

УДК 004:167.7

ПОИСК ИНФОРМАЦИИ В ИНТЕРНЕТЕ (ИСТИНА ГДЕ-ТО РЯДОМ)

С.В. Дудченко

(Центр информационных технологий Межвузовского центра «Крым»,
г. Симферополь, АРК, Украина)

Рассматривается проблема расположения и поиска информации в виртуальном информационном пространстве (Интернете).

Ключевые слова: объективная реальность, субъективная реальность, виртуальная реальность, виртуальное информационное пространство, поиск, информация, индексирование, Интернет

Розглядається проблема розташування і пошуку інформації у віртуальному інформаційному просторі (Інтернеті).

Ключові слова: об'єктивна реальність, суб'єктивна реальність, віртуальна реальність, віртуальний інформаційний простір, пошук, інформація, індексація, Інтернет

The problem of location and information and search in virtual informative space (Internet) is considered in the article.

Keywords: objective reality, subjective reality, virtual reality, virtual informative space, search, information, indexation, Internet

Вместо предисловия

Уже давно Интернет стал своеобразной виртуальной реальностью отличной от реальностей как объективной, так и субъективной. Созданный вроде бы в материальном мире при помощи технических средств Интернет является вместилищем для субъективной реальности человеческого сознания.

Материя, время, пространство образуют объективную реальность, данную нам в ощущениях. Такое несколько измененное определение диалектического материализма может являться основополагающим для философии физики и других точных наук.

Материя - это объективная реальность, существующая вне и независимо от сознания человека и человечества.

Сознание - субъективная реальность, существующая в голове человека, в его восприятии, представлениях, мыслях, как отражение объективной реальности.

Следовательно, «субъективная реальность» – (отраженная) искаженная «объективная реальность». Это искажение зависит от степени информирования субъекта, чем больше степень информирования субъекта: тем меньше искажение объективной реальности.

Также и виртуальная реальность Интернета обладает своей объективной и субъективной реальностью. И если рассматривать Интернет не как множество компьютеров связанных между собой, а как некое информационное пространство¹, то аналогом «материи» в этом виртуальном мире может являться такое понятие как «Информация»

Информация – обработанное мозгом ощущение реальности.

¹ Здесь и далее Интернет рассматривается, только в качестве информационного пространства

Информация 1 типа – это обработанное мозгом ощущение объективной реальности.

Пример: Информация, полученная посредством всех органов чувств и подтвержденная этими же органами чувств. (Резиновый красный мяч находится у человека в руках. Зрением человек может определить то, что он красного цвета и имеет форму шара. Осязанием и обонянием – то, что он резиновый)

Информация 2 типа – это обработанное мозгом ощущение субъективной реальности (сознания) (т.е. реальности, возникающей у субъекта вследствие обработки информации 1 типа – анализ).

Пример: Информация, полученная посредством некоторых органов чувств, в частности, слухом. (Резиновый красный мяч существует на самом деле, но субъект, получает информацию об его существовании от другого субъекта посредством речи и органов слуха. Слово «мяч» дает понять, что предмет имеет форму шара, шар полый, слово «резиновый» говорит о том, что мяч сделан из определенного материала, слово «красный» говорит о том, что мяч определенного цвета. Мозг человека анализирует полученную информацию и человек имеет более или менее точное представление о предмете).

Информация 3 типа – это обработанное мозгом ощущение виртуальной реальности (т.е. реальности, несуществующей в материальном мире)

Пример: Информация, полученная некоторыми из органов чувств, в частности, зрением. (Человек видит на экране компьютера трехмерное изображение мяча в движении, читая текст о том, что мяч резиновый, видя, что он красный и имеет форму шара, человек получает некоторую информацию, анализирует все ее составляющие и получает некоторое представление о предмете, который он видит на экране. Но данного предмета может и не существовать в реальности, хотя и не исключен вариант того, что предмет существует. Ведь слово «виртуальный» означает фактический или возможный, а не так как привыкли теперь все воспринимать, виртуальный значит несуществующий).

Что же такое информация?

