Тема:

Поиск информации в Интернет. Исследование эффективности поиска

.

**Тверь 2023**

Содержание

[Введение 3](#_Toc512416530)

[Основная часть 5](#_Toc512416531)

[1.Изучение теоретической части. 5](#_Toc512416532)

[2.Практическая часть 8](#_Toc512416533)

[ЗАКЛЮЧЕНИЕ 10](#_Toc512416534)

[Список использованных источников 11](#_Toc512416535)

# Введение

Интернет – это объединенные между собой компьютерные сети, глобальная мировая система передачи информации с помощью информационно-вычислительных ресурсов.

Разработка данной системы началась в 1957 году на фоне гонки вооружений. Целью создания такой сети стало решение Министерства обороны США, опасающегося нападения со стороны СССР. В результате была разработана сеть компьютеров, взаимосвязанных друг с другом и способных обмениваться информацией между собой.

Существует 2 типа компьютерных сетей:

* **Локальная сеть** — это два или несколько соединенных между собой компьютеров в одном здании. Обычно это домашние сети, офисные сети.
* **Глобальная сеть**  — состоит из двух или более локальных сетей. Интернет — самая большая существующая Глобальная сеть.

Началось все с сети компьютеров, расположенной в одной комнате, затем сеть расширилась в пределах здания, города, страны… За 15 лет компьютерная сеть разрослась до международных масштабов, объединив передовых ученых всего мира. Рождение Интернета как Всемирной компьютерной сети произошло в 1973 году – к сети ученых с помощью трансатлантического телефонного кабеля подключились английские и норвежские организации.

С тех пор прошло почти 40 лет, компьютерная сеть претерпела ряд существенных изменений – и сейчас Интернет доступен большинству жителей цивилизованного мира.

Интернет сегодня — это огромный мир с неограниченными возможностями. Изначально большинство людей пользовались интернетом только для поиска информации. Но сегодня это постоянно развивающийся «мега-инструмент», который не только содержит множество разнообразных информацией, но также предоставляет возможность учиться, работать, развлекаться, общаться и т.д.

**Цель:** получить основные навыки поиска информации в Интернет, изучить популярные поисковые системы, получить представление об основных критериях эффективности поисковых систем, определить рейтинги поисковых систем Интернет.

**Задачи:**

* изучить теоретический материал о Интернете и релевантности поиска;
* проанализировать несколько поисковых систем;
* найти самую полезную, удобную поисковую систему.

# Основная часть

## 1.Изучение теоретической части.

Информационный поиск – это отрасль знания, которая занимается изучением представления, хранения и доступа к информационным ресурсам.

Информационный ресурс в широком смысле - совокупность данных, организованных для эффективного получения достоверной информации. В более узком смысле информационные ресурсы – это отдельные документы и массивы документов в хранилищах данных информационных систем: библиотеках, архивах, фондах, базах данных, других видах хранилищ данных.

Информационно-поисковая система (ИПС) представляет собой комплекс программных и аппаратных средств, который обеспечивает отбор и представление электронных документов по заданным критериям.

ИПС работают с множествами документов, которые называются также коллекциями документов. Примерами таких множеств могут служить электронная библиотека, русскоязычная часть Интернета, энциклопедия на лазерном диске.

Информационная потребность пользователя

В определенный момент времени у пользователя возникает информационная потребность в документах по той или иной нужной ему тематике, которая обычно достаточно узка.

В начальный момент эта потребность часто не может быть точно выражена словами. При этом любая поисковая система требует достаточно четких запросов. Следовательно, пользователь должен представить свою информационную потребность в виде некоторого выражения, которое может быть воспринято поисковой системой.

Выделяют четыре этапа представления информационной потребности:

1.Реальная информационная потребность – это неосознанная истинная информационная потребность пользователя (потребность в некоторой новой информации при решении стоящей перед пользователем задачи);

3.Осознанная информационная потребность – появляется после осознания пользователем стоящей перед ним проблемы;

4.Выраженная информационная потребность – результат описания осознанной информационной потребности с помощью естественного языка;

5.Формализованная информационная потребность – это результат представления выраженной потребности средствами формального поискового языка ИПС.

