

Ontological spine, localization and multilingual access

Some reflections and a proposal

New Perspectives on Subject Indexing and Classification
in an International Context

International Symposium in Honour of Magda Heiner-Freiling

Frankfurt, 10. April 2008



Problem

How to integrate localized knowledge
into knowledge structures
as classification systems or documentary languages ?



Initial situation

- Universality, internationalization and localization;
 - Meaning of universality,
 - Choosing the proper principle for building hierarchies,
 - Knowledge versus literary warrant,
 - Considering cultural, religious, historical, ... facts and structure.
- Logical validity of classificatory structures, at least in the sense of machine operability along the hierarchies or reference structures;
- Building catalogues not only as tools for searching and finding fully conceptualized information but also as tools for navigation in concept structures;
- Forms of multilingual access;
- Cognitive interpretation of concepts and structure versus formal knowledge representation;
- Top down versus bottom up construction of vocabularies for indexing and retrieval purposes.



Problem

How to integrate localized knowledge into knowledge structures as classification systems or documentary languages ?

Theoretically there are at least two possible solutions:

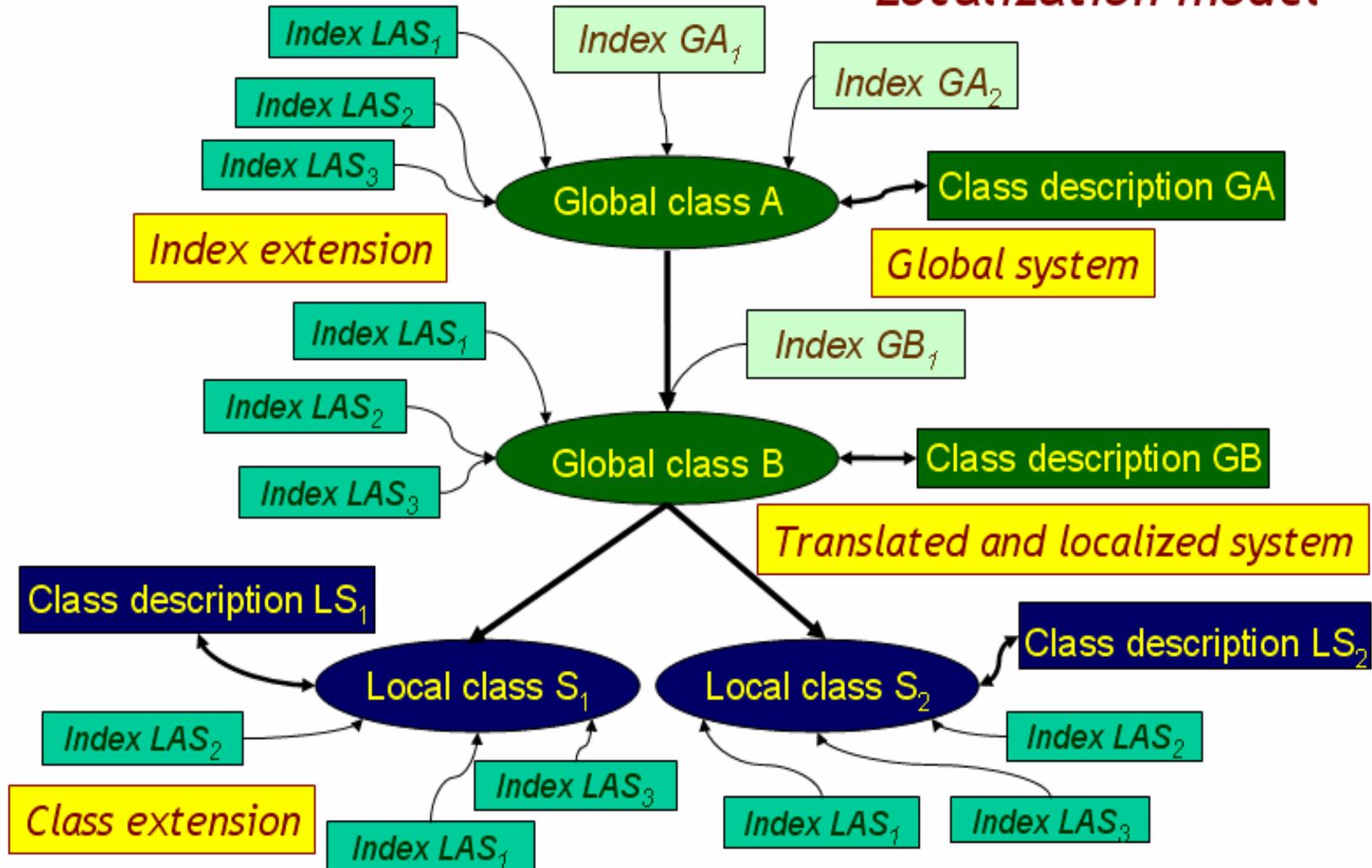
1. Integration of all knowledge - global or localized - into one knowledge structure;
2. Considering different knowledge structures - classification systems and the existing authority files - as parts of a broader system and assigning the representation of global (universal) knowledge to one system (spine) and the localized knowledge to the other systems.

Focusing on multilingual aspects, the second approach seems much more promising.



