

## *Retrieval Concepts and Mapping Strategies:*

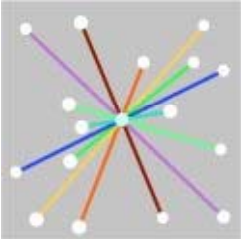
### *The Potential of CrissCross for Improving Access to the DDC*

Jessica Hubrich, M.A., M.L.I.S.  
*Team leader CrissCross project*

Cologne University of Applied Sciences  
Institute of Information Management

Symposium “Dewey goes Europe”,  
Austrian National Library, 28<sup>th</sup> April 2009

## Starting Point



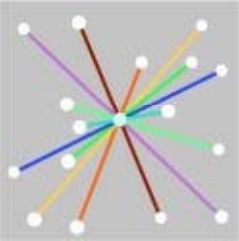
Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

Functionality and efficiency of topical search processes depend on the underlying retrieval concepts and the kind of subject data that is integrated within information retrieval systems. Compared to homogeneous retrieval environments, heterogeneous information spaces require enhanced concepts taking into account the specificity of the information space and the potential of the used distinct indexing data.

### Questions

- How do retrieval concepts influence search functionalities?
- To which extent can the establishment of links between distinct indexing languages improve efficiency of topical queries in heterogeneous information spaces?
- What are the benefits of the linkages produced within the project *CrissCross*?

# Retrieval Concepts (I)

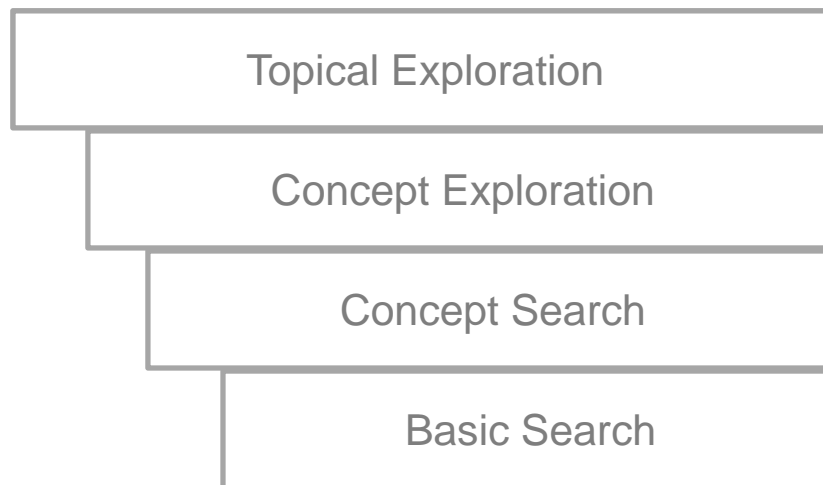


Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

Retrieval concepts aim to support

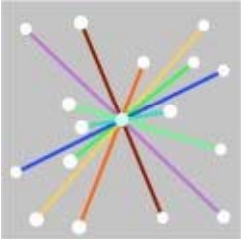
- document retrieval in the narrower sense of the term
- information seekers in finding relevant documents by providing tools for orientation, navigation, exploration

Ideally, retrieval concepts are accompanied by concepts of relevance ranking.



**Central retrieval concepts**  
in respect to topical queries

## Retrieval Concepts (II)

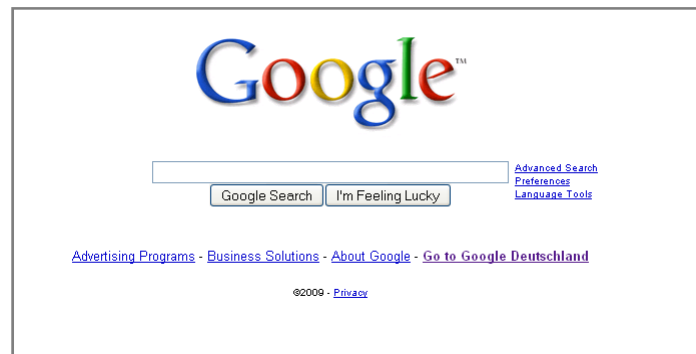


Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

### Basic search based on string matching

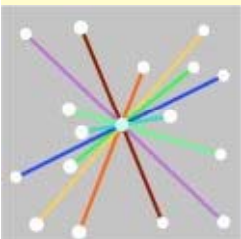
Initial search terms are compared with elements of a generated index and might refer to

- keywords of titles or of abstracts
- main form of subject headings
- notations



Modifications of this search are found in many librarian opacs often combined with the possibility to search within indices.

## Retrieval Concepts (III)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

### Conceptual query based on concept matching

Initial search terms are enhanced and modified in regard to the meant concept. The efficiency of this feature depends on the quality of the integrated controlled vocabulary that identifies synonyms.

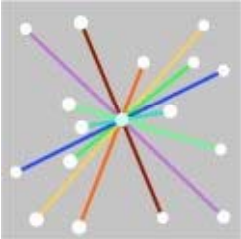
The screenshot shows a search interface titled 'Index Schlagwort'. It includes an input field labeled 'Eingabe' and buttons for '[Anzeigen ab]', 'Start Indexsuche', 'Übernehmen', and 'Zurück'. Below the input field, there are navigation arrows '<<' and '>>'. A list of search results is displayed, each with a checkbox, a term, and a count. The entry 'Malus domestica ==> siehe Apfel' is highlighted with a red circle. Other entries include 'Malus communis ==> siehe Apfel', 'Malus <Gattung> ==> siehe Apfelbaum', 'Malvaceae ==> siehe Malvengewächse', 'Malvasia ==> siehe Monemwasia', 'Malventum ==> siehe Benevent', and 'Malvern <Pa., 2002>'. The counts for these entries are 1, 4, 1, 4, 4, and 5 respectively.

Term	Count
<input checked="" type="checkbox"/> Malus communis ==> siehe Apfel	1
<input checked="" type="checkbox"/> Malus domestica ==> siehe Apfel	4
<input checked="" type="checkbox"/> Malus <Gattung> ==> siehe Apfelbaum	1
<input checked="" type="checkbox"/> Malvaceae ==> siehe Malvengewächse	4
<input checked="" type="checkbox"/> Malvasia ==> siehe Monemwasia	1
<input type="checkbox"/> Malvengewächse	4
<input checked="" type="checkbox"/> Malventum ==> siehe Benevent	4
<input type="checkbox"/> Malvern <Pa., 2002>	5

(Resource: ULB Münster)

This search can be found in many librarian opacs, sometimes combined with the possibility to search within the specific subject index.