Информационная потребность, сформулированная на информационно-поисковом языке, называется запросом. Запрос поступает в ИПС, и система возвращает некоторый ответ. Степень соответствия ответов поисковой системы запросу пользователя, а значит, и его информационной потребности, называется релевантностью.

Релевантность

Релевантность – это фундаментальное понятие теории информационного поиска. Наиболее часто используются следующие виды релевантности:

1.Когнитивная релевантность, или пертинентность– характеризует степень соответствия информации из документа и реальной информационной потребности пользователя. Это истинная и наиболее трудноопределимая релевантность. Все остальные виды релевантности являются ее приближениями.

2.Тематическая (предметная) релевантность – характеризует степень близости предмета (тематики) информационной потребности и найденного документа. Под тематикой здесь понимается область интересов пользователя, в пределах которой существует его информационная потребность.

3.Ситуационная релевантность – определяет полезность информационного ресурса для задачи, решаемой пользователем, с точки зрения временных затрат, способа взаимодействия пользователя с системой и т. п.

4.Системная (алгоритмическая) релевантность – характеризует степень близости между формализованной информационной потребностью (запросом) и найденным поисковой системой документом.

Теоретически релевантность можно измерить некоторым числом в интервале от нуля до единицы (полностью пертинентный документ). На практике же релевантность в основном определяется при оценке просматриваемых документов – подходит ли данный документ пользователю или не подходит. Таким образом, релевантность обычно характеризуется бинарным значением.

Существуют также методы оценки релевантности по многозначной шкале. В этом случае коллектив экспертов просматривает доступные документы и присваивает каждому из них некоторое значение релевантности, которое определяется для каждого запроса. Такая методика используется главным образом при сравнении эффективности нескольких поисковых систем.

Один и тот же документ может быть релевантным согласно одному определению релевантности и нерелевантным согласно другому. Так, документ, релевантный алгоритмически, может оказаться непертинентным.

Качество поиска в информационно-поисковых системах обычно характеризуется двумя критериями – полнотой и точностью. Полнота поиска определяется общим количеством найденных документов, а точность – соотношением между найденными релевантными и нерелевантными документами. Из-за многозначности определения релевантности абсолютного критерия точности поиска не существует, и в результатах любого поиска всегда присутствуют как релевантные, так и нерелевантные документы.

Нерелевантные документы, которые сравниваются с релевантными, иногда называются шумом, по аналогии с теорией передачи информации. Релевантные документы в таком случае можно называть сигналом, а эффективность поиска оценивать по соотношению "сигнал - шум".

Поиск в Интернет

Для поиска нужной информации в Интернете существует отдельный вид сетевых сервисов – поисковые системы (поисковые серверы, поисковые машины, "поисковики"*)*.

Современные поисковые системы предлагают и полнотекстовый поиск, и поиск по категориям, сочетая в себе их достоинства (как, впрочем, и недостатки).

Работа с поисковыми системами состоит в следующем:

В адресной строке браузера набираете его адрес, в строке запроса набираете на нужном языке ключевые слова или фразу, соответствующие ресурсу или ресурсам сети, которые вы хотите найти. Обычно поисковая система выдает результаты поиска небольшими порциями, например, по 10 на одну страницу выдачи.

Формирование запросов является ключевым моментом при работе с поисковыми системами. С первого раза удачно задать вопрос поисковой системе получается не всегда. Если запрос короткий и в нем присутствуют только часто употребляемые слова, может быть найдено очень много документов. Наоборот, если запрос окажется слишком детализированным или в нем будут использованы очень редкие слова, вы увидите сообщение о том, что ресурсов, отвечающих вашему запросу, не найдено. Постепенное сужение или расширение области поиска через увеличение или уменьшение списка ключевых слов, замена неудачных поисковых терминов на более удачные помогут улучшить результаты поиска.

