Задание

Написать конспект (переписать тему и то, что выделено синим цветом), а прочитать внимательно и запомнить всё.

На последнем слайде задание решить письменно (разобранный пример есть на предпоследнем слайде).

Прислать фото конспекта и задания, на полях указывать дату и фамилию.

ПОИСК ИНФОРМАЦИИ

Типы поисковых систем

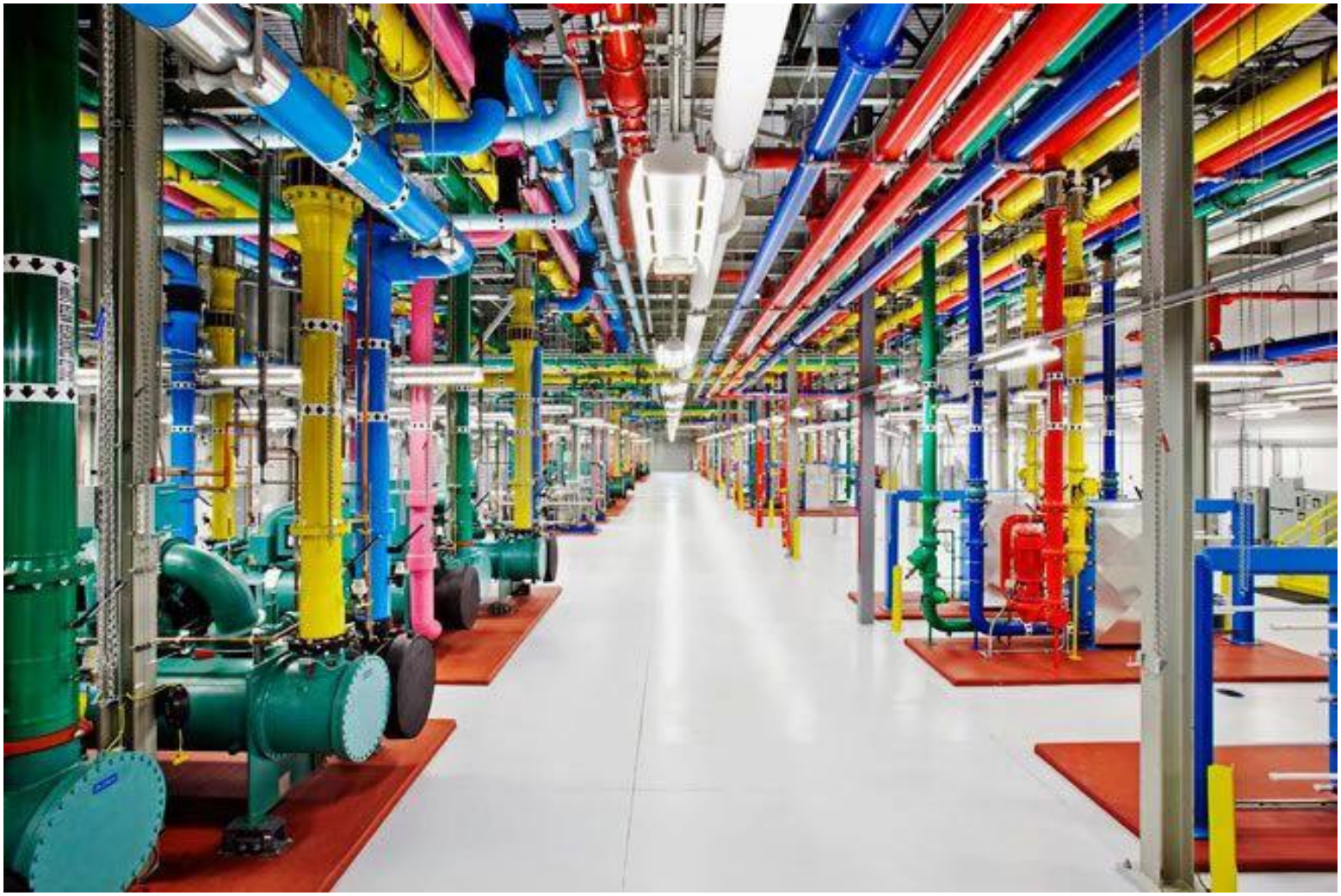
Поисковые системы различаются применяемыми подходами к сбору и обработке информации, организацией баз данных и предоставляемым пользователям возможностями по формулировке запросов и проведению поиска в базах данных. Можно выделить два типа систем.

