Multilingual thesauri and ontologies in cross-language retrieval

Dagobert Soergel

College of Library and Information Services
University of Maryland
College Park, MD 20742
d52@umail.umd.edu

Abstract
This paper sets forth a framework for the use of thesauri and ontologies as knowledge bases in cross-language retrieval. It provides a general introduction to thesaurus functions, structure, and construction with particular attention to the problems of multilingual thesauri. It proposes the creation of environments for distributed collaborative knowledge base development as a way to make large-scale knowledge-based systems more affordable.

Outline
1 Introduction
2 Thesaurus functions
   2.1 User-centered/request-oriented indexing
3 Thesaurus structure
   3.1 Brief review of thesaurus structure principles
   3.2 Special issues in multilingual thesauri
4 Implementing thesaurus functions in retrieval
   4.1 Controlled vocabulary
   4.2 Free-text searching
   4.3 Knowledge-based support of searching
5 Thesaurus construction
6 Affordable implementation of knowledge-based approaches

1 Introduction

A thesaurus is a structure that manages the complexities of terminology in language and provides conceptual relationships, ideally through an embedded classification/ontology. This paper will
• give a tutorial on thesaurus functions and structure;
• present a concept retrieval perspective - concepts bridge languages in retrieval;
• argue the perspective that cross-language retrieval is also cross-cultural retrieval.

The paper covers thesaurus functions beyond retrieval; in retrieval, it considers any kind of object, not just text. We start out with some definitions (Figures 1a and 1b) and then give examples illustrating thesaurus problems and thesaurus structure (Figures 2 and 3).
The example in Figure 2 illustrates the conceptual and terminological problems in aligning the vocabularies of two languages. Concepts lexicalized in one language may not be lexicalized in the other and vice versa, creating significant problems for translation. The complexities of term correspondence are best managed with a conceptual approach, establishing a concept interlingua, so to speak.

Figures 3a and 3b give a first illustration of thesaurus structure, to be discussed more fully in Section 3. The emphasis is on illustrating the concept-based approach to vocabulary management and retrieval, so the examples are drawn from a thesaurus that epitomizes that approach, even though it is monolingual. The reader may want to look at the sample pages from several multilingual thesauri given in the appendix.

Figure 3a presents an excerpt from a thesaurus hierarchy. Through identifying the relevant facets (facet headings EF2, EF4, and EF6) and arranging the concepts within each facet in meaningful order displaying the concept relationships, the hierarchy elucidates the conceptual structure of the domain: The route of administration of drugs can be described by giving the intended scope of drug action, the method of administration, and the body site where the drug is administered. The last two facets have been combined because they are strongly intertwined. The hierarchy gives a logical arrangement for concepts within a facet, allowing the reader to form a clear mental image of methods available for administering drugs. The hierarchy also allows for hierarchic query expansion (whether searching with a controlled vocabulary or free-text).

Figure 3b gives examples of full thesaurus entries. The entry for EF gives many synonyms that can be used for synonym expansion of query terms. The RT cross-references suggest further descriptors that might be useful for searching. The entries for EF2 give scope notes that carefully define each concept. Thus, the thesaurus can serve as a reference. Juxtaposing the scope notes for hierarchical neighbor concepts allows the indexer or searcher to pick the right concept at the right hierarchical level.

Having established a general understanding of thesaurus structure, we can now deal with the functions, structure, and construction of thesauri in more detail.
EF route of administration

EF2 — by scope of drug action
EF2.2 topical and local administration
EF2.2.2 topical administration
EF2.2.4 local drug administration
EF2.4 systemic administration

EF4 — by method or body site
EF4.2 enteral administration
EF4.2.2 oral enteral administration
EF4.2.4 rectal enteral administration
EF4.4 mucosal administration
EF4.4.2 transdermal administration
EF4.4.4 inhalation, smoking, sniffing
EF4.4.4.2 smoking
EF4.4.4.2.2 smoking w/out inhalation
EF4.4.4.2.4 smoking with inhalation
EF4.4.4.4 nasal administration
EF4.4.4.6 pulmonary administration
EF4.4.6 oral mucosal administration
EF4.4.6.2 buccal administration
EF4.4.6.4 sublingual administration
EF4.4.6.8 rectal mucosal administration
EF4.6 parenteral administration
EF4.6.2 intravenous injection
EF4.6.2.2 intravenous infusion
EF4.6.4 intra-arterial injection
EF4.6.6 intraperitoneal administration
EF4.6.8 intracutaneous injection
EF4.6.10 admin. through skin implant
EF4.6.12 subcutaneous injection
EF4.6.14 intramuscular injection
EF4.6.16 CNS injection
EF4.6.16.2 intrathecal injection
EF4.8 skin administration
(The full entry shows Narrower Term cross-references to the more specific methods involving the skin: EF4.4.2, EF4.6.8, EF4.6.10, and EF4.6.12)

EF4.10 oral administration
(NT to EF4.2.2, EF4.4.4.2, and EF4.4.6)

EF4.10 rectal administration
(NT to EF4.2.4 and EF4.4.8)

EF6 drug administration by self vs. others
EF6.2 self administration of drugs
EF6.4 drug administration by others

Fig. 3a. Excerpt from a thesaurus hierarchy

EF route of administration

ST medication route
ST method of delivery of drugs or food
ST mode of substance administration
ST route of drug application
ST route of drug entry
ST route of exposure
BT +EE12 pharmacokinetics
RT +AA2 AOD use
RT +BS AOD substance by route of admin.
RT EE12.2e drug absorption
RT +EE14.4.8 drug effect by location
RT +HR drug therapy
RT MD2.2.2.2 drug paraphernalia

EF2 route of admin. by scope of drug action
SN Use one of these descriptors in combination with a descriptor from +EF4 route of admin. by method or body site.

EF2.2 topical and local administration
SN The application of a substance to a localized area, chiefly for local effects at this site.
NT HU4.2 local anesthesia
RT GH10.2 chemical injury

EF2.2.2 topical administration
SN The application of a substance on the surface of the skin or on a mucous membrane (incl. the gastrointestinal membrane) so that the substance will take effect on the surface or on a localized layer under the surface.
For example, for the administration of a decongestant spray, use EF2.2.2 topical administration combined with EF4.4.4.4 nasal administration.

EF2.2.4 local drug administration
SN The introduction of a substance into a localized area of the skin or other tissue, as through injection.
NT EF4.6.4 intra-arterial injection
NT EF4.6.8 intracutaneous injection
NT +EF4.6.16 CNS injection

EF2.4 systemic administration
SN The introduction of a substance into systemic circulation so that it is carried to the site of effect.
NT +EF4.6.2e intravenous injection
NT EF4.6.10 admin. through skin implant
NT HU4.4 general anesthesia
RT +GH10.4 chemical poisoning

Fig. 3b. Examples of full thesaurus entries
2 Thesaurus functions

A thesaurus with its embedded classification/ontology serves many functions, all of which are significantly affected by multilinguality. Our emphasis will be on thesaurus functions in retrieval (Figure 4a), but from a broader perspective one must not lose sight of the many other functions a thesaurus and its embedded classification/ontology can serve (Figure 4b).

In information retrieval a thesaurus or a classification/ontology without the surrounding terminological structure can be used in two scenarios:

1. knowledge-based support of free-text searching (applicable only to written or spoken text, although the text could point to another object, e.g., retrieving images through a free-text search of image captions or through a search of the text portion of a movie);
2. controlled vocabulary indexing and searching (applicable to any kind of retrieval).

The first two functions apply to either scenario. A user can always profit from looking at a conceptual framework of the domain to clarify the search topic, and the thesaurus then further assists in finding good search terms for the concepts identified. Synonym expansion includes mapping to terms from a different language.

Using a thesaurus as an indexing tool applies only to controlled vocabulary indexing. Indexing, the assignment of a set of descriptors to a document or other object, can be manual or automated. Of particular importance, so often overlooked, is an approach to indexing that places the users, their problems and questions squarely at the center of attention: user-centered or request-oriented indexing. The few empirical studies evaluating user-centered (as opposed to the commonly used document-centered) indexing show a positive effect on retrieval performance (Pejtersen 1983). This approach is central in information retrieval; it will be discussed in Section 2.1.

