


Subject: Library and Information Science

Production of Courseware

 -Content for Post Graduate Courses



Paper No: 07 Information Storage and Retrieval

Module : 02 Basic Concept and Components of IR Systems



Development Team

Principal Investigator
&
Subject Coordinator

Dr. Jagdish Arora, Director
INFLIBNET Centre, Gandhinagar

Paper Coordinator

Prof Devika P Madalli, Professor, Documentation
Research and Training Centre (DRTC), Bangalore

Content Writer

Dr Pijush Kanti Panigrahi,
Dean and Professor, University of Calcutta, Calcutta

Content Reviewer

Prof Devika P Madalli, Professor, Documentation
Research and Training Centre (DRTC), Bangalore

Basic Concepts and Components of Information Retrieval Systems

I. Objectives

Objective of this module is to:

- Introduce the basic concepts of IR systems and their components
- Brief about methods that enable users to find out relevant information from an organized collections of resources.
- Introduce various features of IR systems that help in easy retrieval of documents from interdisciplinary field.
- Introduce different functions of IR system which deals with various format(i.e. text, audio, image, and video) of information .

II. Learning Outcomes

After reading this module:

- The student will gain the knowledge about basic concepts and characteristics of IR.
- The learner will understand the various components of information retrieval processes.
- The reader will gain the knowledge of various tools and technologies used in IR systems.
- The reader will gain the knowledge of different types of information retrieval system.
- The student will understand measures of retrieval efficiency in terms of Precision and recall.

III. Structure

1. Introduction
2. Features of IR System
3. Scope of IR System
4. Types of IR System
5. Functioning of IR System
6. Basic components involved in IR process
 - 6.1. Indexing: Creating Document Representations
 - 6.2. Query Formulation: Creating Query Representations
 - 6.3. Matching the query representation with entity representations
 - 6.4. Selection
 - 6.5. Relevance Feedback and Interactive Retrieval
7. Purpose and Function of IR System
 - 7.1 Purpose
 - 7.2 Function
8. Summary
9. References

1. Introduction

We conceptualize the knowledge system into which an IR system is implanted to consist of three of component parts: a) people in their role as information-processors, b) documents in their role as carriers of information, and c) topics as representations. We are connected with the life cycle of each of these three objects and with the dynamic interactions among them. Thus the objective of an information retrieval system is to enable users to find relevant information from an organized collection of documents. In fact, most information retrieval systems are, truly speaking, document retrieval systems, since they are designed to retrieve information about the existence (or non-existence) of documents relevant to a user query. Lancaster

comments that an information retrieval system does not inform (change the knowledge of) the user on the subject of their enquiry; it merely informs them of the existence (or non-existence) and whereabouts of documents relating to their request. However, this notion of information retrieval has changed since the availability of full text documents in bibliographic databases. Modern information retrieval systems can either retrieve bibliographic items, or the exact text that matches a user's search criteria from a stored database of full texts of documents. Although information retrieval systems originally meant text retrieval systems, since they were dealing with textual documents, many modern information retrieval systems deal with multimedia information comprising text, audio, images and video. While many features of conventional text retrieval systems are equally applicable to multimedia information retrieval, the specific nature of audio, image and video information has called for the development of many new tools and techniques for information retrieval. Modern information retrieval deals with storage, organization and access to text, as well as multimedia information resources.

2. Features of IR Systems

An information retrieval system is developed in order to help users to discovery relevant information from a storehouse containing collection of documents. The idea of information retrieval assumes that there exist several documents or records comprising data that have been arranged in a suitable order for easy retrieval. The storehouse contains many bibliographic information, which is quite different from other kinds of information or data. Let us consider some examples such as if we maintain a database of information about an institution or a supermarket, all we have are the different types of records and related facts, such as, for a college, names of students, faculties, staffs, their positions, qualifications and so on; in the case of a supermarket, names of different commodities, market prices, quantity and so forth. For such scenarios the retrieval system is designed to search for and retrieve specific facts or data, such as the qualification of a particular faculty, or the market price of a certain type of rice. Conventional database management systems, such as Access, Oracle, MySQL, etc, deal with structured data, where the arrangement or structuring of data takes place on the basis of the specific attributes of the data elements. For example, in a database of recipe, the various data elements could be the attributes of specific recipe records, such as recipe instruction, recipe yield, type of recipe, ingredients required, etc. In contrast to this, a database of items sold in a supermarket could be the name of the item with its barcode, manufacturer, supplier, price and so forth. So, the first database in this example will be structured according to the specific attributes of institution, whereas for the other database will be organized in accordance to the attributes of specific commodities. The main objective of these databases is to enable the user to search for specific records that be matched with one or more specific conditions or search criteria, for example, details of a certain recipe containing a particular ingredient; details of a specific product within a specific range of market price; a list of all the faculties that are involved with a specific course; or the products of a particular type grown at certain states in the country, for example basmati rice available in north eastern states in India.

