

## Практикум операторы информационно-поискового языка

### Задание 1. Естественнo-языковoй поиск. Проверьте справедливoсть указанных правил по трактовке слов в поисковых машинах Yandex, Rambler, Google, Aport.

Знаки «+» и «-». Если вы хотите, чтобы слова из запроса обязательно были найдены, поставьте перед каждым из них «+». Если вы хотите исключить какие-либо слова из результата поиска, поставьте перед каждым из них «-».

*Например, запрос «частные объявления продажа велосипедов», выдаст много ссылок на сайты с разнообразными частными объявлениями. А запрос с «+» «частные объявления продажа +велосипедов» покажет объявления о продаже именно велосипедов.*

*Если вам нужно описание Парижа, а не предложения многочисленных турагентств, имеет смысл задать такой запрос «путеводитель по парижу -агентство –тур».'*

Обратите внимание на знак «-». Это именно минус, а не тире и не дефис. Знак «-» надо писать через пробел от предыдущего и слитно с последующим словом, вот так: «рак –гороскоп». Если написать «рак-гороскоп» или «рак – гороскоп», то знак «-» будет проигнорирован.

Придумайте два словосочетания из 3-х слов и продемонстрируйте использование операторов «+», «-»

### Задание 2. Основные операторы. Проверьте справедливoсть указанных выше правил по естественнo-языковoму поиску в поисковых машинах Yandex, Rambler, Google, Aport.

Несколько набранных в запросе слов, разделенных пробелами, означают, что все они должны входить в одно предложение искомого документа. Тот же самый эффект произведет употребление символа '&'.

*Например, при запросе 'лечебная физкультура' или 'лечебная & физкультура', результатом поиска будет список документов, в которых в одном предложении содержатся и слово 'лечебная', и слово 'физкультура'. (Эквивалентно запросу '+лечебная +физкультура')*

Между словами можно поставить знак '|', чтобы найти документы, содержащие любое из этих слов. (Удобно при поиске синонимов).

*Запрос вида 'фото | фотография | фотоснимок | снимок | фотоизображение' задает поиск документов, содержащих хотя бы одно из перечисленных слов.*

Еще один знак, тильда '~', позволит найти документы с предложением, содержащим первое слово, но не содержащим второе.

*По запросу 'банки ~ закон' будут найдены все документы, содержащие слово 'банки', рядом с которым (в пределах предложения) нет слова 'закон'.*

Чтобы подняться на ступеньку выше, от уровня предложения до уровня документа, просто удвойте соответствующий знак. Одинарный оператор (&, ~) ищет в пределах предложения, двойной (&&, ~~) - в пределах документа.

Например, по запросу *'рецепты && (плавленый сыр)'* будут найдены документы, в которых есть и слово *'рецепты'* и словосочетание *'(плавленый сыр)'* (причем *'(плавленый сыр)'* должен быть в одном предложении. А запрос *'руководство Visual C ~~ цена'* выдаст все документы со словами *'руководство Visual C'*, но без слова *'цена'*.

Придумайте два словосочетания из 3-х слов и продемонстрируйте использование основных операторов.

### **Задание 3. Поиск с расстоянием. Проверьте справедливость указанных выше правил по использованию основных операторов в поисковых машинах Yandex, Rambler, Google, Aport.**

Часто в запросах ищут устойчивые словосочетания. Если поставить их в кавычки, то будут найдены те документы, в которых эти слова идут строго подряд.

Например, по запросу *«красная шапочка»* будут найдены документы с этой фразой. (При этом контекст *«а шапочка у нее была красная»* найден не будет.)

Как Yandex адресует слова? Если все слова в тексте перенумеровать по порядку их следования, то расстояние между словами a и b - это разница между номерами слов a и b. Таким образом, расстояние между соседними словами равно 1 (a не 0), а расстояние между соседними словами, стоящими «не в том порядке», равно -1. То же самое относится и к предложениям.

Если между двумя словами поставлен знак '/', за которым сразу напечатано число, значит, требуется, чтобы расстояние между ними не превышало этого числа слов.

Например, задав запрос *'поставщики /2 кофе'*, вы требуете найти документы, в которых содержатся и слово *'поставщики'* и слово *'кофе'*, причем расстояние между ними должно быть не более двух слов и они должны находиться в одном предложении. (Найдутся *"поставщики колумбийского кофе"*, *"поставщики кофе из Колумбии"* и т.д.)

Если порядок слов и расстояние точно известны, можно воспользоваться пунктуацией *'/+n'*. Так, например, задается поиск слов, стоящих подряд.