Например, в «Толковом словаре русского языка» С.И. Ожегова и Н.Ю. Шведовой [1] дается такое определение этого слова:

ИНФОРМАЦИЯ, -и, ж. 1. Сведения об окружающем мире и протекающих в нем процессах, воспринимаемые человеком или специальным устройством (спец.). *Передача информации. Теория информации* (раздел кибернетики, изучающий способы измерения и передачи информации). 2. Сообщения, осведомляющие о положении дел, о состоянии чего-н. *Научно-техническая и. Газетная и. Средства массовой информации* (печать, радио, телевидение, кино). ♦ **Генетическая информация** (спец.) – совокупность наследственных признаков, передаваемых от клетки к клетке, от организма к организму. || прил. **информационный**, -ая, -ое. *Информационное бюро. И. бюллетень.*

Новейший словарь иностранных слов и выражений [2] определяет понятие информации следующим образом:

ИНФОРМАЦИЯ [< лат. informatio разъяснение; просвещение] 1) осведомление; сообщение, сведения о ч.-л.; 2) обмен сведениями между людьми, человеком и автоматом, автоматом и автоматом; обмен сигналами в животном и растительном мире; передача признаков от клетки к клетке, от организма к организму (генетическая И.); 3) в математике и кибернетике – количественная мера устранения неопределенности (энтропии), мера организации системы; теория И. – теория сообщений; область кибернетики, в которой математическими методами исследуются способы измерения количества И., ее сбора, кодирования, преобразования и передачи.

Политехнический словарь [3]:

ИНФОРМАЦИЯ (от лат. informatio — разъяснение, изложение, осведомление) — совокупность к.-л. сведений, данных, передаваемых людьми устно (в форме речи), письменно (в виде текста, таблицы, рисунков, чертежей, условных обозначений) либо др. способом (напр., с помощью звуковых или световых сигналов, электрич. и нервных импульсов, перепадов давления или темп-ры); с сер. 20 в. — общенауч. понятие, включающее обмен сведениями между людьми, человеком и автоматом, автоматом и автоматом, обмен сигналами в животном и растит. мире, передачу наследств. признаков от клетки к клетке, от организма к организму; одно из осн. понятий *кибернетики*. И. может иметь непрерывный (аналоговая) или прерывистый (дискретная) характер. Общие св-ва И. независимо от ее смыслового содержания, способы матем. описания и количеств. оценка, методы кодиро-

вания, передачи, хранения, извлечения и классификации И. являются предметом исследования и изучения *информации теории и информатики*.

Толковый словарь по информатике [4]:

ИНФОРМАЦИЯ [informatio] 1. Совокупность знаний о фактических данных и зависимостях между ними. Является одним из видов ресурсов, используемых человеком в трудовой деятельности и в быту. 2. В вычислительной технике – содержание, присваиваемое данным посредством соглашений, распространяющихся на эти данные; данные, подлежащие вводу в ЭВМ, хранимые в ее памяти, обрабатываемые на ЭВМ и выдаваемые пользователям.

Словарь по кибернетике [5]:

ИНФОРМАЦИЯ (лат. informatio - разъяснение, осведомление) — одно из основных понятий *кибернетики*. Первоначально означало сообщение *данных*, сведений, осведомление и т.п. Кибернетика вывела понятие И. за пределы человек. речи и др. форм коммуникации между людьми, связала его с целенаправл. системами любой природы – биол., тех., социальными. И. выступает в трёх формах: биологической (биотоки в организмах, связи в генетич. механизмах), машинной (сигналы в электронных цепях) и социальной (движение человек. знаний в обществ. системах). С общ. стороны И. – связь в любых целенаправл. системах, определяющая их целостность, устойчивость, уровень функционирования. И. можно выразить математически и измерять с помощью информационной единицы – *бита*. Как отражение явлений реального мира, понятие И. раскрывается указанием действий, в которых она участвует: передачи, преобразования и хранения. Хранение И. предполагает наличие *носителя информации*. Передача И. предполагает наличие передатчика, приёмника и *канала связи*, способного отображать состояние передатчика в состоянии приёмника. Обработка И. - выполнение любого *алгоритма*, исходные данные для к-рого отождествляются с состоянием того или др. носителя. Различают дискретную и непрерывную форму И. Дискретная И. является осн. формой И. и существует (передается, обрабатывается, хранится) и виде конечных совокупностей качественно различных *символов* (печатные, рукописные документы, состояния цифровых *автоматов* и г. п.); непрерывная И. воплощается в образах (зрительных, звуковых и др.) и непрерывно наблюдаемых траекториях процессов (напр., осциллограммах). Как в естест. так и в искусств. процессах, в к-рых участвует И., одни формы её переходят в другие. Изучение общ. свойств И. независимо от её смыслового содержания является предметом *информации теории*. Вопросами, связанными с конкретными видами И., занимаются области науки и техники, использующие их для решения задач. Социальную И. (в виде человек. знаний), включая её семантич. аспекты, роль, механизм воздействия на обществ. прогресс, изучает *информатика*.