To look at thesaurus functions more generally, we first observe that a thesaurus is a knowledge base of concepts and terminology; other such knowledge bases are dictionaries and ontologies developed for AI applications, linguistic systems, or data element definition. Since these different types of knowledge bases — though developed for different purposes — overlap greatly, it would be best to integrate them through a common access system (Soergel 1996). The functions to be served by such a virtual integrated knowledge base of concepts and terminology are listed in Figure 4b.

2.1 User-centered/request-oriented indexing

As summarized in Figures 5a and 5b, user-centered indexing involves analyzing actual and anticipated user queries and interests and constructing a framework, a hierarchically structured controlled vocabulary, that includes the concepts of interest to the users and thus communicates these interests to the indexers or an expert system that can infer user-relevant concepts from text. The indexers then become the "eyes and ears" of the users and index materials from the users' perspective. The indexer uses the structured list of user-relevant concepts as a checklist, applying her understanding of a document (or other object) to judge its relevance to any of these concepts. This process ensures that users will find the documents that they themselves would judge relevant upon examination.

Request-oriented indexing contrasts with document-oriented indexing, where the indexer simply expresses what the document is about or where simply the terms in the text are used. But a document can be relevant for a concept without being about the concept: a document titled The percentage of children of blue-collar workers going to college is not necessarily about intergenerational social mobility, but a researcher interested in that topic would surely like to find it, so it is relevant. Another example: Since users are interested in the biochemical basis of behavior and also in longitudinal studies, these descriptors are in the thesaurus. The indexer examines the document CSF studies on alcoholism and related behaviors and finds that it is relevant to both descriptors. Longitudinal is not mentioned in the document, but careful examination of the methods section reveals the concept.
• Provide a semantic road map to individual fields and the relationships among fields; relate concepts to terms, and provide definitions, thus providing orientation and serving as a reference tool.

• Improve communication and learning generally:
  • Assist writers: suggest from a semantic field the term that best conveys the intended meaning and connotation.
  • Assist readers in ascertaining the proper meaning of a term and placing it in context.
  • Support learning through conceptual frameworks.
  • Support language learning and the development of instructional materials.

• Provide the conceptual basis for the design of good research and practice.
  • Assist researchers and practitioners in exploring the conceptual context of a research project, policy, plan, or implementation project and in structuring the problem.
  • Assist in the consistent definition of variables and measures for more comparable and cumulative research and evaluation results. Especially important for cross-national comparisons.

• Provide classification for action:
  • a classification of diseases for diagnosis,
  • of medical procedures for insurance billing,
  • of commodities for customs.

• Knowledge base to support information retrieval
  (Fig. 4a)

• Ontology for data element definition. Data element dictionary. Consider data processing systems in a multinational corporation

• Conceptual basis for knowledge-based systems.

• Do all this across multiple languages

• Mono-, bi-, or multilingual dictionary for human use. Dictionary/knowledge base for automated language processing - machine translation and natural language understanding (data extraction, automatic abstracting/indexing).

Fig. 4b. Broader functions of a knowledge base of concepts and terminology

This kind of indexing is expensive, unless it can (to a degree) be automated through a knowledge-based system for automated indexing. Is it worthwhile? The worth derived from improved performance depends on the use of the retrieval results.

Construct a classification/ontology (embedded in a thesaurus) based on actual and anticipated user queries and interests.

Thus provide a conceptual framework that organizes user interests and communicates them to indexers.

Index materials from users' perspective:
Add need-based retrieval clues beyond those available in the document. Increase probability that a retrieval clue corresponding to a query topic is available.

Index language as checklist.

Indexing = judging relevance against user concepts.
Relevance rather than aboutness

Implementation: Knowledgeable indexers or an expert system using syntactic & semantic analysis & inference.

Fig. 5a. User-centered / request-oriented indexing

Document
The drug was injected into the aorta
User concept: Systemic administration

Document:
The percentage of children of blue-collar workers going to college
User concept: Intergenerational social mobility

Document:
CSF studies on alcoholism and related behaviors
User concept: Longitudinal study
(Longitudinal not mentioned in the document; determined through careful examination of the methods section.)

Fig. 5b. Request-oriented indexing. Examples

This perspective on indexing has implications for cross-language retrieval: The conceptual framework must be communicated in every participating language to allow a meeting of minds to take place, regardless of the languages of the user and the indexer. This is particularly salient in the context of indexing images with descriptors that capture imponderables, such as the mood of an image: One needs to make sure that, as far as possible, the term used by the indexer in one language communicates the same mood as the term given to the user in another language for searching.
3 Thesaurus structure

After a brief review of general principles (3.1), we discuss issues specific to multilingual thesauri (3.2).

3.1 Brief review of thesaurus structure principles

Thesaurus structure consists of the terminological structure that relates terms to concepts (by establishing synonym relationships and disambiguating homonyms) and conceptual structure. We discuss each in turn.

Terminological structure (Figure 6)

<table>
<thead>
<tr>
<th>Controlling synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
</tr>
<tr>
<td>Alcoholism</td>
</tr>
<tr>
<td>Inheritance</td>
</tr>
<tr>
<td>Ultrasonic cardiography</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Afro-American</td>
</tr>
<tr>
<td>Pregnant adolescent</td>
</tr>
</tbody>
</table>

Disambiguating homonyms

administration 1 (management)
administration 2 (drugs)
Läufer 1 (Sportler) Engl.: runner (athlete)
Läufer 2 (Teppich) Engl.: long, narrow rug
Läufer 3 (Schach) Engl.: bishop (chess)
discharge 1 (From hospital or program) German: Entlassung
discharge 2 (From organization or employment) Preferred synonym: Dismissal German: Entlassung
discharge 3 (Medical symptom) German: Absonderung, Ausfluss
discharge 4 (into a river) German: Ausfluss
discharge 5 (Electrical) German: Entladung (which also means unloading)

The terminological structure is equally important in controlled vocabulary and in free-text searching. In free-text searching, synonym expansion of query terms is important, and homonym indicators can trigger a question to the user on the intended meaning of the query term.

Conceptual structure

A well-developed conceptual structure is a sine qua non for user-centered indexing and is very useful for free-text retrieval as well. The two principles of conceptual structure are facet analysis and hierarchy.

Facets. Semantic factoring or feature analysis

Semantic factoring means analyzing a concept into its defining components (elemental concepts or features). This gives rise to a concept frame with facet slots. See Figure 7 for examples.

| liver cirrhosis               |
| Pathologic process: inflammation |
| Body system:                 | liver                     |
| Cause:                       | not specified             |
| Substance/organism:          | not specified             |

| alcoholic liver cirrhosis    |
| Pathologic process:         | inflammation             |
| Body system:                | liver                    |
| Cause:                      | chemically induced        |
| Substance/organism:         | alcohol                  |

| hepatitis A                  |
| Pathologic process:         | inflammation             |
| Body system:                | liver                    |
| Cause:                      | infection                |
| Substance/organism:         | hepatitis A virus        |

A facet groups concepts that fall under the same aspect or feature in the definition of more complex concepts; it groups all concepts that can be answers to a given question. In frame terminology: The facets listed above are slots in a disease frame; a facet groups all concepts that can serve as fillers in one slot.

Using elemental concepts as building blocks for constructing compound concepts drastically reduces the number of concepts in the thesaurus and thus leads to conceptual economy. It also facilitates the search for general concepts, such as searching for the concept dependence, which occurs in the context of medicine, psychology, and social relations.

Fig. 6. Terminological structure examples

Fig. 7. Facet analysis examples
Facets can be defined at high or low levels in the hierarchy, as illustrated in Figure 8.

### Top-level facets
- organism
- body part
- chemical substances by function
- chemical substances by structure.

### Low-level facets
- route of administration
  - route of administration by scope of drug action
    - (local/topical or systemic)
  - route of administration by body part
  - route of administration by method of application
    - (injection, rubbing on, etc.)
- liver
  - liver tissue (hepatocyte, Kupffer cell, etc.)
  - liver part (hepatic lobule, portal lobule, etc.)