Запрос *'синяя /+1 борода'* означает, что слово *'борода'* должно следовать непосредственно за словом *'синяя'*. (К тому же результату приведет запрос *"синяя борода"*)

В общем виде ограничение по расстоянию задается при помощи пунктуации вида *'/(n m)'*, где 'n' минимальное, а 'm' максимально допустимое расстояние. Отсюда следует, что запись *'/n'* эквивалентна *'/(-n +n)'*, а запись *'/+n'* эквивалентна *'/(+n +n)'*.

Запрос *'музыкальное /(-2 4) образование'* означает, что *'музыкальное'* должна находиться от *'образование'* в интервале расстояний от 2 слов слева до 4 слов справа

Практически все знаки можно комбинировать с ограничением расстояния.

Например, результатом поиска по запросу *'вакансии ~ /+1 студентов'* будут документы, содержащие слово *'вакансии'*, причем в этих документах слово *'студентов'* не следует непосредственно за словом *'вакансии'*.

Когда знаки ограничения по расстоянию стоят после двойных операторов, то употребленные там числа - это расстояние не в словах, а в предложениях. Расстояние в абзацах определяется аналогично расстоянию в словах.

Запрос *'банк && /1 налоги'* означает, что слово *'налоги'* должно находиться в том же самом, либо в соседнем со словом *'банк'* предложении.

Придумайте два словосочетания из 3-х слов и продемонстрируйте использование операторов поиска с расстоянием.

#### Задание 4. Синтаксис языка запросов (строгий поиск) Проверьте справедливость указанных выше правил по ранжированию результатов поиска в поисковых системах Yandex, Rambler, Google, Aport.

Синтаксис	Что означает оператор	Пример запроса
пробел или &	логическое И (в пределах предложения)	лечебная физкультура
&&	логическое И (в пределах документа)	рецепты && (плавленный сыр)
	логическое ИЛИ	фото   фотография   снимок   фотоизображение
+	обязательное наличие слова в найденном документе (работает также в применении к стоп-словам)	+быть или +не быть
()	группирование слов	(технология   изготовление) (сыра   творога)
~	бинарный оператор И НЕ (в пределах предложения)	банки ~ закон
~ или -	бинарный оператор И НЕ (в пределах документа)	путеводитель по парижу ~ (агентство   тур)
/(n m)	расстояние в словах (-назад +вперед)	поставщики /2 кофе; музыкальное /(-2 4) образование; вакансии ~ /+1 студентов
«а»	поиск фразы	"красная шапочка" (эквивалентно красная /+1 шапочка)
&&/(n m)	расстояние в предложениях (-назад +вперед)	банк && /1 налоги

#### Поиск в элементах

Синтаксис	Что означает оператор	Пример запроса
\$title (выражение)	поиск в заголовке	\$title (CompTek)

<b>\$anchor</b> (выражение)	поиск в тексте ссылок	\$anchor (CompTek   Dialogic)
<b>#keywords</b> =(выражение)	поиск в ключевых словах	#keywords=(поисковая система)
<b>#abstract</b> =(выражение)	поиск в описании	#abstract=(искалка   поиск)
<b>#image</b> ="значение"	поиск файла изображения	#image="tort*"
<b>#hint</b> =(выражение)	поиск в подписях к изображениям	#hint=(lenin   ленин)
<b>#url</b> ="значение"	поиск на заданном сайте (странице)	#url="www.comptek.ru*"
<b>#link</b> ="значение"	поиск ссылок на заданный URL	#link="www.yandex.ru*"

### Советы по проведению поиска:

- Можно обойтись без механизма поиска, если то, что вы ищете, вам хорошо знакомо. Достаточно ввести предполагаемый адрес, например [www.cocacola.com](http://www.cocacola.com), [www.harrypotter.com](http://www.harrypotter.com) или [www.billbradley.com](http://www.billbradley.com).
- Экономьте время, ограничив область поиска конкретной категорией.
- Не щелкайте по ссылкам на полученных страницах. Вместо этого щелкните на ссылке правой клавишей мыши и выберите пункт меню **Open in New Window** (*Открыть в новом окне*) или перенесите ссылки мышью во второе окно браузера.
  - Избегайте специальных компьютерных терминов, таких, как *file*, *folder*, *disk* и *memory*, если вы не обозначаете ими компьютерные понятия.
  - При поиске имени собственного используйте режим поиска «точно по фразе» и кавычки, если это возможно.
  - Если в результате было обнаружено слишком мало страниц, переключитесь из режима поиска «точно по фразе» в режим поиска по всем словам, из него – в режим поиска по одному из слов или используйте меньше ключевых слов.
  - Если в результате поиска было обнаружено слишком много страниц, то переключитесь из режима поиска по одному из слов в режим поиска по всем словам или добавьте больше ключевых слов.
  - Для того чтобы узнать ответы на простые вопросы (например, какова высота Эйфелевой башни), обратитесь на узел, воспринимающий вопросы на разговорном английском языке.
  - Следите за правописанием. Придумайте два словосочетания из 3-х слов и продемонстрируйте использование синтаксиса поискового языка.