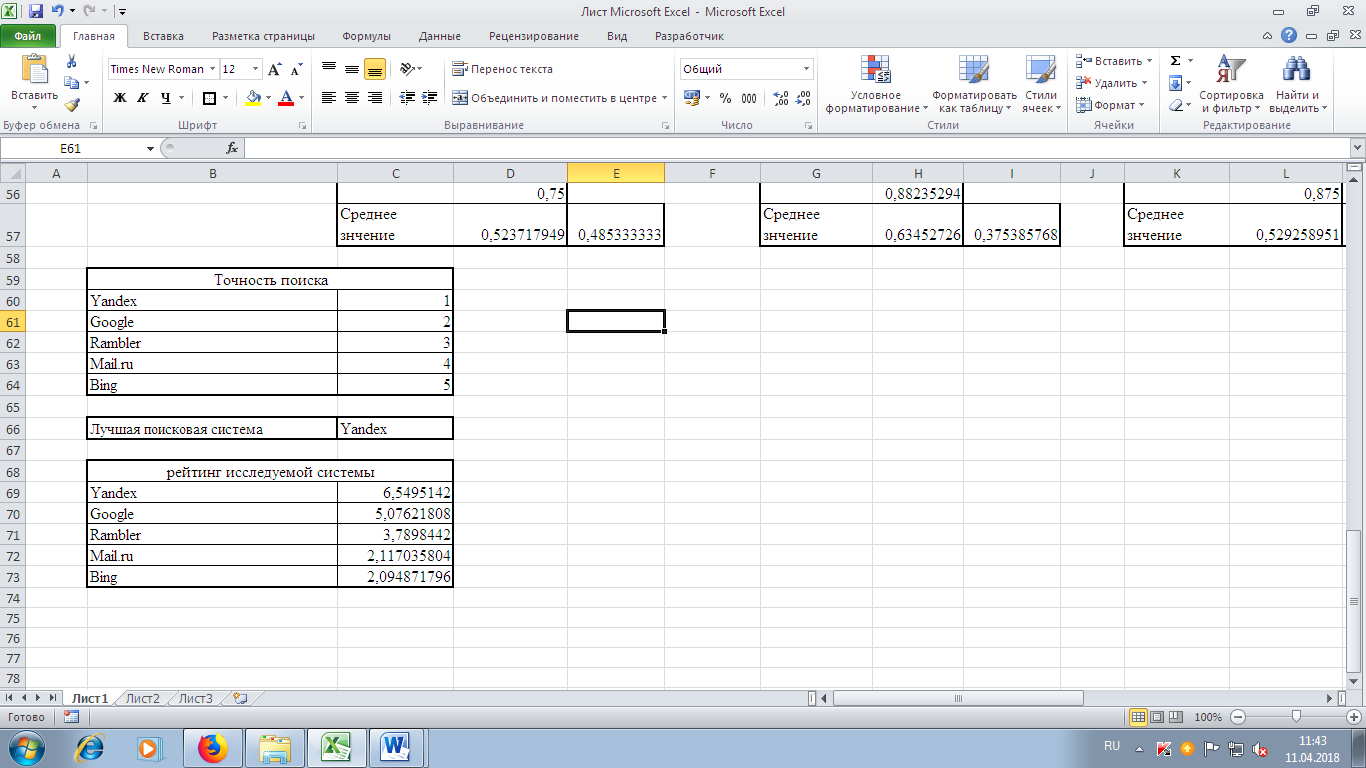
Ключевые слова, составляющие поисковый запрос, обычно просто разделяются пробелами. Необходимо помнить, что различные поисковые системы по-разному интерпретируют это. Некоторые из них отбирают по такому запросу только документы, содержащие все ключевые слова, то есть воспринимают пробел в запросе как логическую связку И. Некоторые интерпретируют пробел как логическое ИЛИ и ищут документы, содержащие хотя бы одно из ключевых слов. При формировании поискового запроса большинство систем позволяют в явном виде указать логические связки (И, ИЛИ, НЕ) и задать некоторые другие параметры поиска. В разных поисковых системах при формировании расширенного поискового запроса используется разный синтаксис – так называемый язык запросов. С помощью языка запроса вы можете указать, какие слова обязательно должны встретиться в документе, каких быть не должно, какие желательны.

Как правило, современные поисковые системы используют при поиске все возможные словоформы использованных слов. Независимо от того, в какой форме вы употребили слово в запросе, поиск учитывает все его формы: например, если задан запрос 'идти', то в результате поиска будут найдены ссылки на документы, содержащие слова 'идти', 'идет', 'шел', 'шла' и т.д.