Fig. 8. Facets at different hierarchical levels

### Hierarchy
A sample hierarchy was presented in Fig. 3a. For another example, consider a search for a broad concept and the more specific concepts that should be included in the query as illustrated in Figure 9.

- groups at high risk of drug use
  - suicidal or physically or mentally disabled
  - persons from unstable or low-cohesion families
  - children of alcoholic or other drug-abusing parents
  - SN Adult or still under age
  - children of single teenage mothers
  - persons subjected to abuse or neglect (now or past)
    - persons subjected to abuse/neglect by parents
      - latchkey children
    - persons subjected to abuse/neglect by spouse
      - single teenage mothers
  - school dropouts or those at risk of dropping out
  - unemployed or in danger of being unemployed
  - economically disadvantaged
  - homeless
  - runaway youth
  - gateway drug users
  - persons engaged in violent or delinquent acts

Fig. 9. Hierarchy for searching

### Uses of facet analysis and hierarchy
Through facet analysis and hierarchy building, the lexicographer often discovers concepts that are needed in searching or that enhance the logic of the concept hierarchy; he then needs to create terms for these concepts. Examples are traffic station as the semantic component common to train station, bus station, harbor, and airport or distinct distilled spirits (as the semantic component common to gin, whiskey, cherry brandy, tequila, etc., the counterpart of the already lexicalized neutral distilled spirits), or analytic psychotherapy as an umbrella term for a host of methods (such as insight therapy, Gestalt therapy, and reality therapy) that all seek to assist patients in a personality reconstruction through insight into their inner selves.

Fig. 10 lists the most important uses of facet analysis and hierarchy. These uses will be more fully discussed in Section 4.

Help to organize the concept space and establish concept relationships.

Assist the user in analyzing and clarifying a search problem both in terms of the facets involved and the hierarchical structure within each facet.

Facilitate the search for general concepts, such as

- inflammation or
- dependence (which occurs in the context of medicine, psychology, and social relations)

Hierarchic query term expansion

These functions are useful in both controlled vocabulary and free-text searching.

Fig. 10. Uses of facet analysis and hierarchy

### 3.2 Special issues in multilingual thesauri
A multilingual thesaurus for indexing and searching with a controlled vocabulary can be seen as a set of monolingual thesauri that all map to a common system of concepts. With a controlled vocabulary, indexing is concept-based; cross-language retrieval is simply a matter of providing designations for these concepts in multiple languages so that queries can be written in multiple languages. However, as the example in Fig. 2 illustrates, conceptual systems represented in the vocabulary of different languages do not completely coincide.
The crux of the matter, then, is which concepts to include. Ideally, the thesaurus should include all concepts needed in searching by any user in any of the source languages. Language differences often also imply cultural and conceptual differences, more so in some fields than in others. We need to create a classification that includes all concepts suggested by any of the languages. At a minimum this includes all relevant concepts lexicalized in at least one of the source languages. Also, different languages often suggest different ways of classifying a domain; the system needs to be hospitable to all of these. The problem that has bedeviled many developers of multilingual thesauri is that a concept lexicalized in one language may not be lexicalized in another and that the terms that do exist often vary slightly in meaning, possibly giving rise to different relationships. Starting from the misguided notion that a thesaurus should include only concepts for which there is a term in the language and that term meanings cannot be adjusted for purposes of the thesaurus, they had difficulty making the system of concepts the same for all languages. But, as we have seen, even in a monolingual thesaurus the lexicographer often discovers concepts needed in searching or to enhance the logic of the concept hierarchy and then needs to create terms for these concepts. In multilingual thesauri this necessity arises more often, particularly when different languages differ in the hierarchical levels at which they lexicalize concepts. The principle proposed here is to establish a common conceptual system, which may require an arduous, and expensive, process of negotiation, and then arrange for the terms in all languages to fit, giving proper definitions, of course.

It is clear that the problems discussed here and illustrated in Fig. 2 and in Section 3.1 have major implications for cross-language free-text searching: Each query term should be mapped from the source language to its multiple equivalents in the target language; each of these equivalents may have other meanings in the target language, presenting potential problems for retrieval. The query term may not have a precise equivalent in the target language; one may need to map to broader or narrower terms, distorting the meaning of the original query.

4 Implementing thesaurus functions in retrieval systems

4.1 Controlled vocabulary

With a controlled vocabulary there is a defined set of concepts used in indexing and searching. Cross-language retrieval simply means that the user should be able to use a term in his own language to find the corresponding concept identifier that is used to retrieve documents (or whatever the retrieval objects are). In the simplest system, this can be achieved through manual look-up in a thesaurus that includes for each concept corresponding terms from several languages and has an index for each language. In more sophisticated systems the mapping from term to descriptor would be done internally. As an example, consider a library catalog using the Library of Congress Subject Headings, for which French and Spanish translations are available. In the VTLS automated library system, each subject heading is identified by a number that is used in the document record. The authority file includes for each subject heading the preferred term and any synonyms; this information can be included in multiple languages. One could use this arrangement for assisting the user in finding subject headings or automatic mapping of user terms to subject headings as follows: Do a free-text search on authority records to find any subject heading for which either the preferred term or any synonym contains the user's query word or phrase in any language. Once appropriate subject headings are found, they can be used to retrieve documents.

Whenever the mapping from user terms to descriptors is done "behind the scenes", transparent to the user, the system should ask the user for clarification whenever the query word or phrase has multiple meanings and cannot be disambiguated automatically. Beyond that, showing the user the descriptor(s) the system came up with in their hierarchical context might improve the accuracy of the query formulation and thus retrieval. The success of this type of interaction depends on the quality of the hierarchy and the interface.

If the user has voice input available, one might even include the spoken form of terms in the thesaurus to enable voice input of query terms which would then be mapped to the appropriate descriptors.

A cross-language retrieval system with controlled vocabulary must also support indexing of documents or other objects, that is the assignment of controlled vocabulary descriptors, in the various languages. For manual indexing, this is accomplished by having thesaurus versions in each of the languages so that each indexer has a version in her own language. But that is not enough. The
thesaurus version in each language must make sure that the indexer in that language fully understands the meaning of a descriptor that originated from another language; otherwise, the indexing of such a descriptor will not be consistent across the database.

Automated indexing with a controlled vocabulary, particularly if it is to take a request-oriented slant, can be accomplished with a knowledge base that (1) allows recognition of important words and phrases (for spoken text this requires the inclusion of spoken forms) and allows for homonym disambiguation and (2) gives mapping rules that lead from the (possibly weighted) set of words and phrases identified for a document to a set of descriptors that should be assigned.

Such mapping rules can take many forms. In their simplest form, they specify a direct mapping from text words or phrases to the appropriate descriptors for each word or phrase (and possibly even word or phrase combinations). To increase accuracy, the mapping can be made dependent on context (Hlava 97). A more complex mapping relies on association strengths between terms (words and phrases) and descriptors. Broadly speaking, the association strength between term T and descriptor D could be seen as the predictive probability that the document containing term T should be indexed with descriptor D. Such association strengths can be computed from a training set of indexed documents. This is the approach often taken in automated text categorization, where often, but not always, the goal is to index each document by only one descriptor (assign it to one of a set of non-overlapping categories). An advanced version of this approach is the use of "topic signatures", profiles consisting of a set of terms with weights; a document is assigned the topic if its terms match the topic signature (Lin 1997). In effect, a topic signature is a query which identifies documents relevant to the topic.

As the foregoing discussion illustrates, the knowledge base needed to support automated indexing is more complex than a thesaurus for manual indexing. It must include more terms and term variants so that the words and phrases important for indexing can be recognized in the text, and it must include information needed for the disambiguation of homonyms (which often requires determining the part of speech of a text word).

