## МИНИСТЕРСТВО ОБРАЗОВАНИЯ

## Белорусский государственный университет

## информатики и радиоэлектроники

Факультет информационных технологий и управления

### Кафедра ИТАС

#### ОТЧЕТ

**по лабораторной работе №1**

**“Технология поиска информации ”**

Выполнил:

студент гр.120602

Будный Р. И.

Проверил:

Стригалёв Л.С.

Минск 2013

1. **Цель работы**

Практическое освоение технологии эффективного поиска информации

1. **Краткая характеристика использованных операторов и конструкций языка запросов**

Поисковый запрос в общем случае может состоять из одного или нескольких слов, логических операторов и знаков препинания. Простые запросы не требуют знания языка поисковых запросов, так что если ввести в поисковую строку несколько слов без знаков препинания и логических операторов, то будут найдены документы, содержащие все эти слова, причем на произвольном расстоянии друг от друга. Знание языка запросов конкретной поисковой машины и правильное его применение позволяет сделать поиск быстрым и эффективным.

* + **AND** — все термины, соединенные AND, должны присутствовать в предлагаемом документе. Некоторые поисковые системы используют значок «&» вместо AND;
  + **OR** — как минимум одно из ключевых слов, соединенных OR, должно присутствовать в искомом документе. Можно использовать значок «|»;
  + **NOT** — ключевое слово (слова), следующее за NOT, не должно появляться в искомом документе. Некоторые поисковые системы используют значок «&!» вместо NOT;
  + **скобки** — могут использоваться для построения вложенных запросов, изменения области действия операторов, а также для изменения принятых по умолчанию приоритетов логических операторов
  + **кавычки** — слова внутри кавычек являются фразой, которая должна быть найдена в пределах документа или файла.

Приведенные выше элементы языка поисковых запросов используются, как правило, всеми поисковыми машинами. При этом поисковые машины могут использовать и другие конструкции, включая метасимволы испециальныеоператоры, так что за деталями следует обращаться к справочной системе конкретной поисковой машины.

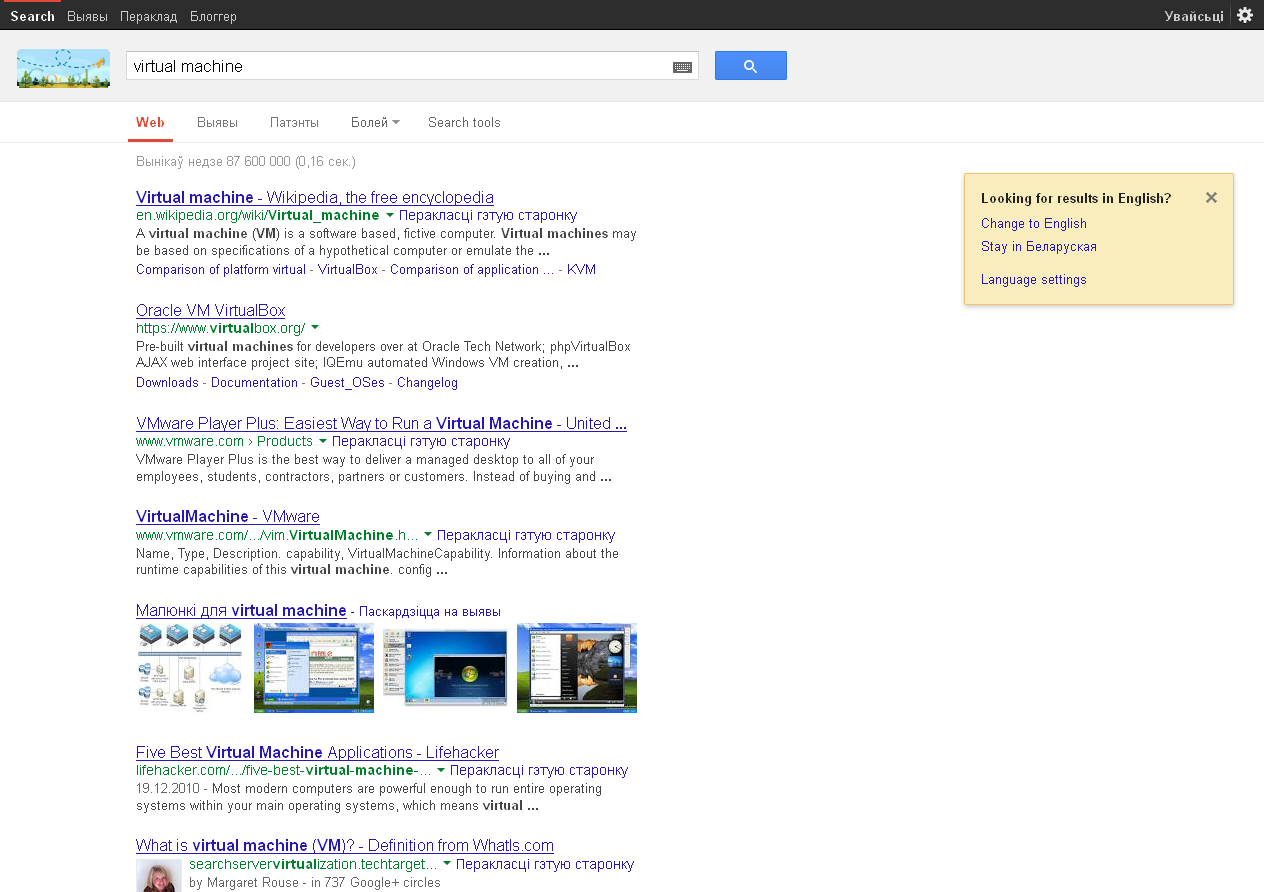
1. **Описание результатов поиска документов по вариантам поискового запроса**

