Information Storage and Retrieval notes, lectures 1 and 2

Overview

Information is a critical business resource and like any other critical resource, it must be properly managed. Constantly evolving technology, is changing the way even the very small businesses manage vital business information. An information or records management system – most often electronic – **designed to capture, process, store and retrieve** information that holds a business together.

Information is increasingly important in our daily lives. We have become information-dependent in the 21st century, living in an on-command, on-demand world, which means, we need information when and where it is required. We access the Internet every day to perform searches, participate in social networking, send and receive e-mails, share pictures and videos, and use scores of other applications. Equipped with a growing number of content-generating devices, more information is created by individuals than by organizations (including business, governments, non-profits and so on). Information created by individuals gains value when shared with others. When created, information resides locally on devices, such as cell phones, smartphones, tablets, cameras, and laptops. To be shared, this information needs to be uploaded to central data repositories (data centers) via networks. Although the majority of information is created by individuals, it is stored and managed by a relatively small number of organizations.

The importance, dependency, and volume of information for the business world also continue to grow at astounding rates. Businesses depend on fast and reliable access to information critical to their success. Examples of business processes or systems that rely on digital

Definitions of key concepts

Storage is a repository that enables users to store & retrieve this digital data.

Information Storage

Information storage is the part of the accounting system that keeps data accessible to the information processors. In other words, an accounting systems, information storage unit is either a hard drive or server that usually contains a database.

Data storage is the collective methods and technologies that capture and retain digital information on electromagnetic, optical or silicon-based storage media. Storage is a key component of digital devices, as consumers and businesses have come to rely on it to preserve information ranging from personal photos to business-critical information.

Storage is frequently used to describe the devices and data connected to the computer through input/output (I/O) operations, including hard disks, flash devices, tape systems and other media types.

Data:-

1. It is a collection of raw facts from which conclusions may be drown.

2. Data in the form of 0's & 1's is called digital data & is accessible by the user only after it is processed by a computer.

3. While the advancement of Computer & Communication technologies, the rate of data generation & sharing has increased exponentially.

Types of data:-

Data can be classified as structured or unstructured based on how it is stored & managed.

Information Storage Systems

They are built by taking into considerations the basis capability of a storage device, such as HDD & adding hardware & software to obtain high performing, reliable & easily managed system.

Information retrieval

Information retrieval (**IR**) in information science is the process of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and searching for the metadata that describes data, and for databases of texts, images or sounds.

Information storage and retrieval

1. is the systematic process of collecting and cataloging data so that they can be located and displayed on request. Computers and data processing techniques have made possible the high-speed, selective retrieval of large amounts of information for government, commercial, and academic purposes.
2. **is** the systematic process of collecting and cataloging **data** so that they can be located and displayed on request. The **data** are **stored** within the computer, either in main **storage** or auxiliary **storage**, for ready access

Information Retrieval Systems

1. Informative retrieval system is a system with a user interface that provides the facility for the user to create, search & modify the data stored in a storage network.

2. This is typically a peer-to-peer network which is operated & maintained by private organizations, however access rights are provided to the public.

3. The access can be performed via Internet from outside the organisation & via Intranet within the organizations.

4. Information Storage & IR are addressed as two sides of the same cover.

5. If a person is able to search the required information, then that information must have already been stored in same format. The format in which the information is often represented to the people can be either techs, images, audios or videos, which makes it different to obtain clear & precise answer to multiple questions which the users may require. Searching a document involves a collection of information which may be either easy or complicated depending on how the collection is organised.

6. Almost all of the IR systems fielded today are either Boolean IR systems for major document collections or text pattern search systems for handling small document collections (for ex: personal collections of files).

7. Text patterns search queries are strings or regular expressions. The grape family of tools, in the UNIX environment is a well-known example of text pattern searchers.

8. In Boolean IR System, documents are represented by sets of keywords, usually stored in an inverted file. In inverted file is a list of keywords & identifiers of the documents in which they occur. Boolean queries are keyword connected with Boolean logically operations (AND,OR,NOT).

9. Conceptual models facet focuses on the performance enhancements of IR systems neither the information associated with statistical distribution of terms.

10. The statistical models such as vector space, probabilistic or clustering models do the statistical distribution of terms where every documents in retrieved collection is allocated with probability of relevance.

Types of Information storage and retrieval systems

There are several basic types of information-storage-and-retrieval systems.