- **поисковые системы (роботы**), которые автоматически по заданному алгоритму обходят Web-серверы и скачивают Web-страницы, используя имеющиеся на них гиперссылки, а затем осуществляют полнотекстное индексирование всех найденных документов и формируют базу данных, в которой хранятся сведения о содержащихся в документах словах и URL-адресов документов. Пользователь, задавая в запросе ключевые слова, получает в результате подборку ссылок на документы, содержание которых удовлетворяет критерию поиска. Заметим, что ведущие системы позволяют формулировать достаточно сложные запросы, используя объединение ключевых слов в логические выражения и введение дополнительных ограничений (например, на дату создания документа, доменные имена серверов).
- **Поисковые каталоги ресурсов** (справочно-поисковые системы) - устроены по тому же принципу, что и тематические каталоги библиотек. Они обычно представляют собой иерархические гипертекстовые меню с пунктами и подпунктами, определяющими тематику сайтов, адреса которых содержатся в данном каталоге, с постепенным, от уровня к уровню, уточнением темы. Поисковые каталоги создаются вручную. Высококвалифицированные редакторы лично просматривают информационное пространство WWW, отбирают то, что по их мнению представляет общественный интерес, и заносят в каталог. В большинстве каталогов все имеющиеся сведения индексируются, что позволяет проводить поиск по ключевым словам.

Виды поисковых систем

- Поисковые системы общего назначения
  - RAMBLER (http://www.rambler.ru)
  - Апорт (http://www.aport.ru)
  - Яндекс (http://www.yandex.ru)
  - Google (http://www.google.ru)
- Специализированные поисковые системы (позволяют искать информацию в других информационных слоях Интернета: серверах файловых архивов, почтовых серверах и др.)
  - http://www.wikipoisk.ru/ - поиск по энциклопедиям
  - http://beemp3.com/ - поисковик музыки
  - http://www.Whowhere.com – поисковая система, позволяющая найти адрес электронной почты по имени человека и наоборот

Так выглядит внутри один из центров обработки данных Google



Как работает поисковая система?

- **Первый этап работы поисковой системы** – это индексирование информации, находящейся в Internet. Сетевой робот поисковой системы просматривает огромное количество страниц и заносит адреса и краткое содержание этих страниц к себе в базу данных, точнее в поисковый индекс. Владельцы сайтов будут очень рады, если их сайт просмотрит сетевой робот. Для того, чтобы робот обязательно просмотрел сайт и внес его в поисковый индекс, владельцы сами регистрируют свои сайты в поисковых системах. Процесс занесения адреса и описания страницы в поисковую систему называется *индексация*. Таким образом, когда вы спросите поисковую систему о чём-то, она не будет искать совпадения в огромной сети Internet, а всего лишь быстро поищет в своем заранее подготовленном поисковом индексе.
- **Второй этап работы поисковой системы** – это выдача накопленной и отобранной, отсортированной и классифицированной информации по запросам пользователей.

Правила составления запроса

- Слова в запросе надо писать грамотно. Ошибка в одной букве может существенно затянуть процесс поиска.
- Для достижения необходимого результата уточняйте запрос, используя ключевые слова. Чем точнее будет составлен запрос - тем выше вероятность найти ответ в первых строчках выдачи.
- Пользуйтесь синонимами. Если запрос "реферат" не принёс желаемого результата, попробуйте заменить его на "курсовая работа".
- Поиск является регистрозависимым. Все запросы желательно вводить в нижнем регистре, т.е. не заглавными буквами. Исключения из этого - названия, пишущиеся с большой буквы.