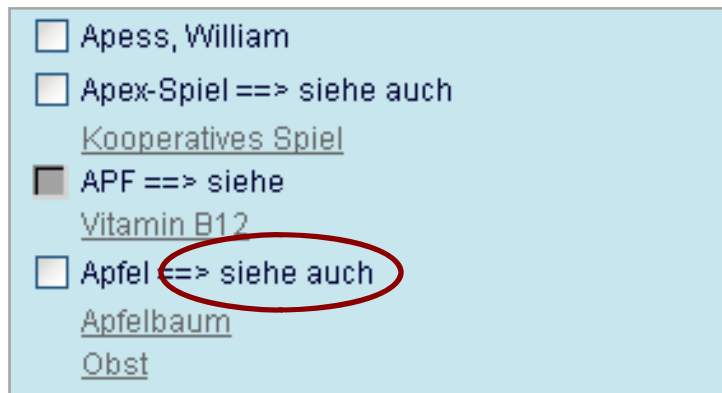
## Retrieval Concepts (IV)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

### Conceptual exploration based on a priori conceptual relations

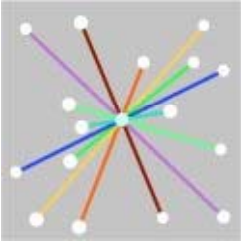
The semantic environment of a concept that corresponds to the initial search term is provided for search modification. The degree of orientation and the efficiency of such a feature depend on the quality and expressiveness of the semantic structure of the knowledge system that is referred to.



(Resource: ULB Münster)

The expressiveness of semantic relations within indexing languages is often restricted. This retrieval concept has not yet been integrated adequately in librarian opacs.

## Retrieval Concepts (V)



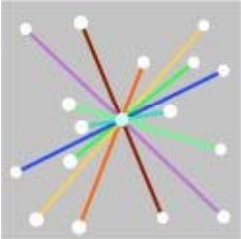
Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

### **Topical Exploration** based on a posteriori conceptual relations

Taking former search results as initial points, this retrieval concept aims to support topical exploration processes to assist information seekers in clarifying their information needs. Expressive a priori semantic relations between concepts of an integrated knowledge organization system as well as syntactical operators are provided that allow qualified statements about a posteriori relations inherent in topics of the specific documents.

A system that adequately supports processes of topical exploration has not been realized yet.

# Relevance Ranking



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

Search concepts in the narrower sense of the term can be supplemented by concepts of **relevance ranking**. Concepts of relevance ranking provide algorithms for ordered display of search results based on specific assumptions concerning the factors that may influence the relevance of a document in respect to the conducted search.

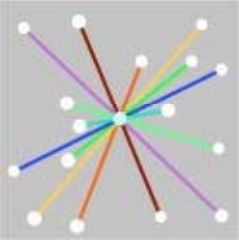
Criteria for topical ranking in librarian catalogues might be

- Uniqueness of search terms within the database
- Proportion of search terms present in a bibliographic record
- Fields in which search terms occur (Subject fields vs. title fields).
- ....

In respect to heterogeneous information spaces, criteria concerning the relevance of embedded data of distinct indexing languages must be developed integrating the potential given with the specific mapping data.



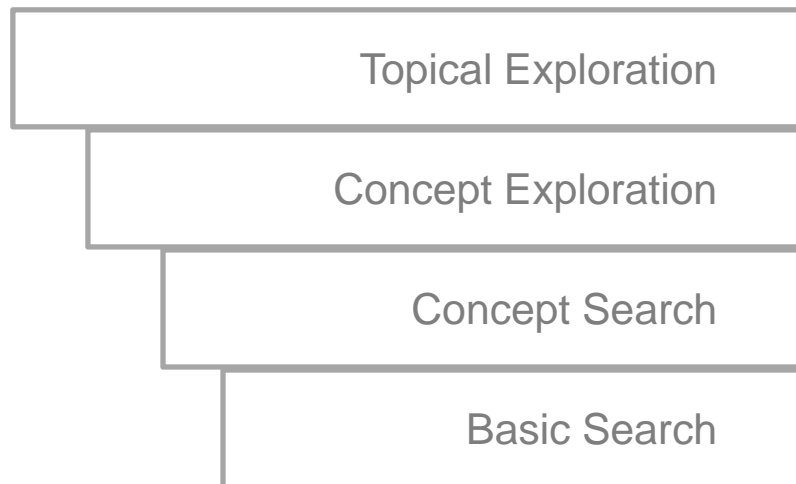
# Retrieval Concepts and Mapping Strategies



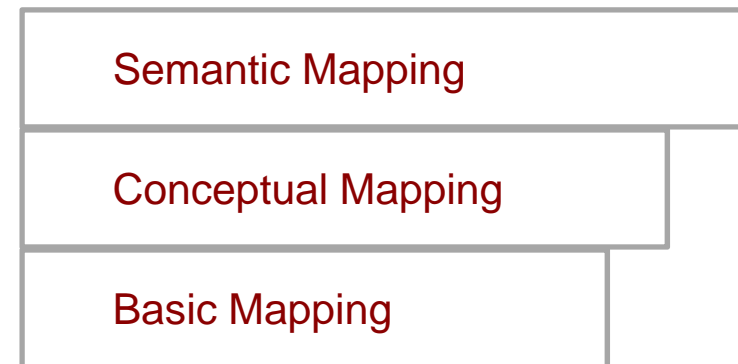
Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

In respect to heterogeneous information spaces, functionality and efficiency of queries can considerably be improved by establishing links between relevant indexing languages. However, their practicability concerning the different retrieval concepts differ according to the specific mapping strategy applied.