Unlike a conventional database management system, an information retrieval system deals with unstructured data also. The main purpose of designing an information retrieval system is to meet the user requirements. It enables in document retrieval in-order to answer to the users' queries. The retrieved information can be in represented in different forms. The database can store abstracts of some bibliographic resources or full texts of documents, such as journal articles, conference proceedings, newspaper articles, textbooks, encyclopedias, legal documents, and statistical records, etc along with audios, graphics, images and videos information. No matter what the database may contain, be it bibliographic resources, full-text documents or multimedia information – the system assumes that there exists a target group of users for whom the system is designed and fulfill their requirements. Users may have certain queries or information needs, and they search for required information, the information retrieval system should be able to fetch the necessary bibliographic references of those documents bearing the required information; some systems also retrieve the actual text, image, table or chart relevant to the information needs of the user.

Let us consider a very simple example to understand the basic functioning of an information retrieval system. Let us consider a simple scenario where a user wants to discovery information about a term, say 'nature', in a book. One approach would be to start with the very first word in the first sentence present in the book, and continue to search for the term 'nature' until we find it or we come to the end of the book. However, in real life, this is not the scenario. Instead, to save our time we preferably use an index – the 'back-of-the-book index' – to look for an ideal match for the search term, and if we find a match then we take note of the corresponding references – the page number(s) where the term occurs – and we move to the specific page(s)

to find the information and the given context. In their simplest form, most information retrieval systems work in this way.

Although historically information retrieval systems were established to help end users find relevant information from bibliographic and textual databases, in this 21st century information retrieval systems is used in almost each and every facet of our daily lives, for example, to retrieve a song on YouTube or e-mail received or sent on a specific date; to find sms sent to or by a particular person; to find a person's entity on the web; to search for an e-book in an online library catalogue or in a digital library; to search for a book available for purchase in Amazon.com and so on.

3. Scope of IR System

- a. **Unstructured Information:** This information either does not have a pre-defined data model or is not organized in a pre-defined order. Unstructured information is typically text-heavy, but may contain datasets such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional computer programs as compared to data stored in fielded form in databases or annotated (semantically tagged) in documents (Wikipedia). Examples of "unstructured data" may include books, journals, documents, metadata, health records, audio, video, analog data, images, files, and unstructured text such as the body of an e-mail message, Web page, or word-processor document. While the primary content being conveyed does not possess a defined structure, it generally comes packaged in objects (e.g. in files or folders or documents, ...) that themselves have some metadata and are thus a combination of structured and unstructured data, but normally it is referred to as "unstructured data".[7] For example, if we consider an HTML web page it is tagged, but HTML mark-up typically serves the purpose of presentation. It is not being able to capture the significance or function of tagged elements in order to assist automated processing of the information content of the page. XHTML tagging does allow machine processing of elements, although it typically does not capture or convey the semantic meaning of tagged terms. There are several techniques such as data mining and text analytics and noisy-text analytics, information visualization which give different methods to search for patterns in, or otherwise interpret from the available unstructured information. The most popular technique for providing structure to several unstructured resources usually involve manual tagging with metadata or part-of-speech tagging for further text mining-based structuring. Unstructured Information Management Architecture (UIMA) provides a common model for processing this information to extract meaning and create structured data about the information [5].
- b. **Structured Information:** It is information that is already structured in fields, such as "name", "age", "gender", "hobby", "address", "profession", "salary". This is the typical example of what we find in a record of a relational database table. When information is organized in a structured form, it is usually relatively easy to search it, since one can directly query the database : give me the list of names whose profession is student in the table PERSON, where age is greater than 25 and name starts with the letter B. Structured data first depends on creating a data model – a model of the types of business data that will be recorded and how they will be stored, processed and accessed. This includes defining what fields of data will be stored and how that data will be stored: data type (string, integer, etc) and any restrictions on the data input (number of characters; restricted to certain terms such as Mr., Ms. or Dr.; M or F).