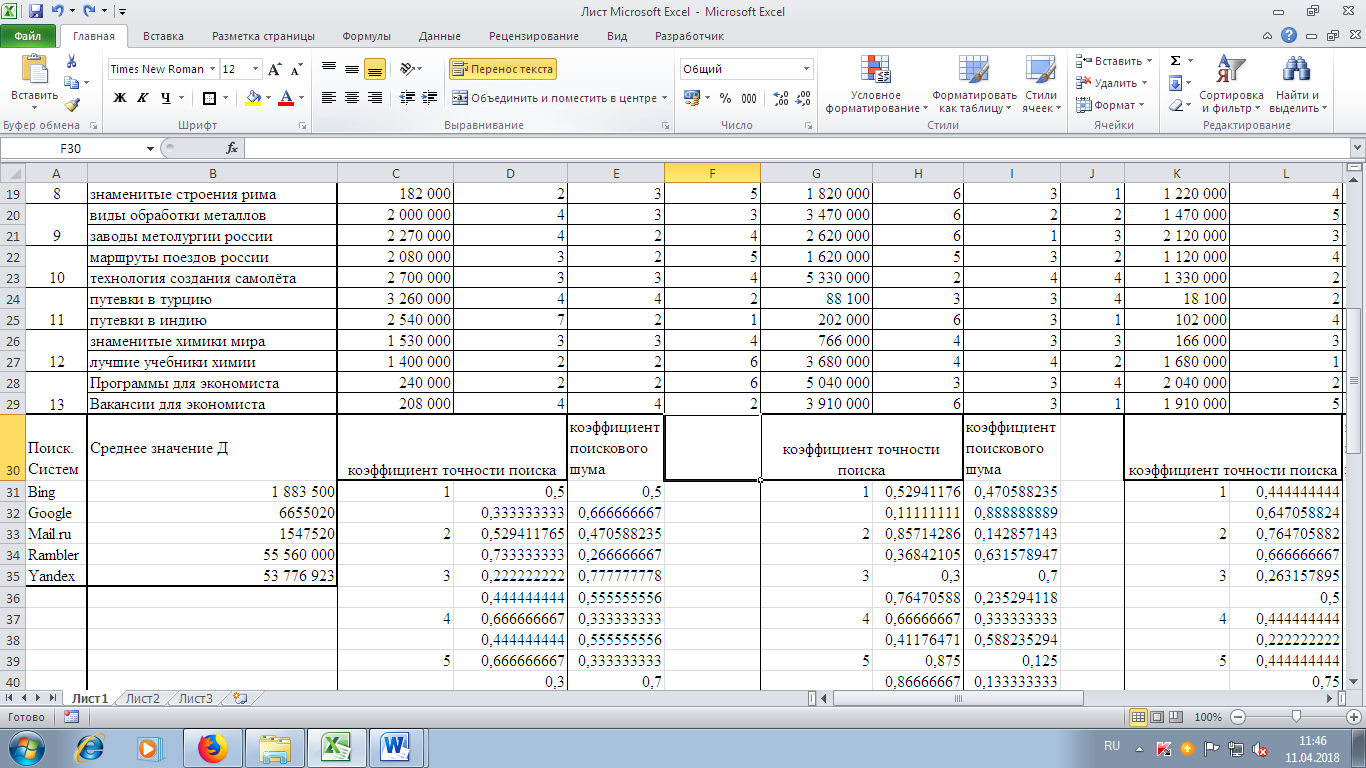
## 2.Практическая часть

Задача стояла в то, что нужно было исследовать на релевантность несколько поисковых систем, по нескольким темам (точнее 13 тем, на каждую тему 2 запроса).

