

^ Непроверенные источники (русск → английск)

With support from UNESCO, IFEX has set up a specially designed website which posts alerts and transmits properly validated [...]  <a href="https://unesdoc.unesco.org">unesdoc.unesco.org</a>	При поддержке ЮНЕСКО ИФЕКС создала специальный веб-сайт, на котором публикуются предупреждения об опасности и на глобальной [...]  <a href="https://unesdoc.unesco.org">unesdoc.unesco.org</a>
[...] best scientific knowledge available and on procedures that are specially designed for evaluating the ethical issues at stake.  <a href="https://unesdoc.unesco.org">unesdoc.unesco.org</a>	[...] основываться на имеющихся новейших научных знаниях и процедурах, специально разработанных для оценки возникающих этических проблем.

# Что получилось

Адрес сайта

[hidden-eyrie-40264.herokuapp.com](https://hidden-eyrie-40264.herokuapp.com)

Коллокации из предложений в базе данных

Поисковый запрос: "specially designed"

На сайте пример системы из 1000 предложений из-за ограничений для бесплатных сайтов

Русский перевод выделяется по англо-русскому глоссарию, полученному автоматически путем Word Alignment в NLTK

[Вернуться к форме поиска](#)

## Поиск в контексте

Сайт содержит около одной тысячи предложений из резолюций СБ ООН.

Примеры поисковых запросов:

specially designed, humanitarian assistance, arms embargo, displaced persons, troop contributing countries, decides remain seized, unmanned aerial vehicles, non civil certified aircraft

Термины для поиска (максимум 4 слова):

Найти!

Вы искали: specially | designed

## Результаты поиска

Vibration test equipment and specially designed parts and components capable of simulating flight conditions of less than 15,000 meters. a.

a.2. Equipment specially designed for ion implantation, ion-enhanced or photo-enhanced diffusion, having any of the following characteristics:

Delete "... radars specially designed for meteorological use ...".

Software specially designed for the development or production of the components or equipment above; b.

a. Any type of telecommunications equipment, specially designed to operate outside the temperature

Вибрационные испытательные установки и специально спроектированные детали и компоненты, способные воспроизводить условия полета на высоте ниже 15 000 метров.

a.2. установки, специально предназначенные для ионов имплантации, ионо- или фотоусиленной диффузии, имеющие любую из следующих характеристик:

Опустить слова <<...радиолокаторы, специально созданные для метеорологических целей...>>.

a. Программное обеспечение, специально разработанное для создания или производства вышеуказанных компонентов или установок;

a. Любой тип телекоммуникационного оборудования, специально разработанного для работы за пределами интервала температур от 218

# Источник текстов

<https://conferences.unite.un.org/UNCorpus/>

Англо-русские тексты ООН.

Три файла, в каждом 23 239 280 строк (текст уже разбит на предложения, каждая строка — предложение):

UNv1.0.en-ru.en - английский текст (3,45 GB)

