**Раздел Простой поиск. Поиск с подстановочными знаками**

Задание 1:

Слово "child" в разных формах (child, children, childhood) встречается 22 раза в тексте книги "Alice's Adventures in Wonderland" by Lewis Carroll.

Слово "child" встречается в тексте Alice’s Adventures in Wonderland 11 раза.

Soon her eye fell on a little glass box that was lying under the table: she opened it, and found in it a very small cake, on which the words “EAT ME” were beautifully marked in currants.

Текст "Alice’s Adventures in Wonderland" by Lewis Carroll представлен в британской орфографии.

Задание 2

|  |  |  |
| --- | --- | --- |
| Задание | Формула поиска | Ответ |
| 1. Найдите в тексте первые пять слов, состоящих из пяти букв | [! ][! ][! ][! ][! ] | Alice, tired, twice, Alice, could |
| Сколько в тексте шестибуквенных слов, начинающихся на букву s и заканчивающиеся на букву r? | s[! ][! ][! ][! ]r | 38 |
| |  |  | | --- | --- | | 3. Найдите в тексте первые пять трёхбуквенных слов, начинающиеся на гласную букву |  | | [eyuioa][! ][! ] | and, and, use, own, and |
| |  |  |  | | --- | --- | --- | | 4. Сколько в тексте слов, состоящих из двенадцати букв? По каким формальным признакам их можно сгруппировать? Приведите пример из каждой группы слов |  |  | | 5. Сколько в тексте слов с суффиксом -tion? Приведите пример использования такого слова в контексте | |  | | [a-z]{12} | 48 слов  Эти слова, можно сгруппировать по частям речи:  Nouns - Hippopotamus  Verbs – Contradicted  Adjectives -  “Do bats eat cats?” for, you see, as she couldn’t answer either question, it didn’t much matter which way she put it.  69 слов |
| |  |  | | --- | --- | | 6. Есть ли в тексте слова, включающие четыре согласные буквы подряд? |  | | [!eyuioa][!eyuioa][[!eyuioa] [!eyuioa] | Да, в тексте "Alice’s Adventures in Wonderland" есть слово "stream" (поток), которое содержит четыре согласные буквы подряд - "strm". |
| |  |  | | --- | --- | | 7. Сколько раз в тексте встречаются пассивные конструкции единственного числа прошедшего времени? |  | | was | 332 раза |
|  |  | Contemptuous  Adverbs - Accidentally  Particles - Nevertheless |

1. Соответствие документа содержанию информационного запроса называют смысловой релевантностью, а соответствие поискового образа этого документа формализованному поисковому предписанию, выражающему данный информационный запрос, — формальной релевантностью. В целом, в то время как формальная релевантность – это простой и понятный подход к поиску, смысловая релевантность может обеспечить более точные и полезные результаты, принимая во внимание контекст и значение поискового запроса.
2. Степень соответствия информационного запроса реальной информационной потребности называется ***пертинентностью***. Способы снижения пертинентности: изменение запроса, использование альтернативных поисковых систем, уточнение запроса, использование дополнительных инструментов.
3. В ***документальных*** ИПС хранятся тексты документов или их описания (рефераты, библиографические карточки и пр.). До последнего времени обычной формой представления данных в документальных ИПС был реферат (или другое краткое описание документа) и его библиографические данные. В этом массиве, который называется первым документальным контуром, и проводился основной поиск. ***Фактографические*** ИПС имеют дело с описанием конкретных фактов, причем не обязательно в текстовой форме. Это могут быть таблицы, формулы и пр. Существуют и смешанные ИПС, включающие как документы, так и фактографическую информацию. В настоящее время фактографические ИПС строятся на основе технологий баз данных (БД).
4. Ручное индексирование подразумевает, что человек (индексатор) просматривает документы и ручным образом определяет ключевые слова, связанную с документом. После этого, индексатор добавляет эту информацию в базу данных ИПС. ***Ручное индексирование*** является более точным, так как люди могут учесть различные контексты и значения слов, но оно более трудозатратное и дорогостоящее. ***Автоматическое индексирование*** основано на использовании алгоритмов и компьютерных программ, которые анализируют содержимое документов и автоматически определяют ключевые слова. Автоматическое индексирование позволяет быстро индексировать большое количество документов, но может быть менее точным, особенно при работе с сложными и многозначными терминами. Некоторые примеры систем ручного индексирования:

* MEDLINE - база данных медицинских статей, вручную индексируемая экспертами-библиографами.
* INSPEC - база данных по инженерии и техническим наукам, которая также вручную индексируется экспертами.