Введем запрос в поисковые системы: google.com, yandex.ru, mail.ru, bing.com

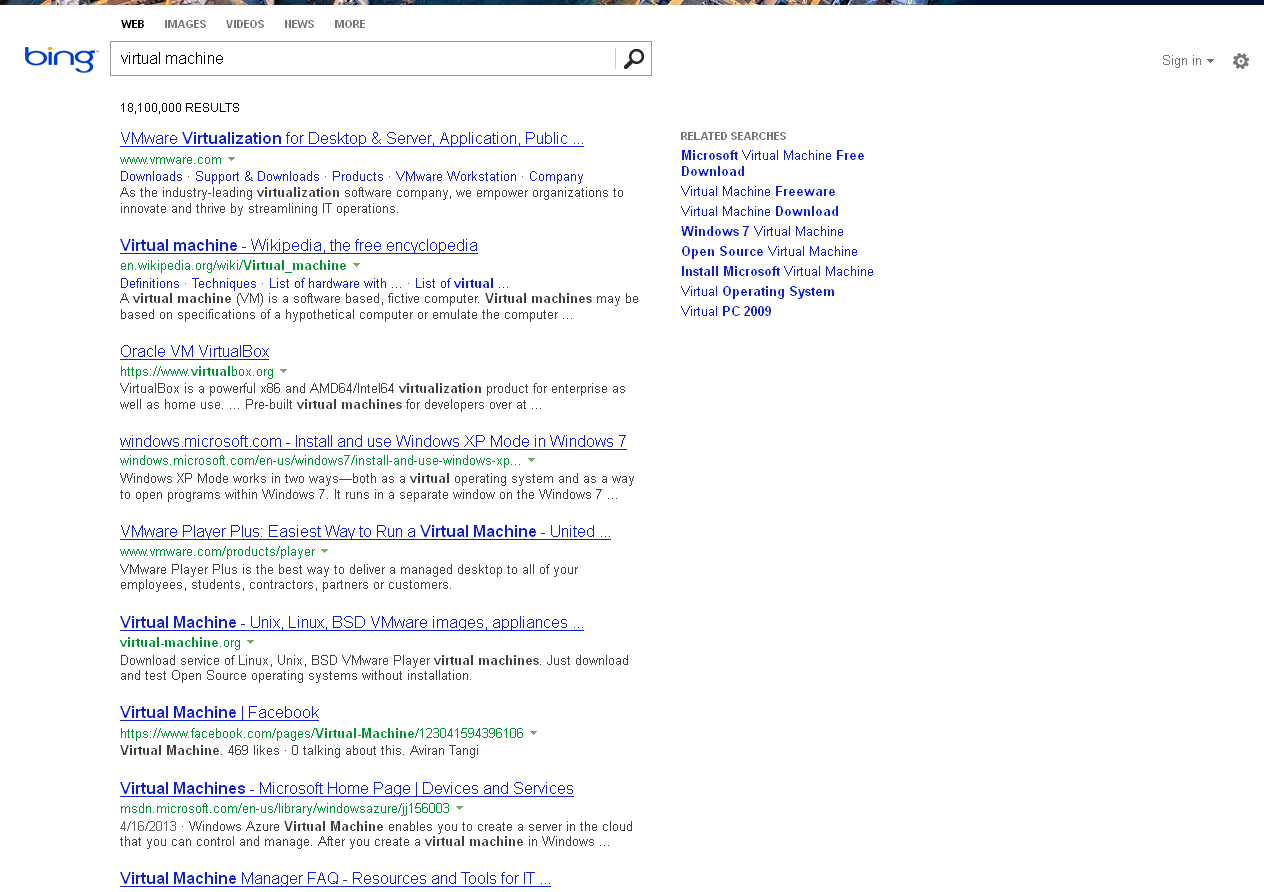
Пусть, например, нас интересует информация, касающаяся виртуальных машин. Будем рассматривать лишь информацию, умещающуюся на одном экране, без прокрутки, тем самым оценивая релевантность выдачи поисковой системы.