Structured data can be handled easily as they can be easily entered, stored, queried and analyzed. Due to the increase in cost and performance limitations of storage, memory and processing, relational databases and spreadsheets using structured data were the only way to effectively manage data. It is a common phenomena where anything that couldn't fit into a tightly organized structure requires to be stored on paper in a filing cabinet. Structured data is often managed using Structured Query Language (SQL) – a programming language created for entering, managing and querying data in relational database management systems. Originally developed by IBM in the early 1970s and later developed commercially by Relational Software, Inc. (now Oracle Corporation).

It is to be noted that information retrieval systems and database systems merely find what is already there: for example, from students databases his/her marks, from the marks it can point to the position of the student

in class. An expert system on the other hand goes beyond just finding facts- it creates new information by inference: it identifies a student and gauges their merit in different subjects and their future prospect.

4. Types of IR System

IR has concentrated more on finding the documents consisting of written text; much IR research focuses more specifically on text retrieval – the computerized retrieval of machine-readable text without human indexing. But it has spread across other interesting areas. Such as:

Speech Retrieval: Speech is an information-rich element of multimedia. Now there exist several techniques where information can be extracted from a speech signal in a number of different ways. Thus there are several well-established speech signal analysis research fields. These fields include speech recognition, speaker identification, voice detection, sentiment analysis and fingerprinting. The information that can be extracted from tools and methods developed in these fields can greatly enhance multimedia systems and help mankind in various aspects.

Cross language information retrieval: It is an application area of information retrieval, which deals with fetching information written in a particular language different from the language of the user's query. E.g., Using Hindi queries to retrieve English documents. It is one of the challenging fields and a lot of research is going on in this area.

Question-answering IR system: It is a computer science discipline within the domains of information retrieval and natural language processing (NLP), which is involved with building systems that automatically answer questions posed by humans in a natural language. A QA implementation, usually a computer program, may construct its answers by querying a structured database of knowledge or information, usually a knowledge base. More commonly, QA systems can pull answers from an unstructured collection of natural language documents. (Wikipedia)

Image Retrieval: It is part of sub-field of information retrieval. It helps the retrieval system for browsing, searching and retrieving images from a large database. The database may contain only digital images, images along with text or may contain other types of resources like graphics, videos, audios along with the image, etc. Most popular and common techniques of image retrieval utilize some method of adding metadata such as use of captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words. The manual process of image annotation is not only time-consuming but is also a laborious and expensive affair; to address this; there has been a large amount of research done on automatic image annotation and image detection. Moreover, with the increase in usage of social networks and a shift in paradigm from web to data web warrants new technology framework have inspired the evolution of several web-based image annotation tools.

Music Retrieval: Music information Retrieval (MIR) is the interdisciplinary field of retrieving useful information from music. MIR, although small yet it is a growing field of research with many real-world applications. Several researchers working in MIR may come from different background which includes computer science, instrumentation, musicology, psychology, academic music study, signal processing, machine learning or some combination of these.

In addition to the above mentioned retrieval systems, IR also deals with any type of entity or object: work of art, software, courses offered at a university, people, products of any kind, etc. Text, speech or images, printed or digital, carry information, hence information retrieval.

5 Functioning of IR System

An information system essentially makes ensure that users should be satisfied with the service. The system will be able to accomplish tasks, solve problems, and make decisions, based on the user needs. In short an information retrieval system should 1) find out the requirement of a target group of users, 2) a collection of relevant documents and other information resources should be made and indexed appropriately, and 3) match documents with user needs in-order to fetch relevant documents. To determine the user needs, it involves in

studying information needs of users in general as a basic for designing responsive system (such as determining what study materials required for library and information science students typically need to do assignments in content management), and actively soliciting the needs of specific users, expressed as query descriptions, so that the system can provide the information. To have a successful retrieval system, it should figure out what information the users require to solve a problem. Query matching involves in mapping a query description with relevant documents in the collection; this is the task of the IR system.

All operations pertaining to information retrieval surround around usefulness and relevance of documents. The use of a document is dependent upon on three major things, topical connectedness, applicability, and originality. A resource is considered to be topically significant for a particular context, question, or task if it consists of information that either instantly provides answer to the query or can be used, in combination with other information, to infer an answer or perform the task. The appropriateness of the answer completely depends upon the user for a given context. It is original if it provides an input to the user's knowledge. Let us consider a simple situation where, a basketball player is important for a team if his abilities and playing style fit the team strategy, applicable if he is compatible with the coach, and possess unique talent if the team is missing a player in his position.

