

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ ИМЕНИ Н. Э. БАУМАНА**  
**Факультет информатики и систем управления**  
**Кафедра теоретической информатики и компьютерных технологий**

Лабораторная работа №1  
по курсу «Информационный поиск»  
**«Извлечение фактов из Википедии»**

Выполнил:  
студент группы ИУ9-21М  
Беляев А. В.

Проверила:  
Лукашевич Н. В.

Москва 2019

## Цель работы

Найти на Википедии в архиве статей проекта «Знаете ли вы» 3 факта. В каждом из фактов есть ссылки на статьи википедии, в которых упомянут этот факт.

Необходимо найти этот факт в «подстатьях» и оценить сложность его поиска автоматическими средствами.

## Ход работы

В ходе работы были найдены следующие 3 факта, содержащие ссылки на «подстатьи» соответственно:

1. В рецензии на компьютерную игру критик пожаловался на то, что смерть заставляет начинать уровень заново.
2. Под стенами осаждённой шведами русской крепости немцы побили шотландцев за пиво.
3. Есть версия, что Джек Потрошитель был женщиной.

---

В рецензии на компьютерную игру критик пожаловался на то, что смерть заставляет начинать уровень заново.

Ссылка на [компьютерную игру] ведет к статье [Earth Shaker]. В ней есть предложение, содержащее факт:

- «Рецензент *Your Sinclair* посчитал раздражающим то, что после потери жизни приходится каждый раз начинать уровень снова.»

Обнаружить этот факт было бы несложно:

- весь факт представлен в единственном предложении
- предложение содержится в исходной форме с точностью до синонимов и пропущенных слов.

---

Под стенами осаждённой шведами русской крепости немцы побили шотландцев за пиво.

Ссылка [осаждённой шведами русской крепости] ведет на статью [Осада Везенберга 1574 года]. Предложение, содержащее факт:

- «17 марта между **немцами** и **шотландцами** дело дошло до потасовки, вызванной неоплаченным **элем** и взаимными оскорблениями, драка быстро переросла в кровавый бой».

Это предложение - наиболее точное упоминание факта. Однако, оно не содержит прямое упоминание объекта С (пиво -> эль) в фрейме «А побил В за С». Участники А

и В содержатся в предложении, но в факте присутствует больше потенциальных участников (шведы, немцы, шотландцы). Локатив, содержащийся в первой половине предложения отсутствует. Таким образом, найти фрагмент будет очень сложно.

Ссылка от слова [пиво] ведет на статью об эле. В статье, при этом, нет упоминания исходного фрагмента.

---

Есть версия, что Джек Потрошитель был женщиной.

Статья о Джеке Потрошителе в разделе «Джек-потрошитель в культуре и искусстве» имеется следующее предложение:

- «*«Собираатель душ» — сериал, 2 сезон 11 серия, в которой представляется версия о том, что **Джек-потрошитель был женщиной.**»*

Это предложение содержит близкий к оригинальному набор слов, слова при этом стоят в той же форме, что и в оригинале. Этот фрагмент является наиболее простым для поиска из всех представленных.

Ссылка со слова [женщина] ведет на статью о Мэри Пирси - англичанке, осужденной за убийство жены своего любовника. В статье содержится предложение:

- «*Сэр Артур Конан Дойл, создатель Шерлока Холмса, полагал в то время, что **Джек Потрошитель** вполне мог **быть женщиной**, поскольку женщина могла бы притвориться акушеркой и появиться на публике в окровавленной одежде, не вызывая подозрений и не привлекая внимания.*»

В этом предложении сохранен порядок слов оригинальной формулировки факта и после нормализации слов фрагмент может быть найден. В дополнение ко всему, статья содержит множество упоминаний Джека Потрошителя, являющегося субъектом фрейма.