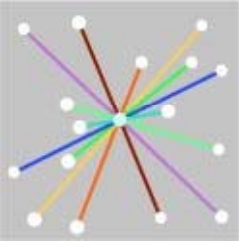
## Retrieval Concepts



## Mapping Strategies



# Mapping Strategies (I)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

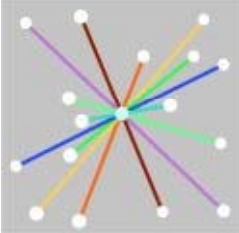
**Basic Mapping** focused on the main representation form of a concept

Crosswalks between indexing languages are established taking the main representation form of a concept as initial point. The semantic relations between the mapped terms are not further described. Generally, the mappings are saved separately from the databases of the knowledge systems.

In retrieval scenarios

- the matching algorithms are extended taking advantage of existing indexing data. *Recall* is improved.
- equivalence links are conceived as term clusters
- controlled access points to other vocabularies are provided in form of main headings; information seeker might use the language he or she is familiar with

## Mapping Strategies (II)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

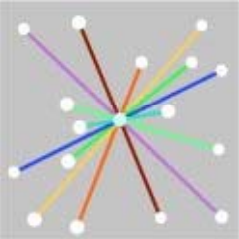
The mapping strategy of *Multilingual Access to Subjects (MACS)* is originally based on this mapping concept.

The screenshot shows the Theische European Library search interface. The search term 'ethik' is entered in the search bar. The results are displayed in three columns:

- Search Results:** Results 1 - 7 of 7 for *ethik*. (0.008 seconds)
  - MACS - Select this item to query in TEL*
    - English: Ethics
    - French: Morale
    - German: *Ethik*
  - MACS - Select this item to query in TEL*
    - English: Merit (Ethics)
    - French: Mérite
    - German: Verdienst (*Ethik*)
- Search SRU urls for TEL:** [Hide]
  - Library: SNL [query=\(subject exact "Ethik"\)](#)
  - Library: BNF [query=\(subject exact "Morale"\)](#)
  - Library: BL [query=\(subject exact "Ethics"\)](#)
  - Library: DNB [query=\(subject exact "Ethik"\)](#)
  - [search in TEL portal](#)
- Selected items to search in TEL:**
  - MACS - Remove this item from the selection*
    - English: *Ethics*
    - French: *Morale*
    - German: *Ethik*

(Resource: <http://lvat.hoppie.nl:8080/portal/en/lvat.html>)

## Mapping Strategies (III)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

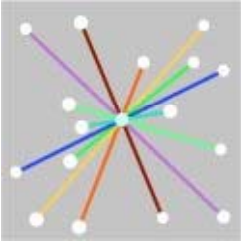
### **Conceptual Mapping** focused on concepts

The mapping strategy aims to establish linkages between concepts of distinct indexing languages taking the whole connotation scope of a concept as initial point and describing exactly the mapping direction wherever necessary. The intersystem relations are further described and are stored together with the identifier of the mapped concept/s within a knowledge organization system.

In retrieval scenarios

- the matching algorithms are further extended taking advantage of existing indexing data. *Recall* is improved.
- conceptual search is supported
- intersystem relations allow to influence *recall* and *precision* and to navigate more effectively between knowledge systems

## Mapping Strategies (IV)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

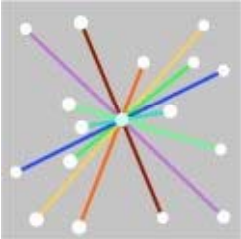
**Semantic Mapping** considering the concepts as well as intraconcept relations

Ideally, mapping relations complement highly expressive and accurately structured relational knowledge systems. The relational structure of the participating systems contribute to the meaning and usage of the individual concepts. Taking the structural and functional setups of these systems into account and additionally erecting expressive, logical valid and specified intersystem relations characterizes the strategy of semantic mapping.

Semantic mapping has not been conducted yet.

However, the additional value would be substantial: In retrieval scenarios all search matching processes would be supported as well as intercultural and international concept exploration.

# CrissCross



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

Project run time: 2006 – 2010

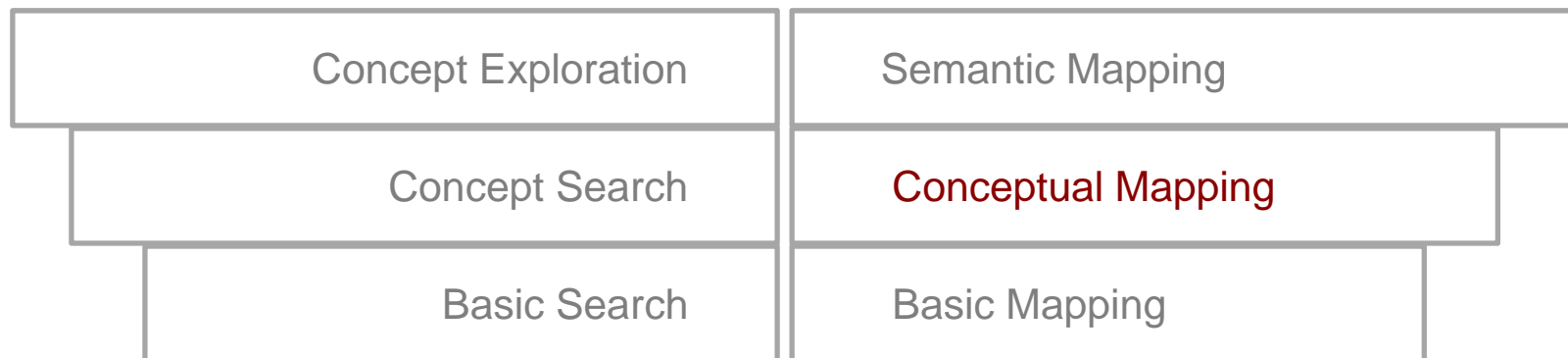
Project Sponsor: German Research Foundation

Cooperation partners: German National Library

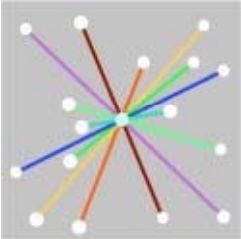
Cologne University of Applied Sciences

Aim: Creation of a thesaurus-based and user-friendly research vocabulary that facilitates research in heterogeneously indexed collections

Central focus: Linking of subject headings of the *German Subject Heading Authority File (SWD)* to notations of the *Dewey Decimal Classification (DDC)*



# CrissCross — Mapping Strategy (I)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

## Characteristica of the CrissCross Conceptual Mapping

- unidirectional: SWD → DDC
- as comprehensive as possible / *One-to-many Mapping*