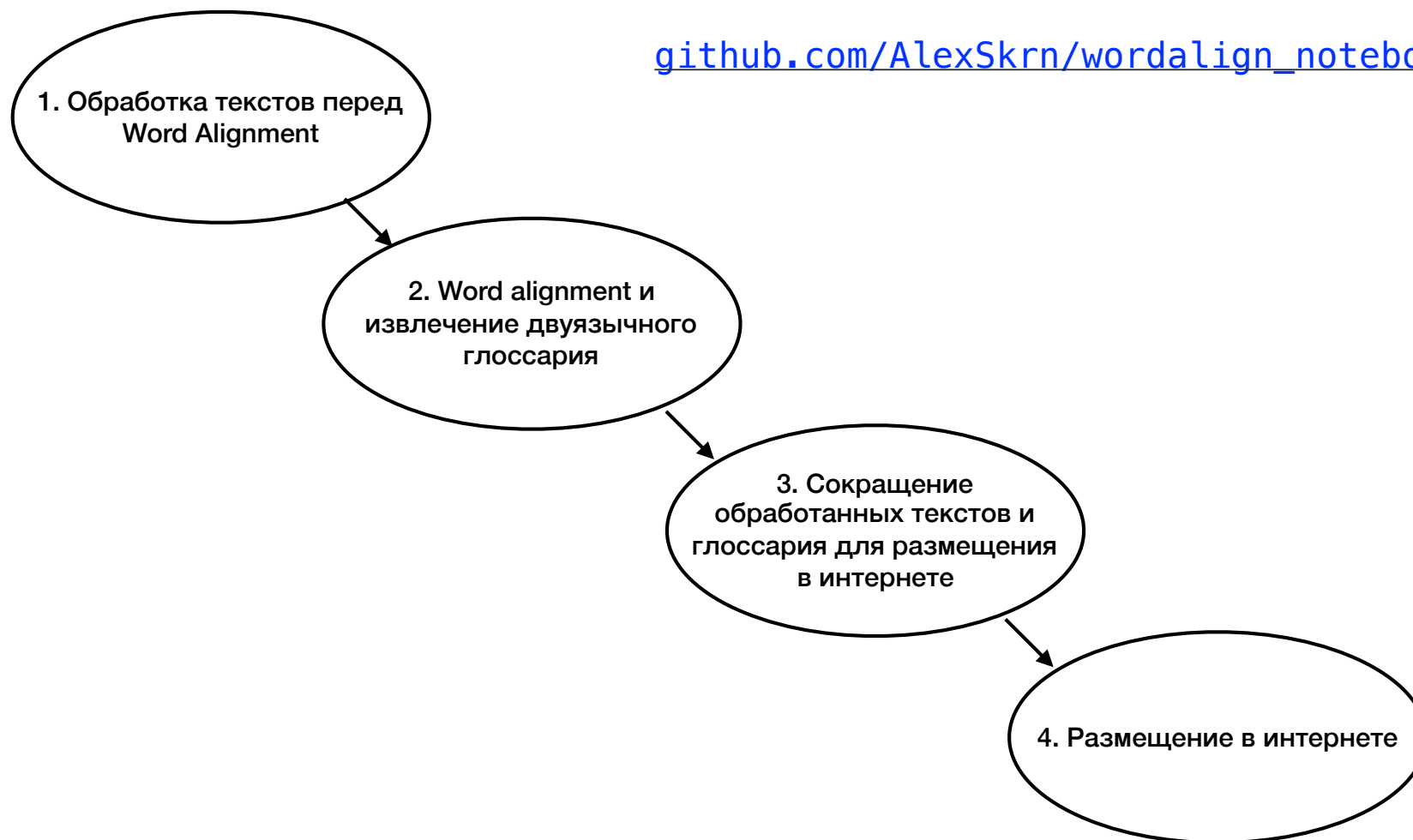
UNv1.0.en-ru.ru - русский текст (6,95 GB)

UNv1.0.en-ru.ids - индекс

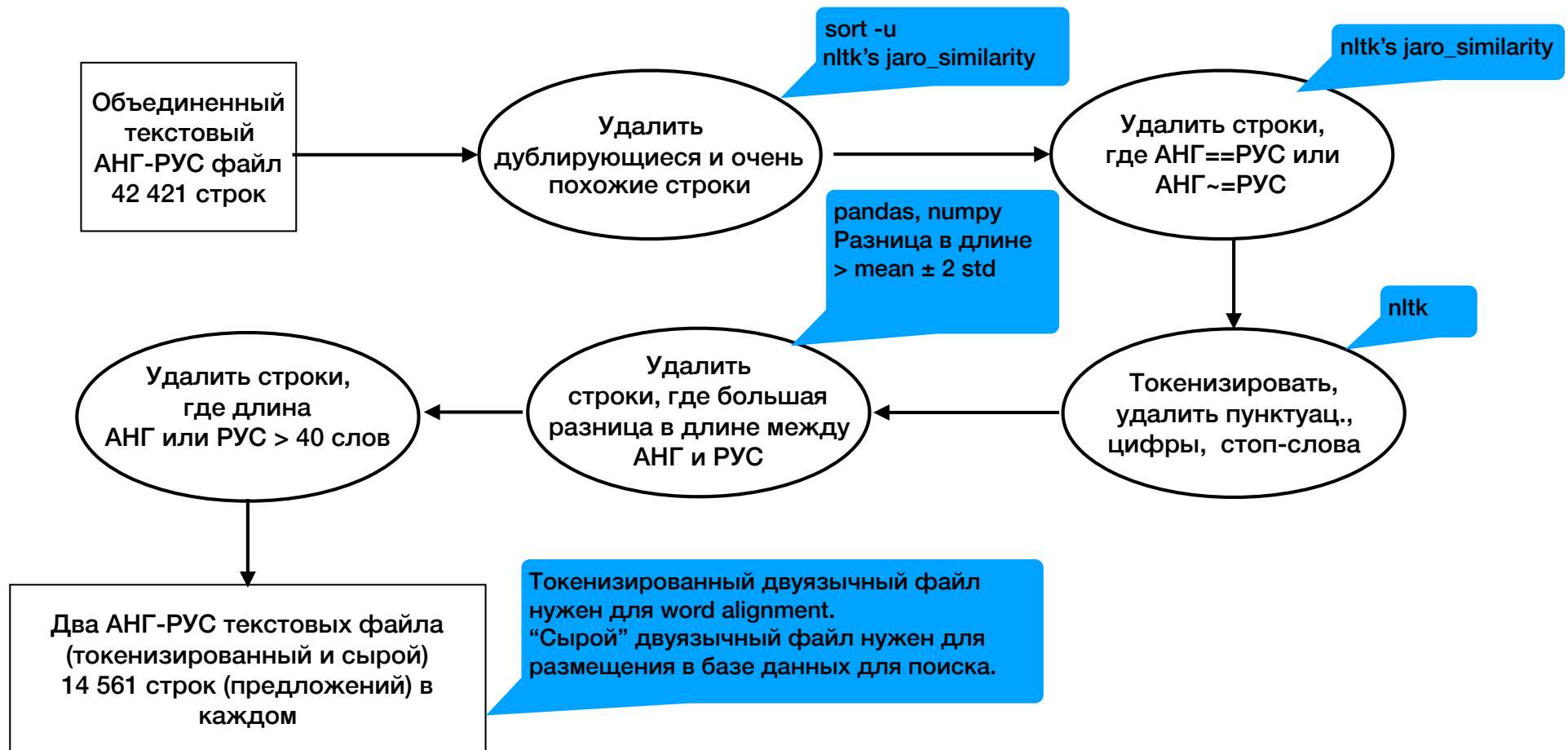
Используя индекс, я отобрал только Резолюции Совета безопасности ООН — 42 421 строк (предложений)