Релевантность

- Поисковая машина обычно производит сортировку найденных документов по принципу релевантности.
- При индексации документов поисковые машины высчитывают так называемый «вес» слова на странице – соотношение количества повторов на странице заданного Вами слова к общему количеству слов на странице документа. Если Вы задаете запрос, состоящий из нескольких слов, то более релевантными будут документы, в которых совокупный вес слов будет максимальный. Однако при подсчете веса не учитывается, рядом или раздельно стоят данные слова, и поэтому нет гарантий, что в первых документах содержится максимальное количество повторений словосочетания. Вполне возможно, что такого словосочетания там вообще не будет.
- Поэтому, если Вы хотите найти заданное словосочетание – задавайте запрос в окне поисковой машины в кавычках. В этом случае будет высчитываться вес словосочетания целиком. Соответственно, гарантируется наличие именно данного словосочетания в найденных документах.

Пример

В таблице приведены запросы к поисковому серверу. Расположите обозначения запросов в порядке возрастания количества страниц, которые найдет поисковый сервер по каждому запросу.

Для обозначения логической операции «ИЛИ» в запросе используется символ |, а для логической операции «И» - .

А	законы & физика
Б	законы | (физика & биология)
В	законы & физика & биология & химия
Г	законы | физика биология

Решение:

Приведем два способа решения, один из которых основан на рассуждении, а второй предполагает использование графического представления операций над множествами. Рассуждая логически, мы видим, что больше всего будет найдено страниц по запросу Г, так как при его исполнении будут найдены и страницы со словом «законы» (в том числе, например, и юридические), и страницы, со словом «физика», и страницы со словом «биология». Меньше всего будет найдено страниц по запросу В, так как в нем требуется присутствие всех четырех слов на искомой странице.

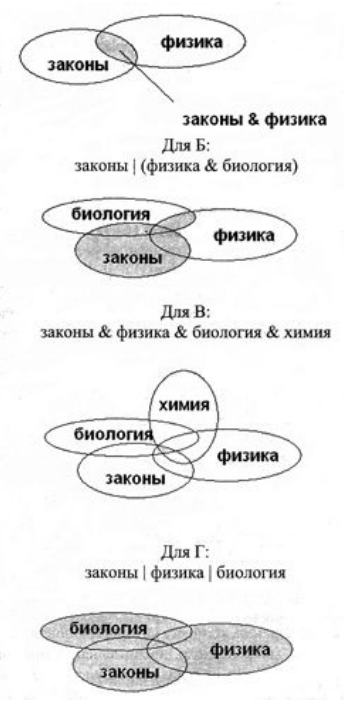
Осталось сравнить запросы А и Б. По запросу Б будет найдены все страницы, соответствующие запросу А, (так как в последних обязательно присутствует слово «законы»), а также страницы, содержащие одновременно слова «физика» и «биология». Следовательно, по запросу Б будет найдено больше страниц, чем по запросу А.

Итак, упорядочив запросы по возрастанию страниц, получаем ответ: ВАБГ.

Для решения вторым способом рассмотрим множества страниц, содержащие каждое из искомых слов. Запросу X Y будет соответствовать пересечение множеств X и Y, а запросу X | Y - их объединение. Воспользуемся графическим представлением действий над множествами. Множество страниц, содержащих некоторое слово, будем обозначать эллипсом. Множество, получившееся в результате запроса будем закрашивать серым цветом.

Диаграммы для запросов будут выглядеть следующим образом:

Упорядочив четыре полученные диаграммы по степени закрашенности, получаем ответ: ВАБГ.



Самостоятельно решить

- Используя данные таблицы, расположите номера запросов в порядке убывания количества страниц, которые найдет поисковый сервер по каждому запросу.
- Для обозначения логической операции «ИЛИ» в запросе используется символ I, а для логической операции «И» — &.

А	волейбол | баскетбол | подача
Б	волейбол | баскетбол | подача | блок
В	волейбол | баскетбол
Г	волейбол & баскетбол & подача