800 |s|Apfel

816 641.3411#3#

+

816 583.73#2#

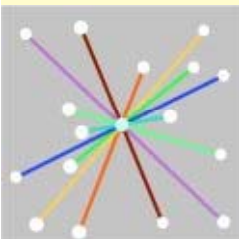
816 634.11#3#

interdisciplinary works on apples –  
located in class for apples as food

works that refer to disciplinary  
aspects of the subject heading  
(botany / agriculture)

- as specific as possible / *Deep Level Mapping*  
Built numbers constructed within the frame of *CrissCross* are stored institutionally in MelvilClass (including number components)

## CrissCross — Mapping Strategy (II)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

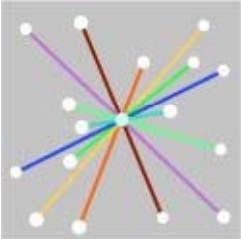
- allocated notations are stored directly in the data record of the specific SWD subject heading

```
005 Ts1
012 /w
021 4002405-2
800 |s|Apfel
808 |a|B 1986
808 |c|Für Biologisches u. Agrarbiologisches zu Teilen der Pflanze verknüpfe z.B. mit Frucht
809 |x|bau *erl
810 32.4;24.3
816 583.73#2#
816 634.11#3#
816 641.3411#3#
830 |s|Malus domestica
830 |s|Gartenapfel
830 |s|Malus communis *B 1986
830 |s|Äpfel
850 |s|Obst
860 |s|Apfelbaum
```

Semantic structure of SWD is  
available with mappings



## CrissCross — Mapping Strategy (III)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

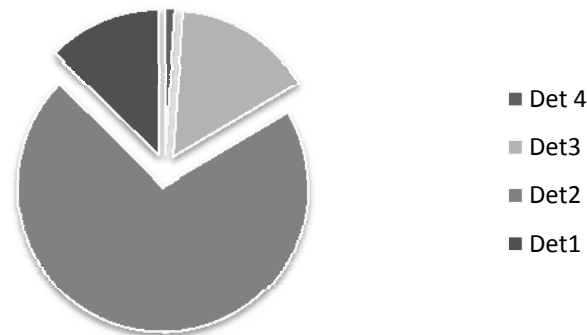
- The different levels of contentual congruence between SWD subject headings and assigned DDC notations are expressed by four so-called *Degrees of Determinacy* which are aligned to the direction of the mapping as well as to the mapping specificity and are - wherever possible - adjusted to the structure of the target classification (esp. instance-class relations)

Det 4: Connotation scope is (nearly) identical

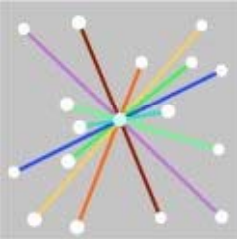
Det 3: Connotation scope approximates the whole

Det 2: Connotation scope reflects a part

Det 1: Connotation scope corresponds to a small part

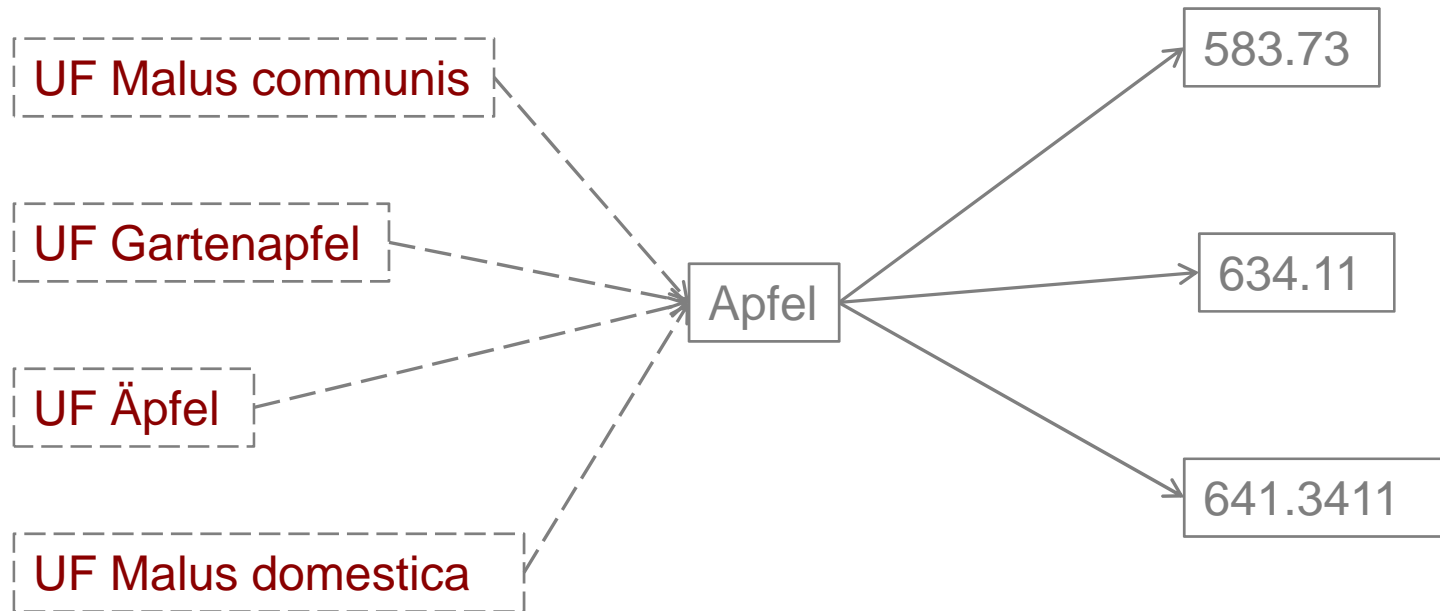


# CrissCross — Retrieval Concepts (I)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

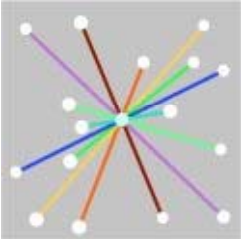
- **String Matching / Concept Matching**