Существует также и ряд довольно кратких определений информации.

«Информация – это содержание сообщения, сигнала, памяти, а также сведения, содержащиеся в сообщении, сигнале или памяти» (Информатика: Энциклопедический словарь для начинающих / Сост. Д.А. Поспелов. – М.: Педагогика. – Пресс, 1994. – 352 с.: ил.)

«Информация – данные, организованные таким образом, что имеют смысл для имеющего с ними дело человека» (Большой толковый словарь компьютерных терминов (Collins Dictionary of Personal Computing, ed. Ian R. Sinclair), изд-во Вече. АСТ. М. 1999).

«Информация – 1) что-то сказанное, новости; знание, полученное любым способом; 2) в информационной теории и теории компьютеров: это точная мера информации, измеренная в битах и охватывающая диапазон от нуля (это когда все известно заранее) и до какого-то максимального значения, когда ничего заранее о содержании сообщения не известно; 3) любые данные, хранящиеся в компьютере» (Webster New World Dictionary of the American Language ed. David Guralnik, Prentice Hall Press, Division of Simon & Schuster).

«Информация – 1) то, что снижает неопределенность (Клод Шэннон); 2) то, что изменяет нас (Грегори Бэйтсон); 3) суть представления факта (или послания) для получателя (Хорнунг)» (Web Dictionary of Cybernetics and Systems).

«Информация» – сведения, воспринимаемые человеком и/или специальными устройствами как отражение фактов материального или духовного мира (Словарь терминов по библиотековедению, библиографоведению и информатике: учебное пособие для студентов. Самарская гос. академия культуры и искусств, 2003 г.).

В самом Интернете довольно легко можно найти огромное количество определений. Для этого лишь необходимо выйти на какую-нибудь поисковую систему, например Яндекс (Рис. 1).