Некоторые примеры систем автоматического индексирования:

* Google - автоматически индексирует миллионы документов в Интернете, используя свои алгоритмы.
* Microsoft Academic - автоматически индексирует миллионы научных статей и других исследований с помощью своих алгоритмов.

В современных ИПС часто используется комбинированный подход, где ручное и автоматическое индексирование используются вместе для достижения наилучшей точности и скорости работы.

1. ***Общий каталог веб-ресурсов*** – это интернет-ресурс, на котором собраны ссылки на различные веб-сайты по разным тематикам, без определенной специализации. Такой каталог помогает пользователям быстро находить нужные сайты в разных категориях, но не специализируется на какой-то одной конкретной теме. Примеры общих каталогов веб-ресурсов: AllTop, Reddit.

***Специализированный каталог веб-ресурсов*** – это каталог, который специализируется на конкретной тематике или отрасли. Такие каталоги помогают пользователям быстро найти ресурсы по определенной теме или в конкретной отрасли. Примеры специализированных каталогов веб-ресурсов: GitHub, Airbnb, Tripadvisor.

1. ***Фасетная классификация*** – это метод классификации информации, основанный на описании документов несколькими независимыми характеристиками, называемыми фасетами. Фасеты позволяют пользователям уточнять поиск, добавляя дополнительные фильтры и условия. Пример фасетов:

Фасет "Бренд": Apple

Фасет "Модель": iPhone 13

Фасет "Цвет": графитовый

Фасет "Ёмкость памяти": 256 ГБ

Фасет "Цена": 71990 рублей

**Google:**

**Запрос:** где и когда появился термин «лингвистика»

**Документ, отвечающий результатам запроса (url):** http://slovar.lib.ru/dictionary/lingvistika.htm#:~:text=Lingua%20-%20язык)%20-%20наука%2C,"Словаря%20французского%20языка"%20Ш.

**Номер этого документа в списке результатов:** 1

**Инф. шум (количество нерелевантных ссылок):** около 4 ссылок отвечало конкретно на мой запрос, однако было много смежных ссылок, на которых был представлен либо более развёрнутый ответ, либо просто описывалось определение лингвистики.

**Полнота (в Рунете всего 4 источника):** буквально по первой выданной ссылке был получен исчерпывающий полный ответ без какой-либо воды.

**Рамблер:**

**Запрос:** где и когда появился термин «лингвистика»

**Документ, отвечающий результатам запроса (url):** https://znachenie-slova.ru/лингвистика

**Номер этого документа в списке результатов:** 3

**Инф. шум (количество нерелевантных ссылок):** только две ссылки с ближайших страниц отвечали конкретно на поставленный вопрос, остальные в основном описывали сам предмет лингвистики.

**Полнота (в Рунете всего 4 источника):** даже одного источника хватило на то, чтобы понять, что информация дана достаточно полная и исчерпывающая, хоть и есть немного лишнего, чего не было в моём запросе.

Однозначно я бы предпочёл ИПС Google, ведь там почти все первые ссылки отвечали моему запросу, было намного меньше информационного шума, нежели в Рамблере. В удобстве выделить никого не могу, задачи и функционал у них одинаковый. Визуально не нравится не один, глаз не цепляет, хотя оформление, запоминаемость и красота – далеко не первая задача поисковых систем.

Слово "child" в разных формах (child, children, childhood) встречается 42 раза в тексте книги "Alice's Adventures in Wonderland" by Lewis Carroll.