# Этапы

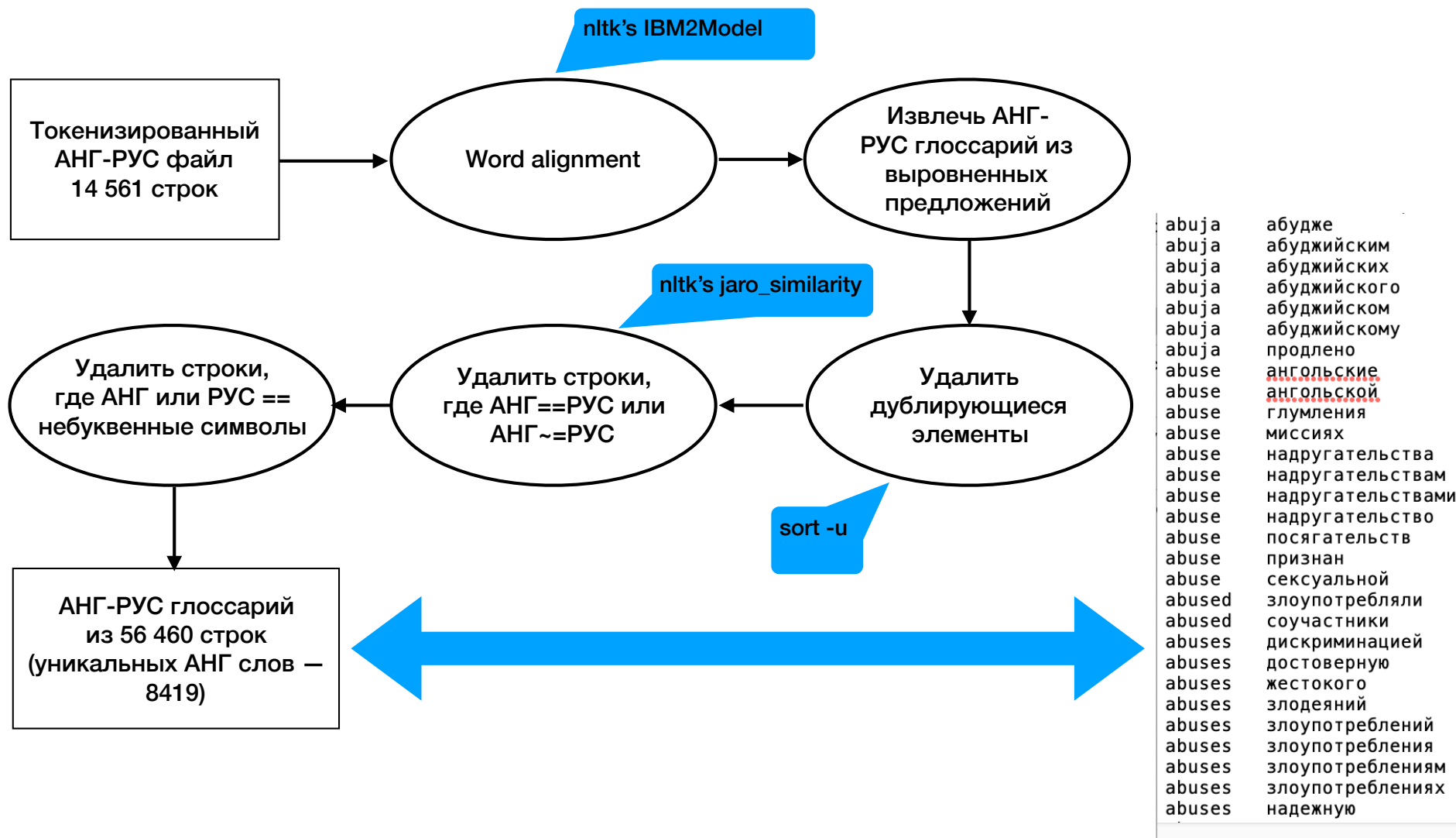
[github.com/AlexSkryn/wordalign\\_notebooks](https://github.com/AlexSkryn/wordalign_notebooks)



