УДК 519.62

# СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ РАСПРЕДЕЛЕННЫЕ СИСТЕМЫ БАЗ ЗНАНИЙ

**Т. Б. Шатовская[1], С. П. Менинник[2]**

[1] ХНУРЭ, г. Харьков, Украина;

[2] ХНУРЭ, г. Харьков, Украина, stas_sam@rambler.ru;

Проведен разноуровневый анализ информационных систем баз знаний. Проанализированы первоначальные предпосылки и концепции создания с дальнейшим анализом динамики и истории развития до современного состояния. Описание и аналитика взаимного соотношения поисковых систем Интернет на рынке на текущее время.

БАЗА ЗНАНИЙ, ИНТЕРНЕТ, ПОИСК, БИЛЛИНГ.

## Введение

На современном этапе развития нового постиндустриального общества, в условиях постоянно и стремительно изменяющейся внешней среды особое значение приобретает развитие компьютерных технологий, неотъемлемой частью которых являются информационные системы, давно и прочно вошедшие в нашу повседневную жизнь. Трудно представить себе какую-либо сферу, в которую не проникли бы информационные технологии. Они применяются в сфере транспорта, культуры, медицины, образования и, в частности, в библиотекарском деле. Определяющим фактором эффективности систем обработки и взаимодействия необходимой информации является использование информационных баз данных с целью облегчения повседневного труда человека, частичной замены или переквалифицирования в оператора персонального компьютера, что приведет к минимизации времени, затрачиваемого им для обработки, поиска и сортировки данных.

Исходя из вышеизложенного, тема применения в области документооборота информационных систем является в наше время невероятно актуальной. Действительно, для работы с любой литературой требуется, в первую очередь, чтобы человек обладал отличной памятью для запоминания местонахождения той или иной литературы, фиксации хотя бы относительного содержания для повышения скорости и эффективности рабочего процесса. Понятно, что сохранение на бумажном носителе не является таким надежным, как сохранение данных в информационной системе, а эффективность зависит в основном от скорости предоставления информации. Поэтому очевидна назревшая необходимость в переходе общества к компьютерным информационным системам. Таким образом, в представляемой информационной системе показываются преимущества, связанные с хранением документов (информации), обеспечением возможности легкого быстродоступа, рассматриваются вопросы эффективности хранения и обработки информации в распределенных системах баз знаний. Также стоит отметить важность в современном мире применения более инновационных технологий и степень обширности поставленной задачи. Ввиду универсальности системы в ней могут и должны использоваться лингвистические технологии, математическая психология, методы информатизации, технологии поиска, на которых отдельно остановимся и рассмотрим на более детальном примере [4].

Цель данной статьи — разъяснить проблематику и методику формирования эффективных информационных систем баз данных как всеобъемлющего механизма с множеством возможностей в современном мире.

Задачей, решаемой в данной статье, является разносторонний анализ современных информационных систем и аналитические предсказания значимости, темпов развития, эффективности их в будущем. Также рассматриваются некоторые концепции защиты информации, поддержка биллинг-системы.

С целью наглядного демонстрирования некоторых аспектов вопроса часто используется унифицированный язык моделирования UML. Он позволяет проектировать варианты использования системы, диаграммы для визуализации функциональных возможностей системы в целом и отдельные модули [5].

## 1. Анализ предметной области

С развитием науки, компьютерных и информационных технологий возникает необходимость в получении все большего количества информации. В современном мире человеку нужно получать как можно больше информации за короткий промежуток времени, поэтому возникает необходимость постоянного присутствия источников информации у себя «под рукой», а также возможность быстрого поиска информации. Долгие годы развития технологий, неутихающие споры о безопасности для здоровья человека, споры о моральном обличии нового мира — вот тот небольшой список проблем из истории, которые предшествовали появлению на свет

современного электронного документооборота. Печатные источники не могут дать таких возможностей, поэтому на смену им пришли электронные издания. Электронные издания могут распространяться посредством электронных носителей информации и компьютерных сетей, количество экземпляров при этом не является ограниченным. Также электронные издания не подвержены повреждениям, поскольку на них не влияет физический фактор, а при необходимости их можно вывести на бумагу. Бесспорным преимуществом электронных изданий является и то, что для их хранения не требуется огромного пространства, так как это происходит в случае традиционных библиотек. Актуальность системы является неоспоримой, поскольку развитие электронных изданий имеет больше преимуществ по сравнению с печатными изданиями, но, к сожалению, не может полностью заменить их.