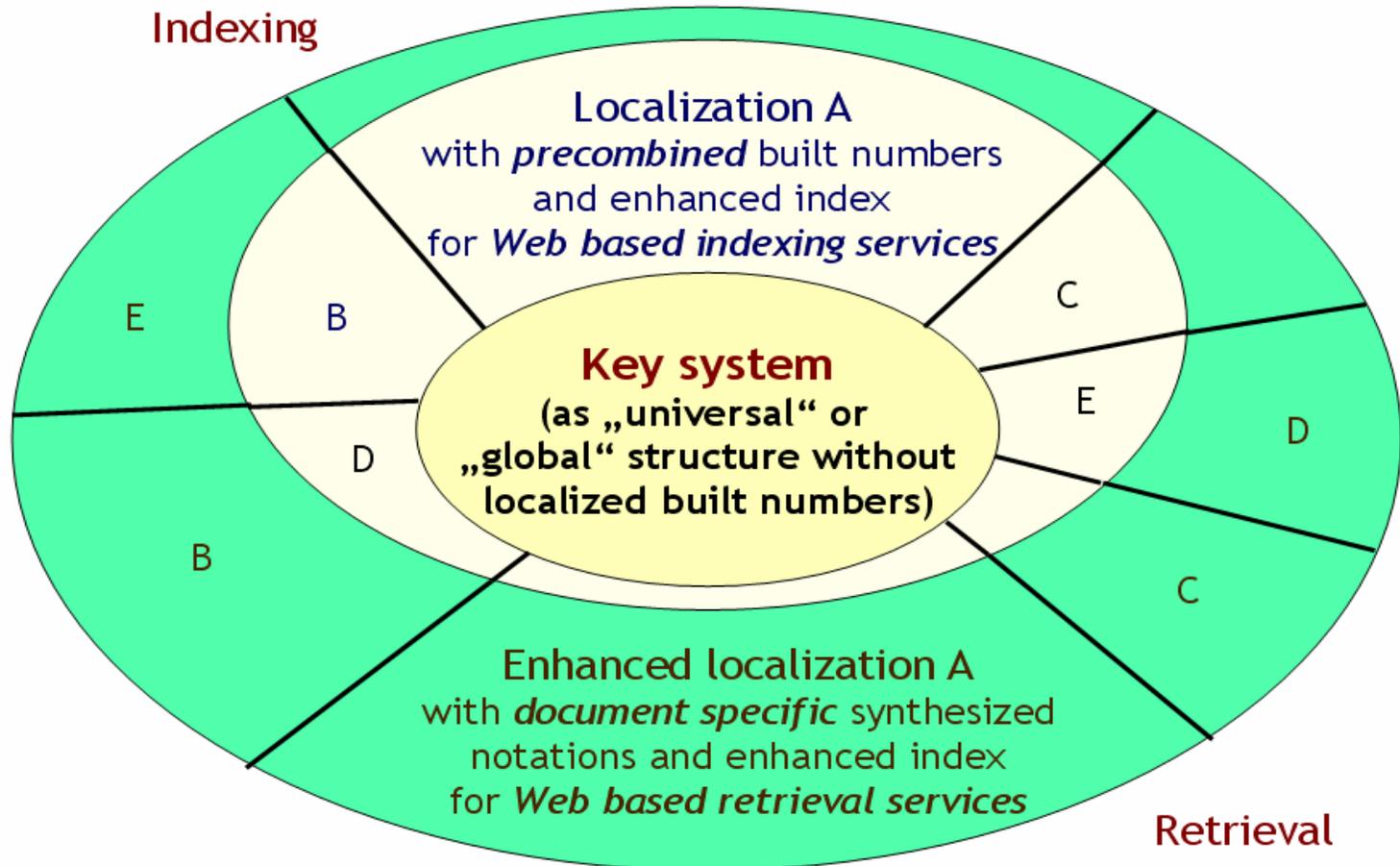
As a first possible answer with respect to the *DDC*, we proposed in 2005 to build up a de-localized version of the *DDC* by considering multilingual and localized needs in different *DDC* translations and corresponding indexes (W. Gödert, M. Preuss).

Localization model



Features of the proposal

- Uncoupling and adding back the local perspective;
- Isolation of localized facts and concepts into linkable elements;
- Localization by systematic actualization of instructions for notational synthesis;
- Transformation of instructions for synthesis as model examples of the localization.



Evaluation of the proposal from today's view

- Insufficient answers to requirements of the Semantic Web regarding the logical validity of the relations used, and the machine processing of the represented concepts as well as their relations;
- Insufficient consideration of multilingual requirements for processing search queries;
- Problems with complete integration of the criterion of localization.

Result

Search for new possibilities connecting classificatory structures with improved verbal access forms, which cope better with the problems mentioned, especially the harmonization of internationalization by localization.



Multilingual ACcess to Subjects (MACS) and CrissCross

The conjunctive aim of two projects is to create a multilingual, thesaurus-based and user-friendly research vocabulary that facilitates research in heterogeneously indexed collections. Subject headings of the *German Subject Heading Authority File (SWD)* are being linked to notations of the *Dewey Decimal Classification*, i.e. its German translation *DDC Deutsch*, as well as to equivalents of the *Library of Congress Subject Headings (LCSH)* and the French indexing vocabulary RAMEAU.

Basic principles of *MACS*

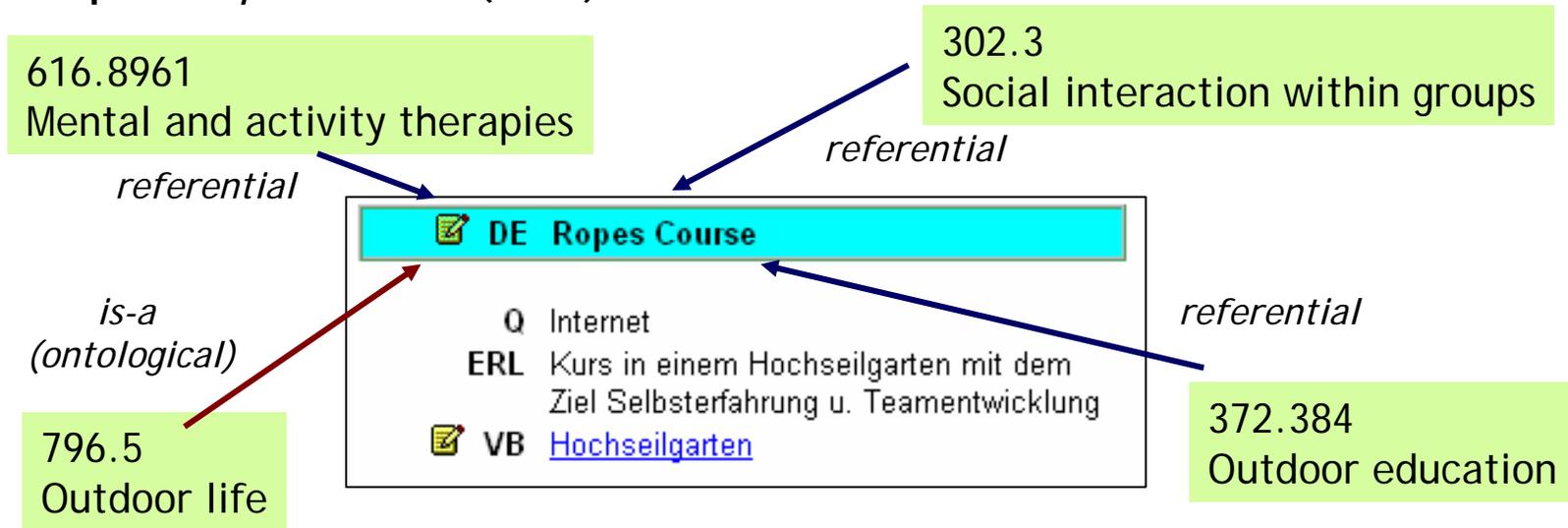
- Equality of languages and Subject Heading Languages (no pivot language) with autonomy of each Subject Heading Language;
- Establishment of equivalences (no translation) between the Subject Heading Languages involved (no new thesaurus);
- Equivalence links conceived as concept clusters.

