

УДК 69

DOI 10.21661/r-466479

П.А. Капитан, И.В. Чугунков

РАЗРАБОТКА МЕТОДОВ И СРЕДСТВ АНАЛИЗА ДАННЫХ НА ОСНОВЕ ОТКРЫТЫХ ИСТОЧНИКОВ

Аннотация: в данной статье рассматривается проблема значимости открытых источников информации, а также создание консолидированной информационно-аналитической системы хранения данных, приведена классификация систем управления базами данных. По мнению авторов, в современном мире нельзя недооценивать значимость информации в нашей жизни.

Ключевые слова: информация, информационные системы, системы управления базами данных, базы данных, информационный поиск.

P.A. Kapitan, I.V. Chugunkov

DEVELOPMENT OF METHODS AND MEANS OF DATA ANALYSIS BASED ON OPEN SOURCES

Abstract: this article is about the importance of open sources of information, and about creation of a consolidated information storage system. Also, the article contains a classification of database management systems. According to the authors, these days you cannot underestimate the importance of information in our lives.

Keywords: information, information systems, database management systems, databases, information retrieval.

В современном мире очень сложно представить жизнь без сети Интернет и высоких технологий. Всё больше и больше информации выкладывается в общедоступные сети и накапливается там. В таком огромном количестве различного рода информации, представленной в различных видах, форматах становится все тяжелее отсеивать и находить необходимую для своих нужд информацию. Создание поисковых машин во многом облегчило задачу поиска в просторах сети Интернет информации, отвечающей заданным условиям поиска.

Информационно-поисковые системы работают по не меняющемуся, установленному алгоритму поиска информации.

Зачастую многим компаниям и организациям приходится работать с большими массивами информации, которые через определенное количество времени обновляются и дополняются. Для каких-либо определенных нужд, работниками компании или организации приходится искать нужную, для их работы информацию в таких огромных массивах. Причем такое может встретиться в самых разнообразных сферах деятельности. В связи с этим возникает необходимость создания большого количества разнообразных хранилищ данных и информации для разного назначения, а также информационные системы, отвечающие всем современным требованиям. В основе информационных систем, как мы знаем, лежит создание базы данных, содержащей сведения и информацию о конкретных объектах реального мира в зависимости от цели создания информационной системы и ее предназначения в заданной предметной области.

Примеров можно привести достаточно много. Допустим у компаний, занимающейся реализацией и продажей мебели есть задача повысить продажи в текущем квартале. Нужно осуществить мониторинг цен на рынке, и ценовой анализ конкурентов. Обычно такую информацию можно найти в открытом доступе, но осуществлять поиск вручную очень долго, и может не привести к желаемому результату. В связи с этим у компании возникает необходимость создания информационно-поисковой системы для внутреннего пользования, в функциях которой присутствовала бы возможность периодического мониторинга цен в нужном секторе. Этого бы можно было добиться, разработав алгоритм программы с заданными условиями поиска в сети Интернет. У компании есть ряд наиболее весомых конкурентов на рынке, соответственно информацию об их коммерческих предложениях и ценах можно найти на их официальных сайтах. Значит можно написать алгоритм программы, которая бы могла собирать нужную информацию о продукции и ценах из списка конкурирующих компаний в базу данных, проходя при этом процесс чистки и структуризации.

Также при этом необходимо иметь систему управления базой данных, для того чтобы пользователь мог работать с ней. Аналитику проще работать с удобной в эксплуатации системой управления базой данных. Соответственно, имея такую систему, аналитик может достаточно за короткие промежутки времени делать анализ цен и разрабатывать стратегии повышения продаж на определенные периоды на основе анализа по информации о ценах и коммерческих предложениях конкурирующих компаний.

Для того чтобы написать такую программу необходимо определить критерии информационного поиска, а также список адресов электронных ресурсов, на которых будет производиться поиск. Под информационным поиском понимается такая последовательность действий / операций, которые выполняются для достижения цели по поиску какой-либо информации или определенных документов, необходимых для решения задачи.

Информационный поиск, включает в себя следующие этапы:

- формулирование информационного поиска;
- определение ряда необходимых информационных источников;
- поиск нужной информации и извлечение ее из источников;
- анализ и оценка полученного результата.

После реализации информационного поиска и процесса стандартизации и структуризации полученных данных можно подумать о создании базы данных. Далее нужно выбрать нужную систему управления базой данных для конкретного случая. Таким образом, на Рисунке 1 представлена схема последовательности работы пользователя.