Данная система призвана помочь многим предприятиям и другим организациям, а также просто отдельным людям в документообороте, получении важной информации. Эффективный поиск реальной информации, хранение большой базы ссылок на внешние и внутренние источники данных — все это дает безусловные преимущества и очевидные выгоды в современном мире электронных технологий.

Существуют системы различного уровня сложности и покрытия, глобальной системой на сегодняшний день является сеть Интернет, о которой и пойдет речь в статье с точки зрения истории, причин возникновения и эффективного использования в качестве универсального обменника полезной информацией. Интернет многообразен и обладает множеством как отрицательных, так положительных свойств, на которых и будем акцентировать внимание.

Основными проблемами данной предметной области являются практическая ненадежность печатной продукции, большой объем для хранения, потребность в посреднике для доступа к книгам. В данном случае возникает необходимость быстрого и эффективного доступа и поиска, возможность сортировки по многим параметрам.

Предметной областью этой статьи являются физическая деятельность, документооборот в какой-либо его форме, распределенные системы баз знаний, а именно: предоставление прямых ссылок на издания или информацию сколь угодно различной тематики, их сортировка и систематизация для более удобного доступа. Анализируя предметную область, оговорим некоторые ограничения и условности модели для рассматриваемой системы. Любое информативное издание, любую информацию будем условно называть книгой — термин, в который вкладывается смысл любого вида текстового, графического и других видов содержания. Книга может фактически относиться к многим разделам науки, и конкретные направления имеют множество книг по своей тематике [4].

При анализе предметной области были рассмотрены основные проблемы, связанные с особенностями инвентаризации литературы в каталоги разных разделов, а также возможности эффективной автоматизации получения требуемой информации, возможность и разновидности доступа к информации с разных физических носителей и серверов, что привело к потребности в биллинг-системе при доступе к данным непосредственно с самого сервера [7].

Анализ предметной области позволил выделить основные понятия (книга, автор, издательство, раздел) и смысловые связи между ними в виде простейшей концептуальной модели предметной области, приведенной на рис. 1. Под книгой, как было оговорено выше, на данном рисунке стоит понимать любую информацию в электронном виде.

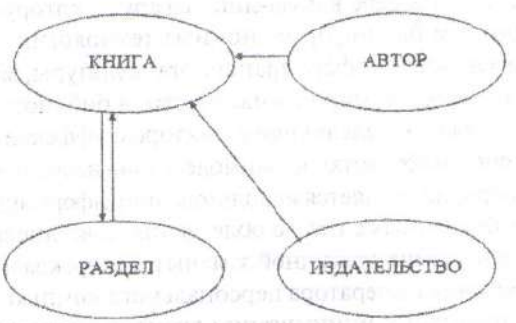


**Рис. 1.** Простейшая концептуальная модель предметной области

## 2. Анализ существующей глобальной системы

Для того, чтобы понять темпы развития, а главное, темпы ускорения развития современных баз знаний, приведу некоторые факты из истории создания глобальной информационной распределенной системы Интернет. Данные факты являются общедоступными и исторически логичными в стремлении человека к информатизации общества.

1967 год. Ларри Робертс (Larry Roberts), практик, воплощающий в жизнь теоретические идеи Ликлидера, предлагает связать между собой компьютеры ARPA. Начинается работа над созданием ARPANET.

1969 год. ARPANET заработал. К нему подключаются компьютеры ведущих, в том числе и невоенных, лабораторий и исследовательских центров США.

1971 год. Рэй Томлисон (Ray Tomlison), программист из компьютерной фирмы Bolt Beranek and

Newman, разрабатывает систему электронной почты и предлагает использовать значок @ («собака»).

1974 год. Открыта первая коммерческая версия ARPANET — сеть Telenet.

1976 год. Роберт Меткалф (Robert Metcalfe), сотрудник исследовательской лаборатории компании Xerox, создает Ethernet — первую локальную компьютерную сеть.

1977 год. Число хостов достигло одной сотни. 1980 год. Писатель и политический аналитик Алвин Тоффлер (Alvin Toffler) опубликовал книгу «Третья Волна» (The Third Wave), в которой описал постиндустриальный мир, где «первую скрипку» играют информационные технологии. Тоффлер, в частности, сумел оценить перспективы развития компьютерных сетей и предположил, что однажды такая сеть сможет объединить весь мир, наподобие того, как все обладатели телевизоров могут смотреть одну и ту же передачу. При этом компьютерная сеть, по прогнозу Тоффлера, даст людям несравненно больше возможностей, чем обычное ТВ.

