Nowadays, modern organizations usually use database management system or information retrieval system in order to store and retrieve their information and data efficiently. Both systems have their own advantage and disadvantage. Database Systems work with relational database but Information Retrieval System works on unstructured data.

The main data structure was used in Database System is relational tables with well define values for each row and column. And IR system uses inverted index as their main data structure. Inverted index is the index of entries for example term and docIDs. In each term there is a corresponding postings list or the list of
documents in which the term is present. Database systems also work on data that are related to each other and have a well defined domain. However IR systems may or may not have such a thing. Database Management System has functions like Data Dictionary Management, Data Storage Management, Data Transformation and Presentation, Security Management, Multiuser Access Control, Backup and Recovery Management, Data Integrity Management, Database Access Languages and Application Programming Interfaces, Database Communication Interfaces and Transaction Management.

The Data Dictionary Management is where the DBMS stores definitions of the data elements and their relationships or metadata. DBMS uses the Data Dictionary Function to look up the required data component structures and relationships. When a program access data in a database they are basically going through the DBMS itself. This function can removes structural and data dependency, then provides the user with data abstraction. This makes things a lot easier for the end user. The Data Dictionary Management is often hidden from the user and is used by Database Administrators and Programmers. Meanwhile, Data Storage Management is used for the storage of data and any related data, such as entry forms or screen definitions, report definitions, data validation rules, procedural code and structures that can handle video and picture formats. However, users usually do not need to know how data is stored and manipulated. Also there is a term called performance tuning that relates to a database’s efficiency in relation to storage and access speed in this structure. It’s make the storage and access process of data much faster and efficient than usual. This is very important function in Data Storage Management, if we want to manage millions of data that is in storage.

The Security Management function is one of the most important functions in DBMS. Security Management can sets rules that determine specific users that are allowed to access the database. These users will have a special username and password. The Security Management function also can sets restraints on what specific data user can see and manage. For the Data Transformation and Presentation, its main existing reason is to transform any data that are entered or key in into the required data structures. By using this function the DBMS can determine the difference between logical and physical data formats. Like Security Management, Multiuser Access Control is the very useful tool in DBMS too. Its basis function is data integrity and data consistency. The Multiuser Access Control can enable multiple users to access the database simultaneously without affecting the integrity of the database. On the same time, Backup and Recovery Management is brought to mind because there is potential outside threats to a database that can damage the DBMS. Especially when there is power blackout. Recovery management is refers to the process to recover the database after the damage occur and backup management is a process to save the data and the integrity before the threats occur. Example, backing up the data on external hard disk and when the damaged occur, we can used it to replace the data that was damaged or lost due to the power shortage. Meanwhile the Data Integrity Management is enforces to reduce the data redundacy and maximizing data consistency. Data redundancy occurred when data is stored in more than one place unnecessarily. According to Dewan Bahasa dan Pustaka, data consistency is an error arising from the use of statements, data, or instructions that are not consistent in the form of a statement that waited or not in the required format. Data Integrity Management will make sure the database is returning the correct or same answer each time for same question asked and avoid any data redundancy and data consistency. The Database Access Languages and Application Programming Interfaces is other important functions for DBMS. Database access language or query language is a nonprocedural language such as SQL (Structured Query Language). It is the most common query language supported by the majority of DBMS vendors nowadays. Using this language, user can easily specify what they want done without the need to think and explain on how to specifically do it. And Application Programming Interfaces or API can define a set of functionalities that are independent of their respective implementation. API then will allow both definition and implementation to vary without compromising each other.

And the Database Communication Interfaces is referring to how DBMS can accept different end user requests through different network environments. For example, a DBMS will provide access to the database using the internet which is through web browser like Internet Explorer or Google Chrome. Lastly, Transaction Management which refers to how a DBMS provide a method that will guarantee all the updates in a given transaction are made or not made. No matter what and how the method of transaction is, it’s must follow what is called the ACID properties. ACID is acronym for Atomicity, Consistency, Isolation and Durability.
Atomicity is a state when a transaction is an indivisible unit that is either performed as a whole and not by its parts, or not performed at all. The consistency is mean that a transaction must alter the database from one constant state o another. While isolation means that a transaction must be executed independently of one another. The part of the transaction in progress also must not be able to seen by another transaction. And durability mean a successfully completed transaction which must be recorded permanently in the database and must not be lost due to failures.