For indexing and searching, a controlled-vocabulary cross-language retrieval system can be seen as a set of monolingual systems, each of which maps the terms from its language to a common system of concepts used in indexing and searching. For manual indexing and query formulation, this is accomplished through a multilingual thesaurus, which may in fact consist of multiple monolingual thesauri linked through common descriptor identifiers (such as Dewey Decimal class numbers). Automated indexing in cross-language text retrieval with texts in multiple languages means mapping from each language to the common conceptual structure represented in the controlled vocabulary. The knowledge base component dealing with identification of words and phrases for automated indexing can be developed independently for each language. Mapping rules that are entirely term-based can also be developed independently for each language. However, some mapping rules, for example rules based on context or topic profiles, may include conceptual elements that could be shared across languages.

There are a number of controlled-vocabulary cross-language retrieval systems based on manual indexing in use in bilingual or multilingual areas such as Switzerland, Belgium, Canada, and areas of the US with large Spanish-speaking populations; in international organizations, such as the European Community; and in international collaborative systems, such as AGRIS. These systems are based on the Universal Decimal Classification, which has been translated into many languages (library of the ETH, Zurich); on the Library of Congress Subject Headings, which have been translated into French; on EUROVOC, an EC thesaurus in 9 languages; and AGROVOC, a thesaurus in three languages created by translation from its original English-only version. There are a large number of thesauri that either have been developed as multilingual thesauri or have been translated into several languages.

4.2 Free-text searching

Cross-language free-text searching, finding texts in one language that are relevant for a query formulated in another language without relying on controlled vocabulary indexing, is a more complex proposition. It requires that each term in the query be mapped to a set of search terms in the language of the texts, possibly attaching weights expressing the degree to which occurrence of a search term in a text would contribute to the relevance of the text to the query term. To assist with this task, a thesaurus must include the mapping information. If the thesaurus includes fine-grained definitions that deal with subtle differences of meaning, distance between such definitions can be used to derive term weights.

A major difficulty of this mapping is that a homonym used in the query gives rise to multiple translations, each corresponding to one of its meanings. The target terms may in turn be homonyms in their language and thus retrieve many irrelevant documents unless text terms are disambiguated. (This problem exists in synonym expansion in one language as well but is exacerbated in cross-language text retrieval.) When the mapping goes to a term that has multiple meanings, the specific meaning should be identified, possibly in interaction with the user. For best retrieval results the terms in the texts should also be disam-
biquated so that only documents that include the term in the right sense score.

The issue of homonymy in retrieval is not as straightforward as it may seem at first glance (Sanderson 1994). First of all, quite a bit of disambiguation may occur "naturally", in that a given term may assume only one of its meanings in the specific domain of the collection and therefore in the queries. Second, in a multi-component query, a document that includes a homonymous term from the first query component in a meaning other than that intended in the query is unlikely to also include a term from another query component, so excluding irrelevant documents may not require disambiguation in either the query or the texts. On the other hand, with single-concept query to a general collection (such as the World Wide Web), disambiguation can be expected to have a beneficial effect on retrieval performance. Failing that, a system might be able to suggest to the user an additional query component that would separate out the documents that include the query term but in a different meaning. Note that information extraction is much more dependent on homonym disambiguation.

In any event, for best support of free-text retrieval a thesaurus should flag homonyms, give their senses, and include rules for disambiguation.

The greater difficulty of free-text cross-language retrieval stems in no small measure from the fact that one must work with actual usage, while in controlled-vocabulary retrieval one can, to some extent, dictate usage.

4.3 Thesauri for knowledge-based search support

Whether searching is by controlled vocabulary or by free text, it is often helpful to the user to browse a well-structured and well-displayed hierarchy of concepts, preferably with the option of including definitions. A more sophisticated system may guide a user through a facet analysis of her topic. These aids provided by the system enable the user to form a better idea of her need and to locate the most suitable descriptors or free-text search terms. The guidance through facets and their hierarchical display must be available in the language of the user. These suggestions are based on the assumption that browsing a hierarchy is natural to most users and that users will appreciate the structure provided. This assumption rests on the belief that people try to make sense of the world and that guided facet analysis and browsing well-structured hierarchies help them do so. There is anecdotal evidence to support this assumption, but it needs to be investigated by building prototype systems and studying users' success (see, for example, Pollitt 1996).

This is one example of using a thesaurus as a knowledge base to make searching more successful. The assistance provided does not require that the user be an expert in classification and thesauri. This is even more true for "behind-the-scenes" assistance. There is no need to teach users about following a cross-reference from a synonym to a descriptor if the system searches for the descriptor automatically. There is no need to tell the user to look under narrower terms also if the system can do a hierarchically expanded search. There is no need to tell the user about strategies of broadening the search if the system, in response to a user input that not enough was found, can suggest further descriptors to be searched based on cross-references in the thesaurus. Sophisticated retrieval software can make the use of thesauri in retrieval independent of the user’s knowledge and thereby can get much more mileage out of the investment in thesauri.

5 Thesaurus construction

Building a thesaurus, especially a multilingual thesaurus, takes a lot of effort. Some term relationships can be derived by statistical analysis of term occurrence in corpora, but this will not result in the kind of well-structured conceptual system described above. Developing such a structure requires intellectual effort.

A common method for thesaurus construction in a single language is to work bottom-up: One collects a list of terms (words and phrases), preferably from search requests, but also from documents, free-term indexing, and other thesauri. These terms are then sorted into increasingly fine-grained groups, until a group contains only synonyms or terms that, for purposes of the thesaurus, can be considered synonyms. In this process at least some homonyms will be detected; they must be disambiguated into several senses, each expressed by its own (possibly newly coined) term having one meaning and being grouped accordingly. A group of synonyms can be considered to represent a concept; usually a preferred term to designate the concept is selected, but some other concept identifier can be used. A first rough hierarchy of concepts emerges from this process.

This is followed by conceptual analysis, especially facet analysis at various levels, resulting in a well-structured faceted hierarchy. Next, one needs to write definitions (scope notes), in the process of which one may rethink the hierarchy, and introduce relationships between concepts that complement the hierarchy.

The development of a multilingual thesaurus is, naturally, an even more complex undertaking; the basic approaches are summarized in Figure 12. The ideal way to develop a multilingual thesaurus is to start from a pool of terms in all covered languages and carry out the process without regard to the language of the terms. This will
bring together terms from different languages that have the same meaning into one group. This process gives all languages an equal chance to contribute concepts and concept relationships. It also forces a careful analysis of the meaning of each term in each language to determine the degree of equivalence, making it possible to develop the fine-grained structure of definitions that has the potential of providing powerful support to free-text cross-language retrieval.

Of course, this process would require a lexicographer knowledgeable in the subject matter of the thesaurus and fluent in all covered languages, not a very practical requirement. A more practical variation that still maintains the spirit of this approach is to start with two languages and develop the conceptual structure — a bi-lingual lexicographer is needed in any event. Definitions should be written in both languages. One would then work on a pool of terms in a third language and fit it into the structure, creating new concepts as necessary. This is not at all the same as translating the thesaurus into the third language. This requires a lexicographer fluent in one of the starting languages and the third language. Following the same principle, one can now add other languages.

The result of such a process is a conceptual system that brings the conceptual structures embedded in the different languages under one roof, so to speak.

The most common approach to the construction of a multilingual thesaurus is to translate an existing monolingual thesaurus into one or more languages. But this approach is problematic: The original language and its vocabulary determine the conceptual structure, and one merely looks for equivalent terms in the second language without covering its terminological richness. In some multilingual thesauri, only one term in the target languages is provided, making the thesaurus unsuitable for query term expansion in free-text searching.

In between is an approach in which one starts with a monolingual thesaurus as the center and fits terms from one or more other languages into the structure of this central thesaurus without changing the concepts or the hierarchy. EuroWordNet (Gillaranz 1997) takes an improved variation of this approach, working with the English WordNet as its central thesaurus. In EuroWordNet, separate and independent word nets are constructed in each language in parallel efforts, each identifying synonym sets in that language (A synset can be considered a concept). Each individual language project then independently maps its synsets to the corresponding WordNet synsets; no changes are made to WordNet. In addition to identity, this mapping allows for hyponym and hypernym relationships, thus indicating that the concept identified in the language being worked on is not included in WordNet, but giving at least the hierarchical location. EuroWordNet also uses a

Requirements
Must cover all concepts of interest to the users in the various languages, at a minimum all domain concepts lexicalized in any of the participating languages.
Must accommodate hierarchical structures suggested by different languages.