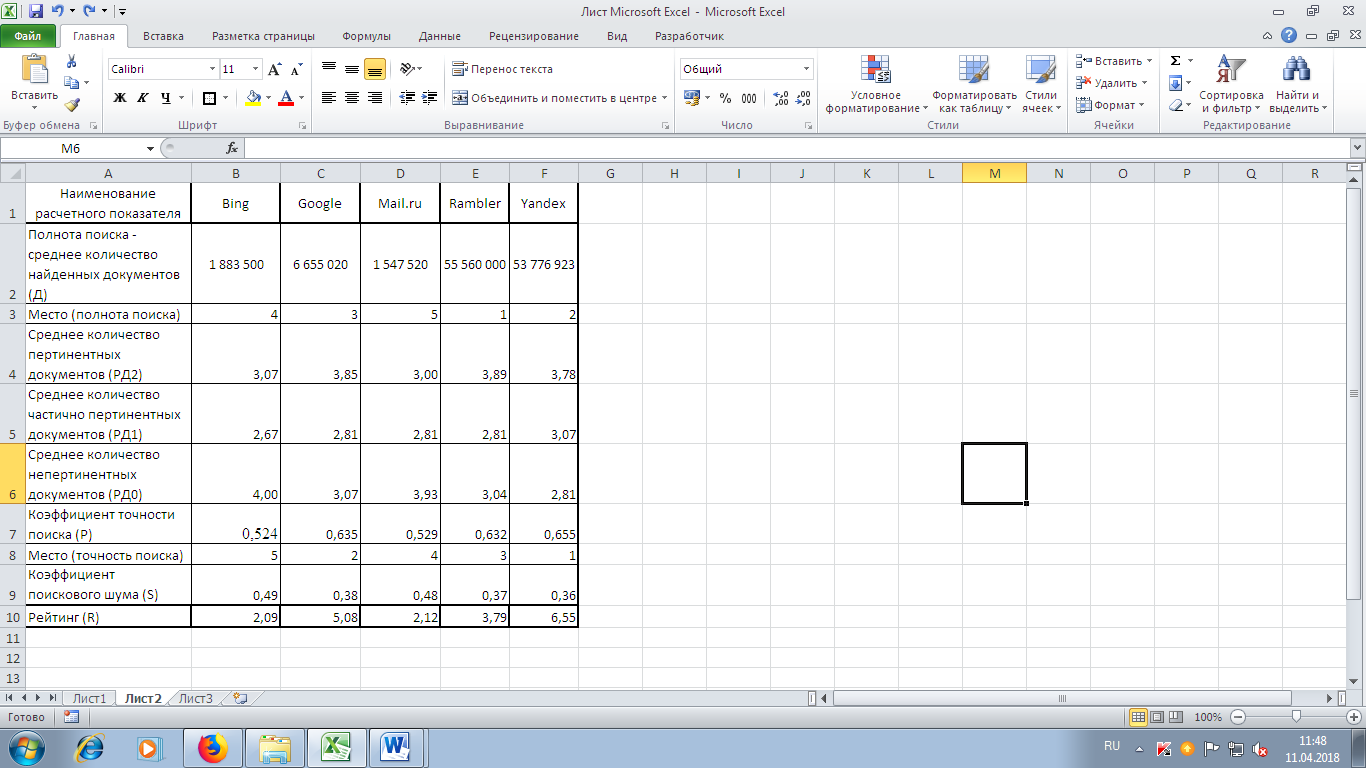
Выбрано было 5 поисковых систем: Bing, Google, Mail.ru, Rambler, Yandex. Проведено исследование на точность поиска по 10 первым сайтам:



И также по полноте поиска(количество найденных сайтов):



И проведен анализ на лучшую поисковую систему. Все это была систематизировано и выглядит так:



# ЗАКЛЮЧЕНИЕ

Bing – наименьшее воздействие поискового шума

Rambler – наиболее полный поиск по теме.

Проведя все вышеперечисленные действия можно выделить лучшую поисковую систему: Yandex.

# Список использованных источников

1. Введение в информационно-поисковые системы: метод. указания по изучениюдисциплины "Мировые информ. ресурсы" для студентов спец. "Прикл.информатика (в эконом.)" и "Информ. сист. и технол.". Ч. 1 / Иванов, В.К.,Иванов, К.В. - Тверь: ТГТУ, 2005. - 36 с.

2. Введение в информационно-поисковые системы: метод. указания по изучениюдисциплины "Мировые информ. ресурсы" для студентов спец. "Прикл.информатика (в экономике)" и "Информ. сист. и технол.". Ч. 2 / Иванов, В.К.,Иванов, К.В. - Тверь: ТГТУ, 2005. - 36 с.

3. http://procomputer.su/osnovy-interneta/54-poiskovaya-sistema