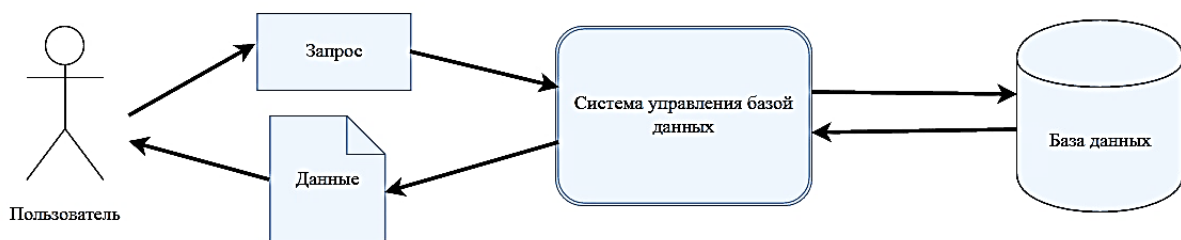


Рис. 1. Схема последовательности работы пользователя с созданной БД

Системы управления базами данных можно классифицировать по типу модели данных следующим образом:

- иерархические СУБД (данные распределяются по иерархической структуре, имеются различные уровни);

- сетевые СУБД (состоит из различных наборов экземпляров определенного типа записей и наборов экземпляров конкретного типа связей между этими самыми записями);

- реляционные СУБД (данные представляются в виде таблиц, которые состоящих из строк и столбцов, между ними могут быть связи вида «один-к-одному» или «один-ко-многим»);

- объектно-ориентированные СУБД (данные представляются как абстрактные объекты с различными свойствами и использующие методы для взаимодействия с другими объектами БД);

- объектно-реляционные (более расширенные реляционные СУБД, где есть возможность повторного и совместного использования компонентов).

От выбора системы управления базой данных зависит скорость и качество работы пользователя. От того насколько адаптирована система зависит ряд нюансов. Например, необходимость принятия на работу человека, обладающего специальными знаниями работы с системами управления базами данных, умение написания поисковых запросов, знание языка SQL и т. д. В таблице, приведенной ниже, представлены самые популярные системы управления базами данных на сегодняшний день.

Таблица 1

Список самых популярных систем управления баз данными

№	Название	Разработчик	Год выпуска
1	MySQL	Oracle	1995
2	PostgreSQL	Сообщество PostgreSQL	1995
3	MS SQLServer	Microsoft	1988
4	SQL Lite	Richard Hipp	2000
5	Oracle Database	Oracle	1979

Можно ограничиться той информационной системой, что была создана, состоящей из программы, осуществляющей информационный поиск, базой данной, получившейся в результате программы и выбранной системой управления базой данных, но для более упрощенного варианта можно также разработать пользовательский интерфейс к созданной базе данных. Это следует сделать в том случае, если нет необходимости нанимать на работу квалифицированного специалиста, обладающего навыками обращения с системами управления базам данных и написанием поисковых вопросов.

Таким образом, из информации, находящейся в открытом доступе можно сделать удобную консолидированную информационную систему хранения данных для внутреннего пользования.

Если взять во внимание все вышесказанное, можно понять что из информации, которая находится в открытом доступе, а именно в сети Интернет удобно собирать и создавать, при необходимости для реализации определенных целей, информационные системы полезные компаниям для своих нужд.

Заключение. В эпоху современного времени нельзя недооценивать значимость информации в нашей жизни. В связи с этим возникает потребность в ее сборе, формализации, правильном хранении, обработке и анализе. Все эти действия могут быть реализованы в зависимости от целей и потребностей компаний и организаций, для повышения качества работы и получения прибыли.

Список литературы

1. Учебное пособие «Системы управления базами данных» / М.В. Баканов, В.В. Романова, Т.П. Крюкова [Электронный ресурс]. – Режим доступа: <http://textarchive.ru/c-1903056-pall.html>
2. Учебное пособие «Базы данных» / В. Швецов.
3. Фуфаев Э.В. Базы данных. – 7-е изд. / Э.В. Фуфаев, Д.Э. Фуфаев [Электронный ресурс]. – Режим доступа: http://www.proklondike.com/books/dbobshee/fufaevi_DB_2012.html

4. Статья «Информационно-поисковые системы (ИПС) и их виды» Р. Чуешов [Электронный ресурс]. – Режим доступа: <http://romanchueshov.ru/raznoe/informatsionno-poiskovyie-sistemi.html>

References

1. Bakanov, M. V., Romanova, V. V., & Kriukova, T. P. Uchebnoe posobie "Sistemy upravleniia bazami dannykh". Retrieved from <http://textarchive.ru/c-1903056-pall.html>
2. Shvetsov, V. Uchebnoe posobie "Bazy dannykh".
3. Fufaev, E. V., & Fufaev, D. E. Bazy dannykh. Retrieved from http://www.proklondike.com/books/dbobshee/fufaevi_DB_2012.html
4. Stat'ia "Informatsionno-poiskovyie sistemy (IPS) i ikh vidy" R. Chueshov. <http://romanchueshov.ru/raznoe/informatsionno-poiskovyie-sistemi.html>

Капитан Полина Александровна – студентка ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ», Россия, Москва.

Kapitan Polina Alexandrovna – student at National Research Nuclear University MEPHI, Russia, Moscow.

Чугунков Илья Владимирович – канд. техн. наук, доцент, преподаватель ФГАОУ ВО «Национальный исследовательский ядерный университет «МИФИ», Россия, Москва.

Chugunkov Ilya Vladimirovich – PhD in technical sciences, associate professor, lecturer at National Research Nuclear University MEPHI, Russia, Moscow.