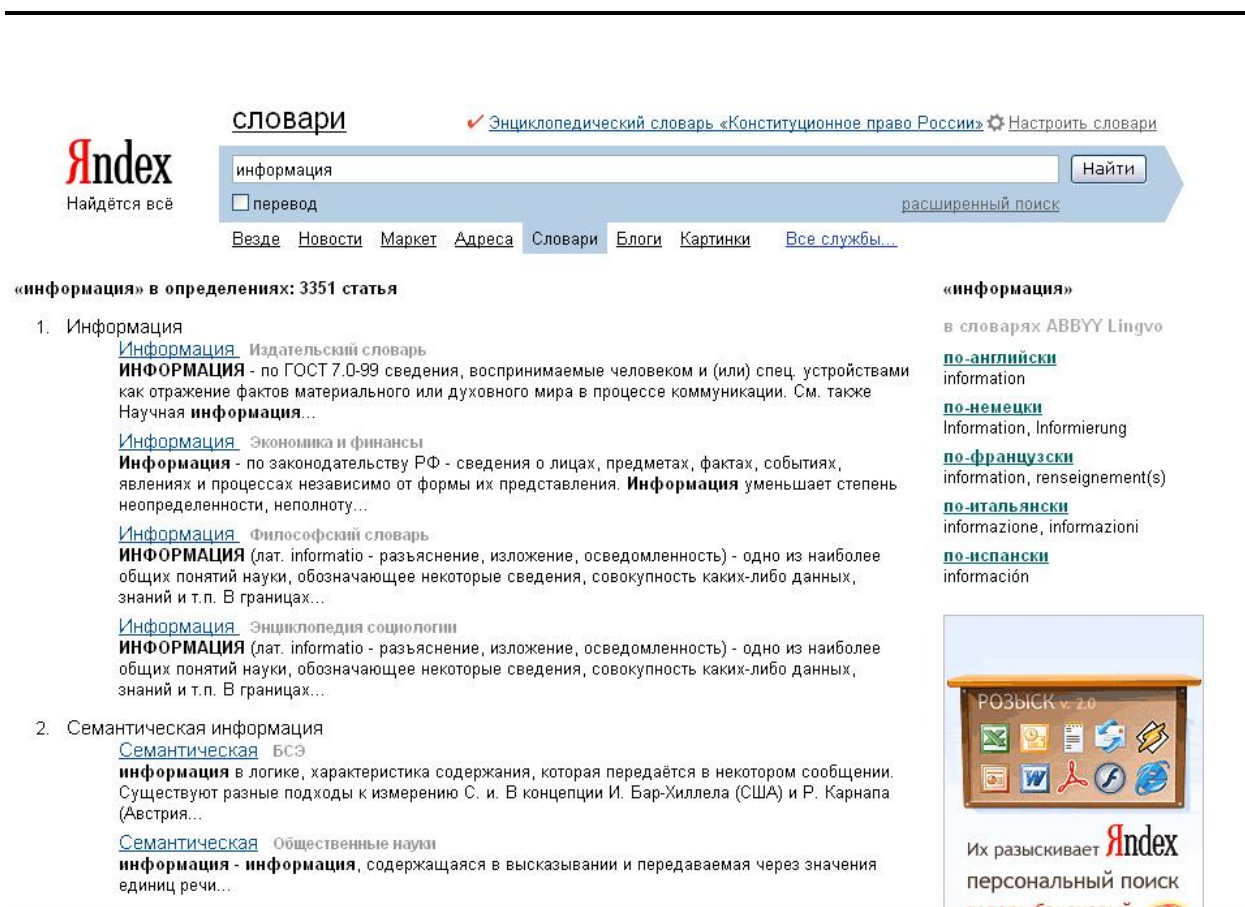


Рис. 1

Существуют также довольно-таки оригинальные определения, например [6]:

information информация 1. Сведения, неизвестные до их получения. 2. Значение, приписанное данным. 3. Данные. *См. т.ж.* data (accounting information, housekeeping information, state information)

«Сведения, неизвестные до их получения». Так к большому сожалению, при поиске в Интернете большинство сведений вовсе не получаются а, следовательно так и остаются неизвестными.

Поиск в Виртуальном информационном пространстве

Откуда берется информация? Очевидно, что она берется из информационного пространства. А что же является информационным пространством? Скорее всего, это окружающий нас мир, вселенная и т.п. И виртуальная реальность, к которой принадлежит Интернет, является лишь частью этой вселенной.

Вселенная бесконечна, но не безгранична, поэтому для того, чтобы показать место Интернета в информационном пространстве, воспользуемся простейшей плоской, но вполне удовлетворяющей данному условию, геометрической моделью. В плоском случае все информационное пространство (объективная реальность) может быть представлено в качестве окружности, а виртуальное информационное пространство (виртуальная реальность) – в качестве окружности, находящейся внутри окружности общего информационного пространства (Рис. 2).



Рис. 2

Исходя из представленной схемы, очевидно, что если информация, которую необходимо найти, находится за пределами окружности «Информационное пространство» (Виртуальная реальность), то отыскать ее при помощи компьютера в принципе невозможно.