Landry, P.: *MACS: multilingual access to subject and link management: Extending the Multilingual Capacity of TEL in the EDL Project.*
In: <http://www.edlproject.eu/workshop/programme.php>.



Michael Panzer has demonstrated that problems arise when constructing mappings from concepts without clear semantic boundary (e.g. the concepts of the *SWD* or the other authority files) to a classificatory structure (e.g. the *DDC*) as it is done in the CrissCross project:

Example: *Ropes course* (*SWD*)



- Referential vs. ontological relations;
- Mapping as framing of the semantic boundary of the subject term - Consequences for retrieval purposes;
- *DDC* as metalanguage providing context for the framing of the subject term.

Panzer, M.: *DDC in Germany: Recent Developments and Current Activities*.
In: http://www.oclc.org/dewey/news/conferences/ala_crissx_jan2007.pdf.



Criteria for mappings

We have to face the following characteristics of our authority files (e.g. *SWD, LCSH, Rameau*):

- Normally no explicit semantic precision by indication of definitions, but only by cognitive interpretation;
- Transfer of semantic meaning only by 3 types of relations, whose logical validity became insufficient secured with the production of the vocabulary.
- What criteria can be given for a decision between the different types of relations ?
- Is it justified to prefer any one of the relations over the others ?

In summary: Is it justified to speak of semantic equivalence of mappings between concepts of documentary languages without semantic frames ??

Conclusion

- Dubiety of the conceptual precision of the individual term in the monolingual context;
- Dubiety of semantic mappings between the terms of two or more authority files in the multilingual context.



Looking for alternative approaches

First realization: *ETHICS*

“Information retrieval can be improved by using multilingual thesaurus terms based on an intermediate or switching language to search with. Universal classification systems in general can play the role of switching languages.

1. Why a universal classification system and not another thesaurus ?

Because the UDC like most of the classification systems uses symbols. Therefore, it is language independent and the problems of compatibility between such a thesaurus and different other thesauri in different languages are avoided.

2. Why not assign running numbers to the descriptors in a thesaurus and make a switching language out of the resulting enumerative system ?

Because of some other characteristics of the UDC: hierarchical structure and terminological richness, consistency and control. One big problem to find an answer to is: can a thesaurus be made having as a basis a classification system in any and all its parts? To what extent this question can be given an affirmative answer? This depends much on the attributes of the universal classification system which can be favourably used to this purpose.

Those classes of UDC are best fitted for building a thesaurus structure out of them which are both hierarchical and faceted.”

Frâncu, V.: *Multilingual access to information using an intermediate language.*

Antwerpen: Faculteit Taal- en Letterkunde, Germaanse Taal- en Letterkunde 2003. VII; 195 S.



What does *localization* mean for our discussion ?

Localization is a well-known concept in software engineering, understood as

Adaptation of computer software for non-native environments, especially other nations and cultures or the process of translating a product into different languages or adapting a language for a specific country or region.

This understanding is insufficient for our purposes. We will understand by localization

not only translating a concept into different languages or adapting a language for a specific country or region

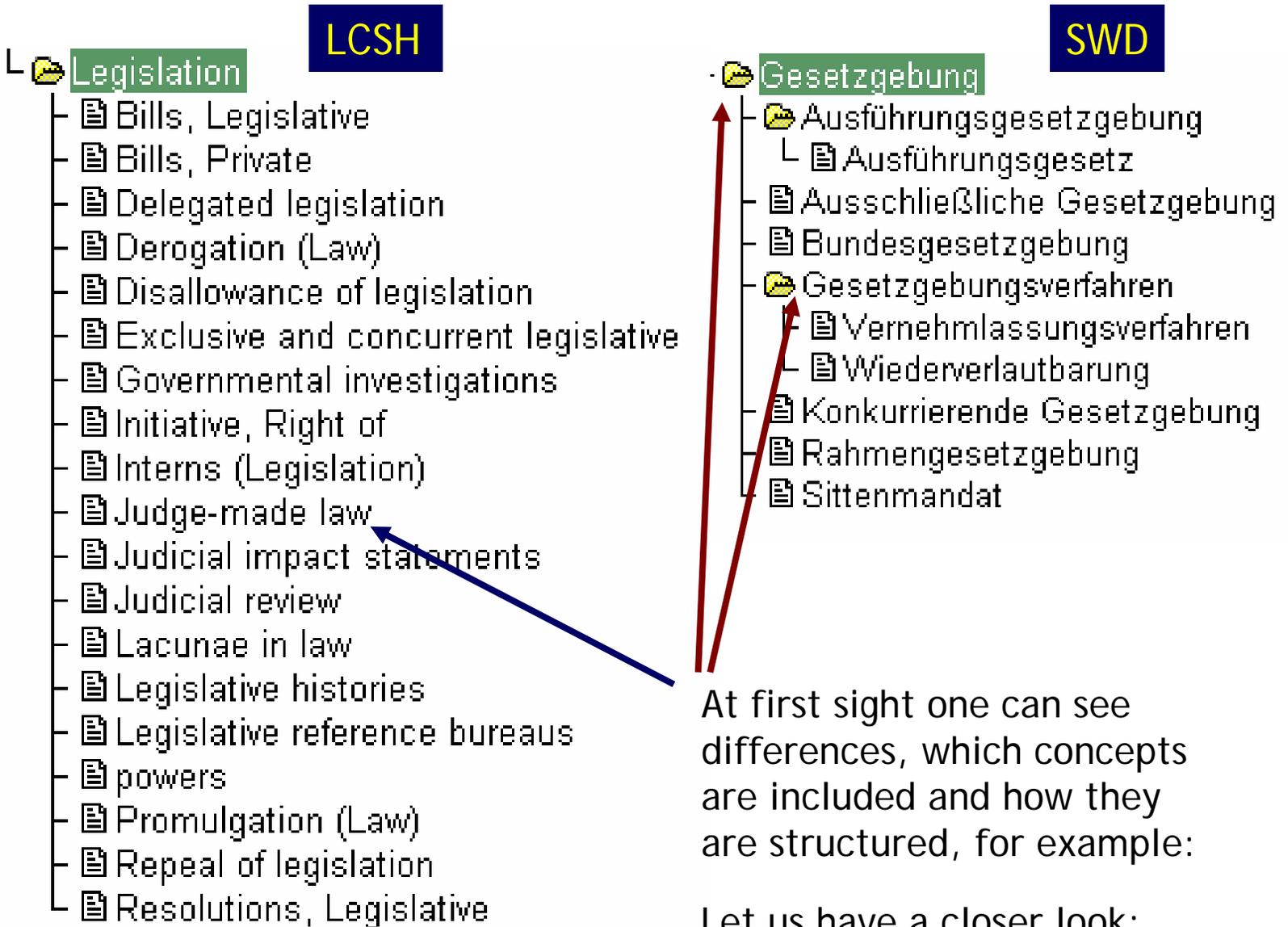
but also

the representation of concepts and their semantical relations for their native environments, especially other cultures, history, or nations with their political and social structures.