1, *Document-retrieval systems*- store entire documents, which are usually retrieved by title or by key words associated with the document. In some systems, the text of documents is stored as data. This permits full text searching, enabling retrieval on the basis of any words in the document

2. *Database systems* - store the information as a series of discrete records that are, in turn, divided into discrete fields (e.g., name, address, and phone number); records can be searched and retrieved on the basis of the content of the fields (e.g., all people who have a particular telephone area code). The data are stored within the computer, either in main storage or auxiliary storage, for ready access.

3. *Reference-retrieval systems*- store references to documents rather than the documents themselves. Such systems, in response to a search request, provide the titles of relevant documents and frequently their physical locations. Such systems are efficient when large amounts of different types of printed data must be stored. They have proven extremely effective in libraries, where material is constantly changing.

The Importance of Information Storage & Retrieval Systems in an Organization

An information or records management system – most often electronic – designed to capture, process, store and retrieve information is the glue that holds a business/organisation together.

The benefits/importance include:

1. Regulatory Compliance

Unlike a public company, a privately held business isn't subject to most federal and state government compliance requirements. Despite this, many choose to comply voluntarily, both to provide transparency and enhance the business's public image. In addition, small-business owners must store and maintain tax information so, in case of an audit, the information is readily accessible. A well-organized information storage and retrieval system that follows compliance regulations and tax record-keeping guidelines significantly increases a business owner's confidence the business is fully complying.

2. Efficiency and Productivity

Any time a business/organisation owner or employees spend searching through stacks of loose files or spend trying locate missing or misfiled records is inefficient, unproductive and can prove costly to a small business. A good information storage and retrieval system, including an effective indexing system, not only decreases the chances information will be misfiled but also speeds up the storing and retrieval of information. The resulting time-saving benefit increases office efficiency and productivity while decreasing stress and anxiety.

3. Improve Working Environment

It can be disheartening to anyone walking through an office area to see vital business documents and other information stacked on top of file cabinets or in boxes next to office workstations. Not only does this create a stressful and poor working environment, but if customers see this, it can cause them to form a negative perception of the business. Contrast

this with an office area in which file cabinets, aisles and workstations are clear and neatly organized to see how important it is for even a small business to have a well-organized information storage and retrieval system.

4. Electronic vs. Manual System

Although a very small business may choose to institute a manual system, the importance of electronic information storage and retrieval systems lie in the fact that electronic systems reduce storage space requirements and decrease equipment and labor costs. In contrast, a manual system requires budgetary allotments for storage space, filing equipment and administrative expenses to maintain an organized filing system. Additionally, it can be significantly easier to provide and monitor internal controls designed to deter fraud, waste and abuse as well as ensure the business is complying with information privacy requirements with an electronic system.

History of Information Storage and Retrieval

Students are expected to report back on this topic

Database Management systems Lecture 3 and 4

History of DBMS

Here, are the important landmarks from the history:

- 1960 - Charles Bachman designed first DBMS system
- 1970 - Codd introduced IBM'S Information Management System (IMS)
- 1976- Peter Chen coined and defined the Entity-relationship model also know as the ER model
- 1980 - Relational Model becomes a widely accepted database component
- 1985- Object-oriented DBMS develops.
- 1990s- Incorporation of object-orientation in relational DBMS.
- 1991- Microsoft ships MS access, a personal DBMS and that displaces all other personal DBMS products.
- 1995: First Internet database applications
- 1997: XML applied to database processing. Many vendors begin to integrate XML into DBMS products.

Definitions

A database

A database is a collection of related data which represents some aspect of the real world. A database system is designed to be built and populated with data for a certain task.

A database management System

1. **A database management system (DBMS**) is a software package designed to define, manipulate, retrieve and manage data in a database. A DBMS generally manipulates the data itself, the data format, field names, record structure and file structure. It also defines rules to validate and manipulate this data.
2. **Database Management System (DBMS)** is a software for storing and retrieving users' data while considering appropriate security measures. It consists of a group of programs which manipulate the database. The DBMS accepts the request for data from an application and instructs the operating system to provide the specific data. In large systems, a DBMS helps users and other third-party software to store and retrieve data.

DBMS allows users to create their own databases as per their requirement. The term "DBMS" includes the user of the database and other application programs. It provides an interface between the data and the software application. Database management systems are set up on specific data handling concepts, as the practice of administrating a database evolves.

Example of a DBMS

For instance, a simple example of a university database. This database is maintaining information concerning students, programmes, course units, and grades in a university environment. The database is organized as five files:

- The STUDENT file stores data of each student
- The COURSE file stores contain data on each course.
- The SECTION stores the information about sections in a particular course.
- The GRADE file stores the grades which students receive in the various sections
- The TUTOR file contains information about each professor.