### Задание 5. Поиск в зонах. Проверьте справедливость указанных выше правил по использованию в запросе скобок в поисковых системах Yandex, Rambler, Google, Aport.

Можно искать информацию в «зонах» - заголовках (имя «зоны»: **Title**), ссылках (имя «зоны»: **Anchor**) и адресе (имя «зоны»: **Address**).

Синтаксис: **\$имя\_зоны (поисковое выражение)**.

*Запрос **'\$title CompTek'** ищет в заголовках документов слово **'CompTek'**.*

*Запрос **'\$anchor (CompTek | Dialogic)'** находит документы, в ссылках внутри которых есть одно из слов **'CompTek'** или **'Dialogic'**.*

### **Поиск в определенных элементах.**

Можно ограничить поиск информации списком серверов или наоборот исключить сервера из поиска (url). Можно также искать документы, содержащие ссылки на определенные URL (link), и файлы картинок (image). Если вы хотите работать не с конкретным URL (image), а со всеми, начинающимися с данной последовательности символов, используйте "\*".

Синтаксис: **#имя\_элемента=«имя\_файла (URL)»**.

*По запросу **'CompTek ~ #url=«www.comptek.ru\*»'** будут искажаться упоминания компании **'CompTek'** везде, кроме ее собственного сервера (www.comptek.ru). А запрос **'#link=«www.comptek.ru\*»'** покажет все документы, которые сослались на сервер компании.*

*Запрос **'#image=«tort\*»'** даст ссылки на документы с изображениями тортов (хотя, возможно, найдется и портрет черепахи Тортилли).*

Можно также искать по ключевым словам (keywords), аннотациям (abstract) и подписям под изображениями (hint).

Синтаксис: **#имя\_элемента=(поисковое выражение)**.

*Запросу **'#keywords=(поисковая система) | #abstract=(поисковая система)'** будут искажаться все страницы, в meta тегах которых есть эти слова.*

*По запросу **'#hint=(кино)'** будут найдены документы, содержащие изображение с такой подписью.*

### **Ранжирование результата поиска.**

При поиске для каждого найденного документа Яндекс вычисляет величину релевантности (соответствия) содержания этого документа поисковому запросу. Список найденных документов перед выдачей пользователю сортируется по этой величине в порядке убывания. Релевантность документа зависит от ряда факторов, в том числе от частотных характеристик искомых слов, веса слова или выражения, близости искомых слов в тексте документа друг к другу и т.д. Пользователь может повлиять на порядок сортировки, используя операторы веса и уточнения запроса.

Задание веса слова или выражения применяется для того, чтобы увеличить релевантность документов, содержащих «взвешенное» выражение.

Синтаксис: **слово:число** или **(поисковое\_выражение):число**

*По запросу **'поисковые механизмы:5'** будут найдены те же документы, что и по запросу **'поисковые механизмы'**. Разница состоит в том, что наверху списка найденного окажутся документы, где чаще встречается именно слово **'механизмы'**.*

*Запрос 'поисковые (механизмы | машины | аппараты):5' равнозначен запросу 'поисковые (механизмы:5 | машины:5 | аппараты:5)'.*

Задание уточняющего слова или выражения применяется для того, чтобы увеличить релевантность документов, содержащих уточняющее выражение.

Синтаксис: <- слово или <- (уточняющее\_выражение)

*По запросу 'компьютер <- телефон' будут найдены все документы, содержащие слово 'компьютер', при этом первыми будут выданы документы, содержащие слово 'телефон'.*

*Если ни в одном документе со словом 'компьютер' нет слова 'телефон', результат запроса будет эквивалентен запросу 'компьютер'.*

Придумайте два словосочетания из 3-х слов и продемонстрируйте использование поиска в зонах.

**Задание 6. Скобки. Проверьте справедливость указанных выше правил по поиску с расстоянием в поисковых системах Yandex, Rambler, Google, Aport.**

Вместо одного слова в запросе можно подставить целое выражение. Для этого его надо взять в скобки.

*Например, запрос '(история, технология, изготовление) /+1 (сыра, творога)' задает поиск документов, которые содержат любую из фраз 'история сыра', 'технология творога', 'изготовление сыра', 'история творога'.*

Придумайте два словосочетания из 3-х слов и продемонстрируйте использование скобок в поисковых языках.