It is questionable whether this could be done within the context of only one knowledge system or documentary language.



Illustration: Subject headings for a corresponding concept from two authority files (LCSH, SWD) in a hierarchical view:

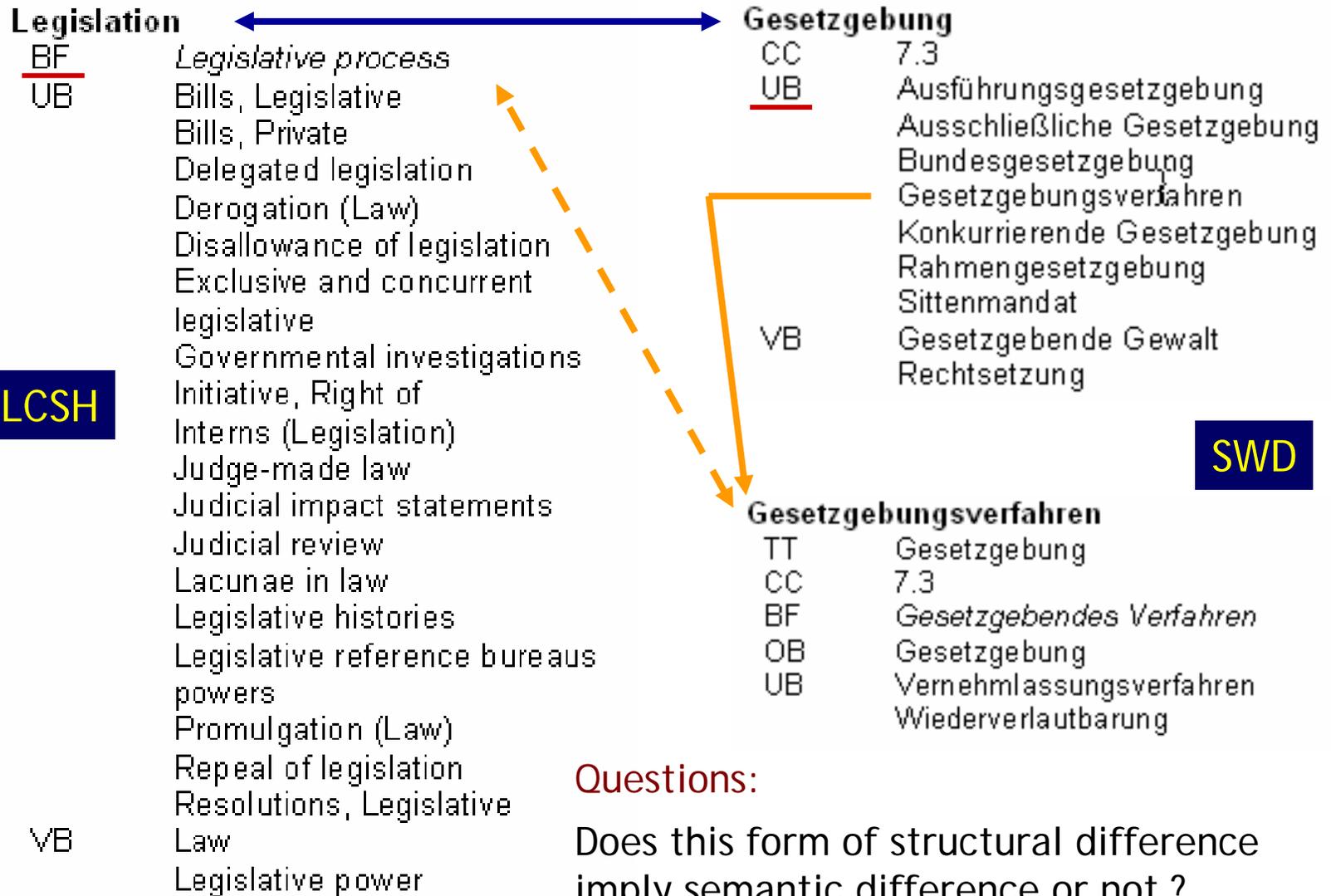


At first sight one can see differences, which concepts are included and how they are structured, for example:

Let us have a closer look:



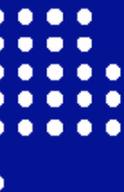
The same concepts seen in the standard thesaurus format:



Questions:

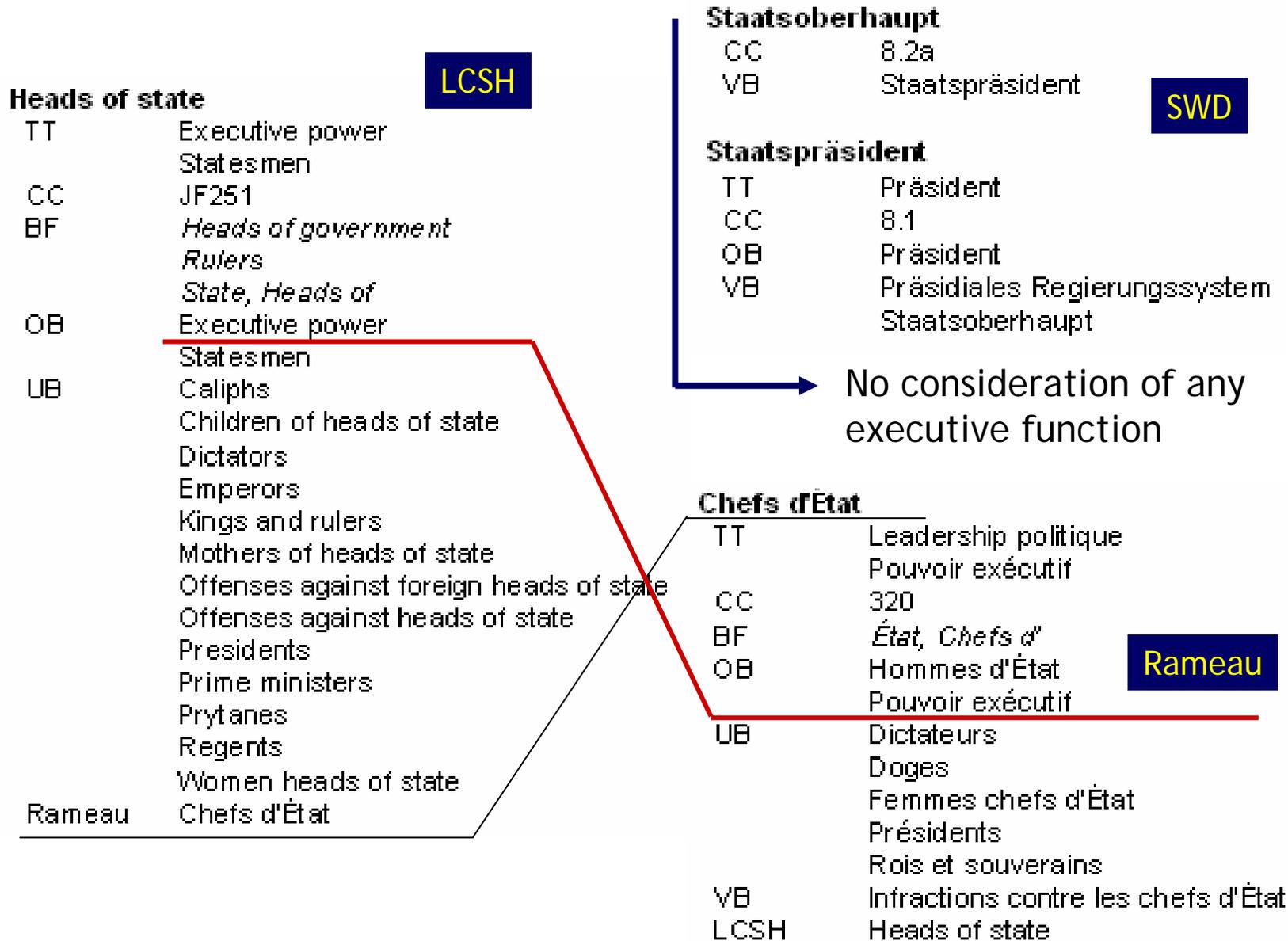
Does this form of structural difference imply semantic difference or not ?

Is it justified to speak of *Legislation* and *Gesetzgebung* as semantically equivalent ?



A further example:

Relation between *Heads of state* and *Executive power*



By analyzing more examples one can come to the conclusion:

Conceptual relations which are results of localized aspects cannot be represented in a one-to-one correspondence by mappings between subject headings of authority files or mappings between subject headings and classes of an ontology:

- a sufficient semantic correspondence between the individual semantic containers does not exist;
- the respective conceptual structure cannot be represented within the mapping process.

Taken a kind of information retrieval as focus point that permits conceptual exploration and navigation in the conceptual structure, it is even more desirable to preserve the individual structure of each localization and to make benefit of them for retrieval purposes.

A realization of such ideas is possible by using semantic technologies for constructing conceptual networks along the requirements of knowledge representation.

I will give an example out of a thesis just finished by one of my students. He has studied the potential of expanding and typifying the existing relations of the *SWD* for retrieval purposes.



This work can be seen as continuation of studies undertaken by the *ALCTS/CCS Subject Analysis Committee, Subcommittee on Subject Relationships/Reference Structure* in 1997 to investigate:

1. the kinds of relationships that exist between subjects, the display of which are likely to be useful to catalog users;
2. how these relationships are or could be recorded in authorities and classification formats;
3. options for how these relationships should be presented to users of online and print catalogs, indexes, lists, etc.

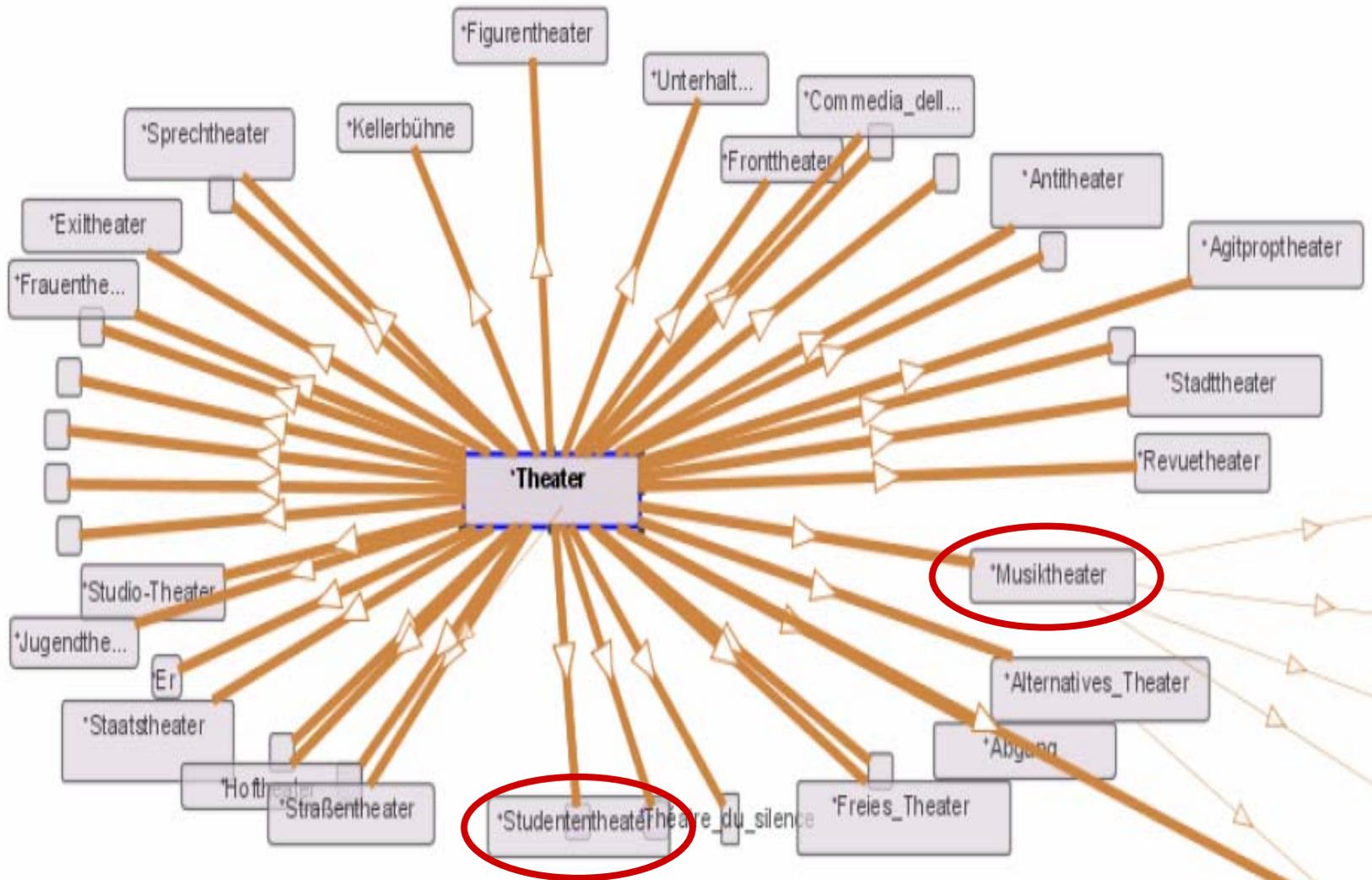
Final Report to the ALCTS/CCS Subject Analysis Committee. June 1997