SWD main headings as additional access points to the DDC

SWD concepts as additional access points to the DDC

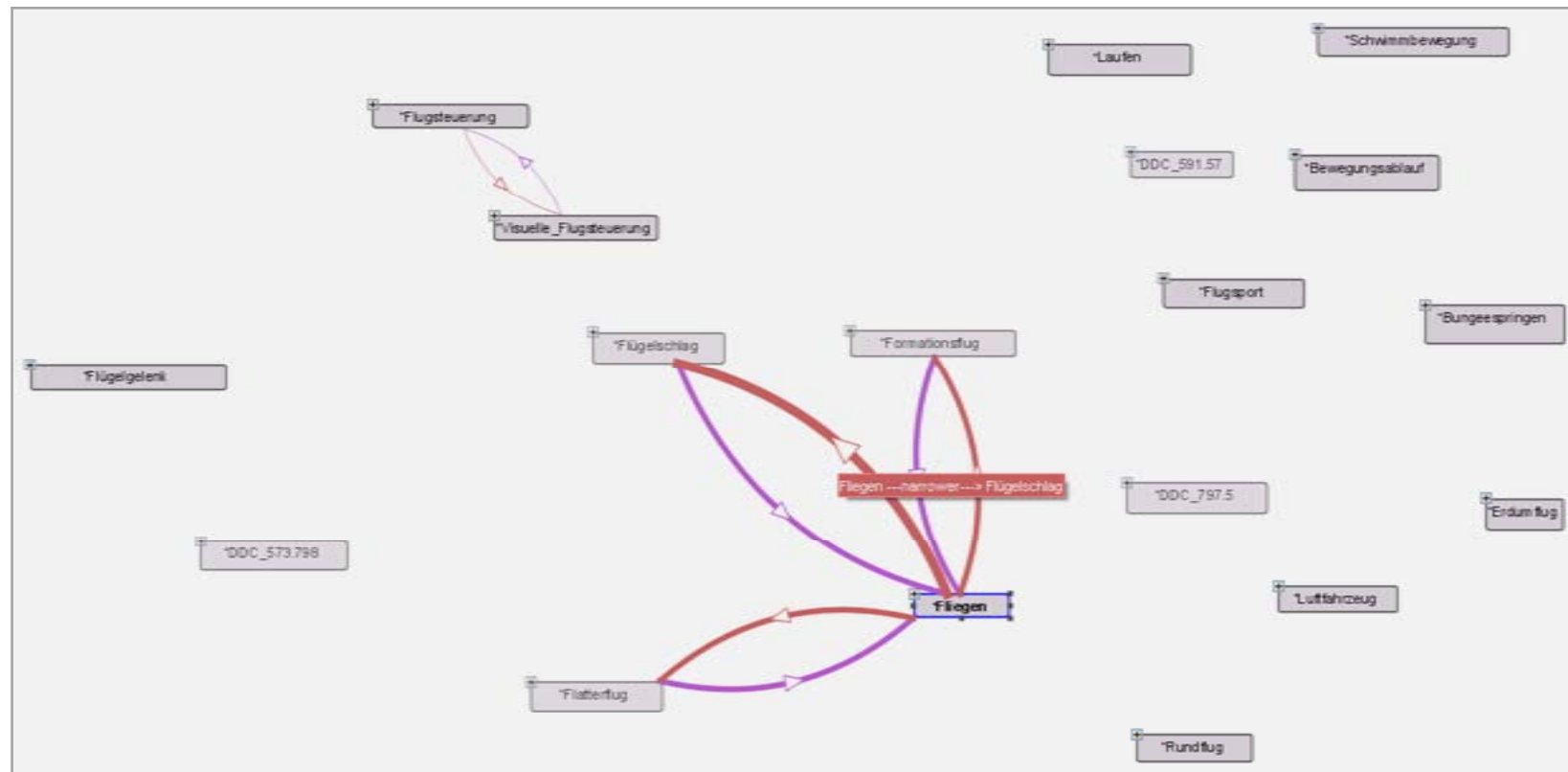
## CrissCross — Retrieval Concepts (II)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