In contrast information retrieval (IR) systems use a simpler data model than database system which will organized information as a collection of documents. And the documents usually unstructured and has no schema. Information retrieval will locates relevant documents, on the basis of user input such as keywords or example of documents. For the example, to find documents containing the words ‘database systems’. IR can be used even on textual descriptions provided with non??textual data such as images. And web search engines are the most familiar example of IR systems.

Other differences between IR and DBMS from database systems is, IR systems don’t deal with transactional updates, including concurrency control and recovery. But database systems deal with structured data and with schemas that define the data organization. Its also deal with some querying issues not generally addressed by database systems. IR usually does approximate searching by keywords and then ranking the retrieved answers by estimated degree of relevance.

In full text retrieval, all the words in each document are considered to be keywords search. Information retrieval systems typically allow query expressions formed using keywords and the logical connectives and, or, and not. Ands are implicit, even if not explicitly specified. The ranking of documents on the basis of estimated relevance to a query is critical. The relevance ranking using term is based on factors such as term frequency. And term frequency is a frequency of occurrence for query keyword in document. The similarity based retrieval function of IR on other hand, can retrieve documents which similar to a given document and can be used to refine answer set to keyword query. For example when user selects a few relevant documents from those retrieved by keyword query, and system finds other documents similar.

Meanwhile web search engine like web crawlers are programs can locate and gather information on the web. Its recursively follow hyperlinks present in known documents, to find other documents. Also information retrieval systems originally treated documents as a collection of words. Information extraction systems usually infer structure from documents such as extraction of house attributes like size, address, number of bedrooms from a text advertisement. Documents also can reside in multiple places in a hierarchy in an information retrieval system, since physical location is not important. Classification hierarchy is thus Directed Acyclic Graph (DAG) which knows as classification DAG. And a web directory is just a classification directory on Web pages like Yahoo! Directory and Open Directory project.

1. The Differences Between Data And Information

There is a subtle difference between data and information. Data are the facts or details from which information is derived. For data to become information, data needs to be put into context. Data refers to the lowest abstract or a raw input which when processed or arranged makes meaningful output. It is the group or chunks which represent quantitative and qualitative attributes pertaining to variables. Information is usually the processed outcome of data. More specifically, it is derived from data. Information also is a concept and can be used in many domains.

Information can be a mental stimulus, perception, representation, knowledge, or even an instruction. The examples of data can be facts, analysis, or statistics. In computer terms, symbols, characters, images, or numbers are data. These are the inputs for the system to give a meaningful interpretation. In other words, data in a meaningful form is information. Information can be explained as any kind of understanding or knowledge that can be exchanged with people. It also can be about facts, things, concepts, or anything relevant to the topic concerned.

If data is at the lowest level in the series, information is placed at the next step. Data can be in the form of numbers, characters, symbols, or even pictures. A collection of these data which conveys some meaningful idea is information. It may provide answers to questions like who, which, when, why, what, and how. The raw input is data and it has no significance when it exists in that form. When data is collected or organized into something meaningful, it gains significance. This meaningful organization is called information. Data is often obtained as a result of recordings or observations. For example, the temperature of the days is data. When
this data is to be collected, a system or person monitors the daily temperatures and records it. Finally when it is to be converted into meaningful information, the patterns in the temperatures are analyzed and a conclusion about the temperature is arrived at. So information obtained is a result of analysis, communication, or investigation.

Data and information are interrelated. Data usually refers to raw data, or unprocessed data. It is the basic form of data, data that hasn’t been analyzed or processed in any manner. Once the data is analyzed, it is considered as information. Information is a sequence of symbols that can be interpreted as a message. It provides knowledge or insight about a certain matter. Data and information are interrelated. In fact, they are often mistakenly used interchangeably. Data is considered to be raw data. It represents ‘values of qualitative or quantitative variables, belonging to a set of items.’ It may be in the form of numbers, letters, or a set of characters. It is often collected via measurements. In data computing or data processing, data is represented by in a structure, such as tabular data, data tree, a data graph, etc. Data also usually refers to raw data, or unprocessed data. It is the basic form of data, data that hasn’t been analyzed or processed in any manner. Information can be recorded as signs, or transmitted as signals.

Basically, information is the message that is being conveyed, whereas data are plain facts. Once the data is processed, organized, structured or presented in a given context, it can become useful. Then data will become information, knowledge. Data in itself is fairly useless, until it is interpreted or processed to get meaning, to get information. In computing, it can be said that data is the computer’s language. It is the output that the computer gives us. Whereas, information is how we interpret or translate the language or data. It is the human representation of data. Other differences between data and information are that data is the raw material used as input for the computer system but information is the product or output of data. Data is also unprocessed facts figures which doesn’t carry a meaning and doesn’t depend on information whereas information is processed data that carry a logical meaning and must depend on data. Moreover data is not specific and single unit. But information is a group of data which carries news or meaning and is specific unlike data.