In:

<http://web2.ala.org/ala/alctscontent/CCS/committees/subjectanalysis/subjectrelations/finalreport.cfm>.

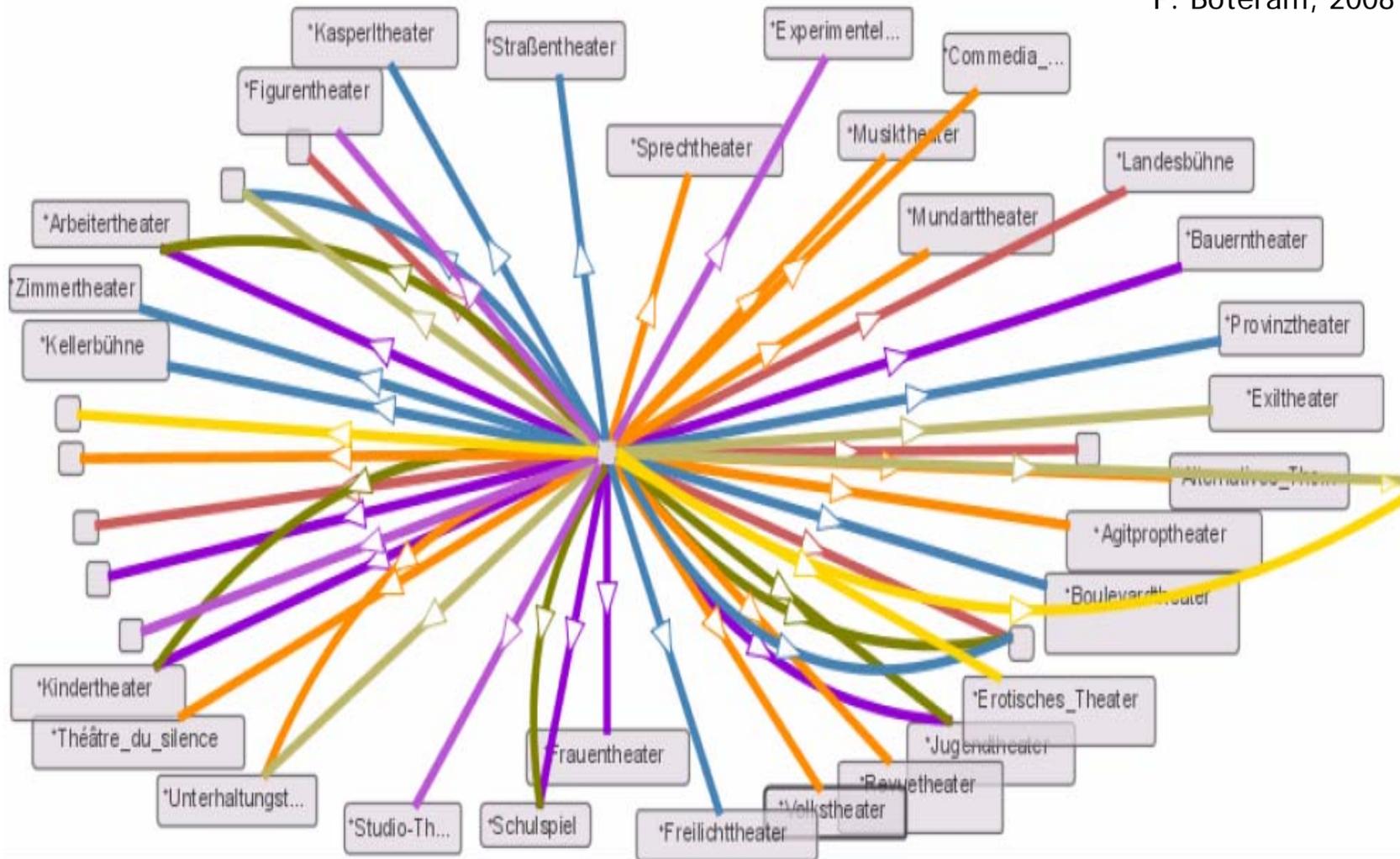


Example



Visualization of the subject domain *Theater* of the *SWD* considering only the existing relation *BT / NT*. The state as it is now.





