Features of Using Thesaurus in Modern Information Search Systems

Prenov Barlikbay Barakbaevich[1], Nurmaxanov Kayrat Erpayizovich[2], D.Djumabaev[3], Allaberganov Baxram Ismailovich[4]

[1]Doctor of physical and Mathematical Sciences, professor
[2]Doctor of philosophy in Pedagogical Sciences (PhD), docent
[3]University of Exact and Social Sciences, Tashkent, Uzbekistan, Email: djumabaevd@mail.ru
[4]Senior Lecturer, Nukus State Pedagogical Institute named after Azhiniyaz, Nukus city, Uzbekistan

ABSTRACT

This article describes the different approaches to the formalization of the thesaurus, information retrieval thesauri are described particularly with thesauri in information systems, as well as problems arise, to implement the requirements of the thesaurus. The peculiarities and differences of classifiers resources and conventional terminology and linguistic thesauri. We also consider the organization of user interfaces for use with a thesaurus and use them when searching for the information system, as well as thesauri administration interfaces.

Keywords: information retrieval thesauri are described particularly with thesauri in information systems, as well as problems arise, to implement the requirements of the thesaurus.

INTRODUCTION

To describe any subject area, a certain set of terms is always used, each of which denotes or describes a concept from this subject area. A set of terms describing a given subject area, indicating the semantic relations (links) between them, is a thesaurus. Such relations in a thesaurus always indicate the presence of a semantic (semantic) connection between terms.

The main relationship (link) between terms in a thesaurus is the connection between broader (more expressive) and narrower (more specialized) concepts. Two subtypes of this relationship are often distinguished:

One term denotes a concept that is part of the concept denoted by another term (for example, "science" and "mathematics", "mathematics" and "number theory")

One term denotes an element of a class denoted by another term ("mountain regions" and "Caucasus").

This relation on a set of terms is a partial order relation, i.e. a set of terms with such connections forms an acyclic graph, or a polyhierarchical structure. There are also other connections between terms. For example, one concept or notion can be designated by several terms that are synonyms. Some terms can be antonyms for others. Often, among the terms related to one concept, a single (for each thesaurus language) most preferred (most suitable) term is distinguished, which best characterizes or designates this concept. The remaining terms are less preferred (less suitable).

In addition to the above, there may also be other associative links between terms if the concepts denoted by these terms are somehow related to each other in their meaning, with the exception of the hierarchical links described above.

In multilingual thesauri, there are also equivalence links between terms in different languages. There is full (strict) equivalence and several types of partial (non-strict) semantic equivalence of terms in different languages. A thesaurus often contains comments on terms that reveal the meaning of the term to the user and explain how to use it.

Thesauri are used primarily for classifying and searching for information resources. In this case, each resource may be associated with one or more concepts described by the terms in the thesaurus during classification, and the user conducting the search can use the thesaurus to find the concepts of interest in a given subject area, as well as all the terms that characterize them. That is, based on the thesaurus links, the search query is expanded (the query words are expanded with synonymous, more general or more specific terms). Navigating the thesaurus links helps formulate the query itself more clearly. There are a number of thesauri whose main task is not resource indexing, but their classification. In this case, the main objects of such thesauri (classifiers) are not

terms, but concepts and, often, unique identifiers (classification codes) that identify them. Relations in such a thesaurus are not semantic links between terms, but relationships between concepts (headings) that characterize the logic of the described subject area. Examples of such thesauri include subject classifiers in various branches of science.

The structure of the classifier corresponds to the structure of a regular thesaurus, since the connections between its headings are the same in meaning as between the terms of the thesaurus, and the classifier is its special case. However, when classifying, it is not the terms that are assigned to the resources, but the concepts they denote. Therefore, in the data scheme of the information system, the concepts of the thesaurus must be distinguished as independent objects. This means that such a scheme must have a structure different from the standards described above, in which concepts do not act as separate objects, but only terms and connections between them. At the same time, the scheme must allow working with thesauri described in accordance with these standards, i.e. be compatible with them. Thus, it will be more natural in the thesaurus data scheme for information systems to bind the language to terms, rather than to concepts. Moreover, such an approach is the only possible one for classifiers, in which it is language-independent concepts that classify other resources. Usually such classifiers are initially created as monolingual, and only later are translations into other languages made for them. In this case, only strict equivalence takes place between terms in different languages, since during translation for each term its strict equivalent is given (which is equivalent by definition, in the context of this classifier, even if in fact the translation is not entirely accurate). Linking a language to a concept would mean the need to make a separate copy of the same concept for each language, and to make a separate connection between each copy of the concept and the resource it classifies. Linking a language to a term links all equivalent terms in different languages to the same concept.

However, in thesauri, where there are many partial equivalence relations between multilingual terms, and there are also different hierarchies for terms in different languages, even fully equivalent terms may end up in different hierarchies, and therefore cannot be tied to one concept. All this means that to support multilingual thesauri, the data schema must provide for the equivalence relations between terms in different languages described in ISO and GOST standards as relations between concepts. At the same time, for each thesaurus, depending on its specifics, it is necessary to make a choice on how to implement the full equivalence relation between different terms:

1. Assign terms to different concepts, and set a full equivalence relation between concepts.
2. Assign terms to the same concept.

Obviously, for classifiers it is necessary to use the second approach, and for multilingual thesauri with different hierarchies in different languages - the first. It should be noted that a thesaurus that has a relation of incomplete equivalence already implies the presence of different hierarchies in different languages, which means that the first approach is necessary when implementing them. Another important attribute of a term in a thesaurus is a comment to it (ScopeNote). In classifier thesauri, where, in fact, the concept is primary, not the term, the comment, as a rule, also characterizes the concept. However, in other thesauri, a comment may relate specifically to a term. For example, to describe cases of preferential use of this synonym over others. Thus, in different thesauri, comments may relate to both concepts and terms. The choice depends on a specific thesaurus. A universal data scheme in an information system should allow both options for using comments.

References

1. Arkhangelskaya V.A., Bazarnova S.V. Information retrieval thesaurus on economics and demography [Text] / V.A. Arkhangelskaya, S.V. Bazarnova – 2001, pp. 24-32.
2. Beloozerov, V.N. Experience of developing a dictionary with an extensive system of thesaurus links [Text] / V.N. Beloozerov, Yu.P. Kosarskaya // NTI. Ser.2. Information processes and systems – 2001. – No. 8. – P. 28 – 31.
3. Lavrenova O.A. Modeling semantic texts of scientific and technical content in connection with the automation of information processes. Dissertation of the candidate of philological sciences: 10.02.21
4. Lavrenova O.A. Methodology of development of information retrieval thesaurus [Text] / O.A. Lavrenova – Moscow, 2001-p.5-7.
5. Lukashevich, N.V. Modifiers of conceptual relations in the thesaurus for automatic indexing [Text] / N.V. Lukashevich, B.V. Dobrov // NTI. Ser.2. Information processes and systems – 2001. – No.4. – P. 21 – 28.
6. Mikheeva E.V. Information technologies in professional activity: study guide.- M.: TK VELBI, Prospect Publishing House, 2007.- 448s.