2. The Concepts And Components Of Database Management System And Information Retrieval System

CONCEPTS AND COMPONENTS OF DBMS

The DBMS or Database Management System is a database program. It is a software system that uses a standard method of catalogue, retrieving and running queries on data. The DBMS can manage incoming data, organizes it and provides ways for the data to be modified or extracted by users or other programs. DBMS main role is to acts as an interface between the user and the database. The user will request the DBMS to perform various operations such as insert, delete, update and retrieval on the database. The DBMS software is partitioned into several modules. Each module or component is assigned a specific operation to perform. Some of the functions of the DBMS are supported by operating systems to provide basic services and DBMS is built on top of it. For example the physical data and system catalogue are stored on a physical disk. Access to the disk is controlled primarily by, which schedules disk input or output. The components of DBMS perform these requested operations on the database and give necessary data to the users. The components of DBMS are shown in the figure of DBMS structured below.

Figure of DBMS Structured

DDL or Data Description Language is syntax similar to a computer programming language. Its function is to defining data structures especially database schemas. Whereas DDL Compiler function is to processes schema definitions specified in the DDL. It includes metadata information such as the name of the files, constraints, data items, mapping information, storage details of each file and many more. DML or Data Manipulation Language is a family of syntax elements too, which is similar to a computer programming language. But it was used for selecting, inserting, deleting and updating data in database. Thus DML will have commands such as insert, update, delete and retrieve. And the functions of DML Compiler are to translate DML statements in a query language into low-level instruction. It will help the query evaluation engine understanding.

For the Query Optimizer its function is to try determining the most efficient way to execute a given query by considering the possible query plans. For example, when the application program is sending commands to the DML compiler, the DML Compiler then compilation it into object code for database access. The object code is then optimized in the best way to execute a query by the query optimizer and then send to the data
The Data Manager or Database Control System is the central software component of the DBMS. Its main function is to convert operations in user's Queries coming from the application programs or combination of DML Compiler and Query optimizer which is known as Query Processor from user's logical view to physical file system. Data Manager also can controls DBMS information access that is stored on disk, handling buffers in main memory and controls the backup and recovery operations. Its also can enforce constraints to maintain consistency and integrity of the data, and synchronizes the simultaneous operations performed by the concurrent users.

Data Dictionary is a repository of description of data in the database. And it stored information and data such as names of the tables, names of attributes of each table, length of attributes, and number of rows in each table. Its also has data and information about relationships between database transactions and data items referenced which is useful in determining which transactions are affected when certain data definitions are changed. The detail information on physical database design such as storage structure, access paths, files and record sizes also included. More than that, its also contain constraints on data for example range of values permitted, Access Authorization such as the description of database users, their responsibilities and their access rights and usage statistics such as frequency of query and transactions.

Actually, Data Dictionary is also used to control the data integrity, database operation and accuracy. With Data Dictionary, the control of DBA over the information system and user’s understanding of use of the system is improves. Data Dictionary also helps in documentation the database design process by storing documentation of the result of every design phase and design decisions. Its helps in searching the views on the database definitions of those views and provides great assistance in producing a report of which data elements such as data values are used in all the programs. More over Data Dictionary also promotes data independence by addition or modifications of structures in the database application program which not affected. Data Files contains the data portion of the database whereas the DML complier converts the high level Queries into low level file access commands known as compiled DML.

CONCEPTS AND COMPONENTS OF IR SYSTEM

Information retrieval is a problem-oriented discipline, concerned with the problem of the effective and efficient transfer of desired information between human generator and human user. In modern history, the ‘information overload’ problem is much older than you may think. Its origins is in period immediately after World War because there are tremendous scientific progress during the war. Nowadays, information retrieval systems are everywhere such as Web search engines, library catalogues, store catalogues, cookbook indexes, and so on.

Information retrieval (IR), also called information storage and retrieval (ISR or ISAR) or information organization and retrieval. Traditionally, IR has concentrated on finding whole documents consisting of written text. Many IR research focuses more specifically on text retrieval which is the computerized retrieval of machine readable text without human indexing. Utility and relevance underlie all IR operations. A document’s utility depends on three things, topical relevance, pertinence, and novelty. A document is topically relevant for a topic, question, or task if it contains information that either directly answers the question or can be used. Its possibly in combination with other information and to derive an answer or perform the task. It is pertinent with respect to a user with a given purpose. If, in addition, it gives just the information needed which is compatible with the user’s background and cognitive style. Then he can apply the information gained authoritatively. And it is novel if it adds to the user’s knowledge.