Utility can be measured in monetary terms: "To what extent the document is useful for the user?" "What is the role of the player for a team?" "What is the recall and precision of the search engine"? From the literatures point of view, the term "relevance" is used for different purpose; it can indicate utility or topical relevance or pertinence. Many IR systems focus on finding topically relevant documents, leaving further selection to the user.

Relevance is a matter of degree; some documents are highly relevant and indispensable for the user as it serves the purpose of the users' need; others may not contribute much to the users' requirements. For example if a user seeks information for 'orange' which is a fruit, all the documents about the fruit orange are relevant. Other documents may have the word 'orange' but might not indicate about the fruit (see ranked retrieval in the section on Matching). From relevance assessments; measures of retrieval performance can be computed such as

$$\text{recall} = (\text{relevant items correctly retrieved}) / (\text{all relevant items in the collection}).$$

$$\text{discrimination} = (\text{irrelevant items correctly rejected}) / (\text{all irrelevant items in the collection})$$

$$\text{precision} = (\text{relevant items retrieved}) / (\text{all items retrieved})$$

Evaluation studies normally use recall and precision or a combination of both; but there exists a lot of argument whether these can be considered as the best measures for information retrieval systems.

6. Basic components involved in IR process

An IR system performs retrieval operation by indexing documents and designing queries, thereby leading to representation of documents and representation of queries, respectively; the system then matches the indexed documents with that of user query and displays the matched documents found and the user selects the relevant items. These operations are tightly intertwined and are directly dependent on each other. The search process often goes through several iterations: several cases feature similarity measurement is used in order to distinguish the relevant documents from irrelevant ones and thereby it is used to improve the query or the indexing (relevance feedback).

6.1 Indexing: Creating Document Representations

Indexing (from the library science point of view can also be referred as cataloging, metadata assignment, or metadata extraction) is the manual or automated process creating indexes for record collections. Having indexes allows researchers to more quickly find records for specific individuals; without them, researchers might have to look through hundreds or thousands of records to locate an individual record. We focus here on subject indexing – act of describing or classifying a document by index terms or other symbols in order to indicate what the document is about, to summarize its content or to increase its find ability. In other words, it is about identifying and describing the subject of documents. Indexes are constructed, separately, on three

distinct levels: terms in a document such as a book; objects in a collection such as a library; and documents (such as books and articles) within a field of knowledge (Wikipedia). Indexing can also be document-oriented – the indexer captures what the subject of the document, or request-oriented – the indexer assesses the document's relevance to other features of interest to users; for example, indexing the recipes in a cookbook in accordance to the course-type or meal or primary ingredients, etc making the resource interesting for the users. Abstracting is related to indexing – act of providing a summary of the full document giving the main content of the document or sometimes it may also include important results (informative abstract, summary). A lot of researchers have their interest on designing algorithms for building automatic summarization.

Automatic indexing begins with feature selection and extraction, this demands in extracting all the words from a text, this is followed by elimination of stop-words (words which are filtered out prior to, or after, processing of natural language data (text)). There is not one definite list of stop words which all tools use and such a filter is not always used. Some tools specifically avoid removing them to support phrase search), stemming (the process for reducing inflected words to their stem, base or root form—generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root), counting (using only the most frequent words), and mapping to concepts using a thesaurus or ontology (Wikipedia). In case of images, extractable features include color distribution, texture or shapes detection. For music, extractable features comprises of frequency of occurrence of notes or chords, harmonies, melody, main pitch, beats per minute or rhythm in the piece.

Features are generally processed further for retrieval. The system makes use of a classifier that links the raw or refined features with that of a descriptor from a pre-established index language. A classifier can be built manually by making each descriptor act as a query description and building a query formulation for it. Moreover a classifier can be built automatically by making use of training sets, for example, the list of documents for biotechnology, for machine learning of what features predict what descriptors. There exist several techniques that enable prediction of different words and word combinations by using the same descriptor, thereby making it easier for users to find all relevant documents on a given context. The process of assigning documents to (mutually exclusive) classes of a classification is also known as text categorization. Analyzing the documents having similar features and clustering them in one group lead to identification of unique classes in which the documents belong. These are some initial steps of document classification.