Characteristics of Database Management System

- Provides security and removes redundancy
- Self-describing nature of a database system
- Support of multiple views of the data
- Sharing of data and multiuser transaction processing
- DBMS allows entities and relations among them to form tables.
- DBMS supports multi-user environment that allows users to access and manipulate data in parallel.

Advantages

A Database Management System (DBMS) is defined as the software system that allows users to define, create, maintain and control access to the database. DBMS makes it possible for end users to create, read, update and delete data in database.

Some of these advantages are given below −

- *Reducing Data Redundancy*

  The file based data management systems contained multiple files that were stored in many different locations in a system or even across multiple systems. Because of this, there were sometimes multiple copies of the same file which lead to data redundancy. This is prevented in a database as there is a single database and any change in it is reflected immediately. Because of this, there is no chance of encountering duplicate data.

- *Sharing of Data*

  In a database, the users of the database can share the data among themselves. There are various levels of authorisation to access the data, and consequently the data can only be

shared based on the correct authorisation protocols being followed. Many remote users can also access the database simultaneously and share the data between themselves.

- *Data Integrity*

  Data integrity means that the data is accurate and consistent in the database. Data Integrity is very important as there are multiple databases in a DBMS. All of these databases contain data that is visible to multiple users. So it is necessary to ensure that the data is correct and consistent in all the databases and for all the users.

- *Data Security*

  Data Security is vital concept in a database. Only authorised users should be allowed to access the database and their identity should be authenticated using a username and password. Unauthorised users should not be allowed to access the database under any circumstances as it violates the integrity constraints.

- *Privacy*

  The privacy rule in a database means only the authorized users can access a database according to its privacy constraints. There are levels of database access and a user can only view the data he is allowed to. For example - In social networking sites, access constraints are different for different accounts a user may want to access.

- *Backup and Recovery*

  Database Management System automatically takes care of backup and recovery. The users don't need to backup data periodically because this is taken care of by the DBMS. Moreover, it also restores the database after a crash or system failure to its previous condition.

- *Data Consistency*

  Data consistency is ensured in a database because there is no data redundancy. All data appears consistently across the database and the data is same for all the users viewing the database. Moreover, any changes made to the database are immediately reflected to all the users and there is no data inconsistency.

Limitations of Database Management System

Database Management System is quite useful compared to the file based management system. However, it does have some disadvantages. Some of those are as follows −

More Costly

Creating and managing a database is quite costly. High cost software and hardware is required for the database. Also highly trained staff is required to handle the database and it also needs continuous maintenance. All of these ends up making a database quite a costly venture.

High Complexity

A Database Management System is quite complex as it involves creating, modifying and editing a database. Consequently, the people who handle a database or work with it need to be quite skilled or valuable data can be lost.

Database handling staff required

As discussed in the previous point, database and DBMS are quite complex. Hence, skilled personnel are required to handle the database so that it works in optimum condition. This is a costly venture as these professionals need to be very well paid.

Database Failure

All the relevant data for any company is stored in a database. So it is imperative that the database works in optimal condition and there are no failures. A database failure can be catastrophic and can lead to loss or corruption of very important data.

High Hardware Cost

A database contains vast amount of data. So a large disk storage is required to store all this data. Sometimes extra storage may even be needed. All this increases hardware costs by a lot and makes a database quite expensive.

Huge Size

A database contains a large amount of data, especially for bigger organisations. This data may even increase as more data is updated into the database. All of these leads to a large size of the database.

The bigger the database is, it is more difficult to handle and maintain. It is also more complex to ensure data consistency and user authentication across big databases.

Upgradation Costs

Often new functionalities are added to the database.This leads to database upgradations. All of these upgradations cost a lot of money. Moreover it is also quite expensive to train the database managers and users to handle these new upgradations.

Cost of Data Conversion

If the database is changed or modified in some manner, all the data needs to be converted to the new form. This cost may even exceed the database creation and management costs sometimes. This is the reason most organisations prefer to work on their old databases rather than upgrade to new ones.