Many IR systems focus on finding topically relevant documents, leaving further selection to the user. Relevance is a matter of degree for example when some documents are highly relevant and indispensable for the user's tasks. And others contribute just a little bit and could be missed without much harm. For ranked retrieval, performance measures are more complex. All of these measures are based on assessing each document on its own, rather than considering the usefulness of the retrieved set as a whole. IR is a component of an information system and an information system must make sure that everybody it is meant to serve has the information needed to accomplish tasks no matter where that information is available. An information system must actively find out what users need, acquire documents which then resulting in a collection and match documents with needs.

An IR system prepares for retrieval by indexing documents and formulating queries, resulting in document representations and query representations, respectively. The system then matches the representations and
displays the documents found and the user selects the relevant items. These processes are closely intertwined and dependent on each other. Indexing is a process to making statements about a document’s subjects. Indexing can be document-oriented. The indexer captures what the document is about. Or it can be request-oriented which is a situation when the indexer assesses the document’s relevance to subjects and other features of interest to users. Related to indexing is abstracting which is a process to creating a shorter text that describes what the full document is about or even includes important results.

Automatic summarization has attracted much research interest. Automatic indexing begins with raw feature extraction, refinements, counting and mapping using a thesaurus. A program can analyze sentence structures to extract phrases and images, extractable features include colour distribution or shapes. For music, extractable features include frequency of occurrence of notes or chords, rhythm and melodies. The refinements process is including transposition to a different key. Raw or refined features can be used directly for retrieval or they can be processed further. The system also can use a classifier that combines the evidence from raw or refined features to assign descriptors from a pre-established index language. A classifier can be built by hand by treating each descriptor as a query description and building a query formulation for it as described in the next section. Or a classifier can be built automatically by using a training set.

Many different words and word combinations can predict the same descriptor, making it easier for users to find all documents on a topic. Assigning documents to classes of a classification process are also known as text categorization. Its the process when the query description is transformed, manually or automatically into a formal query representation or query formulation. And this query representation can combines features that predict a document’s usefulness. The query expresses the information need in terms of the system’s conceptual schema, ready to be matched with document representations. A query then can specify text words or phrases or look for or any other entity feature. A query also can simply give features in an unstructured list or combine features using Boolean operators. The Boolean query specifies three ANDed conditions, all of which are necessary for contribution to the document score. Stating the information need and formulating the query process often go hand-in-hand.

An IR system also can show a subject hierarchy for browsing and finding good descriptors or its can ask the user a series of questions and from the answers its can construct a query. The system also can suggest synonyms and narrower and broader terms from its thesaurus. Its then matching the query representation with entity representations. The match then uses the features specified in the query to predict document relevance. In exact match the system finds the documents that fill all the conditions of a Boolean query when it predicts relevance as 1 or 0. To enhance recall, the system has to used synonym expansion and hierarchic expansion or inclusive searching. Since relevance or usefulness is a matter of degree, many IR systems rank the results by a score of expected relevance or ranked retrieval.

In other hand, the selection process occur when user examines the results and selects relevant items. Results then can be arranged in rank. Users may need assistance with making the connection between an item found and the task at hand. Once the user has assessed the relevance of a few items found, the query can be improved. The improvement will help the system assist the user in improving the query by showing a list of features found in many relevant items and another list from irrelevant items. Or the system can improve the query automatically by learning which features separate relevant from irrelevant items and thus are good predictors of relevance.

Moreover, IR systems can evaluated a view to improvement or with view to selecting the best IR system for a given task of summative evaluation. IR systems also can be evaluated on system characteristics and on retrieval performance. The requirements for recall and precision vary from query to query, and retrieval performance varies widely from search to search, making meaningful evaluation difficult. Standard practice evaluates systems through a number of test searches. Its computing for each a single measure of goodness that combines recall and precision and then averaging over all the queries. The system also has ability to adapt to the specific recall and precision requirements of each individual query. The most important evaluation efforts today are TREC and TDT.

4. The Differences between Structured and Non Structured Data

Data that resides in a fixed field within a record or file is called structured data. Structured data first depends on creating a data model of the types of business that will be recorded, stored, processed and accessed. This includes data contained in relational databases and spreadsheets. Structured data has the advantage of
being easily entered, stored, queried and analyzed. At one time, because of the high cost and performance limitations of storage, memory and processing, relational databases and spreadsheets using structured data were the only way to effectively manage data. Anything that couldn’t fit into a tightly organized structure would have to be stored on paper in a filing cabinet.