Что же такое виртуальное информационное пространство, и почему в качестве модели пространства была выбрана окружность, а не круг? Потому, что область внутри окружности пока еще не заполнена (хотя стремится к заполнению), т.к. еще не все компьютеры (которые являются средством доступа к виртуальному информационному пространству) связаны друг с другом. А пока виртуальное информационное пространство состоит из множества подпространств, больших и маленьких (локальных компьютерных сетей и отдельно стоящих компьютеров). Самым большим подпространством в виртуальном информационном пространстве является Интернет – глобальная информационная сеть, которая постоянно расширяется. Остальные подпространства находятся в самом различном отношении к подпространству Интернета. Рассмотрим различные типы сочетания информационных подпространств по отношению к Интернету и расположение информации в этих подпространствах (Рис. 3).

Если искомая информация попала в зону Интернета, то удаленный пользователь может ее получить в полном объеме, используя поисковики Интернета. В случае, если информация попала в зону «Электронные библиотеки, проиндексированные поисковиками Интернета», удаленный пользователь также может получить доступ к информации в полном объеме, но уже используя не только поисковики Интернета, но и специализированные поисковые системы, созданные не по стандартам организации поиска принятой в Интернете (электронные библиотеки со своей систематизацией имеющейся информации). Как видно из схемы (Рис. 3) «Электронные библиотеки, проиндексированные поисковиками Интернета» находятся в виртуальном информационном пространстве и соприкасаются с подпространством Интернета в одной точке, что характеризует то, что через поисковики Интернета можно найти лишь «вход» в электронную библиотеку, информацию же о содержательном наполнении данной электронной библиотеки можно получить только после того как пользователь «зайдет» в эту электронную библиотеку.

«Локальные электронные библиотеки» образуют отдельные информационные подпространства, не связанные как с Интернетом, так и друг с другом. Информацию о них из Интернета получить невозможно: для получения информации необходимо находиться за компьютерами локальных сетей или отдельно стоящими компьютерами, где располагается поисковая информация.

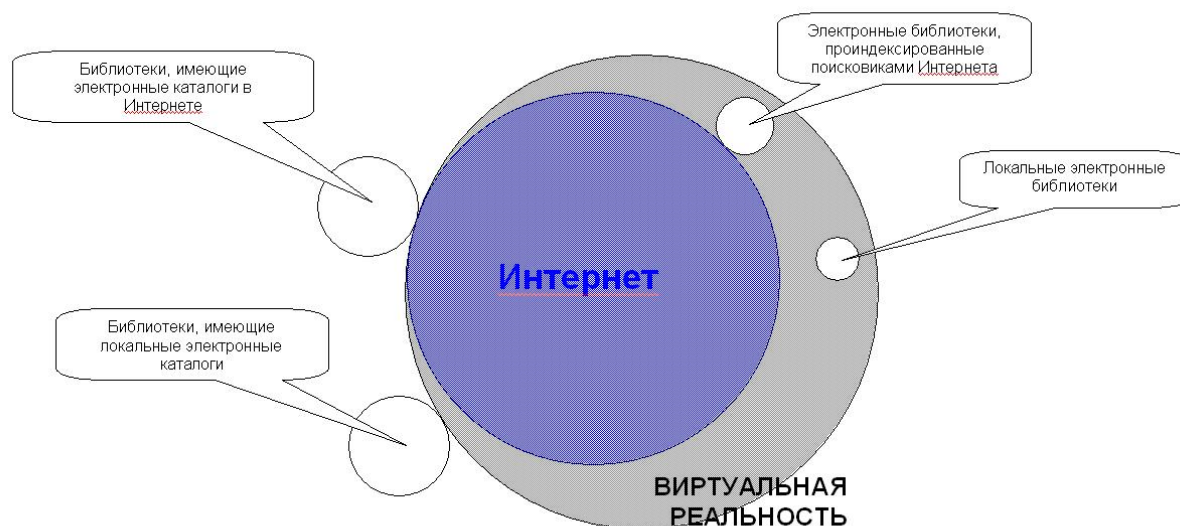


Рис. 3.

«Библиотеки, имеющие электронные каталоги в Интернете» (Рис. 3.) также соприкасаются с подпространством Интернета в одной точке, но информационное пространство такого рода библиотек находится за пределами Виртуального информационного пространства (Виртуальной реальности). Удаленный пользователь может получить только информацию о том, что в некоторой библиотеке имеется интересующая его информация, но ни о каком удаленном доступе в данном случае не может идти и речи.

