



University of Mosul
College of Art
Dept. of information and knowledge Techniques

Course Name: Organization
Information containers
Stage: Masters
Lecturer Name: Prof. Rafal N. A.
School Study: 2025-2026

Thesaurus

It is a list of authority descriptors or indexing terms used in an information system. It also serves as the tool relied upon by both the indexer—to obtain the appropriate terms or descriptors for describing the content of a document—and the researcher or end-user of the system during the retrieval process. Thus, one can say that the link between the indexer and the researcher is the thesaurus, which provides a high degree of efficiency in storage and retrieval, especially after traditional subject headings proved inadequate for this purpose.

There are several definitions of the thesaurus, among the most important being those by Surace and Rolling. Surace defined it as "a list of agreed terms, or descriptors, that serve specialists and identify the concepts found in publications, which, when organized and displayed in a certain way, show relationships of a semantic or hierarchical nature." Rolling defined it as "a controlled language built for use in information storage and retrieval systems." The International Organization for Standardization (ISO) indicated that "functionally, a thesaurus is a means of controlling the terms used to translate from the natural language of documents, indexers, or users to a more complex system language (documentation language); structurally, it is a controlled, dynamic language consisting of terms interconnected semantically or relationally, covering a branch of knowledge."

We find that the thesaurus imposes the terms to be assigned to documents and suggests terms that the indexer may have considered without assistance—that is, it imposes the language that the researcher must use, guiding them from a non-standard term to the controlled term. The user is referred to more commonly used terms through references such as "USE" and "Narrower Term (NT)." To facilitate retrieving a set of related documents, the thesaurus highlights terms that are conceptually linked, much like the use of "SEE ALSO" references in alphabetical lists.

Therefore, the descriptors permitted by the thesaurus for indexing document contents to be stored in a computer are, in fact, terms that represent concepts or combinations of concepts, or terms denoting single entities, such as subjects, personal names, geographic names, etc.

Thus, it can be said that thesauri resemble traditional subject heading lists in that they are alphabetical lists of terms, but they differ in form, structure, and the way terms are displayed. The primary function of a thesaurus is to control the specific indexing language—both manually and electronically—in information retrieval systems. In recent years, the process of thesaurus construction has evolved significantly to



University of Mosul
College of Art
Dept. of information and knowledge Techniques

Course Name: Organization
Information containers
Stage: Masters
Lecturer Name: Prof. Rafal N. A.
School Study: 2025-2026

include combinatorial methods based on analytical techniques, which will be discussed under the development of thesauri.

Functions and Tasks of the Thesaurus

It is worth noting that thesauri are designed primarily for the indexer first, then the researcher, and finally the end-user. The most important features and functions of thesauri are:

1. They allow the indexer to represent the subject content of documents in a consistent, standardised, and more complete manner, at varying levels of generality and from multiple technical perspectives.
2. They serve as the tool used by the indexer, researcher, or system user to select appropriate, standardised descriptors that match their needs and are consistent with the system's descriptors and those chosen during the indexing process.
3. They provide the researcher or user with the means to modify their search strategy in order to achieve high recall and precision ratios according to the requirements of the search.
4. They offer explanatory notes to resolve problems faced by either the indexer or the end-user, along with providing important definitions.
5. They provide references in accordance with thesaurus rules, in addition to offering broad entry points that guide the end-user.

Shalmica, in his study on the thesaurus, adds that "it can serve two aspects:

- First: imposition – meaning imposing the terms that should be given to documents.
- Second: suggestion – meaning suggesting terms to be considered instead of, or in addition to, those considered by the indexer."

It must be noted here that the thesaurus encompasses both aspects: it is prescriptive in one respect and suggestive in the other.



University of Mosul
College of Art
Dept. of information and knowledge Techniques

Course Name: Organization
Information containers
Stage: Masters
Lecturer Name: Prof. Rafal N. A.
School Study: 2025-2026

Sections of the Thesaurus

Generally, a thesaurus is divided into the following main sections:

Section One: The Hierarchical List – This list represents a system of concepts and hierarchical relationships among them, sometimes referred to as categories. It typically includes broad headings for fields and major facets of the thesaurus.

Section Two: The Alphabetical List of Descriptors – This is the main section of the thesaurus, as it contains references between terms and related terms, or references from unused terms to commonly used terms.

Section Three: The Faceted Arrangement of Descriptors – Also known as the classificatory arrangement of terms. It is similar to a faceted classification scheme, as it reassembles terms that were arranged alphabetically in the second section, grouping them according to the categories or groups to which they belong.

Section Four: The Rotated or Permuted List of Descriptors – In this section, terms are arranged alphabetically, and under each term, all other descriptors in which that term appears are listed, along with an indication of the category to which the descriptor belongs.

Thesaurus Construction

The proper performance of a thesaurus depends on the availability of standardised terms suitable for use in the information system, and on displaying the various relationships between these terms—what is known as thesaurus construction. The two methods used in constructing a thesaurus are:

A. The Theoretical or Deductive Method – This method determines the main subjects and their subdivisions by collecting terms from the top down, through special committees assigned to this task.



University of Mosul
College of Art
Dept. of information and knowledge Techniques

Course Name: Organization
Information containers
Stage: Masters
Lecturer Name: Prof. Rafal N. A.
School Study: 2025-2026

However, a drawback of this method is the difficulty of envisioning all necessary hierarchical categories, and it is not certain that the property indicating the division of a genus will lead to a classification that meets user needs. For example, the class "adult games" might be subdivided by material, colour, user age, animation style, etc., and one characteristic (perhaps animation style) may be more important to a game designer than other characteristics.