# 1. Обработка текстов перед Word Alignment



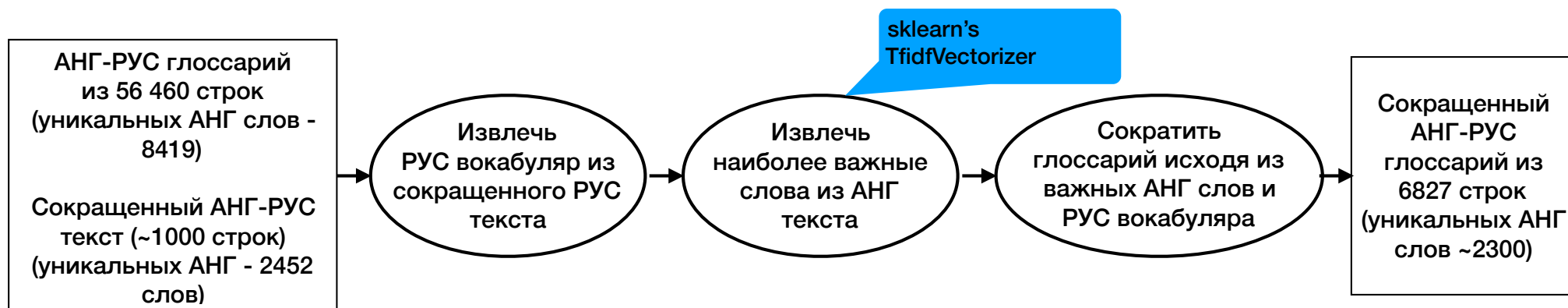
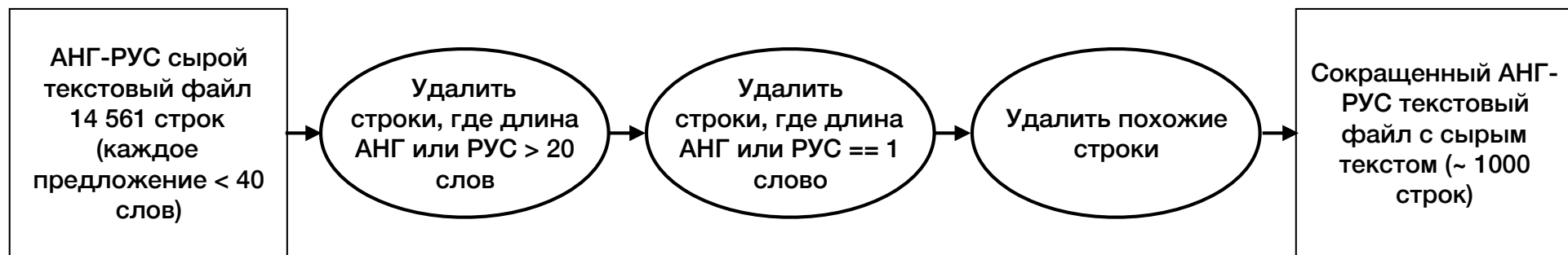
## 2. Word alignment и извлечение двуязычного глоссария





### 3. Сокращение обработанных текстов и глоссария для размещения в интернете

Платформа Heroku имеет ограничение в 10,000 строк базы данных для бесплатных сайтов. Поэтому для размещения на сайте мне нужно было сократить текст до примерно 1 тысячи строк, а глоссарий до примерно 9 тысяч строк



# 4. Размещение в интернете

Адрес сайта: [hidden-eyrie-40264.herokuapp.com](https://hidden-eyrie-40264.herokuapp.com)

- Веб-фреймворк для написания сайта — Flask
- База данных PostgreSQL, менеджер бд - питоновский модуль рееее
- Примеры поисковых запросов на веб-сайте — некоторые из наиболее частых n-граммов из английского текста
- Платформа для размещения сайта — Heroku

# Чего не хватает

- Поиска в направлении русский -> английский
- Функции AutoSuggest в поле поиска
- ...