DBMS vs. Flat File

DBMS	Flat File Management System
Multi-user access	It does not support multi-user access
Design to fulfill the need for small and large businesses	It is only limited to smaller DBMS system.
Remove redundancy and Integrity	Redundancy and Integrity issues
Expensive. But in the long term Total Cost of Ownership is cheap	It's cheaper
Easy to implement complicated transactions	No support for complicated transactions

Popular DBMS Software

Here, is the list of some popular DBMS system:

- MySQL
- Microsoft Access
- Oracle
- PostgreSQL
- dBASE
- FoxPro
- SQLite
- IBM DB2
- LibreOffice Base
- MariaDB
- Microsoft SQL Server etc.

Application of DBMS

Sector	Use of DBMS
Banking	For customer information, account activities, payments, deposits, loans, etc.
Airlines	For reservations and schedule information.
Universities	For student information, course registrations, colleges and grades.
Telecommunication	It helps to keep call records, monthly bills, maintaining balances, etc.
Finance	For storing information about stock, sales, and purchases of financial instruments like stocks and bonds.
Sales	Use for storing customer, product & sales information.
Manufacturing	It is used for the management of supply chain and for tracking production of items. Inventories status in warehouses.
HR Management	For information about employees, salaries, payroll, deduction, generation of paychecks, etc.

Four Types of DBMS systems are:

- Hierarchical database
- Network database
- Relational database
- Object-Oriented database

Hierarchical DBMS

In a Hierarchical database, model data is organized in a tree-like structure. Data is Stored Hierarchically (top down or bottom up) format. Data is represented using a parent-child relationship. In Hierarchical DBMS parent may have many children, but children have only one parent.

Network Model

The network database model allows each child to have multiple parents. It helps you to address the need to model more complex relationships like as the orders/parts many-to-many relationship. In this model, entities are organized in a graph which can be accessed through several paths.

Relational model

Relational DBMS is the most widely used DBMS model because it is one of the easiest. This model is based on normalizing data in the rows and columns of the tables. Relational model stored in fixed structures and manipulated using SQL.

Object-Oriented Model

In Object-oriented Model data stored in the form of objects. The structure which is called classes which display data within it. It defines a database as a collection of objects which stores both data members values and operations.

Additional for your information

Advantages of DBMS

- DBMS offers a variety of techniques to store & retrieve data
- DBMS serves as an efficient handler to balance the needs of multiple applications using the same data
- Uniform administration procedures for data

- Application programmers never exposed to details of data representation and storage.
- A DBMS uses various powerful functions to store and retrieve data efficiently.
- Offers Data Integrity and Security
- The DBMS implies integrity constraints to get a high level of protection against prohibited access to data.
- A DBMS schedules concurrent access to the data in such a manner that only one user can access the same data at a time
- Reduced Application Development Time

Disadvantage of DBMS

DBMS may offer plenty of advantages but, it has certain flaws-

- Cost of Hardware and Software of a DBMS is quite high which increases the budget of your organization.
- Most database management systems are often complex systems, so the training for users to use the DBMS is required.
- In some organizations, all data is integrated into a single database which can be damaged because of electric failure or database is corrupted on the storage media
- Use of the same program at a time by many users sometimes lead to the loss of some data.
- DBMS can't perform sophisticated calculations

Information Retrieval Processes, Techniques and Evaluation

Overview of concepts

1. The tracing and recovery of specific information from stored data.

"an information retrieval system"

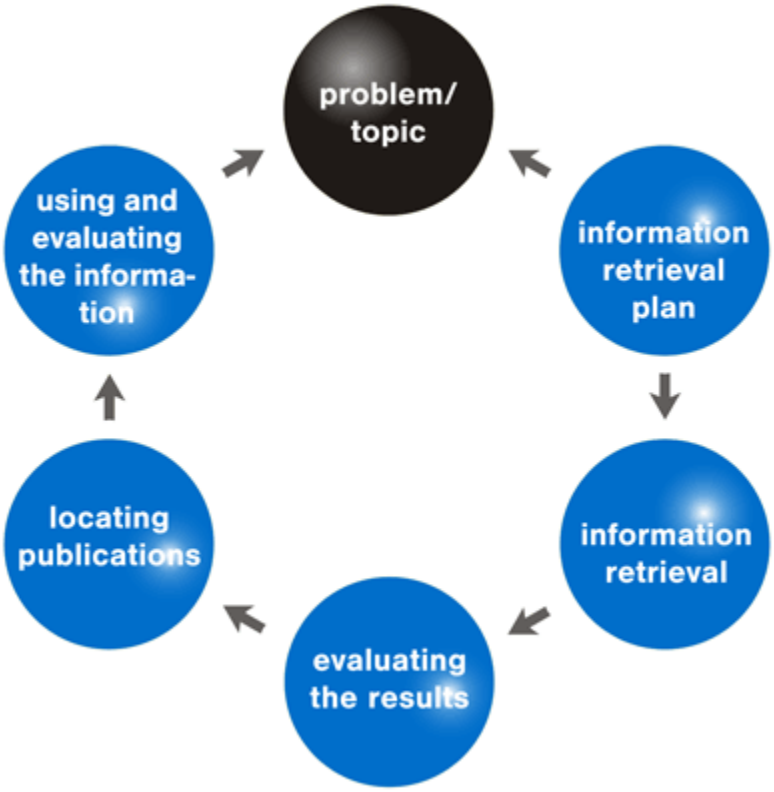
2. Information retrieval in computing and information science is the process of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing.