«Библиотеки, имеющие локальные электронные каталоги» (Рис. 3.) также как и «Локальные электронные библиотеки» образуют отдельные информационные подпространства, не связанные как с Интернетом, так и друг с другом, но их основной информационный ресурс находится за пределами Виртуального информационного пространства.

В связи с тем, что виртуальное информационное пространство и его подпространства расширяются, требуется как-то систематизировать информацию для ускорения доступа к ней.

В различных локальных информационных подпространствах это могут быть различного рода поисковые системы, в электронных библиотеках – электронные каталоги, но все они могут обслуживать лишь свое локальное информационное подпространство.

В Интернете для поиска информации существует ряд поисковых сервисов, помогающих пользователю найти необходимую информацию, но по ряду причин поиск они осуществляют далеко не идеально (по ряду как объективных, так и субъективных причин). Чтобы хоть как-то облегчить процесс поиска в поисковиках Интернета, пользователю необходимо знать принципы работы поисковых систем или, другими словами, знать, как нужно искать информацию в Интернете.

Как же надо искать информацию в Интернете?

Поиск в Интернете надо осуществлять, используя какую-либо поисковую систему.

А что такое поисковая система? Ответ очевиден: система, предназначенная для поиска информации в Интернете, называемая поисковыми серверами или сервисами. Поисковые сервисы занимаются сканированием Интернета, сохранением так называемого среза информации на конкретную дату и систематизацией данных этого среза (индексированием).

ИНДЕКСИРОВАНИЕ [< лат. index указатель; список; заголовок; описание] выражение главного содержания текста к.-л. документа в терминах языка информационно-поисковой системы (т.е. в виде индекса б); применяется для облегчения поиска нужного документа среди множества других [2].

Конечно, выбор поисковой системы важен, но для того, чтобы максимально эффективно производить поиск в Интернете, недостаточно пользоваться какой-либо одной поисковой системой по причине того, что никакая отдельно взятая поисковая система неспособна охватить все пространство Интернета целиком. Этому существует ряд причин:

1. Пространство Интернета постоянно расширяется. На поиск новых информационных ресурсов уходит достаточно много времени, многие новые информационные подпространства не регистрируют себя в поисковых системах.
2. Способы сканирования Интернета, практически у всех поисковиков различны по языковому, национальному, тематическому, по повторениям запросов или по какому-либо еще признаку.
3. Несовершенство алгоритмов поиска информации и поисковых форм. Очень велик процент так называемого информационного мусора как по вине пользователя производящего поиск, так и по вине создателей сайтов и Интернет-сканеров.
4. Порядочность создателей Интернет-сканеров. В чистом виде любой «хозяин» сайта может запретить индексирование своего сайта, но поскольку в большинстве случаев это лишь команда в файле robots.txt, то для создателей Интернет-сканеров обойти эту защиту не представляет труда.
5. Нестабильность работы Интернета. Сайты то пропадают, то появляются, отключаются по техническим и другим причинам.

Несмотря на объективные и субъективные причины, чтобы получить максимальный результат, пользователь должен знать, что пока не существует ни одного поискового сервиса Интернета, который бы смог охватить все Интернет-пространство. Если пользователь, произведя поиск, обнаружил, что поисковая система Google по запросу обнаружила 100 000 документов с результатами поиска, а поисковая система Яндекс только 2 000 документов, то это не значит, что Google нашел все то, что смог найти Яндекс.

Охват поиска поисковыми сервисами Интернет-пространства приближенно показан на Рис. 4. Конечно, количество Интернет-поисковиков в Интернет-пространстве не ограничивается представленными пятью, но для того, чтобы понять принцип организации поиска, их вполне достаточно.

Так, если искомая информация попала в зону 1 (Рис. 4), то она будет найдена всеми поисковыми системами (в данном случае Google, Rambler, Alta-Vista, Яндекс, Апорт). Если искомая информация попала в зону 2, то она будет найдена лишь некоторыми поисковиками (в данном случае Google, Alta-Vista, Яндекс).