Запрос вида “Virtual Machine”:

Google.com

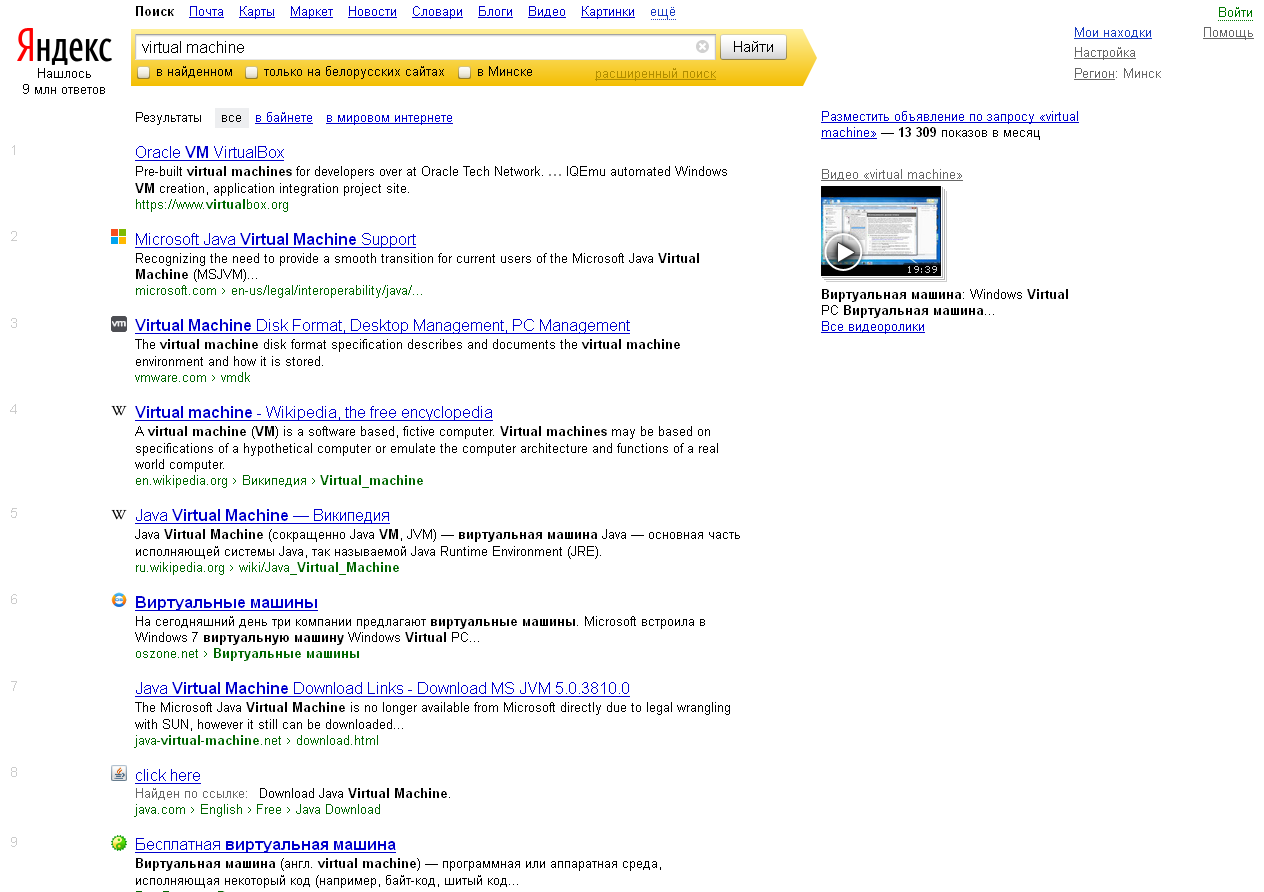


Первой идет статья с Wikipedia.org, где находится некоммерческая статья о виртуальных машинах в целом. Далее находятся ссылки на коммерческие виртуальные машины.Bing.com