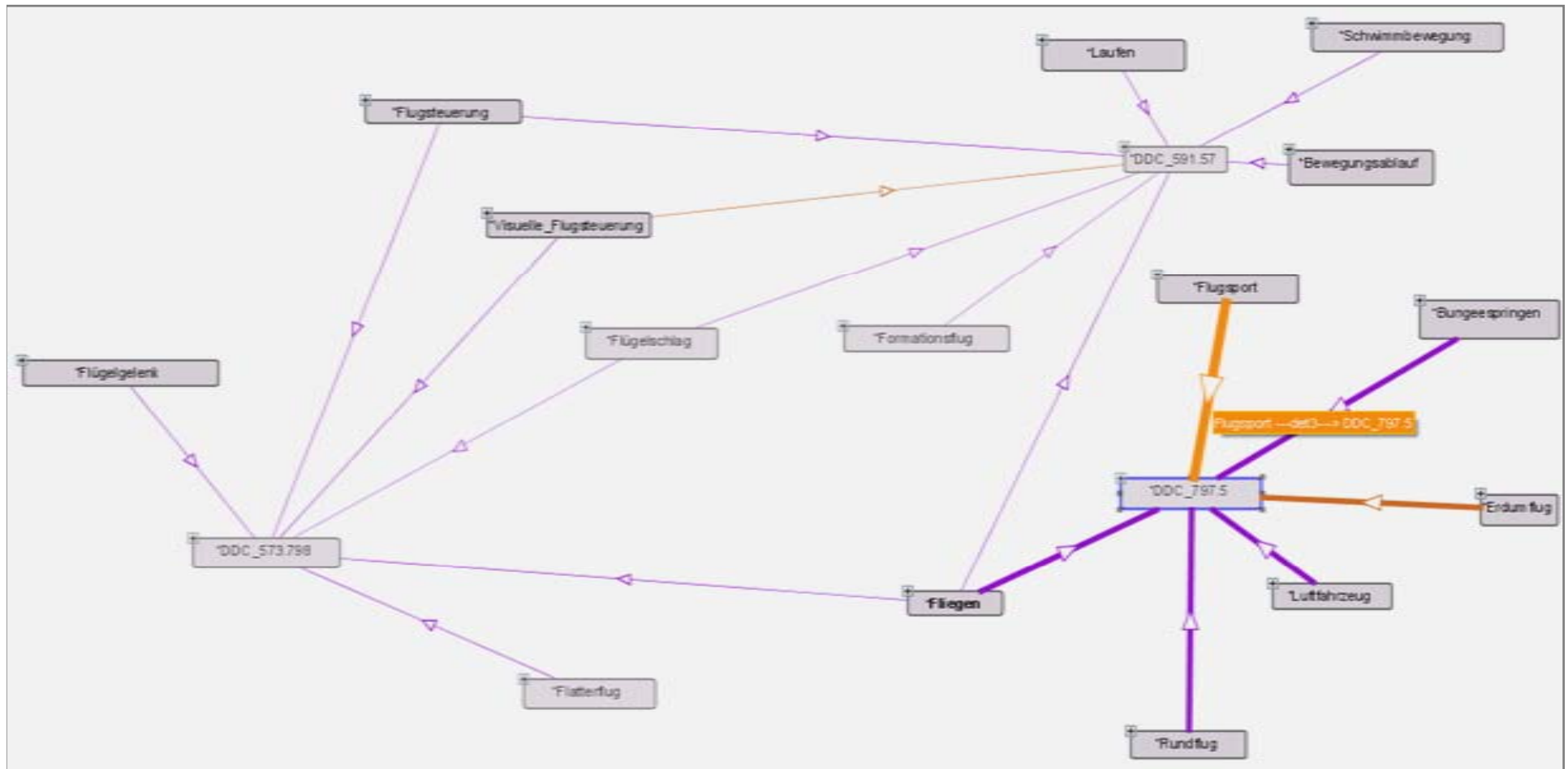
- **Conceptual Exploration**

- based on semantic structure of the DDC (primarily hierarchical)
- based on semantic structure of the SWD (BT, NT, RT)



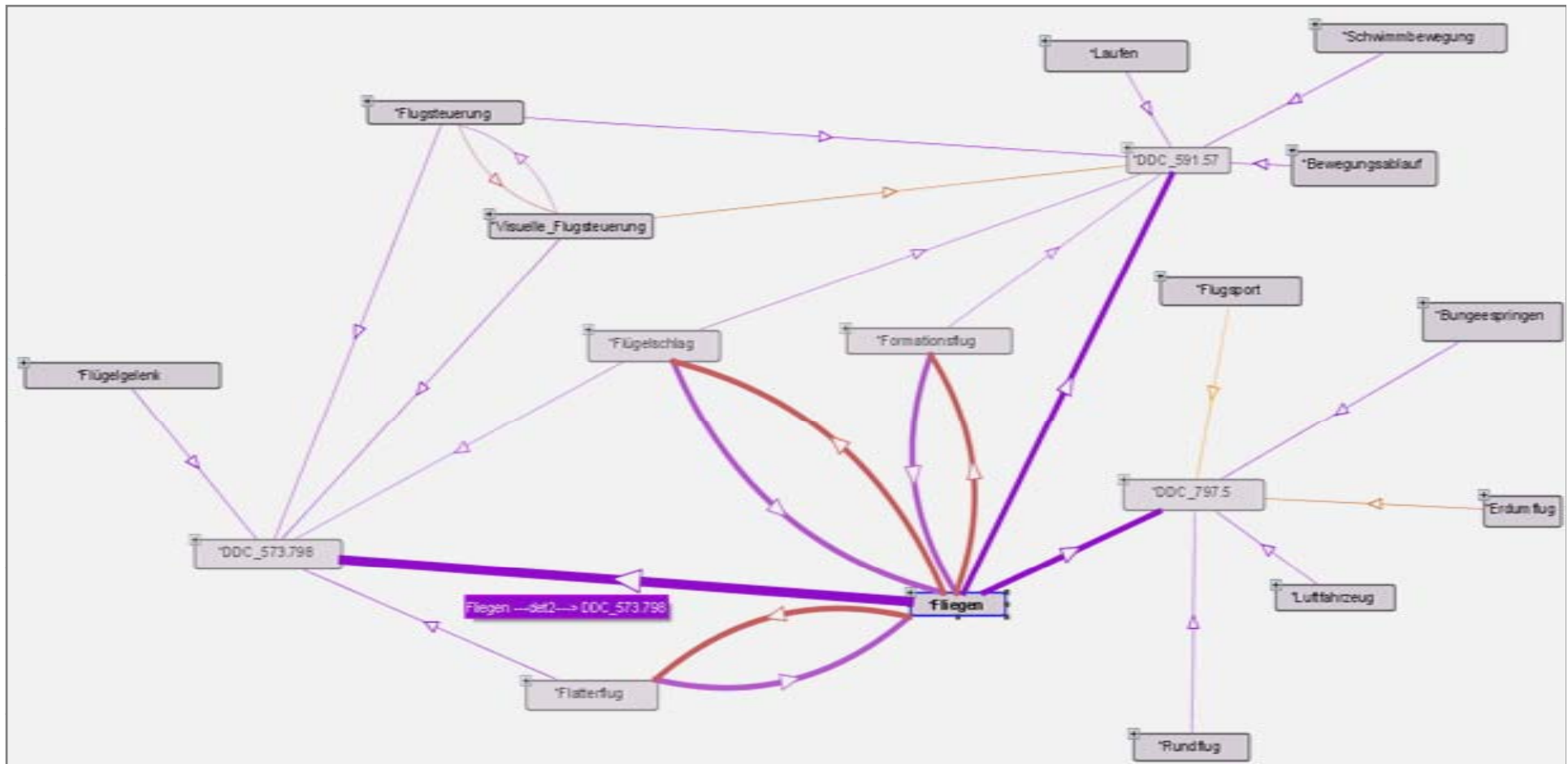
# CrissCross — Retrieval Concepts (III)

## Conceptual Exploration based on CrissCross

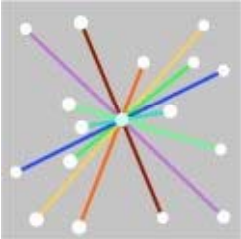


# CrissCross — Retrieval Concepts (IV)

## Conceptual Exploration based on SWD and CrissCross



# CrissCross — Relevance Ranking (I)



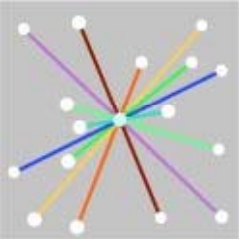
Retrieval Concepts and Mapping Strategies : The Potential of CrissCross

