Interoperability in distributed archives with authority-controlled Ontologies

Alfons Ruch

University of Passau

December 2009

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Overview

1. Motivation
2. Authority Control
3. Interoperability
4. Conclusion and Future Work

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Overview

1 Motivation

2 Authority Control

3 Interoperability

4 Conclusion and Future Work

Motivation

Numerous digital archives exist in many different application domains.

Application domains

- cultural heritage preservation

- national libraries

- earth observation

Motivation

Characteristics of the archives

- independent,
- autonomous and
- ever growing.

These facts together result in a semantical divergence.

Motivation

Archives may define arbitrary annotation schemes for the documents depending on their application domain.

Annotation schemes

- Tags
- Taxonomies
- Ontologies

But these are specialized to the specific needs of the application domain.

Motivation

Archives may define arbitrary annotation schemes for the documents depending on their application domain.

Annotation schemes

- Tags
- Taxonomies
- Ontologies

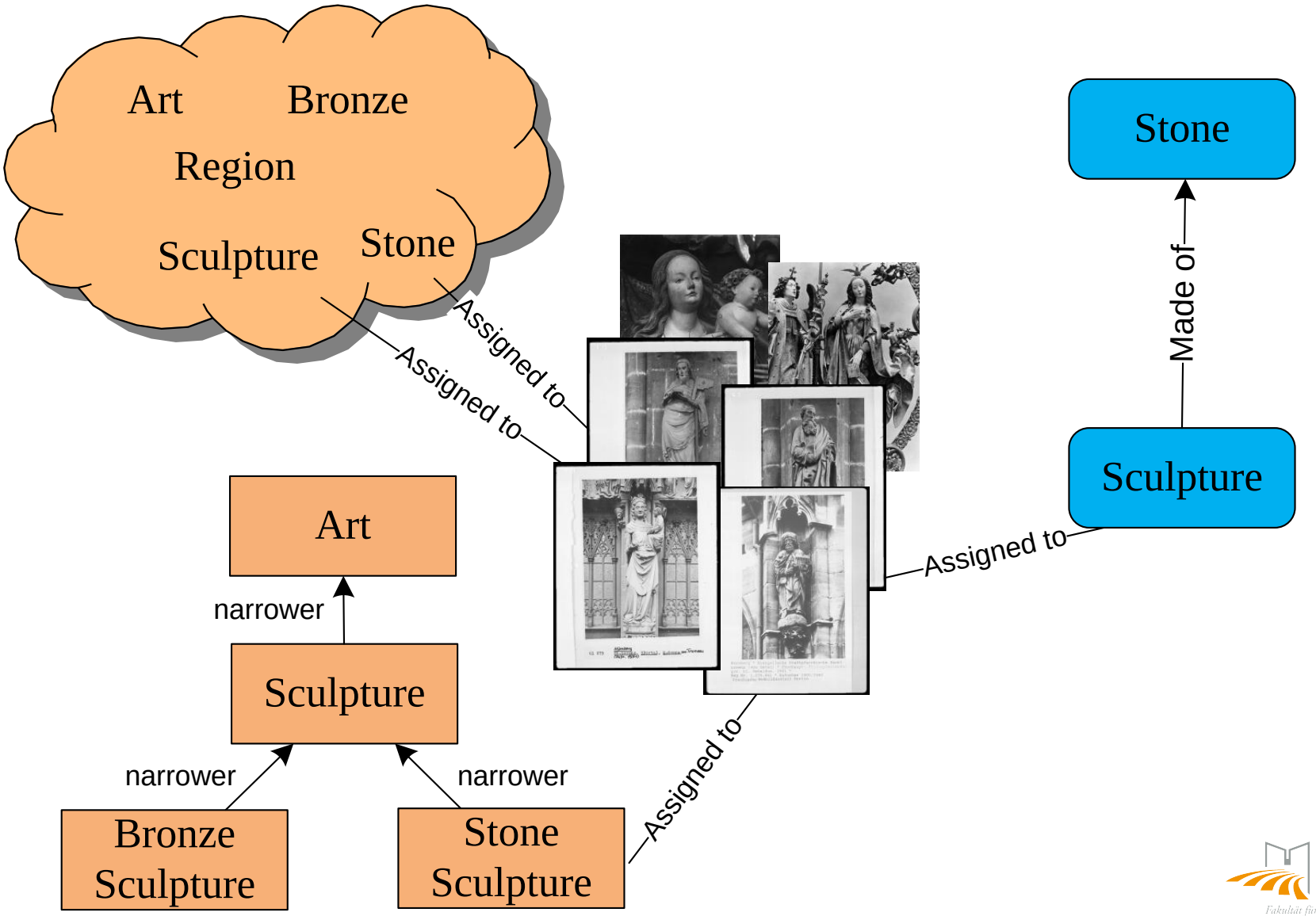
But these are specialized to the specific needs of the application domain.