3. **Information retrieval**, recovery of information, especially in a database stored in a computer. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links.

The Information Retrieval Process

Information retrieval is often a continuous process during which you will consider, reconsider and refine your research problem, use various different information resources, information retrieval techniques and library services and evaluate the information you find. The figure below implies that the stages follow each other during the process, but in reality they are often active simultaneously and you usually will repeat some stages during the same information retrieval process.

The different stages of the information retrieval process are:



- **Problem / topic**: an information need occurs when more information is required to solve a problem
- **Information retrieval plan**: define your information need and choose your information resources, retrieval techniques and search terms
- **Information retrieval**: perform your planned information retrieval (information retrieval techniques)
- **Evaluating the results**: evaluate the results of your information retrieval (number and relevance of search results)
- **Locating publications**: find out where and how the required publication, e.g. article, can be acquired

- **Using and evaluating the information**: evaluate the final results of the process (critical and ethical evaluation of the information and information resources)

Information Resources

1. Books and journals

Traditional resources such as books and articles published in journals have gone through an acceptance process prior to their publication, so they can be considered as trusted sources of information.

As the publication process of journals is generally quicker than that of books, scientific articles in journals often contain the most current information available in a scientific field. The bibliographies of relevant books are also a good place to look for new information resources for your research.

2. Reference works

Reference works (dictionaries, encyclopedias, handbooks, registers and standards) provide factual information, including relevant terminology and concepts, and help you form a general picture of your research topic. Some reference books are multidisciplinary, whereas others are specialised in a certain field.

3. Theses

When starting your final thesis project, you should check whether any research has been done on your topic. Relevant theses from previous students provide a good starting point and example. However, you should not use them as the only information resource because the quality of the theses can vary.

4. Reports and conference publications

Research reports, report serials and conference publications by different organisations also offer information on new research.

5. Archive materials

The archives of different organisations maintain and arrange documents, objects and images accumulated by the mother organisation that can be used for research purposes.

Databases

Databases can be categorised for example as follows:

1. **Reference databases** point towards the required information, but you will still have to retrieve the information from the original document (e.g. library collection databases).
2. **Full-text databases** contain full-text versions of the required materials (e.g. electronic journals and books).
3. **Image databases** include databases managed by memory organisations (museums, archives and libraries), image banks by image agencies and open access images available on the Internet.
4. **Factual databases** contain the required information so that retrieving the original document is not necessary (e.g. statistical databases).

Evaluating information

1. **Are your resources up to date**

You should look for the most current information in your field because new research is sometimes produced quickly, leaving earlier research in need of updating. Information changes in different time frames in different scientific fields. For example, philosophic information is updated less frequently than that in e.g. technical sciences.

2. **The reliability of information**

It is important that you recognise the difference between popular and scientific information.

The quality of articles in scientific publications is guaranteed by peer review that is part of the publication process. The scripts of peer reviewed articles are reviewed by experts in the field, and they also decide on their publication. A peer

review group consists of experts who review the content of a publication objectively.

Evaluating electronic resources

When using Internet search engines (e.g. Google) instead of databases established in your field, you should be especially critical towards the information you find. As Internet content is not monitored by any authorised party who can guarantee the validity, accuracy or quality of the information, it is your responsibility as an information-seeker to evaluate the information critically.

Information Retrieval Techniques

Full-text or word search

Full-text or word search is most suited for the beginning stages of your research on a topic. Full-text search is also helpful if the field of research is new, the terminology has not yet been established in subject terms, or the topic is very specific.

Any words of natural language can be used as search terms in a full-text search. The search query can include multiple words and synonyms which will give comprehensive search results. The search terms can be in different languages, especially if the search is done in a database that contains information in multiple languages.