Due to the qualitative mapping strategy that is adjusted to the participating knowledge systems, CrissCross provides several possibilities for relevance ranking:

- Ranking of documents that are assigned a specific DDC number based on the *Degrees of Determinacy as the Degrees on Determinacy* describe how a subject heading „fits“ into a class

	<u>DDC-Übersicht</u>	<b>SWD-Schlagwörter</b>	<b>Relevanz</b>
<u>500</u>	<u>Naturwissenschaften</u>	<u> s Rosengewächse</u>	<b>hoch</b>
<u>570</u>	<u>Biowissenschaften; Biologie</u>	<u> s Rosenpflanzen</u>	
<u>579-590</u>	<u>Naturgeschichte einzelner Arten von Organismen</u>	<u> s Adenostoma</u>	<b>mittel</b>
<u>580-590</u>	<u>Pflanzen und Tiere</u>	<u> s Apfel</u>	
<u>580</u>	<u>Pflanzen (Botanik)</u>	<u> s Apfelbaum</u>	
<u>583-588</u>	<u>Einzelne taxonomische Pflanzengruppen</u>	<u> s Apfelgewächse</u>	
<u>583</u>	<u>*Magnoliopsida (Zweikeimblättrige)</u>	<u> s Aprikose</u>	
<u>583.7</u>	<u>*Rosidae</u>	<u> s Arktische Brombeere</u>	
<b><u>583.73</u></b>	<b><u>*Rosales</u></b>	<u> s Aronia</u>	
<u>583.734</u>	<u>*Rosa (Rosen)</u>	<u> s Birnbaum</u>	
		<u> s Birne</u>	

## CrissCross — Relevance Ranking (II)



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

- Ranking of documents with different DDC numbers based on the *Degrees of Determinacy*

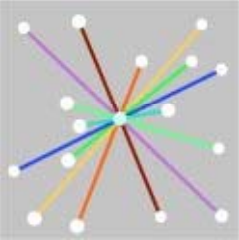
As the *Degrees of Determinacy* are adjusted to the relations between topics and classes like they are displayed in the DDC and the latter are based on literary warrant, it is likely that more relevant literature concerning the concept described by the subject heading can be found within a set of documents that are assigned a DDC number that is mapped with a higher *Degree of Determinacy*.

800 <u>s Apfel</u>	1	2	816 <u>641.3411#3#</u>
	3		816 <u>583.73#2#</u>

Retrieval tests conducted so far could prove this assumption.

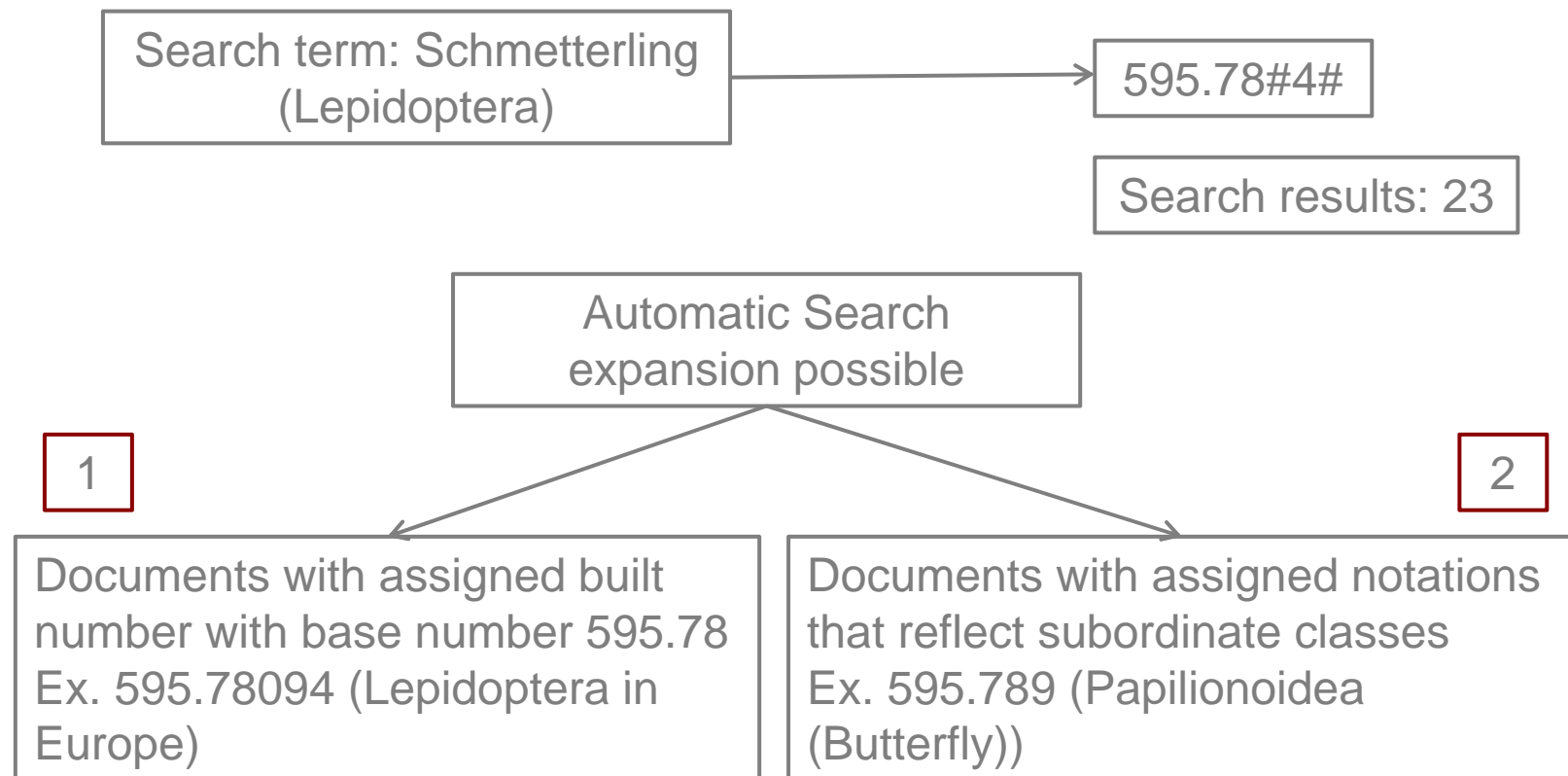
If the integration of the mapping data leads to an unmanageable search result set, the *Degrees of Determinacy* can likewise be used to control *recall* (and *precision*)

## CrissCross — Relevance Ranking (III)



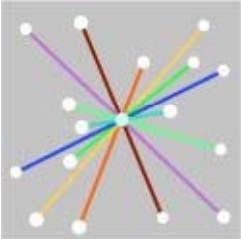
Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

- Even in respect to displaying search results in subsequence to a search expansion integrating a posteriori concepts, the *Degrees of Determinacy* give hints to which assigned DDC numbers might be of higher relevance.