1982 год. Рождение современного Интернета. ARPA создала единый сетевой язык TCP/IP.

1984 год. Число хостов превысило тысячу. И это всего навсего за 15 лет. Сравните темпы роста. За первые 8 лет рост составил 100 хостов, тогда как за последующие 7 лет — уже 900 хостов, что явно указывает на огромную динамику развития отрасли.

1986 год. Национальный Фонд Науки США (The National Science Foumdation) создал NSFNET, связавшую центры с «суперкомпьютерами». Эта сеть доступна лишь для зарегистрированных пользователей, в основном, для университетов.

1989 год. Число хостов превысило 10 тысяч. Т. е. за 5 лет рост составил 9000, сохраняя и увеличивая темпы.

1991 год. Европейская физическая лаборатория CERN создала известный всем протокол — www — World Wide Web. Эта разработка была сделана, прежде всего, для обмена информацией среди физиков. Появляются первые компьютерные вирусы, распространяемые через Интернет.

1993 год. Создан первый интернет-броузер Mosaic, созданный Марком Андреесеном (Marc Andreesen) в Университете штата Иллинойс (University of Illinois). Число интернет-хостов превысило 2 млн., в Сети действует 600 сайтов.

1996 год. Началось соревнование между браузерами Netscape, созданным под руководством Марка Андреесона, и Internet Explorer, разработанным компанией Microsoft. В мире существует 12,8 млн. хостов и 500 тыс. сайтов.

1999 год. Впервые предпринята попытка цензуры Интернета (популярен принцип: «Интернет никому не принадлежит»). В ряде стран (Китай, Саудовская Аравия, Иран, Египет, страны бывшего СССР) государственными органами предприняты серьезные усилия, чтобы технически блокировать доступ пользователей к определенным серверам и сайтам политического, религиозного или порнографического характера.

2002 год. Сеть Интернет связывает 689 млн. человек и 172 млн. хостов. Разрабатываются новые технологии Интернета, которые должны заменить «старый Интернет», расширить его функции или создать национальные компьютерные сети [1].

**Таблица 1**

| Период, гг. | Количество хостов в конце периода | Темп роста в среднем за год |
|---|---|---|
| 1969 | 0 | – |
| 1969–1977 | 100 | – |
| 1977–1984 | 1000 | 128 % |
| 1984–1989 | 10000 | 180 % |
| 1989–1993 | 2 млн | 4500 % |
| 1993–2002 | 172 млн | 860 % |

Как можно заметить, где-то в период 1989–1993 гг. в мире разразился информационный бум с соответствующими последствиями для истории и всей современной жизни. В целом сегодня темпы роста немного замедляются ввиду того, что система фактически охватила большинство сфер жизни в развитых странах, и единственным мощным источником развития и роста сети Интернет как мощнейшей информационной распределенной системы баз знаний являются слаборазвитые страны. Также стоит отметить, что на базе Интернет существует возможность создания полнофункционального искусственного интеллекта из-за слабой вычислительной способности одной единицы вычислительной техники, но весьма внушительной мощности миллиардной армии вычислительных машин по всему миру.

Существует огромная проблематика засоренности и высокого процента бесполезности информации из всемирной глобальной сети, изначально созданной для научно-исследовательских потребностей. Анализируя современное состояние сети Интернет, можно с уверенностью сказать, что она перестала быть научным инструментом в той мере, в которой планировалось, и больше похожа на зону коммерческих войн. Также следует развеять миф о том, что можно найти абсолютно все. По-настоящему интересной и нужной информации скорее всего вы не найдете по множеству причин. Одна из них — эффективность поиска, что составляет огромную проблематику в современном мире. Проблема именно в том, чтобы быстро найти требуемые данные, по пути отсеяв всю рекламу, мусор и другие,

возможно, не имеющие фактического отношения к тематике или разделу, но маскирующиеся под них с целью получения более высоких рейтингов. Вся глобальная система состоит из множества электронных документов как статических, так и динамических.

Интернет – это не только собрание технологий, но и собрание сообществ. Успехи Интернет в значительной степени объясняются как способностью удовлетворить основные социальные потребности, так и умением эффективно использовать общественность для развития инфраструктуры.