B. The Applied or Empirical Method – This is the opposite of the previous method, determining main subjects and their subdivisions by collecting terms from the bottom up—known as the inductive method (through actual document indexing). This method is more reliable than the former because the terms appearing in the subject field are collected from various sources, such as specialised dictionaries, encyclopaedias, periodical indexes, abstracts, and also by consulting subject specialists or potential users.

This method is distinguished by the fact that it does not allow the formation of hierarchical categories for terms until their usefulness to the target group for which the thesaurus is built has been confirmed. For instance, subdividing "adult games" by colour might be theoretically possible, but it may not matter to the game designer for whom the thesaurus was constructed. It is noted that in this method, two essential elements must be provided when collecting terms: the literary warrant and the user warrant. The first element involves gathering subject-field terms by referring to reference sources that may contain a high concentration of terms, requiring a reasonable degree of currency. The second element emphasises that collected terms should genuinely represent user needs, requiring the identification of terms indicative of their subject interests, in addition to terms extracted from the literature that complement—not replace—them.

In this case, document titles and abstracts can be considered valuable sources of terms, along with questionnaires sent to all relevant users or a sample thereof to gather data indicating their interests.

Development of Thesauri

When discussing thesauri, one must mention Roget's Thesaurus, the most famous thesaurus known since 1852 under the title Thesaurus of English Words & Phrases, which continues to be published to this day. Subsequently, various types and forms of thesauri have appeared. When we speak of types of thesauri, we refer to subject thesauri, which may be either general and broad in scope—such as the TEST thesaurus in engineering and technological sciences, or the UNESCO Thesaurus—or specific and specialised, such as the ISO Thesaurus, the ERIC Thesaurus, and other foreign thesauri.



University of Mosul
College of Art
Dept. of information and knowledge Techniques

Course Name: Organization
Information containers
Stage: Masters
Lecturer Name: Prof. Rafal N. A.
School Study: 2025-2026

As for Arabic thesauri in this field, although they lag far behind foreign ones and suffer from a clear shortage in producing such tools—much like Arabic subject heading lists, some of which are little more than translations of foreign thesauri—the Arab scene has recently witnessed tangible progress in producing thesauri. These include general ones such as the League of Arab States Thesaurus (1987), the Al-Faisal Thesaurus (1994), and the Expanded Thesaurus (1995), as well as specialised ones like the Arabic Thesaurus of Earth Sciences (1988) and the AGRIVOC Agricultural Thesaurus (1993). It should be noted that the aforementioned Arabic thesauri are among the most important and recent available in the Arab arena, as they are subject to updates.

As a result of developments in online information retrieval technologies, and after the production of these thesauri on computers in various forms—microform or CD-ROM—some researchers proposed using microthesauri to provide a basic classification structure or a broader subject term to meet the specific needs of sub-fields. Today, thesauri are available on the Internet in electronic form. Prominent examples include Roget's Thesaurus, the Cancer Information Thesaurus (CIT)—whose terms were taken from the Medical Subject Headings (MeSH) list—and the African Libraries Thesaurus, which relies on a list of subject headings specific to African libraries, the UNESCO Thesaurus, and the Library of Congress thesauri, including the Art and Architecture Thesaurus (AAT), among others. As for Arabic thesauri available online, the most notable are the Expanded Thesaurus, the University Thesaurus, and the AGRIVOC Agricultural Thesaurus.

Perhaps the most important feature offered by online thesauri, such as the LC Thesaurus, is providing subject heading pathways or threads to link secondary subdivisions related to the main subject. This helps the researcher or user avoid errors or incorrect retrieval when the relationship between two or more subjects has been previously identified by the researcher but turns out not to meet their needs. Using subject heading threads provides the researcher with a single index entry covering multiple subjects distributed across different places and dates. They are also useful for guidance when the results of a keyword search are too lengthy, which is a key advantage of using pre-coordinated terms, as they easily incorporate the main subject term and its related aspects.

Comparison Between Thesauri and Subject Heading Lists

1. In terms of use: Subject heading lists are used in both manual and automated cataloguing, whereas thesauri are generally used in indexing.



University of Mosul
College of Art
Dept. of information and knowledge Techniques

Course Name: Organization
Information containers
Stage: Masters
Lecturer Name: Prof. Rafal N. A.
School Study: 2025-2026

2. In terms of content: Subject heading lists contain subject headings in their various forms and references, while thesauri contain descriptors linked by relationships with one another. A thesaurus typically uses a set of analytical and synthetic terms according to rules for constructing subject forms.
3. In terms of subject coverage: Subject heading lists may be general and comprehensive in covering fields of knowledge, or specialised in covering a specific field or subject. Thesauri, on the other hand, often cover specialised subject areas.
4. In terms of structure and composition: Subject heading lists contain subject headings and various references showing subject relationships among the headings in the list. Thesauri, however, do not concern themselves with all subject relationships between descriptors in a general sense, but rather with the necessary relationships depending on the type of documents indexed. They also avoid inverted phrases and tend to use direct terms in general, while their references are more precise and rigorous than those in subject heading lists. Thesauri rely fundamentally on hierarchical relationships, moving from broader to narrower terms.
5. In terms of creation and arrangement: The creation of subject heading lists depends on the intellectual output acquired by the library, while the creation of a thesaurus relies on authoritative scientific references to restrict the descriptors used in a specific subject field. Validation is carried out by a committee of specialists in that subject area. As for arrangement, subject headings in lists are usually arranged alphabetically, while thesauri may be organised either alphabetically or by classification.
6. In terms of scope of application: It is observed that information environments dealing with precise, specialised subjects cannot do without designing a thesaurus tailored to the nature and needs of their users. Subject heading lists, however, are widely used in various types of libraries.