# CrissCross — Future Prospects



Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

## CrissCross and the Semantic Web

*Simple Knowledge Organization Language (SKOS)* as quasi-standard for publishing knowledge organisation systems on the *Semantic Web*

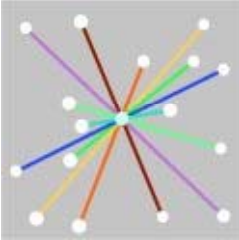
*but*

- not adjusted to classifications and to mappings between typological distinct knowledge systems
- CrissCross relations cannot adequately be represented in *SKOS* mapping relations

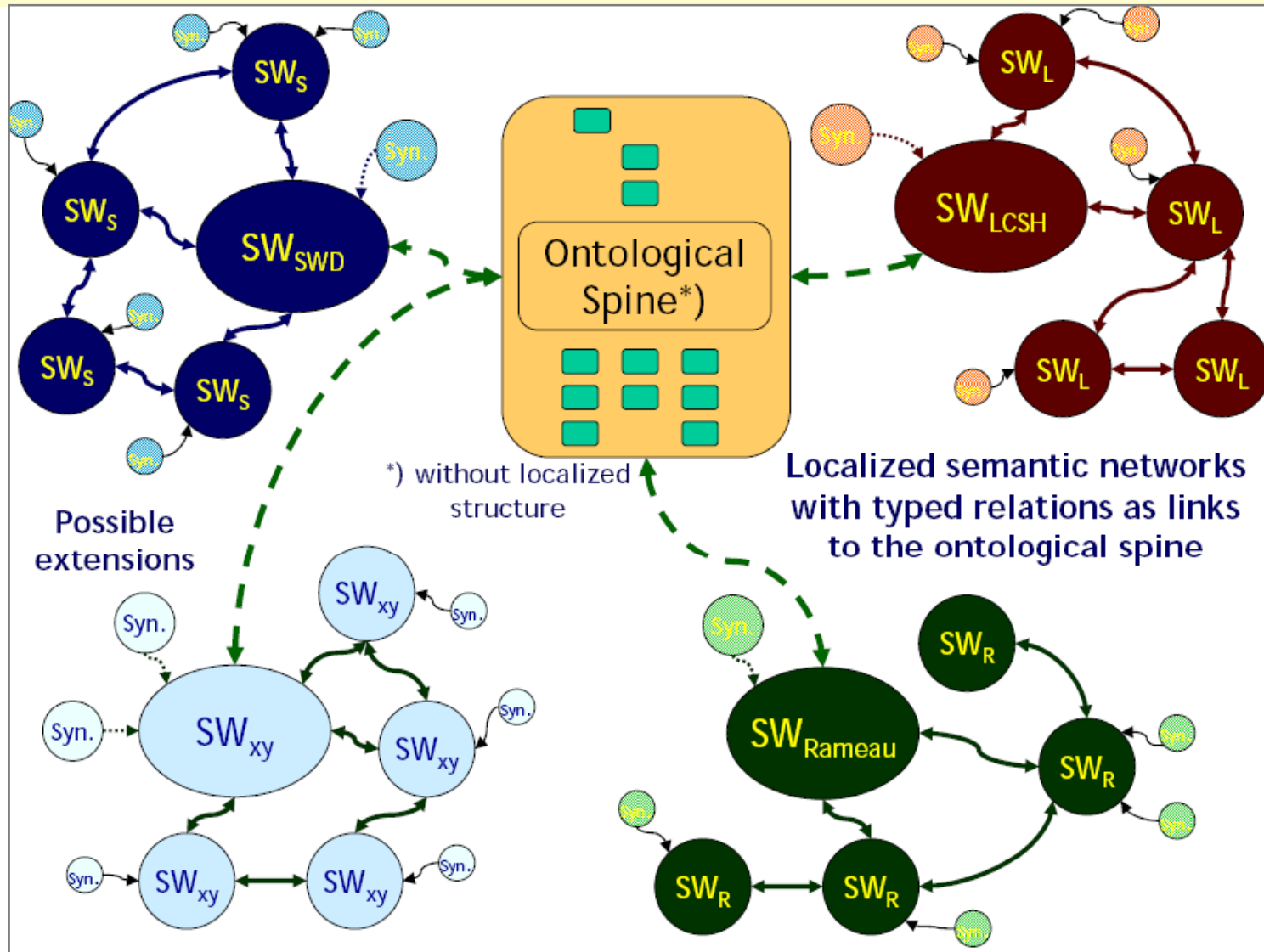
→ Solution:

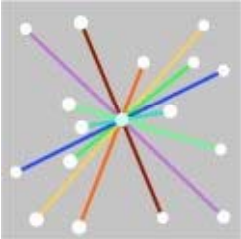
Using *SKOS* and *OWL (Web Ontology Language)*, constructing adequate RDF representation

# CrissCross — Future Vision?



Retrieval Concepts and Mapping Strategies : The Potential of CrissCross





Retrieval Concepts and Mapping  
Strategies : The Potential of CrissCross

Thank you for your attention!

Homepage *CrissCross* project  
[http://linux2.fbi.fh-koeln.de/crisscross/index\\_en.html](http://linux2.fbi.fh-koeln.de/crisscross/index_en.html)

Jessica Hubrich, M.A., M.L.I.S  
*Team Leader CrissCross project*  
jessica.hubrich@fh-koeln.de