Full-text searches often result in a lot of unnecessary data because it looks for your terms anywhere in the record. The broader the database or your topic, the harder it is to find relevant results using word searches.

Subject term search

A subject term search gives more accurate results than a full-text search, and limits the search to the subject term field. However, keep in mind that using just the subject term search function can narrow your search results too much.

A thesaurus is an alphabetical list of subject terms used to describe the content of a document. Thesauruses often have a hierarchical structure and suggest broader, narrower and related terms for each subject term listed.

The thesauruses of international databases are most often specific to the database or scientific field, and are available through the database.

Combining search terms

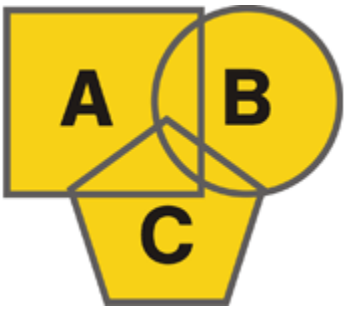
Boolean logic is an essential tool in information retrieval and allows you to combine search terms. When you need more than one word to describe your search problem, you can combine multiple search terms with Boolean operators. The most common operators are **AND**, **OR** and **NOT**.

The **AND** operator is used to retrieve results that contain all of the search terms used.

Therefore AND narrows your search results.

E.g. phones AND marketing

The **OR** operator is used to retrieve results that contain all or any of the search terms used. OR is commonly used when the search terms are very similar in meaning, or when the search terms are foreign language equivalents. Therefore the OR operator broadens the search results.



E.g. phone OR cell OR cellular

The **NOT** operator is used to retrieve results that do not include a certain word. You should be careful when using the operator as it can exclude a lot of useful results.

E.g. students NOT pupils

Phrase search

If you want to search a phrase (two or more words in a specific order), use quotation marks around the words. Without quotation marks, many databases are looking for search words individually, and the results will contain a lot of irrelevant material.

E.g. ”artificial intelligence”, "sustainable design"

Term truncation

Term truncation is particularly useful for languages with case endings, as it allows you to include all the inflected forms of the search term in the search. The symbol used for term truncation varies between databases, most commonly used are *, ? and #.

E.g. the word ***processor*** can be truncated as follows:

process* will search for process as well as variations such as *processor, processing, processed, process, processes* and *processors*

Additional help

Most databases include a guide on how to retrieve information from the database. The search functions in different databases are often very similar, but it is nevertheless recommended that you familiarise yourself with the guide before you start using a new database.

Information Retrieval Tools and Methods.

There are quite a number of information retrieval tools. (Nnadozie, 2007) listed the library information retrieval tools to include: reading list, index, abstract, library catalogue, search engine, OPAC, bibliography, shelf guides, web-based information retrieval systems which are presently at students' disposal to aid them in accessing information. The majorly used Information Retrieval tools include Catalogues, Classification Schemes, Indexes, Abstracts, and Bibliographies.

Other Information Retrieval tools in the library include the following: Encyclopedia, Directories, Dictionaries, Almanacs, Handbooks, Atlases, Periodicals, and Concordances among others which are explained below:

Indexes

Indexes are retrieval tools which provide access to the analyzed contents of information resources (articles in journals, short stories in collections, papers in conference proceedings (Taylor, 2009). An index is an alphabetical list of terms that occurs at the conclusion of a book and includes the page number where the term appears in the book. Subject indexing is a type of classification. We put related subject materials together (sometimes literally, as in the classified arrangement on library shelves) and segregate them from other documents based on their subject matter. Then, in order to keep the established groupings and refer to them, we label these classes of connected documents. To put it another way, we give the classes names, and the names we give them are the index terms.

Bibliography

Bibliographies are systematic listing of information resources on a given subject and or by a given author. University of Calabar library information guide, (unpublished), identifies information resources to include books, periodicals, pictures, maps, manuscripts and any other media of communication that exist. Bibliographies are not only valuable, but also essential, in that they provide direction to students to embark on their research needs. (Inyokwe, 2015)

(Aquolu, 2002)maintained that the primary users of information resources in University libraries are students of the universities. And that any realistic bibliographic organization designed to provide direction to the effective use of library resources must depend on the librarian's knowledge of the needs of these diversified users (students). The authors agree that librarians as mediators between students and the information environment require special skills to design bibliographic systems that will link students to the use of information resources of their needs.

There are hundreds of different types of bibliographies prepared for various purposes, and your reference librarian can tell you which one might be relevant in the topic area of your study using bibliographies.