Approaches (by increasing complexity and quality)
(1) Start from monolingual thesaurus and translate. This approach does not capture concepts lexicalized only in another language and is biased to the conceptual structure underlying the starting language. May not produce all synonyms in the second language.

(2) Start from a monolingual thesaurus as the center. Collect terms from a second (third, ...) language and establish correspondences of these terms to the central thesaurus. Suffers from similar bias toward the starting language as (1), but may cover more synonyms in the other languages.

(3) Work with a central thesaurus as in (2), but after collecting terms from a second language first group them into synsets, that is, derive concepts each of which is represented by a set of terms, and then map each concept to the corresponding concept in the central thesaurus or indicate that the concept is new and give the nearest broader or narrower concept in the central thesaurus. Note that the central thesaurus remains unchanged.

(4) As (2), but add concepts not in the starting thesaurus. This mitigates bias, but the central thesaurus now becomes a moving target.

(5) Start from a pool of terms from all participating languages and organize them into a conceptual framework, establishing term correspondence in the process. This approach results in a true "conceptual interlingua" not biased to any one language, but offering a home to multiple conceptual perspectives. This approach requires most effort.

Fig. 12. Building multilingual thesauri

very weak variation of approach 5: The participants developed a "top ontology", which presumably reflects and integrates perspectives from their individual cultures. In addition to being mapped to WordNet, the individual language synsets are also mapped to this top ontology.
Affordable implementation of knowledge-based approaches

The effort needed for constructing and maintaining any knowledge base, especially a well-structured multilingual thesaurus using method 4, is often forbidding, whence the attempts at constructing thesauri by statistical analysis of corpora. Fortunately, there is another way to reduce the effort, often drastically: Capitalize on the intellectual effort already available in a multitude of existing thesauri and dictionaries by automatically merging term relationships from many sources, as is done in UMLS (Unified Medical Language System) or analyzing dictionary definitions to extract term relationships (Ahlswede 1988). Learn from the structure of text by creating hypotheses on the part of speech and semantic features of words during parsing (Sonnenberger 1995), or deriving term relationships from user queries. A further expansion of this approach calls for collaborative development of thesauri and more comprehensive databases of concepts and terms made possible by computer technology as proposed in Soergel 1996. These approaches are summarized in Fig. 13.

Knowledge-based approaches require major investment for constructing the knowledge base.

Solutions

Use what is available (e.g. WordNet).

Reformat and integrate available sources into structured knowledge bases.

Use machine learning techniques based on text or query analysis for building or adding to a knowledge base, perhaps followed by human editing.

Provide integrated access to multiple sources and an environment for distributed collaborative knowledge base development.

Fig. 13. Implementation of knowledge-based approaches

Conclusion

It was the intent of this paper to present a high-level review of the contribution knowledge-based systems can make to cross-language retrieval, as exemplified in particular through the structure and function of thesauri and ontologies. Many of the ideas presented have been applied in operational or experimental systems, even though empirical results need to be interpreted with caution (Soergel 1994); others await application and testing.

References


Lin, Chin-Yew 1997 (cyl@isi.edu, personal communic.)


<table>
<thead>
<tr>
<th>Language</th>
<th>Title</th>
<th>Description</th>
<th>Pages</th>
<th>Binding</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>KERAMISCHES WOERTERBUCH. (Ger-En)</td>
<td>131p.</td>
<td>Ppbk.</td>
<td>$ 29.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FACHWOERTERBUCH FUER DIE GLASINDUSTRIE. DICTIONARY FOR THE GLASS INDUSTRY. (Ger-En-Ger).</td>
<td>1963. 160p. Bound.</td>
<td>$ 125.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>CERAMIC TILES DICTIONARY. (It-En).</td>
<td>1983. 562p. Bound.</td>
<td>$ 195.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>GLOSSARY OF GLASS-MAKING. (Sp-Fr-It-Sw).</td>
<td>1970. 1,519p. Bound.</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td>DICTIONARY OF GLASS-MAKING. (En-Sw).</td>
<td>1980. 562p. Bound.</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multilingual</td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING.</td>
<td>(Dutch supplement keyed to M-8392).</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING. (Ita-En).</td>
<td>1983. 64p. Ppbk.</td>
<td>$ 29.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING. (Nor-Dan-Ger-En-Finn-Fr-Sw).</td>
<td>Each term is defined in Norwegian. 1982. 64p. Ppbk.</td>
<td>$ 29.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING.</td>
<td>(Spanish supplement keyed to M-14381).</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING.</td>
<td>(Russian supplement keyed to M-14381).</td>
<td>$ 250.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING.</td>
<td>(Spanish supplement keyed to M-14381).</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING.</td>
<td>(German supplement keyed to M-14381).</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING.</td>
<td>(French supplement keyed to M-14381).</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Commission on Glass. DICTIONARY OF GLASS-MAKING.</td>
<td>(Japanese supplement keyed to M-14381).</td>
<td>$ 45.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

28. Chemistry

<table>
<thead>
<tr>
<th>Language</th>
<th>Title</th>
<th>Description</th>
<th>Pages</th>
<th>Binding</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>ANGLO-CHINESE CHEMICAL DICTIONARY.</td>
<td>1,222p. Bound.</td>
<td>$ 95.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENGLISH-CHINESE DICTIONARY OF CHEMISTRY AND CHEMICAL ENGINEERING.</td>
<td>1978. 1,458p. Bound.</td>
<td>$ 95.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ENGLISH-CHINESE DICTIONARY OF INDUSTRIAL CHEMISTRY.</td>
<td>1977. 81p. Ppbk.</td>
<td>$ 19.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czechoslovakian</td>
<td>RUSKO-SESKY CHEMICKY SLOVNIK. (Russ-Cze).</td>
<td>1978.</td>
<td>PNS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>DICTIONARY OF CHEMICAL ENGINEERING AND LABORATORY EQUIPMENT.</td>
<td>(En-Dut-En). 1986. 120p. Ppbk.</td>
<td>$ 49.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

212
# DETAILED CONTENTS OF PARTS I AND II

## Part I: BIBLIOGRAPHY OF UNIVERSAL CLASSIFICATION SYSTEMS

| 041 | Library Classification Systems (by countries) | 1 |
| 041.8 | Library Subject headings lists (by countries) | 1 |

### 042 Universal: Decimal Classification (UDC)

| 042.1 | Full Editions (by languages) | 2 |
| 042.1-cs | Czechoslovakian Full Edition | 2 |
| 042.1-de | German Full Edition | 3 |
| 042.1-en | English Full Edition | 4 |
| 042.1-es | Spanish Full Edition | 7 |
| 042.1-fr | French Full Edition | 7 |
| 042.1-hu | Hungarian Full Edition | 7 |
| 042.1-it | Italian Full Edition | 8 |
| 042.1-ja | Japanese Full Edition | 9 |
| 042.1-pl | Polish Full Edition | 9 |
| 042.1-pt | Portuguese Full Edition | 10 |
| 042.1-ru | Russian Full Edition | 10 |
| 042.1-sk | Serbo-Croatian Full Edition | 11 |

| 042.2 | Medium Editions | 11 |
| 042.3 | Abridged Editions | 12 |
| 042.4 | Proposals and Supplements to the UDC | 13 |
| 042.5 | Extensions and Corrections to the UDC | 13 |
| 042.6 | Specialized UDC Editions | 14 |
| 042.7 | Special Purpose-Oriented UDC Editions | 18 |
| 042.8 | UDC Indexes and Thesauri | 18 |
| 042.9 | Concordances, Correlations | 18 |

### 043 Dewey Decimal Classification (DDC)

| 043.1 | DDC Full Editions | 18 |
| 043.3 | Abridged DDC Editions | 19 |
| 043.4 | Proposals and Supplements to the DDC | 19 |
| 043.5 | Notes and Decisions | 19 |
| 043.6 | Specialized DDC Editions | 19 |
| 043.7 | Special Purpose-Oriented DDC Editions | 20 |