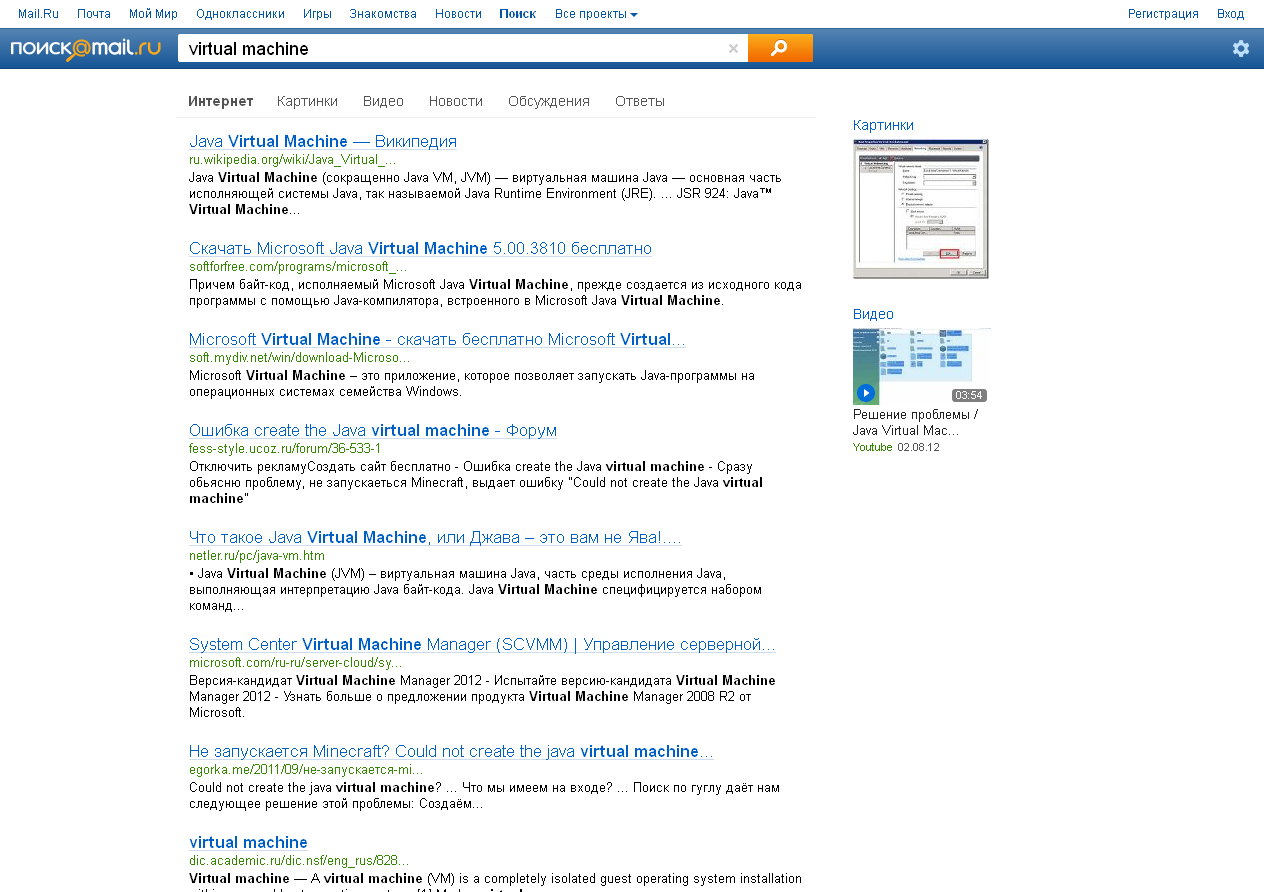
Что характерно, поисковая система от Microsoft первой выдает ссылку на проприетарную виртуальную машину, затем – на Wikipedia.org.