A bibliography, on the other hand, will inform students that a work exists but may not be housed by that particular library. Following the completion of a journal article or a book, the author frequently refers to the sources he used. The source could be a book or an article that he used or supplemented his own knowledge, which is why the list is called a reference or bibliography.

E-Catalogues

A library e-catalogue is a register of all bibliographic items found in a library or group of libraries, such as a network of libraries at several locations. A catalogue for a group of libraries is also called a union catalogue. A bibliographic item can be any information entity that is considered library material or a group of library materials or linked from the catalogue as far as it is relevant to the catalogue and to the users (patrons) of the library. (Highsmith, 2009)

Catalogues are the doors to the library's collection. It is also a systematic organization of objects in alphabetical or other logical order, including a brief explanation. For years, library customers have been familiar with the card catalogue. However, it has been essentially supplanted by the online public access catalog (OPAC). A single library's catalogue may be offered in many physical versions to span different historical periods.

Nevertheless, the precise purposes of the library catalogue vary, they can be broadly summarized as;

To help a library user to find a book which:

- The author is known.
- The title is known.
- The subject is known.

To show the library holdings:

- By a given author.
- On a given subject and related subject.
- In a given form of literature.

To help in the choice of books:

- As to its edition.
- As to its other characteristics.

Classification schemes

Nnadozie (2007) defined classification scheme as a system created for the division into categories of the universe of human knowledge into broad subjects and narrow topics. The original rationale behind the creation of classification schemes was to help librarians map the universe of knowledge such that documents can be put aside for future consideration into specific locations for easy identification and retrieval.

Subject Headings/ List of Headings

In traditional library practice, when a controlled vocabulary is set up in the form of an alphabetical listing of index terms, the individual terms are known as subject heading and the controlled vocabulary as a list of subject headings. Subject headings lists are useful to understand the relationship among concepts to a certain degree, besides their application in indexing. Subject headings lists are highly valuable for indexing. Subject headings are provided in the catalogue entries to provide subject access to information. Cataloguers depend on Lists of Subject Headings from which they can assign subject headings to the catalogued documents. (Chen & Larsen, 2016).

Digitisation Notes

Definitions:

Digitisation is the process of converting information from a physical format into a digital one.

For example: Nicholas scans a signed MoU and saves this scan as a PDF. He then saves this PDF to his C drive.

Digitalisation?

Digitalisation is the process of leveraging digitisation to improve business processes.

For example: Eve scans a signed MoU and saves this scan as a PDF. She then uploads this PDF to the cloud so she can access the MoU anytime and anywhere.

Difference between Digitisation and Digitalisation

In the above examples, both Nicholas and Eve have digitised a part of their business process, but Eve's business process is more digitalised. She is leveraging digitisation to improve business processes to a greater degree.

While digitalisation carries implicit positive connotations, digitisation does not. Digitisation is an act that may enable digitalisation, but the latter always requires the former.

The digitization process

The actual digitisation process can be divided into three areas, each with its own activities:

a. ***Data capture and creation***: the images are digitised, any associated text (metadata) that describes the images and their content is added and the data (i.e. the digital records: image plus associated text) are stored or archived for future re-use. Activities include: image handling and preparation; image capture; hardware and software; file formats and compression; copyright, intellectual property rights (IPR), ethics and data protection; metadata: image description, indexing and cataloguing.

b. ***Data access and delivery*****:** the image archive owners must plan and implement a delivery mechanism that will ensure the user communities can access the image archive. Activities include: search and retrieval; user issues and access management.

c. ***Managing the digital collection*****:** the production of a high-quality digital image collection that adheres to standards and good practice will be ensured by the co-ordination of 1 and 2. Activities include: database creation and system design; workflow and procedures management; quality assurance; project management; digital preservation and storage.[45]

The knowledge and skills required for creating digital collections

Management

Good management is vital for a successful digitisation project because of the unpredictable changes in the technology, the complex nature of the digitisation process, and the high-level skills and training required from project staff. A manager or a management team does not need hands-on experience of all aspects of a digitisation project but must understand the process and be able to lead the project in the correct direction while keeping it within the time and cost constraints.