Дух коллективизма, содружества в Интернет имеет глубокие корни, как было описано выше, он зародился еще в начале работ над ARPANET. И, развившись до наших дней, дух коллективизма сформировал современное информационное сообщество. В связи с чем еще одной особенностью является то, что глобальные информационные системы превращаются в универсальные коммуникаторы современного общества по отношению его к себе самому и новым технологиям [2].

## 2. Биллинг-системы

Однако сама система Интернет не является самодостаточной и опирается на материальную основу. Фактически вся глобальная сеть — это связанные между собой различные вычислительные устройства, которые могут ломаться, потреблять электроэнергию и другие расходные материалы, что приводит к требованиям финансовой безубыточности. Даже простая передача простого сигнала по кабелю уже заложена в стоимость услуг системы. С большим развитием информационной инфраструктуры данные затраты явно теряют свою актуальность с приносимой пользой и, как результат, чей-то конечный доход. В случае, когда система слабо развита, существует необходимость оплаты расходов на физическое поддержание. Одним из таких механизмов является биллинговая система.

*Биллинговая система* — программный комплекс, осуществляющий учет объема потребляемых абонентами услуг, расчет и списание денежных средств в соответствии с тарифами компании.

Исходя из задач и запросов бизнеса, можно набросать схему системы. Чтобы не обсуждать какого-то абстрактного сферического коня в вакууме, будем рассматривать типовой пример оператора связи, продающего трафик абонентам.

Услуги могут быть разными (например — VPN-доступ, dial-up пул, обычный неинкапсулированный трафик, Proxy, VoIP etc), надо обеспечить доставку ядру в единообразном виде информации о том, какой тип услуги, какой абонент, в каком объеме и в какое время потребил. В худшем случае для каждого из типов услуг прийдется разрабатывать свой коллектор, но, если повезет, что-то удастся унифицировать. Технологии, которые могут помочь при создании коллекторов, — SNMP, Radius, NetFlow.

Принятие решения о блокировке абонента при окончании средств на его счету на практике происходит не мгновенно, этот факт тоже надо учитывать. Например, если блокировка срабатывает раз в минуту — при скорости соединения 1 Мбит/с абонент может скачать лишних 7,5 мегабайт в худшем случае. Данная система учета финансовых средств пока что актуальна для Украины в частности, но медленно сдает свои позиции в пользу безучетных услуг передачи информации за единый установленный платеж, который уже косвенно ложится как бремя на конечного пользователя информационной системы баз данных в виде отдельного платного контента или в иной форме. Также существуют варианты с физической за счет пожертвований и на общественных началах, которые изначально не несут в себе серьезных финансовых ресурсов [6, 7].

## 3. Представители прогрессивных поисковых систем Интернет

Для начала скажем, что современные поисковые системы базируются на финансовой основе, применяют самые современные методики и технологии. Та же лингвистическая технология давно используется в поисковых машинах. Применение новаторских идей отдельных людей в свое время привело таких гигантов, как Google и другие, к вершине сегодня. Самые прогрессивные являются самыми удачными, успешными и как результат — доходными.

В настоящее время рынок поиска уже достаточно развит, и мы можем смело выделить основных игроков (а их не так уж и много). Львиная доля всего мирового поиска приходится на такие поисковые системы: Google Search, Yahoo! Search и Microsoft Live Search. Эти три американские компании конкурируют между собой в борьбе за главный приз, то есть пользователей, которые пришли за помощью в Интернет.

Чтобы оценить доли присутствия этих и других компаний, было бы неплохо, если бы поисковые системы сами давали подробные отчеты о своей работе в свободный доступ, однако это маловероятно, так как подобная информация является коммерческой тайной.

Но все-таки существуют компании, которые как раз и занимаются сбором статистики по подобным вопросам. Наиболее авторитетной в этом плане является компания NetRatings, Inc. Именно она постоянно информирует Интернет-сообщество о результатах своих исследований в областях интернет-коммерции, маркетинга, онлайн-рекламы и поисковых технологий.

По данным последнего опубликованного отчета об объемах поисковых запросов за декабрь 2006 года, на американском рынке мы можем наблюдать ситуацию, отображенную в табл. 2.