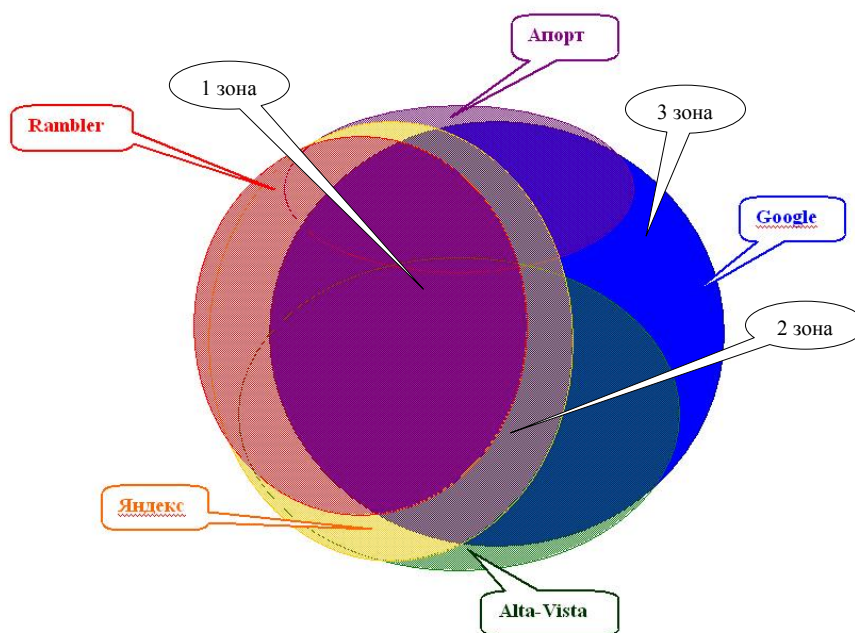


Рис. 4.

В случае, если информация попала в 3 зону, то найдена она может быть только лишь каким-либо определенным поисковиком (в данном случае поисковой системой Google).

И всегда в Интернет-пространстве существует информация, которая может быть найдена лишь каким-либо определенным поисковиком. А поскольку пользователь в большинстве случаев ищет информацию наугад, то он не может заранее знать, в зону поиска какого Интернет-сервиса она попадет. Следовательно, пользователь должен производить поиск в как можно большем количестве Интернет-поисковиков.

Каждый Интернет-поисковик обладает своим синтаксисом поиска, а также своими свойствами расширенного поиска. Чем больше свойств различных поисковых систем Интернета знает и использует пользователь, тем с большей вероятностью он найдет интересующую его информацию.

Подведем итоги

Чтобы продуктивно искать информацию в Интернете необходимо:

1. знать адреса как можно большего количества поисковых сервисов Интернета;
2. не ограничиваться одним поисковым сервисом;
3. знать синтаксис поиска различных поисковых сервисов;
4. правильно формулировать поисковый запрос, использовать синонимы при неудачном поиске;
5. учитывать языковой аспект поиска;
6. использовать опции расширенного поиска.

Список использованных источников

1. Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка: 80 000 слов и фразеологических выражений – 4-е изд., дополненное, – М.: Азбуковник, 1999. – 944 с.
2. Новейший словарь иностранных слов и выражений. – М.: Современный литератор, 2003. – 976 с.
3. Политехнический словарь / Редкол.: А.Ю. Ишлинский (гл. ред.) и др. – 3е изд., перераб. и доп. – М.: Советская энциклопедия, 1989. – 656 с. – ил.
4. Толковый словарь по информатике / В.И. Першиков, В.М. Савинков. – М.: Финансы и статистика, 1991. – 543 с.
5. Словарь по кибернетике / под. ред. В.С. Михалевича. – 2-е изд. – Киев: Гл. ред. УСЭ им. М.П. Бажана, 1989. – 751 с.
6. Борковский А.Б. Англо-русский словарь по программированию и информатике (с толкованиями): Ок. 6000 терминов / А.Б. Борковский. – М.: Московская международная школа переводчиков, 1992. – 335 с.