6.2 Query Formulation: Creating Query Representations

Information Retrieval means making use of the available information in-order to anticipate the extent to which a given document is significant or useful for a particular users' information need as outlined in a free-form query description, also called topic description or query statement. A user's query can be transformed, manually or automatically, into a formal query representation (also called query formulation) when combined with features, it helps to predict the usefulness of a document with respect to the query. The information need of the users can be identified by analyzing the query in terms of the system's conceptual schema, ready to be matched with document collected in the database. A query may be in search of text words or phrases that the system should acknowledge and search (free-text search) or any other entity feature, such as descriptors assigned from a controlled vocabulary, an author's organization, or the title of the journal where a document was published. A query can simply give features in an unstructured list (for example, a "bag of words") or combine features using Boolean operators (structured query). Examples: The Boolean query specifies three conditions, AND, OR, NOT. If a query contains AND operator it indicates narrow search and retrieve records containing all of the words it separates. Similarly if the query includes OR operator it broadens the search and retrieves records containing any of the words it separates. The symbol '|' can be used instead of 'or' (e.g., 'mouse | mice | rat' is equivalent to 'mouse or mice or rat'). Lastly NOT operator indicates narrow search and retrieve records that do not contain the term following it. If there exists some relevant documents, the system can use them as a training set to build a classifier with two classes: relevant and not relevant. These relevant and non-relevant documents will lead to the measurement of recall and precision. The requirement for the information need and formulating the query often acts as a cup and plate as they move together, directly dependent upon each other. An IR system can show a subject hierarchy for browsing and finding good descriptors, or it can ask the user a series of questions and from the

answers construct a query. For buying an online food item, the system might ask the following three questions:

- What kind of food do you prefer (vegetarian, non-vegetarian, ...)?
- Are you allergic to any particular ingredient (prawn, carrot, cumin seeds,...)?
- What kind of cuisine you prefer (Italian, Indian, American,...)?

The system should help the users by suggesting synonyms and narrower and broader terms from its thesaurus. This will help the users to visualize all the features to consider, without which it would not have been feasible. Throughout the search process, users further clarify their information needs as they read titles and abstracts.

6.3 Matching the query representation with entity representations

The document relevance is predicted by identifying the relevant features of the query with that of the document. In case of an exact match the system is able to mark the documents that satisfy all the conditions of a Boolean query (it predicts relevance as 1 or 0). In-order to improve recall, the system can make use of elaborating the synonyms (if the query asks for dessert, it finds sweets as well) and hierarchic expansion or inclusive searching (it finds dairy product as well). Since relevance or adequacy is a matter of degree, many information retrieval systems (including most Web search engines) rank the retrieved results by a score of expected relevance (ranked retrieval). Consider the query “Study of concept analysis in information retrieval”. In this case each term's contribution is a product of three weights: The weight of the query term (the significance of the term to the user), the term frequency (tf) (the number of occurrences of the term in the document, synonyms count also), and the infrequency of the term in context of the document or inverse document frequency (idf) on a logarithmic scale is measured. If document frequency = .01 (1 % or 1/100 of all documents, the term is to be included), then $idf = 100$ or 10^2 and $\log(idf) = 2$.

6.4 Selection

The user searches for the most relevant result and selects the appropriate items. Results can be organized in rank order (the search process can be stopped once the users' need is fulfilled); in case of groupings the documents based on subject, automatic classification scheme or clustering techniques (similar items can be examined side by side) can be applied. The display of titles along with the abstract with key terms highlighted is considered to be the most useful (as title alone is too short, the full text too long). For certain scenarios users may require assistance while making the connection between an item found and the task at hand.

6.5 Relevance Feedback and Interactive Retrieval

Once the user has evaluated the significance of a few items found, the query can be made better. The system can thereby provide assistance for the users in enriching the query by displaying a list of features (assigned descriptors; text words and phrases, and so on) found in many relevant items and another list from irrelevant items. In some cases the system can automatically improve the query by identifying those unique features which can distinguish between relevant from irrelevant items and thus are good predictors of relevance.

7. Purpose and Function of IR System

7.1 Purpose

An information retrieval system serves the purpose to retrieve the resources or information required by the target audience. It is also important that the right information should reach to the right people at right time. Thus, the main aim of an information retrieval system is to collect and organize information in one or more fields in order to help the users to access the retrieved resources. The use of information retrieval systems can be explained considering a simple scenario:

- A writer puts forth his/her idea on a document using some concepts for a given context.

- Somewhere around the globe there might be some target audience or a person who is in need of that unique idea but is unable to find it; in other words, some people is ignorant of the ideas put forward by the author in their work.
- Here Information retrieval systems bridge the gap by matching the writer's ideas expressed in the document with that of the users' requirements or demands for that idea.

