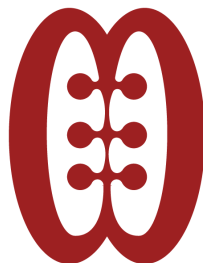




Technical Brief 3:

Tezauri Eurovoc-Africa



Symbol of friendship and interdependence of Akan people of West Africa

Concept note

What is a thesaurus?

A thesaurus is a structured list of expressions intended to represent in unambiguous fashion the conceptual content of the documents stored in a documentary system.

The ordinary language used by the authors of documents and by users who want to find these documents is in fact often ambiguous. It may be possible to express the same concept using a number of synonyms (e.g. agri-business, agro-industry, etc), and a document indexed using one synonym would not be found on the basis of a query that uses a different one. One of the purposes of the thesaurus is to eliminate the problems of **synonymity**.

One term can have various meanings (for example, "press" can mean the newspaper industry, journalists, printing machinery or various machine tools); a query using a term with homonyms would result in the retrieval of documents bearing no relation to the subject the user wants. One of the purposes of the thesaurus is to eliminate the problems of **homonymity**.

A **multilingual thesaurus** is then an added critical feature in the context of the African continent, because it makes it possible to establish correspondence between identical concepts expressed in different languages, letting the user interrogate the documentary system in their own language and retrieve documents irrespective of the language in which they have been indexed.

Why is a thesaurus needed?

Most information retrieval systems nowadays use full-text searching because are fast and very good results can be achieved. However, the use of full-text indexes has its limitations, and this is particularly true in a multilingual context. Full-text searching is not a solution for information access requirements such as the need for efficient document navigation, categorisation, abstracting and other means by which the document contents can be quickly understood and compared with others.

The use of a thesaurus complements full-text indexing because it allows much more focussed and refined search and cross-lingual information access.

With full-text search the discrimination of results becomes a non-trivial operation due to the large number of entries returned. Most of times the title and/or the text referring to the retrieved item is not explanatory enough, but it is necessary to go through the whole document to verify the information we are interested in. Though nobody disputes the usefulness of full text search engines, these search tools are still not good enough in some cases and totally inappropriate in others.

The first kind of less-successful cases is the World Wide Web. Old approaches using purely full-text-based engines were abandoned, since the quality of results provided was declining with the growth of the collection. Therefore, new techniques arose with the aim of filtering and re-qualifying the rank.



The thesauri were thought of as tools to improve the two main measures in information retrieval: **precision**, the number of relevant documents retrieved over the total number of documents retrieved, and **recall**, the number of relevant documents retrieved over the total number of relevant documents in the collection.

As opposed to full-text based search engines, thesauri can give a good performance for both measures, relying on setting relationships between words in a controlled vocabulary, where unambiguous index terms are assigned to a specific database or subject area. On a given query we can improve recall (by expanding to related terms) or precision (by narrowing with less generic terms).

Searching through the use of an electronic thesaurus is very helpful as it allows for:

- **Query expansion:** the search for a term, for example “car”, may be expanded to other synonymous keywords (“cars”, “autos”, “automobile”...) that the user can either choose instead of or in addition to the query.
- **Cross-lingual searching:** when the thesaurus used for indexing is multilingual, users can enter search terms in one language and find relevant documents written in other languages.
- **Multi term search:** concepts which can be expressed in several synonymous ways, such as "Chemotherapy" and "Drug Therapy", are conflated to only one form ("Drug Therapy"), while phrases can be treated as single concepts.
- **Directory or subject tree search:** when searching in highly specialized domains, keywords can be used as a directory or subject tree for the user who is not able to make his/her information needs explicit as a set of query terms. The user can so drill down through the categories narrowing the search space.

Indexing of document collections with controlled vocabulary keywords is complementary to full-text indexing and that provides both human users and machines with the means to analyse, navigate and access the contents of document collections in a way full-text indexing would not permit.

Eurovoc

The Eurovoc thesaurus is the multilingual and multidisciplinary thesaurus of the European Parliament and the Office for Official Publications of the European Commission. It provides a means of indexing the documents in the documentation systems of the European institutions and of their users.

Eurovoc is a thesaurus focussed on the laws and legislation of the European Union (EU) and more generally in the areas of law, legislation and politics. Eurovoc is a specialised thesaurus tuned to the needs and works of Parliaments with more than 20 years of experience and tools and mechanisms to update it and keep the different languages aligned.