### 044 Library of Congress Classification (LCC)

| 044.1 | LCC Full Edition | 20 |
| 044.1.8 | Indexes to the LCC | 21 |
| 044.3 | Outline of the LCC | 21 |
| 044.5 | Additions and Changes of the LCC | 21 |
| 044.7 | Special Purpose-Oriented LCC Editions | 21 |
| 044.8 | LC Subject Headings (LCSH) | 21 |

### 045 Bliss Bibliographic Classification (BBC)

| 045.1 | BBC Full Editions | 21 |
| 045.3 | BBC Abridged Edition | 22 |

### 046 Ranganathan Colon Classification (RCC)

| 046.1 | RCC Full Edition | 22 |
| 046.3 | Outline of RCC | 22 |
| 046.6 | Special RCC Schedules | 22 |
| 046.9 | RCC Concordances and Correlations | 26 |
AGROVOC

SOLANACEAE (cont.)
NT2 solanum quitoense
NT2 solanum tuberosum
NT1 solanum
NT2 solanum lycopersicum
rt tubiflorae
Fr solanaceae
Es solanaceae

SOLANINE
BT1 glycoalkaloids
BT2 alkaloids
BT2 glycosides
BT3 carbohydrates
BT1 phytosterols
BT2 steroids
BT3 carbohydrates
Fr solanine
Es solanina

SOLANUM
BT1 solanaceae
NT1 solanum aethiopicum
NT1 solanum benzoinum
NT1 solanum chaccone
NT1 solanum hypothroidium
NT1 solanum khasianum
NT1 solanum laciniatum
NT1 solanum melongena
NT1 solanum quitoense
NT1 solanum tuberosum
Fr solanum
Es solanum

SOLANUM AETHIOPICUM
BT1 solanum
BT2 solanaceae
Fr solanum aethiopicum
Es solanum aethiopicum

SOLANUM BERTHAULTII
BT1 solanum
BT2 solanaceae
Fr solanum berthaultii
Es solanum berthaultii

SOLANUM CHACCOENSE
BT1 solanum
BT2 solanaceae
Fr solanum chaccoense
Es solanum chaccoense

SOLANUM HYPONOIDIIUM
uf cocco
BT1 solanum
BT2 solanaceae
rt fruit crops
rt tropical fruits
Fr solanum hypothroidium
Es solanum hypothroidium

SOLANUM KHASIANUM
BT1 solanum
BT2 solanaceae
rt drug plants
Fr solanum khasianum
Es solanum khasianum

SOLANUM LACINIATUM
BT1 solanum
BT2 solanaceae
rt drug plants
Fr solanum laciniatum
Es solanum laciniatum

solanum lycopericum
USE lycoperon esculentum

SOLANUM MELOXGENA (cont.)
rt vegetable crops
Fr solanum melongena
Es solanum melongena

SOLANUM MELONGENA
uf pepino
BT1 solanum
BT2 solanaceae
rt fruit crops
rt tropical fruits
Fr solanum melongena
Es solanum melongena

SOLANUM NIGRUM
BT1 solanum
BT2 solanaceae
rt drug plants
Fr solanum nigrum
Es solanum nigrum

SOLANUM QUITOENSE
uf lulo
uf naranjilla
BT1 solanum
BT2 solanaceae
rt fruit crops
rt tropical fruits
Fr solanum quitoense
Es solanum quitoense

SOLANUM TUBEROUS
BT1 solanum
BT2 solanaceae
rt potatoes
rt root vegetables
rt starch crops
rt vegetable crops
Fr solanum tuberosum
Es solanum tuberosum

SOLAR ENERGY
uf solar activity
USE solar energy

SOLAR COLLECTORS
BT1 equipment
rt heating
rt solar energy
Fr capteur solaire
Es collecteurs solaires

SOLAR DRYING
USE natural drying

SOLAR ENERGY
uf solar activity
uf solar radiation
BT1 renewable energy
BT2 energy sources
BT2 renewable resources
BT3 natural resources
rt daylight
rt greenhouse effect
rt solar collectors
Fr energie solaire
Es energie solaire

SOLAR RADIATION
USE solar energy

SOLARIMATE
USE actinometers

SOLE CRIPPPING
(one crop grown alone in pure stands)
uf single crop
BT1 cropping systems
rt monoculture
Fr culture pure
Es culture puro

sole, Dover
→ soles
sole, gray
→ witch (fish)