Yandex.ru

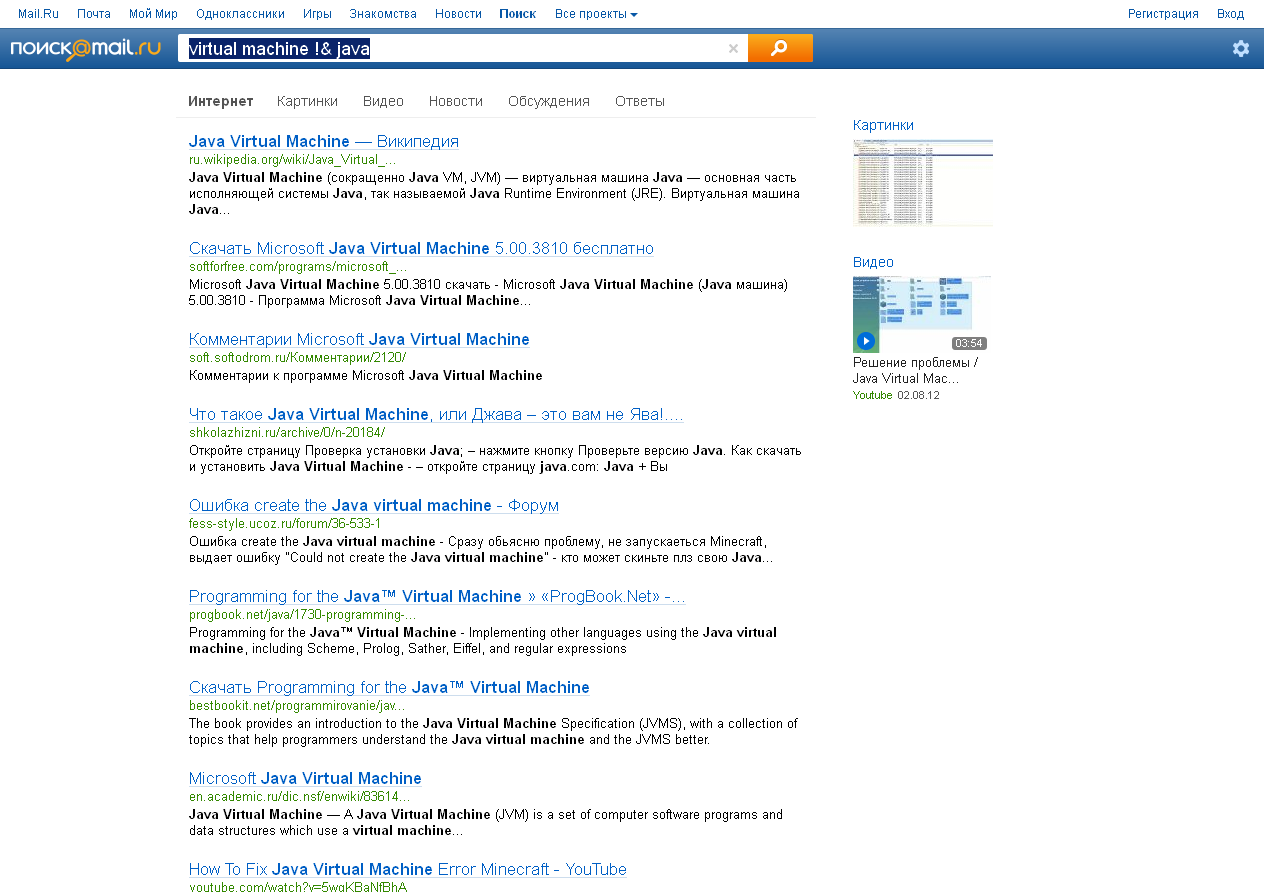


Здесь википедия всего на 4-том месте, кроме неё в выдаче находятся ссылки на коммерческие виртуальные машины и несколько нерелевантных ссылок вроде ”бесплатная виртуальная машина”.

Mail.ru



Первой идет статья на википедии, правда, не совсем релевантная, так как нас интересуют виртуальные машины вообще, а не Java Virtual Machine.  
Прочие ссылки также нельзя назвать релевантнтыми, так как они не соответствуют нашему запросу.

Уточним запрос, используя отрицание java:  
  
Как видно, теперь в выдаче находятся словосочетания из “java”, “virtual”, “Machine” — можно заключить, что поисковая система mail.ru не распознаёт языка поисковых запросов, элементы которого были перечислены в методичке по лабораторной работе.

Подтверждение этим соображениям обнаружим в: http://help.mail.ru/go-help/extended/query\_lang.

1. **Анализ полученных результатов**

Для сложных вопросов .имеет смысл использовать язык запросов. Наиболее эффективным оказался google.com.

1. **Выводы по работе**

В ходе лабораторной работы я использовал один поисковый запрос для сравнения релевантности выдачи пяти поисковых систем. В одном случае (mail.ru) пришлось применить язык запросов, который, как выяснилось, не поддерживается данной поисковой системой.