The following knowledge and skills are the key qualities for managing digitisation projects:

i. **strong leadership capability** and clear understanding of goals of and issues related to creating digital collections;

ii. **effective communication and interpersonal** skills for working with project staff that may be diverse in expertise, location and temperament;

iii. **overall understanding of the digital collections** infrastructure, including processes, standards and technology;

iv. **knowledge of current digitisation technologies,** standards and best practices;

v. **knowledge of policies, procedures and intellectual property rights** involved in digitisation;

vi. **knowledge of requirements analysis,** project planning and tracking, task prioritisation and workflow management;

vii. **understanding of the costs involved** in digitisation and knowledge of funding opportunities available for projects;

viii. **understanding of the** importance of project documentation and statistics reporting;

ix. **marketing skills to disseminate**, promote and publicise digital collections.

Selecting material for digitization

Technical feasibility

Technical feasibility is one of the most important criteria for selecting material for digitisation. The physical characteristics of source material and the project goals for capturing, presenting and storing the digital surrogates dictate the technical requirements. Libraries must evaluate those requirements for each project and determine whether they can be met with the resources available. If the existing staff, hardware and software resources cannot meet the requirements, then the project will need funding to upgrade equipment or hire an outside conversion

agency. If these resources are not available, or if the technology does not exist to meet the requirements, then it is not technically feasible to digitise that material. Considerations for technical feasibility include:

- *Image capture*. Image capture requires equipment, such as a scanner or a digital camera. Different types of material require different equipment, and different equipment produces images of differing quality. When selecting materials for digitising, technical questions that need to be addressed include: does the original source material require high resolution to capture? Are there any oversized items in the collection? Are there any bound volumes in the collection? What critical features of the source material must be captured in the digital product? In what condition are the source materials? Will they be damaged by the digitisation process?

- *Presentation*. Presentation refers to how the digitised materials will be displayed online. Consider the following questions to determine the technical feasibility of presenting the digitised material:
  Will the materials display well digitally?

  How will users use the digital versions?

  How will users navigate within and among digital collections?

  Do the institutionally supported platforms and networked environment have the capability for accessing the images and delivering them with reasonable speed to the target audience?

  Do the images need to be restricted to a specified community?

  Do the images need special display features such as zooming, panning and page turning?

- *Description*. Some archival and special collections have been catalogued for public use and contain detailed finding aids with descriptions about each item and the collection as a whole. Other collections may not have been reviewed and documented in detail and do not have much information on individual items. Those collections will require more time, human resources and significant additional expense to research the materials, check the accuracy of the information obtained, and write appropriate descriptions to aid in discovery and use of the digital items. Typewritten documents, like

the Drew Pearson columns described above, can have reasonably accurate OCR applied to them to replace, for some uses, the detailed descriptions required for discovery of hand-written or picture materials. The selection criteria should clearly state whether the items and collections that do not contain descriptions should be considered for digitisation.

*Human resources***.** When selecting materials for digitisation, the library should consider whether it has the staff and skill sets to support the digitisation, metadata entry, user interface design, programming and search engine configuration that is required for the project to implement the desired functionality. For large collaborative projects, dedicated staff are usually required from each partner. Digital collections also require long-term maintenance, which needs to be considered and planned for. If a project does not have the necessary staff and skills in-house, but funding is available, outsourcing may be a good choice.

Benefits of digitisation

More and more businesses are adding value to their collections through developing digitisation initiatives. Each organisation has different priorities but the benefits of digitisation include:

Access

The information in documents can be published in a number of ways and made available to global audiences, so that access is no longer restricted to those able to visit the physical location, saving time and travel costs.

There is also the ability to access existing resources previously limited by their format, such as large maps and materials stored on microfilm.

Generating income

Many documents contain information that can have a commercial value when presented to the right audiences. Digitisation unlocks this potential to create new income streams.

Brand

Opening up your archives can benefit your brand by inspiring new audiences and raising the profile of the institution. Many modern brands are rediscovering the value of their ‘brand heritage’.

Searchability

Capturing the right descriptive data from a digitised document makes finding relevant content much easier, and helps maximise research efficiency.

Preservation

Following digitisation, physical documents will not need to be accessed as often, therefore reducing potential damage caused by handling.

Interaction

Digitised content is versatile and can be used to communicate with customers across a range of channels. It is a particularly strong tool for social media, opening dialogue with customers and encouraging activities like crowdsourcing and blogging.

Integration

Records can be integrated with digital systems and made readily discoverable within digital catalogues.

Disaster recovery

Paper records are vulnerable to many risks. Digital copies ensure information remains retrievable if a site disaster leads to loss of analogue formats.

Disadvantages/ Challenges

- Skills & knowledge
- Digital infrastructure
- Funds
- Copyright
-