Structured data is also information such as text files, which is displayed in titled columns and rows that can easily be ordered and processed by data mining tools. This could be visualized as a perfectly organized cabinet where everything is identified, labelled and easy to access. Many organizations are likely to be familiar with this form of data and already using it effectively. Spreadsheets can be considered as structured data, which can be quickly scanned for information because it is properly arranged in a relational database system. For the most part, structured data refers to information with a high degree of organization. For example, a relational database is seamless, readily and simply searchable in straightforward search engine algorithms or other search operations. Thus it is a structured data. Whereas unstructured data is essentially the opposite. The unstructured data is lack of structure so it’s wasting time and energy to consuming task.

However, structured data is akin to machine language, its makes information much easier to deal with using computers. Whereas unstructured data is usually for humans, who don’t easily interact with information in strict, database format.

The unstructured data is the binary data that is proprietary. It has no identifiable internal structure. Its can be imagined as a massive unorganized conglomerate of various objects. They are worthless until being identified and stored in an organized fashion. Once we used specialized software to organize the unstructured data, the items then can be searched and categorized. The data mining tools might not be equipped to parse information in email messages. But the data still can be organized and we may have a very good reason to collect and categorize data from this source. That’s why the plausible breadth of unstructured data is important. There are many types of unstructured data such as emails, word processing files, PDF files, spreadsheets, digital images, video, audio and social media posts.

Unstructured data usually is raw and unorganized. And usually the organizations tend to store all of it. We can convert unstructured data into structured data. Unfortunately, this would be costly and consuming much time. Furthermore, not all types of unstructured data can be easily converted into a structured model. For example, an email holds information such as the time sent, subject, and sender. But the content of the message is not easily broken down and categorized. This can introduce some compatibility issues with the structure of a relational database system. What is interesting is the fact that all types of unstructured data can be stored and managed without the format of the file being understood by the system. Which mean they can be stored in an unstructured fashion because the contents of the files are unorganized unlike structure data.

The phrase “unstructured data” usually refers to information that doesn’t reside in a traditional row and column database. As we might expect, it’s the opposite of structured data which is the data that stored in fields in a database. Unstructured data files often include text and multimedia content. Examples include e-mail messages, word processing documents, videos, photos, audio files, presentations, web pages and many other kinds of business documents. Although these sorts of files may have an internal structure, they are still considered “unstructured” because the data contain doesn’t fit neatly in a database.

About Essay Sauce

EssaySauce.com is a completely free resource to help students research their academic work and learn from great essays!

View all posts by Essay Sauce

...(download the rest of the essay above)
About EssaySauce, the student essay site:

EssaySauce.com is a free resource for students, providing thousands of example essays to help them complete their college and university coursework. Students can use our free essays as examples to write their own.

Latest student essays:

- Causes of terrorism
- The death penalty
- Political Thought of Al Mawardi
- LES INIUSTA NON EST LEX (AN UNJUST LAW IS NOT LAW AT ALL)
- IDEOLOGICAL TERRORISM, BOKOHARAM AS A CASE STUDY
- Terrorism and the law
- Difference between force majeure and hardship in international contracts
- Harnessing energy through knowledge – business development strategy of e-commerce companies
- Minimizing of power losses for distribution system
- Translating the Biggles Stories for Czech Readers: A Case of Moderate Transposition
Student essay categories:

Accounting essays
Architecture essays
Business essays
Economics essays
Education essays
Engineering essays
English language essays
English literature essays
Environmental studies essays
Finance essays
Health essays
History essays
Information technology essays
International Relations
Law essays
Literature essays
Management essays
Marketing essays
Miscellaneous essays
Music Essays
Photography and arts essays
Politics essays
Project management
Psychology essays
Religious studies and Theology essays
Science essays
Sociology essays
Zoology essays

Average review:

Overall rating: 0 out of 5 based on 0 reviews.

Q: Is EssaySauce.com free?
Yes! EssaySauce.com is a completely free resource for students. You can view our terms of use here.

Why use Essay Sauce?

The brightest students know that the best way to learn is by example! EssaySauce.com has thousands of great essay examples for students to use as inspiration when writing their own essays.

Is Essay Sauce completely free?

Yes! EssaySauce.com is a completely free resource for students. You can view our terms of use here.

Info:

About
Content policy
Essay removal request
Privacy
Terms of use