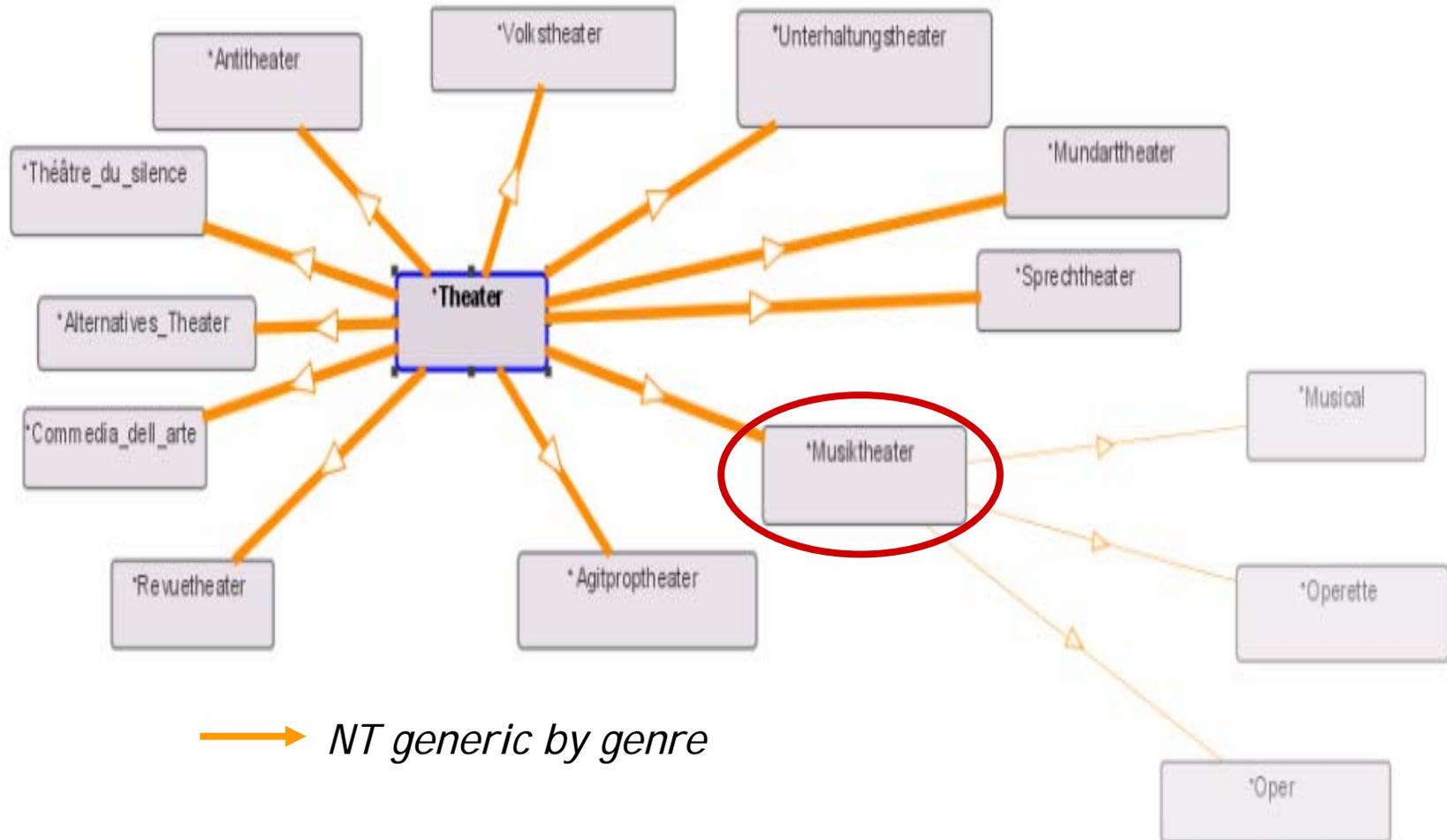
Visualization of the subject domain *Theater* of the *SWD* considering a set of typed semantic relations (coloured edges).

Although this may look a little bit confusing, there is potential for selecting a specific type of relation for improving the retrieval result.

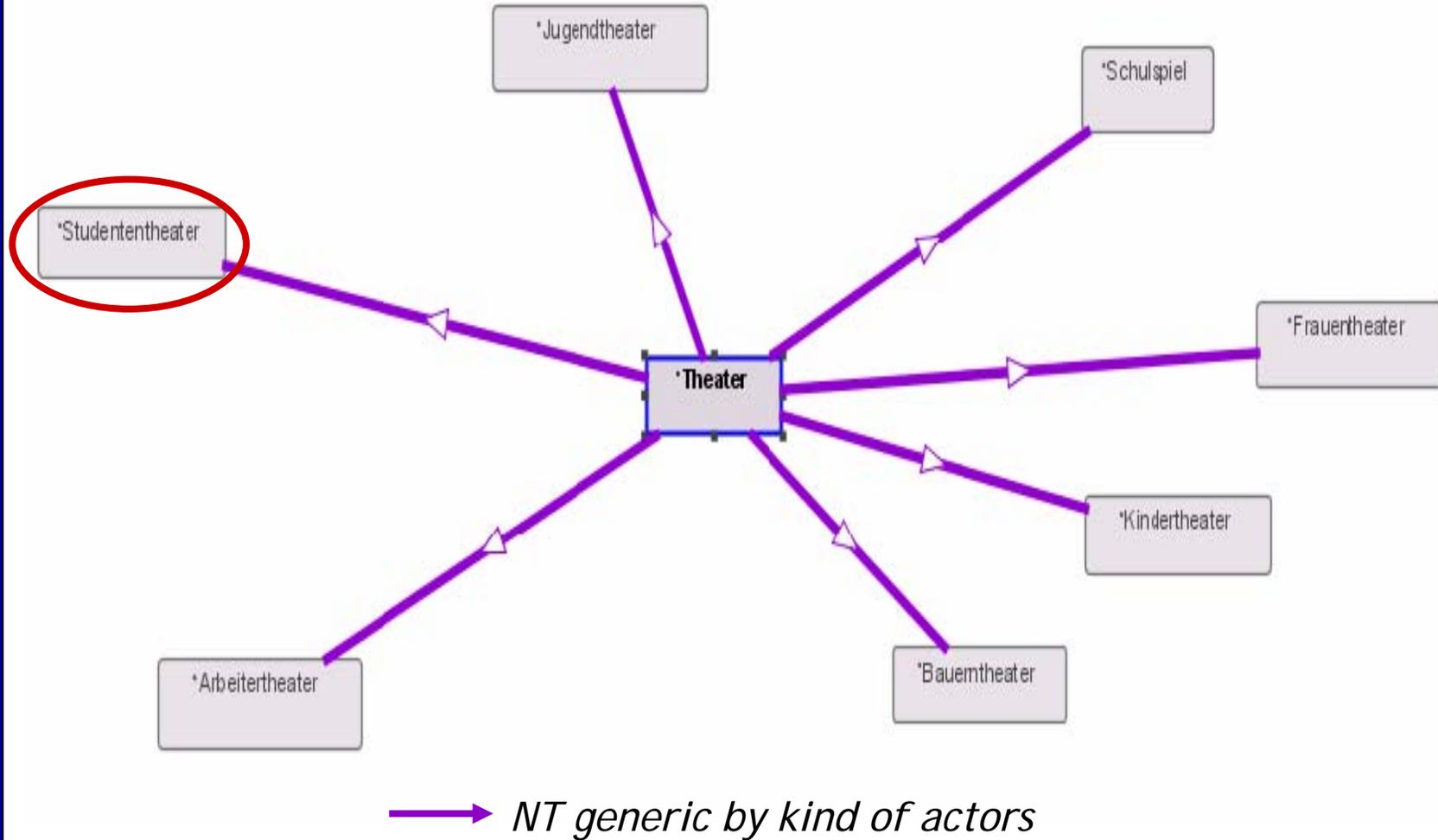


Potential benefit for retrieval purposes

1. Selection of concepts corresponding to *Theater* by choosing the relation *NT generic by genre*.



2. Subnetwork of concepts corresponding to *Theater* by choosing the relation *NT generic by kind of actors*.



Second proposal: Ontological spine with localized semantic networks

Constituents of such a proposal are

- Development of an ontological spine with precise and logically valid relationships between the classes, especially focusing on heredity for hierarchical relations;
- Development and use of an inventory of typed and logically valid relations in the corresponding semantic network(s) representing localized knowledge structures;
- Development and use of clear criteria for connecting the terms of the multilingual concept networks with classes of the ontological spine.