**Таблица 2**

| | |
|---|---|
| Google Search | 50,80% |
| Yahoo! Search | 23,60% |
| MSN/Windows Live Search | 8,40% |
| AOL Search | 6,10% |
| My Way Search | 2,40% |
| Ask.com Search | 2,10% |
| EarthLink Search | 0,50% |
| Others | 6,10% |

Судя по этой статистике, сейчас более 80% американского рынка поиска делят между собой те самые три компании, о которых мы говорили ранее. Оставшиеся 20% распределились между другими менее популярными поисковыми системами и порталами. Некоторые из них используют поисковые технологии первых трех компаний. Даже с учетом того, что, казалось бы, 20% рынка — это тоже очень много, какого-то серьезного игрока, который смог бы отобрать себе этот кусок, на сегодняшний день пока не наблюдается.

Поисковые системы Рунета — так называемой области Интернет русскоязычного населения мира, имеют свое место.

Назовем наиболее популярные из них: Yandex, Mail.ru (на данный момент транслируют выдачу Yandex), Rambler, Google, Meta.ua, Bigmir.net (на данный момент транслируют выдачу Yandex).

Можно назвать еще некоторых представителей, таких как Webalta или Aport. Но о Webalta пока нет смысла говорить, так как эта поисковая система лишь на заре своей деятельности, и ее будущее доселе неизвестно. Вторая же уже угасает [3].

Оценить эффективность мы можем по количеству передаваемого поискового трафика (по России). Поисковые системы можно расположить так, как указано на рис. 2.

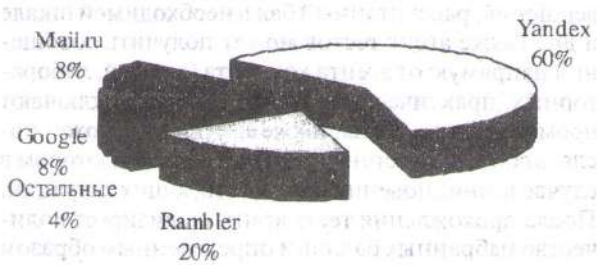


**Рис. 2.** Количество передаваемого поискового трафика

В целом, как видно из рисунка, лидирующее место занимает система Yandex. Стоит отметить, что она имеет свою собственную уникальную методику поиска и ранжирования необходимой информации.

Статистика по Украине показана на рис. 3.

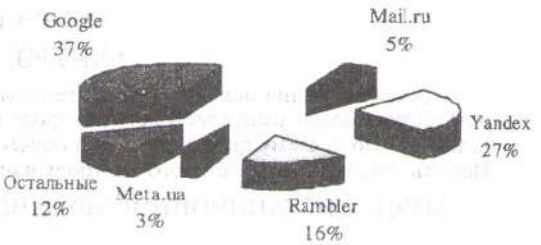


**Рис. 3.** Количество передаваемого поискового трафика по Украине

**Выводы**

Самой современной информационной распределенной системой баз знаний является сеть Интернет, которая уже далеко ушла и от первоначальной задачи документооборота, и от первоначальной концепции в качестве простого электронного хранилища данных, перейдя от простой сети немногочисленных вычислительных машин, созданной в научных целях, до обширной системы с невероятно широкой областью применения. Это система, растущая и развивающаяся огромными темпами, которые не замедляются даже в современном мире. Это современная система, она постоянно эволюционирует и вбирает в себя все новые технологии в области лингвистики, статистики, искусственного интеллекта и др. По результатам моих исследований информационная система баз знаний Интернет, обросшая мощными инструментами эффективного поиска запрашиваемой информации, займет главенствующие позиции во всех сферах жизни.

**Список литературы:** 1. *Гусев В.С.* «Освоение Internet. Краткое руководство». — Диалектика, 2005. — 288 с. 2. *Гусев В.С.* «Internet: учеба, работа, полезные ресурсы. Краткое руководство». — Диалектика, 2005. — 256 с. 3. *Гусев В.С.* «Google: эффективный поиск информации в Интернет. Краткое руководство». — Диалектика, 2005. — 240 с. 4. *Клушанов С.В., Ламотько Д.В.* «Базы данных» — 2002. 5. Rational Unified Process Made Easy: A Practitioner's Guide to the RUP, The By Per Kroll, Philippe Kruchten Publisher Addison Wesley, Pub Date April 11, 2003, ISBN 0-321-16609-4, Pages 464. 6. *Tom Kait* «Oracle expert one on one». 7. *Turganov A.G.* Some Problems of Corporative Information Systems Design // Proc. of the 6th International Workshop on Computer Science and Information Technologies CSIT'2005, Ufa, Russia, 2005, pp. Pages 275.

***Поступила в редколлегию 16.05.07***