

## Отчет к заданию №1 студента 620 группы Бугаевского Владимира

В качестве источника для фактов из Википедии был взят [архив](#) рубрики «Знаете ли вы» за период с 19 по 22 июня 2018 года.

### Факт 1:

[Лётчику](#) понадобилось два десятка лет, чтобы повторно стать подполковником.

Информацию о том, что Николай Александрович Мульков был летчиком, а так же подполковником Российской империи и полковником Красной Армии, легко можно получить из краткой информации о персоналии в начале статьи, а также по еще более краткой таблице справа на странице статьи.

Николай Александрович Мульков	
	
Дата рождения	15 (27) ноября 1890
Место рождения	Гродненская губерния, Российская империя
Дата смерти	после 31 декабря 1951
Принадлежность	<div><div></div> Российская империя</div> <div><div></div> РСФСР</div> <div><div></div> СССР</div>
Род войск	авиация
Годы службы	1910—1951
Звание	подполковник Российской императорской армии полковник Красной армии
Сражения/войны	Первая мировая война Гражданская война в России

Николай Александрович Мульков (15 [27] ноября 1890 — не ранее 1952) — военный **лётчик**. **Подполковник** Российской императорской армии (1917), участник Первой мировой войны. Кавалер ордена Святого Георгия 4-й степени (1915) и Георгиевского оружия (1917). После Октябрьской революции вступил в Красную армию, служил на различных должностях, связанных, преимущественно, с преподавательской деятельностью в области военной авиации. Член Всесоюзной Коммунистической партии (большевики) с 1939 года. В 1945 году получил звание **полковника**.

Однако этой информации недостаточно, чтобы полностью получить изложенный факт.

Остальные вхождения слова «подполковник» в статье указаны вместе с датами:

1 вхождение: 1917 год;

19 мая 1916 года Николай Мульков был произведён в чин штабс-капитана, 17 августа 1917 года в капитаны, а на следующий день — в **подполковники**.

2 вхождение: 1940 год.

11 сентября 1940 года получил звание **подполковника**, а 21 февраля 1945 года — полковника.

**Больших трудностей** с поиском информации не возникло, т. к. ключевое слово «подполковник» достаточно уникально.

## Факт 2:

Не вызвавший первоначально интереса экспертов «**маленький рыжеволосый мальчик**» (на илл.) выставлен на аукционе со стартовой ценой в миллион евро.

Оригинальное название картины: «Ребёнок Иисус, стоящий на коленях перед орудиями Страстей». Название «Маленький рыжеволосый мальчик», согласно статье упоминалось только при разговоре.

Она связалась по поводу картины с владельцем аукционного дома Rouillac и экспертом Апелляционного суда Филипом Руйяком и его сыном Эмериком Руйяком, прежде известным во Франции журналистом, с 2010 года вместе с отцом работающим в сфере торговли произведениями искусства. Владелица увидела Руйяков по телевидению; в разговоре по телефону она называла картину «**Маленький рыжеволосый мальчик**».

Найти информацию о том, что картина первоначально не вызвала интереса экспертов оказалось достаточно легко (повезло, что в статье используется слово «интерес»).

Она сообщила, что ряд парижских аукционистов, к которым она уже обращалась некоторое время назад, не за**интерес**овались картиной, ссылаясь на слабый **интерес** рынка к картинам на религиозный сюжет.

Согласно статье цена картины варьируется. Ключевое слово, по которому удалось найти данный факт – «евро», однако формулировка сильно видоизменилась:

«миллион евро» → «один миллион евро» → «1 млн. евро»

– появилось числительное, записанное в форме цифр, а «миллион» было заменено на сокращение «млн.».

Она оценивается от 3 до 5 млн евро (по другим данным, только в 1 млн евро) — в три раза больше, чем картина «Святой Иероним», которая должна быть выставлена в апреле 2018 года на аукционе «Кристис» (1 млн долларов).

**Основные трудности**, с которыми мы столкнулись:

- картина имеет другое официальное название;
- неявное использование числительных;
- использование сокращений.

**Факт 3:**

Американский фотограф, известный съёмками высшего общества, был также личным фотографом знаменитого мафиозо.

Из краткого описания персоналии в начале статьи легко получить информацию о том, что Ааронс Слим был «американским фотографом, известным съемками высшего общества».

Джордж Аллен «Слим» Ааронс (29 октября 1916 — 30 мая 2006) — американский фотограф, известный съемками «красивой жизни» высшего общества, светских персон, кинозвёзд и прочих знаменитостей.

Ключевые слова «личный» и «мафиозо» не найдены на странице.

Так как формулировка факта в статье могла сильно видоизмениться, попробуем воспользоваться синонимами. Для этого воспользуемся словарями синонимов:

1. электронные версии бумажных словарей:
  - <http://slovari.ru/search.aspx?p=3068>
  - <http://gramota.ru/slovari/dic/>
2. электронные словари:
  - <http://slova.zkir.ru/>
  - <https://synonymonline.ru/>

Основная проблема бумажных словарей в том, что они составляются людьми, поэтому число слов, для которых приводятся синонимы крайне не велико.

Рассмотрим слово «мафиози». К нему приводятся лишь 4 синонима:

- бандит;
- преступник;
- мафиози;
- крестный отец.

Из общего ряда «гламурных» персонажей, которых фотографировал Слим Ааронс, резко выделялся **гангстер** Чарльз Лучано по прозвищу «Lucky» («Счастливчик») — один из «боссов» итало-американской организованной **преступности**. Ааронс следовал за ним повсюду, в том числе в Италию, и сделал множество снимков **«крестного отца»**.

Формулировка «крестный отец» встречается в статье, но в видоизмененной форме «крестного отца».

Также удастся добиться успеха при поиске слов с префиксом «преступ».

Легко заметить, что в абзаце нужное ключевое слово – «гангстер». Получим еще больше синонимов для каждого из 4 синонимов: в итоге суммарно 43 синонима, и «гангстер» есть среди них.

**Главная трудность**, с которой мы столкнулись, – использование слов синонимов, а так же их большое число для ручного перебора.

## Выводы

В общем случае качественная поисковая система представляет собой сложный набор алгоритмов. На примере этого задания можно убедиться с обилием сложностей, с которыми сталкивается разработчик таких систем:

- поиск наиболее важных слов в запросе;
- необходимость нормализции слов;
- грамотное построение дерева запросов с учетом синонимов, возможных сокращений.