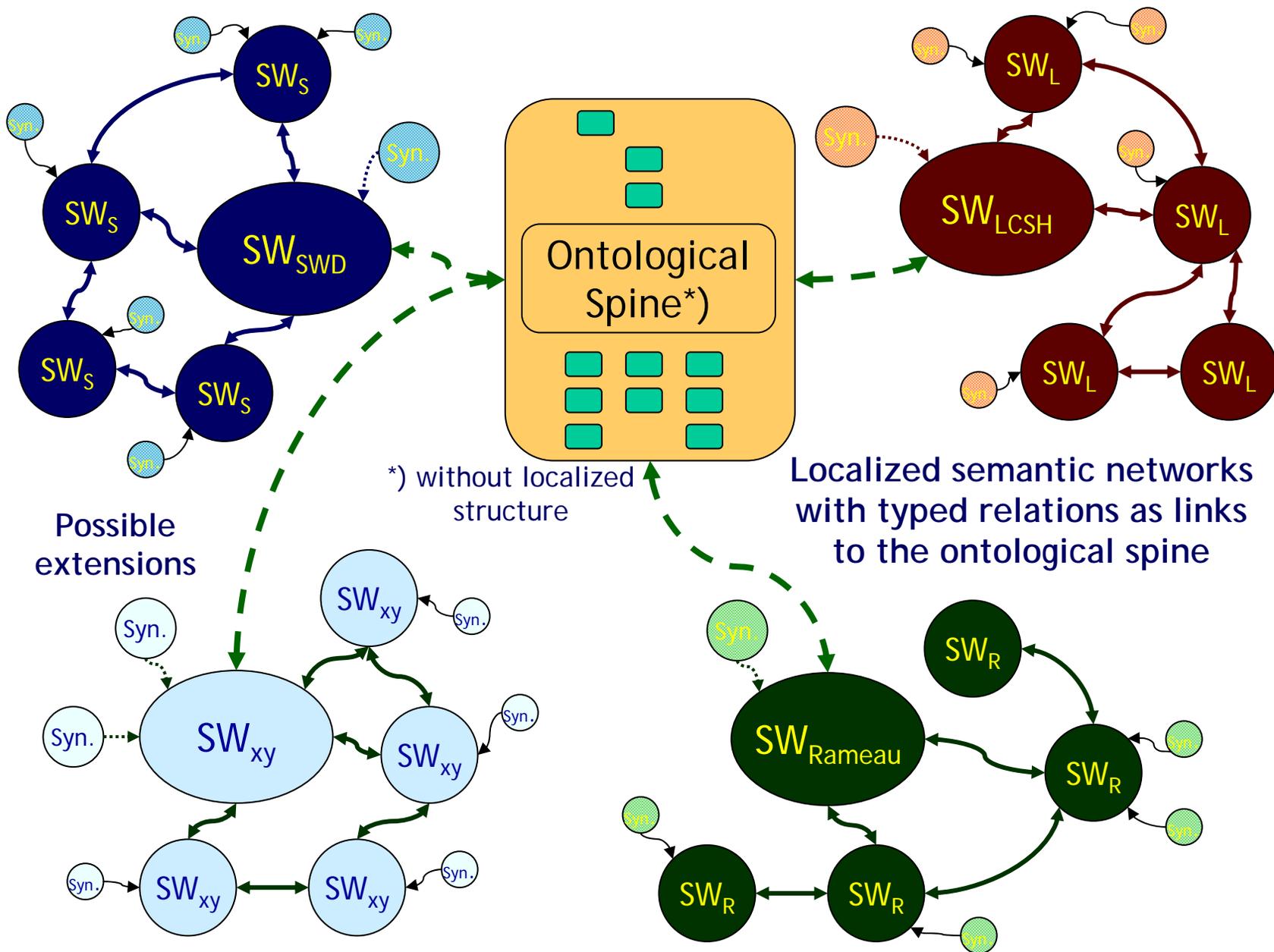
The result of such an approach can be described as an ontological spine with multilingual satellites of localized concept networks. Each network is connected with the spine in order to navigate between the networks and support insight into the respective conceptual context.

This may look like:



Second proposal

Ontological spine without localization as upper ontology and links to localized semantic networks with typed relations



Advantages

- All relations are logically valid;
- Concepts and relations cannot only be interpreted cognitively but are also machine processible (e.g. in form of inferences along the edges of the network);
- All aspects of localization can be retained within the context of the semantic networks;
- Specification of different knowledge contexts within the retrieval process can be done by selection of different types of relations;
- Conceptual navigation processes can be designed on the basis of the elaborated relations;
- The backbone can be seen as a gateway for a user to enter a subject / a thematic context in a knowledge field in not well familiar language and localization when coming from a more familiar language and localization;
- The terms of the semantic networks have the function of an entry vocabulary for the classes of the ontological spine;
- Addition of new languages is very easy and does not have any implications on the spine or any of the semantic networks.



Steps to realize such a proposal

- Development of an ontological spine as backbone and gateway for semantic networks representing localized knowledge structures;
- Development of an inventory of typed and logically valid relations for the networks;
- Transforming the authority files into semantic networks structured by this extended set of relations instead of building a new one;
- Creation of mappings between the spine and the networks along clear criteria for connecting the terms of the multilingual concept networks with classes of the ontological spine;
- Development of representation models for the conceptual entities and structures of the spine and the networks regarding the requirements of Semantic Web standards (e.g. *RDF*, *SKOS*, *OWL*);
- Development of retrieval facilities making use of the terms of the networks as entry vocabulary and the relations within the spine and the networks as navigational tools;
- Development of corresponding Web services.

Is it realistic to believe in such developments ?





Thank you for your attention.

winfried.goedert@fh-koeln.de