Eurovoc is the thesaurus used by the European Parliament and 20 National Parliaments and already developed in 16 official languages of the European Union (Spanish, Czech, Danish, German, Greek, English, French, Italian, Lithuanian, Hungarian, Dutch, Portuguese, Slovak, Slovene, Finnish and Swedish). In addition to these versions, it has been translated by the parliaments of a number of countries ([Albania](#), [Croatia](#), Latvia, [Poland](#), Romania and [Russia](#)). The history of Eurovoc shows the integrating role that it as played as a common tool to access European parliamentary databases, at the national level and that of the European Parliament.

Eurovoc has established itself as the most diffused Parliamentary thesaurus, it has a wide-user base and has acquired a long and valuable experience and tools for both the management of its development and also the translation into many languages.

The Eurovoc thesaurus covers the following fields:

- politics,
-



- international relations,
- European Communities,
- law,
- economics,
- trade,
- finance,
- social questions,
- education and communications,
- science,
- business and competition,
- employment and working conditions,
- transport,
- environment,
- agriculture, forestry and fisheries,
- agri-foodstuffs,
- production, technology and research,
- energy,
- industry,
- geography,
- international organizations.

Some fields are more highly developed than others because they are more closely involved with the European Union's centres of interest. Thus, for example, the names of the regions of each Union Member State are in Eurovoc but not, for example, those of African countries.

Tezauri Eurovoc-Africa

The opportunity for African Parliaments to use the rich and diversified experience gained in more than 20 years with Eurovoc is very strategic because it will allow the Pan African Parliament to jumpstart the development of its own flavour of the Eurovoc with micro-thesauri specific to the African Continent, while benefiting of previous lessons learned and experience.

The European Parliament has granted UN/DESA a licence to pilot implementations and begin the development of "Tezauri Eurovoc-Africa" sub-thesauri relevant to Africa and to the activities of Parliaments in Africa. Tezauri Eurovoc-Africa is already in three of the most widely used languages in Africa (English, French and Portuguese). the feasibility of further translations in indigenous languages will be evaluated.

Tezauri Eurovoc-Africa thesaurus should be used for indexing all internal and external documentation of the African Parliaments with the objective to:

- Enhance the efficiency of searching for information in the information systems of the Pan-African Parliament and National Parliaments;
- Provide cross-lingual search;
- Provide significant support to the harmonization of legal and other specialized terminology using the standard terms of the thesaurus.

Since the Eurovoc Thesaurus, because of its own nature, does not sufficiently cover specific African concepts, topics and institutions or geographical and ethnic terms, Tezauri Eurovoc-Africa will create an "Addendum" with the list of new index terms (descriptors) in order to incorporate them in the main thesaurus structure.

The Africa Addendum will include a hierarchically structured micro thesauri of the African government bodies, local and regional administration, African languages and ethnic groups, geographical names, agencies operating in Africa, political parties, et.



All descriptors included in the Africa Addendum will be marked by a special symbol and coded as AFRIVOC.

The Africa Addendum will constantly be updated and fed by relevant new concepts specific to the political, legal, historical and cultural features of Africa marketing use of the same tools and mechanism devised for Eurovoc.

Tezauri Eurovoc-Africa Addendum Maintenance Committee

The Pan African Parliament with the support of the European Parliament and UN/DESA will also set up a Tezauri Eurovoc-Africa Addendum Maintenance Committee that will benefit from the experience, know-how and tools of Eurovoc.

Like any thesaurus, Tezauri Eurovoc-Africa Addendum will have to be continually adapted to take account on the one hand of the developments in the fields in which African institutions are active and on the other hand of the changes in the language.

The special feature of any thesaurus is that it aims and is able to meet the needs expressed by its users while keeping in mind the objective of covering in a homogeneous way the fields of activity of the African Parliaments.

To this end, Tezauri Eurovoc-Africa Addendum will be subject to a maintenance procedure based on both internal and external needs, and all users are invited to contribute to the development of it by making proposals for changes.

A Tezauri Eurovoc-Africa Addendum Maintenance Committee composed by members drawn from the various institutions will have to be set up and it will formulate proposals and judge the proposals submitted by its users. By means of a vote, the Maintenance Committee will decide on the changes to be made to the thesaurus which will be then submitted to a larger Steering Committee for ratification.