Thus, an information retrieval system functions as a bridge between the world of creators or generators of information and the users who are ignorant of that information. Hence some researchers state that information retrieval is an information-communication system.

7.2 Function

An information retrieval system deals with different sources of information on one hand and on the other hand it has to cater to several users' requirements. It must:

- available contents are to be analyzed in the information sources as well as the users' queries, and then
- then the user queries are matched with the available document in-order to retrieve the relevant resources.

The different functions of information retrieval systems are as follows:

- To identify the information (sources) relevant to the areas of interest of the target users' community; this is a challenging job especially in the web environment where virtually everybody in the world can be the potential user of a web-based information retrieval system.
- To analyze the contents of the sources (documents); this is becoming increasingly challenging as the size, volume and variety of information sources (documents) is increasing rapidly; web information retrieval is carried out automatically using specially designed programs called spiders.
- To represent the contents of analyzed sources in a way that matches users' queries; this is done by automatically creating one or more index files, and is becoming an increasingly complex task due to the volume and variety of content and increasing user demands.
- To analyze users' queries and represent them in a form that will be suitable for matching the database; this is done in a number of ways, through the design of sophisticated search interfaces including those that can provide some help to users for selection of appropriate search terms by using dictionary and thesauri, automatic spell checkers, a predefined set of search statements and so forth.
- To match the search statement with the stored database; a number of complex information retrieval models have been developed over the years that are used to determine the similarity of the query and stored documents.
- To retrieve relevant information; a variety of tools and techniques are used to determine the relevance of retrieved items and their ranking.
- To make continuous changes in all aspects of the system, keeping in mind the rapid developments in information and communication technologies (ICTs) relating to changing patterns of society, users and their information needs and expectations. (Chowdhury)

8. Summary

In short, IR involves in finding some desired information which is stored in a storehouse commonly called database. A typical IR system should meet the following functional and nonfunctional obligations. It must enable the user to create, insert, modify and delete, documents in the database. It should be a platform for the users to search for documents by entering queries, and examining the retrieved documents. An IR system will typically need to support large databases, some in the megabyte to gigabyte range, and retrieve relevant documents in response to queries interactively--often within 1 to 10 seconds. This field has come out with several path-breaking results as several research labs are working to develop many modern techniques for better precision.

9. References

1. Sparck Jones, K. and Willett, P., Overall Introduction. In Sparck Jones, K. and Willett, P. (eds) *Readings in Information Retrieval* , San Francisco, Morgan Kaufmann Pub. Inc., 1997, 1–7.
2. Parsaye, K., Chignell, M., Khosafian, S. and Wong, H., *Intelligent Databases: object- oriented, deductive hypermedia technologies* , New York, John Wiley, 1989.
3. Lancaster, F. W., *Information Retrieval Systems* , New York, John Wiley, 1968.
4. Belkin, N. J., *Anomalous States of Knowledge as a Basis for Information Retrieval*, *Canadian Journal of Information Science* , 5 , 1980, 133–43.
5. Meadow, C. T., Boyce, B. R., Kraft, D. H. and Barry, C., *Text Information Retrieval Systems* , 3rd edn, London, Academic Press, 2007.
6. Lancaster, F. W., *Information Retrieval Systems: characteristics, testing, and evaluation* , 2nd edn, New York, John Wiley, 1979.
7. Kent, A., *Information Analysis and Retrieval*, 3rd edn, New York, Becker and Heys, 1971.
8. Vickery, B. C., *Techniques of Information Retrieval* , London, Butterworth, 1970.
9. Vickery, B. and Vickery, A., *Information Science Theory and Practice* , London, Bowker-Saur, 1987.
10. Rowley, J., *The Basics of Information Systems* , 2nd edn, London, Library Association Publishing, 1996.
11. Liston, D. M. and Schoene, M. L., *A Systems Approach to the Design of Information Systems*. In King, D. W. (ed.) *Key Papers in the Design and Evaluation of Information Systems*, New York, Knowledge Industry, 1978, 327–34.
12. Voorhees, E. and Harman, D. (eds), *Text Retrieval Conference* , Cambridge, MA, MIT Press, 2005.
13. Sparck Jones, K., *What's the Value of TREC: is there a gap to jump or a chasm to bridge?* *SIGIR Forum* , 40 (1), 2006, 10–20.
14. Hearst, M., *Search User Interfaces* , Cambridge, Cambridge University Press, 2009
15. Chowdhury, G. G., *Introduction to Modern Information Retrieval*, 3rd ed. London, 2010.