SOLES
in saecap group b-31
uf Dover sole
uf lascar (fish)
uf lemon sole
BT1 flounder
BT2 saltwater fishes
BT3 fishes
rt quceans
rt soles
Fr sole
Es lenguado
<table>
<thead>
<tr>
<th>FRANÇAIS</th>
<th>ITALIANO</th>
<th>ENGLISH</th>
<th>NEDERLANDS</th>
<th>PORTUGUÊS</th>
</tr>
</thead>
<tbody>
<tr>
<td>pays de l'EAMA</td>
<td>SAMA-paesi membri</td>
<td>AAMS countries</td>
<td>GASMA-landen</td>
<td>países EAMA</td>
</tr>
<tr>
<td>Aarhus</td>
<td>Aarhus</td>
<td>Aarhus</td>
<td>Aarhus</td>
<td>Aarhus</td>
</tr>
<tr>
<td>enfant abandonné</td>
<td>bambino abbandonato</td>
<td>abandoned child</td>
<td>wondeling</td>
<td>criança abandonada</td>
</tr>
<tr>
<td>terre abandonnée</td>
<td>terreno abbandonato</td>
<td>abandoned land</td>
<td>verlaten land</td>
<td>terra abandonada</td>
</tr>
<tr>
<td>accord ABM</td>
<td>accordo ABM</td>
<td>ABM Agreement</td>
<td>anti-rekto-overeenkomst</td>
<td>Acordo ABM</td>
</tr>
<tr>
<td>suppression des droits de douane</td>
<td>abolizione dei dazi doganali</td>
<td>abolition of customs duties</td>
<td>afschaffing van de douanerechten</td>
<td>supressão dos direitos advaneiros</td>
</tr>
<tr>
<td>avortement</td>
<td>aborto</td>
<td>abortion</td>
<td>abortus</td>
<td>aborto</td>
</tr>
<tr>
<td>Abruzzes</td>
<td>Abruzzo</td>
<td>Abruzzi</td>
<td>Abruzzi</td>
<td>Abruzos</td>
</tr>
<tr>
<td>majorité absolue</td>
<td>maggioranza assoluta</td>
<td>absolute majority</td>
<td>absolute meerderheid</td>
<td>maggioria assoluta</td>
</tr>
<tr>
<td>abstentionnisme</td>
<td>astensionismo</td>
<td>abstentionism</td>
<td>stemontoughing</td>
<td>abstencionismo</td>
</tr>
<tr>
<td>Abu Dhabi</td>
<td>Abu Dhabi</td>
<td>Abu Dhabi</td>
<td>machtmissbruk</td>
<td>Abu Dhabi</td>
</tr>
<tr>
<td>abus de pouvoir</td>
<td>abuso di potere</td>
<td>abuse of power</td>
<td>machtmissbruk</td>
<td>abuso de poder</td>
</tr>
<tr>
<td>liberté de l'enseignement</td>
<td>libertà di insegnamento</td>
<td>academic freedom</td>
<td>vrijheid van onderwijs</td>
<td>liberdade de ensino</td>
</tr>
<tr>
<td>accès à la justice</td>
<td>accesso alla giustizia</td>
<td>access to the courts</td>
<td>toegang tot het beroepsleven</td>
<td>toegang tot professie</td>
</tr>
<tr>
<td>adhésion à un accord</td>
<td>adesione ad un accord</td>
<td>accession</td>
<td>toegang tot de communautaire</td>
<td>adesione a un accord</td>
</tr>
<tr>
<td>adhésion à la Communauté</td>
<td>adesione alla Comunità europea</td>
<td>accession to the Community</td>
<td>toegang tot de European Gemeenschap</td>
<td>adesione alla Comunità europea</td>
</tr>
<tr>
<td>accident domestique</td>
<td>infortunio domestico</td>
<td>accident in the home</td>
<td>ongeval thuis</td>
<td>acidente doméstico</td>
</tr>
<tr>
<td>prévention des accidents</td>
<td>prevenzione degli infortunii</td>
<td>accident prevention</td>
<td>ongevallenpreventie</td>
<td>prevenção de acidentes</td>
</tr>
<tr>
<td>compte</td>
<td>conto</td>
<td>account</td>
<td>rekening</td>
<td>conta</td>
</tr>
<tr>
<td>comptable</td>
<td>contabile</td>
<td>accountant</td>
<td>boekhouder</td>
<td>contabilista</td>
</tr>
<tr>
<td>comptabilité</td>
<td>contabilità</td>
<td>accounting</td>
<td>boekhouding</td>
<td>contabilità</td>
</tr>
<tr>
<td>imputation comptable</td>
<td>imputazione contabile</td>
<td>accounting entry</td>
<td>boekhoudkundige afschrijving</td>
<td>imputação contabilistica</td>
</tr>
<tr>
<td>système de comptabilité</td>
<td>sistema di contabilità</td>
<td>accounting system</td>
<td>boekhoudkundige afschrijving</td>
<td>sistema contabilistico</td>
</tr>
<tr>
<td>acculturation</td>
<td>acculturazione</td>
<td>acculturation</td>
<td>boekhoudkundige afschrijving</td>
<td>acculturación</td>
</tr>
<tr>
<td>pluie acide</td>
<td>pioggia acida</td>
<td>acid rain</td>
<td>zure regen</td>
<td>chuva ácida</td>
</tr>
<tr>
<td>acoustique</td>
<td>acustica</td>
<td>acoustic</td>
<td>geluidsteer</td>
<td>acústica</td>
</tr>
<tr>
<td>pays ACP</td>
<td>ACP-paesi membri</td>
<td>ACP countries</td>
<td>ACS-landen</td>
<td>países ACP</td>
</tr>
<tr>
<td>comité des ambassadeurs ACP-CE</td>
<td>comitato degli ambasciatori</td>
<td>ACP-EC Committee of Ambassadors</td>
<td>Comitato Sottosegretari ACP-EC</td>
<td>Comité dos Embaixadores ACP-CE</td>
</tr>
<tr>
<td>convention ACP-CE</td>
<td>convenzione ACP-CE</td>
<td>ACP-EC Convention</td>
<td>ACS-EG-conventie</td>
<td>convenção CE</td>
</tr>
<tr>
<td>Conseil des ministres ACP-CE</td>
<td>consiglio dei ministri ACP-CE</td>
<td>ACP-EC Council of Ministers</td>
<td>Comité dei Ministri ACS-EG</td>
<td>Comité dos Ministros ACP-CE</td>
</tr>
<tr>
<td>institution ACP-CE</td>
<td>istituzione ACP-CE</td>
<td>ACP-EC Institution</td>
<td>ACS-EG-institutio</td>
<td>Comité dos Ministros ACP-CE</td>
</tr>
<tr>
<td>Assemblée paritaire ACP-CE</td>
<td>assemblea paritaria ACP-CE</td>
<td>ACP-EC Joint Assembly</td>
<td>Paritaria Vergadering ACP-EG</td>
<td>Assembleia Paritaria ACP-CE</td>
</tr>
<tr>
<td>comité paritaire ACP-CE</td>
<td>comitato paritario ACP-CE</td>
<td>ACP-EC Joint Committee</td>
<td>paritaria comité ACS-EG</td>
<td>Comité Paritário ACP-CE</td>
</tr>
<tr>
<td>acquis communautaire</td>
<td>acquis comunautaire</td>
<td>acquis communautaire</td>
<td>communautaire verworvenheden</td>
<td>acervo comunitário</td>
</tr>
<tr>
<td>acquisition of the property</td>
<td>acquisto della proprietà</td>
<td>acquisition of property</td>
<td>eigendomsverkrijging</td>
<td>aquisição de propriedade</td>
</tr>
<tr>
<td>recours contentieux administratif</td>
<td>ricorso contenzioso amministrativo</td>
<td>action brought before an administrative court</td>
<td>klacht inzake bestuurlijke geschillen</td>
<td>recurso contencioso administrativo</td>
</tr>
<tr>
<td>recours contentieux communautaire</td>
<td>ricorso contenzioso comunitario</td>
<td>action brought before the EC Court of Justice</td>
<td>klacht inzake communautaire geschillen</td>
<td>recurso contencioso comunitário</td>
</tr>
<tr>
<td>recours du personnel</td>
<td>ricorso del personale</td>
<td>action by staff</td>
<td>beroep door het personeel</td>
<td>recurso del personal</td>
</tr>
<tr>
<td>recours en annulation</td>
<td>ricorso per annullamento</td>
<td>action for annulment</td>
<td>beroep tot nietigverklaring</td>
<td>recurso de anulación</td>
</tr>
<tr>
<td>recours en annulation CE</td>
<td>ricorso per annullamento CE</td>
<td>action for annullment of an EC decision</td>
<td>vordering tot nietigverklaring</td>
<td>recurso de anulación CE</td>
</tr>
<tr>
<td>recours en carence</td>
<td>ricorso per carezza</td>
<td>action for failure to act</td>
<td>beroep wegens onvermogen</td>
<td>acção por omissão</td>
</tr>
<tr>
<td>recours en manquement</td>
<td>ricorso per inadempenza</td>
<td>action for failure to fulfill an obligation</td>
<td>beroep wegens niet-nakomen</td>
<td>acção por incumprimento</td>
</tr>
<tr>
<td>programme d'action</td>
<td>programma d’azione</td>
<td>action programme</td>
<td>actieprogramma</td>
<td>programa de acção</td>
</tr>
<tr>
<td>recours en responsabilité administrative</td>
<td>ricorso per accertamento di responsabilità amministrativa</td>
<td>action to establish liability on the part of an administration</td>
<td>vordering tot administratieve aansprakelijkheid</td>
<td>recurso de responsabilidade administrative</td>
</tr>
<tr>
<td>ESPAÑOL</td>
<td>DANSK</td>
<td>ENGLISH</td>
<td>DEUTSCH</td>
<td>ELΛΗΝΙΚΑ</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>países de la EAMA</td>
<td>AASM-lande</td>
<td>AAMS countries</td>
<td>AASM-Staaten</td>
<td>χώρες της ΕΑΜΑ</td>
</tr>
<tr>
<td>Aarhus</td>
<td>Århus Amt</td>
<td>Aarhus</td>
<td>Århus</td>
<td>Άρχος</td>
</tr>
<tr>
<td>hijo abandonado</td>
<td>forladt barn</td>
<td>abandoned child</td>
<td>ausgesetztes Kind.</td>
<td>εγκαταλελεμένο τέκνο</td>
</tr>
<tr>
<td>tierra abandonada</td>
<td>forladt jord</td>
<td>abandoned land</td>
<td>Sozialbrache</td>
<td>εγκαταλελεμένη για</td>
</tr>
<tr>
<td>Acuerdo ABM</td>
<td>AASM-afale</td>
<td>A0M Agreement</td>
<td>A0M-Vertrag</td>
<td>Συμμετοχή Πειραιώς</td>
</tr>
<tr>
<td>supresión de los derechos de</td>
<td>ophævelse af toldafgifter</td>
<td>abolition of customs duties</td>
<td>Abschaffung der Zolle</td>
<td>Αντιβαλλοστικών Παράλληλων</td>
</tr>
<tr>
<td>aduana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aborto</td>
<td>abort</td>
<td>abortion</td>
<td>Abtreibung</td>
<td></td>
</tr>
<tr>
<td>Abuzos</td>
<td>Abuzzi</td>
<td>Abruzzi</td>
<td>Abruzzen</td>
<td></td>
</tr>
<tr>
<td>abusamiento</td>
<td>arbejdsforførselsse</td>
<td>absenceism</td>
<td>Fernbleiben von der Arbeit</td>
<td></td>
</tr>
<tr>
<td>mayoría absoluta</td>
<td>absolut fiart</td>
<td>absolute majority</td>
<td>absolute Mehrheit</td>
<td></td>
</tr>
<tr>
<td>abstencionismo</td>
<td>manglende valgdeltagelse</td>
<td>abstentionism</td>
<td>Wohlenhaftung</td>
<td></td>
</tr>
<tr>
<td>Abu Dhabi</td>
<td>Abu Dhabi</td>
<td>Abu Dhabi</td>
<td>Amtsmtbrauch</td>
<td></td>
</tr>
<tr>
<td>abuso de poder</td>
<td>embedsmisbrug</td>
<td>abuse of power</td>
<td>Freiheit der Lehre</td>
<td></td>
</tr>
<tr>
<td>libertad de enseñanza</td>
<td>undervisningsfrifhed</td>
<td>academic freedom</td>
<td>Zugang zum Baruf</td>
<td></td>
</tr>
<tr>
<td>acceso a la profesión</td>
<td>adgang til erhverv</td>
<td>access to a profession</td>
<td>Zugang zu</td>
<td></td>
</tr>
<tr>
<td>acceso a la información</td>
<td>adgang til EU-oplysninger</td>
<td>access to Community information</td>
<td>Zugang zu</td>
<td></td>
</tr>
<tr>
<td>comunitaria</td>
<td></td>
<td></td>
<td>Gemeinschaftsinformationen</td>
<td></td>
</tr>
<tr>
<td>acceso a la educación</td>
<td>adgang til uddannelse</td>
<td>access to education</td>
<td>Zugang zur Bildung</td>
<td></td>
</tr>
<tr>
<td>acceso a la información</td>
<td>adgang til information</td>
<td>access to information</td>
<td>Informationszuzag</td>
<td></td>
</tr>
<tr>
<td>acceso a la justicia</td>
<td>adgang til retvæsenet</td>
<td>access to the courts</td>
<td>Zugang zur Rechtspflege</td>
<td></td>
</tr>
<tr>
<td>adhesión a un acuerdo</td>
<td>tiltrædle al en aftre</td>
<td>accession</td>
<td>Abkommensbeiritin</td>
<td></td>
</tr>
<tr>
<td>adhesión a la Comunidad Europea</td>
<td>tiltræde af EF</td>
<td>accession to the Community</td>
<td>Eingemeinschaft</td>
<td></td>
</tr>
<tr>
<td>accidente doméstico</td>
<td>uykke i hjemmet</td>
<td>accident in the home</td>
<td>Unfall im Haushalt</td>
<td></td>
</tr>
<tr>
<td>prevención de accidentes</td>
<td>forebygelse af uykker</td>
<td>accident prevention</td>
<td>Unfallverhütung</td>
<td></td>
</tr>
<tr>
<td>cuenta</td>
<td>regnskab</td>
<td>account</td>
<td>Konto</td>
<td></td>
</tr>
<tr>
<td>contable</td>
<td>regnskabsforer</td>
<td>accountant</td>
<td>Buchhalter</td>
<td></td>
</tr>
<tr>
<td>contabilidad</td>
<td>regnskabsvæsen</td>
<td>accounting</td>
<td>Buchführung</td>
<td></td>
</tr>
<tr>
<td>imputación contable</td>
<td>kontering</td>
<td>accounting entry</td>
<td>Verbuchung</td>
<td></td>
</tr>
<tr>
<td>sistema de contabilidad</td>
<td>regnskabsystem</td>
<td>accounting system</td>
<td>Kontenrahmen</td>
<td></td>
</tr>
<tr>
<td>aculturación</td>
<td>tilpasning til andet kulturmiljø</td>
<td>acculturation</td>
<td>Akkultration</td>
<td></td>
</tr>
<tr>
<td>ácido</td>
<td>syre</td>
<td>acid</td>
<td>Sauer</td>
<td></td>
</tr>
<tr>
<td>lluvia ácida</td>
<td>sur regn</td>
<td>acid rain</td>
<td>sauer Regen</td>
<td></td>
</tr>
<tr>
<td>acústica</td>
<td>akustik</td>
<td>acoustics</td>
<td>Akustik</td>
<td></td>
</tr>
<tr>
<td>países ACP</td>
<td>AVS-lande</td>
<td>ACP countries</td>
<td>AKP-Staaten</td>
<td></td>
</tr>
<tr>
<td>Comité de Embajadores ACP-CE</td>
<td>Ambassaderudvalget AVS-EF</td>
<td>ACP-EC Committee of ACP-CE</td>
<td>Botschatherausschuß ACP-EG</td>
<td></td>
</tr>
<tr>
<td>convenio ACP-CE</td>
<td>AVS/EF-konvention</td>
<td>ACP-EC Convention</td>
<td>Abkommen ACP-EG</td>
<td></td>
</tr>
<tr>
<td>Consejo de Ministros ACP-CE</td>
<td>Ministerådet AFS-FS</td>
<td>ACP-EC Council of Ministers</td>
<td>Ministerrådet AFS-EF</td>
<td></td>
</tr>
<tr>
<td>institución ACP-CE</td>
<td>AVS/EF-institution</td>
<td>ACP-EC Institution</td>
<td>ACP-EF-Institution</td>
<td></td>
</tr>
<tr>
<td>Asamblea paritaria ACP-CE</td>
<td>Den Partisitiske Forsamling</td>
<td>ACP-EC Joint Assembly</td>
<td>Partisitisk Forsamling AFS-EF</td>
<td></td>
</tr>
<tr>
<td>Asamblea paritaria ACP-CE</td>
<td>Den Partisitiske Forsamling</td>
<td>ACP-EC Joint Assembly</td>
<td>Partisitisk Förbundsavtale AFS-EF</td>
<td></td>
</tr>
<tr>
<td>comité paritario ACP-CE</td>
<td>Det Parititiske Udvalg AVS-EF</td>
<td>accu communautaire</td>
<td>Besitzstand der Gemeinschaft</td>
<td></td>
</tr>
<tr>
<td>acervo comunitario</td>
<td>gældende fællesskabseret</td>
<td>Eigentumsvererb</td>
<td>Kträger Kumiftareit</td>
<td></td>
</tr>
<tr>
<td>adquisición de la propiedad</td>
<td>erhvervelse af ejendom</td>
<td>acquisition of property</td>
<td>Klage vor dem Verwaltungssgericht</td>
<td></td>
</tr>
<tr>
<td>recurso contencioso-administrativo</td>
<td>forvaltningsager</td>
<td>action brought before an administrative court</td>
<td>Klage in EG-Streitsachen</td>
<td></td>
</tr>
<tr>
<td>recurso contencioso comunitario</td>
<td>fællesskabseretlige sager</td>
<td>action brought before the EC Court of Justice</td>
<td>Verwaltungsbeschwerden der Bedienstehen</td>
<td></td>
</tr>
<tr>
<td>recurso del personal</td>
<td>klageadgang for ansatte</td>
<td>action by staff</td>
<td>Anfechtungsclage</td>
<td></td>
</tr>
<tr>
<td>recurso de anulación</td>
<td>annullationsseggmål (EF)</td>
<td>action for annulment of an EC decision</td>
<td>EG-Nichtigkeitsklage</td>
<td></td>
</tr>
<tr>
<td>recurso de anulación CE</td>
<td>annullationsseggmål</td>
<td>action for annulment of an EC decision</td>
<td>Untätigkeitsklage</td>
<td></td>
</tr>
<tr>
<td>recurso por omisión</td>
<td>passivitetsseggmål</td>
<td>action for failure to act</td>
<td>Klage wegen Vertragsverletzung</td>
<td></td>
</tr>
<tr>
<td>recurso por incumplimiento</td>
<td>traktatbruddag</td>
<td>action for failure to fulfill an obligation</td>
<td>Aktionsprogramm</td>
<td></td>
</tr>
<tr>
<td>programa de actuación</td>
<td>handlingsprogram</td>
<td>action programme</td>
<td>Amtshaftungsclage</td>
<td></td>
</tr>
<tr>
<td>recurso por responsabilidad</td>
<td>erstatningsseggmål i forvaltningsager</td>
<td>action to establish liability on the part of an administration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>administrativa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

216