Use Case: Digital Archives for Monumental buildings

The MonArch Project (Monumental buildings Archive network) has the following objectives:

- Make archival items digitally accessible
- Connect existing isolated and autonomous archives
- Connect to semantically different digital archives
- Develop a web-based platform

Use Case: Different Annotations



Use Case: MonArch Application Domains

In the MonArch Project different application domains are involved.

Application Domains

- Architectural buildings documentation and archiving
- Cultural heritage preservation

Overview

1 Motivation

2 Authority Control

3 Interoperability

4 Conclusion and Future Work

Authority Control and Authority Files

Definition (Authority Control)

Authority Control means to use and maintain the forms of names and subjects consistently. Authority Control consists of two main parts:

- Name Authority Control is the procedure serving to maintain a consistent use of the names.
- Subject Authority Control represents the concepts used for the subject heading of the records.

Authority Control and Authority Files

Definition (Authority Record)

Authority Record is the information about a name, subject or place. An authority record contains a heading, cross references and statement of justification.

Definition (Authority File)

Authority File is the catalog of authority records.

Authority Record coded in MAB

Example (Subject “Sculpture”)

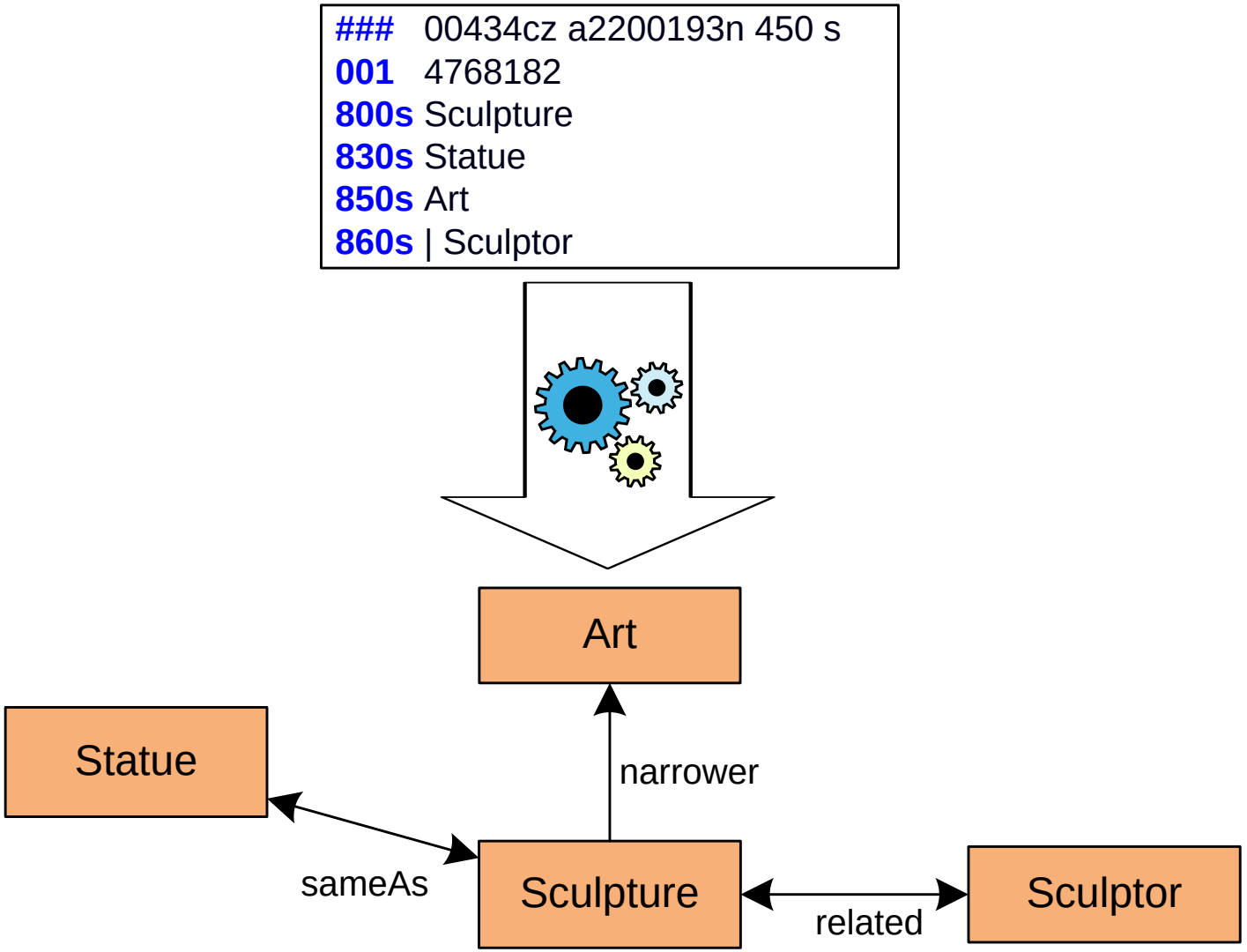
```
###  00434cz a2200193n 450 s
001  4768182 (Identifier)
800s Sculpture  (name of the concept)
830s Statue (synonym of the concept)
850s Art (a broader concept)
860s |Sculptor  (related concept)
```

Authority Control and Authority Files

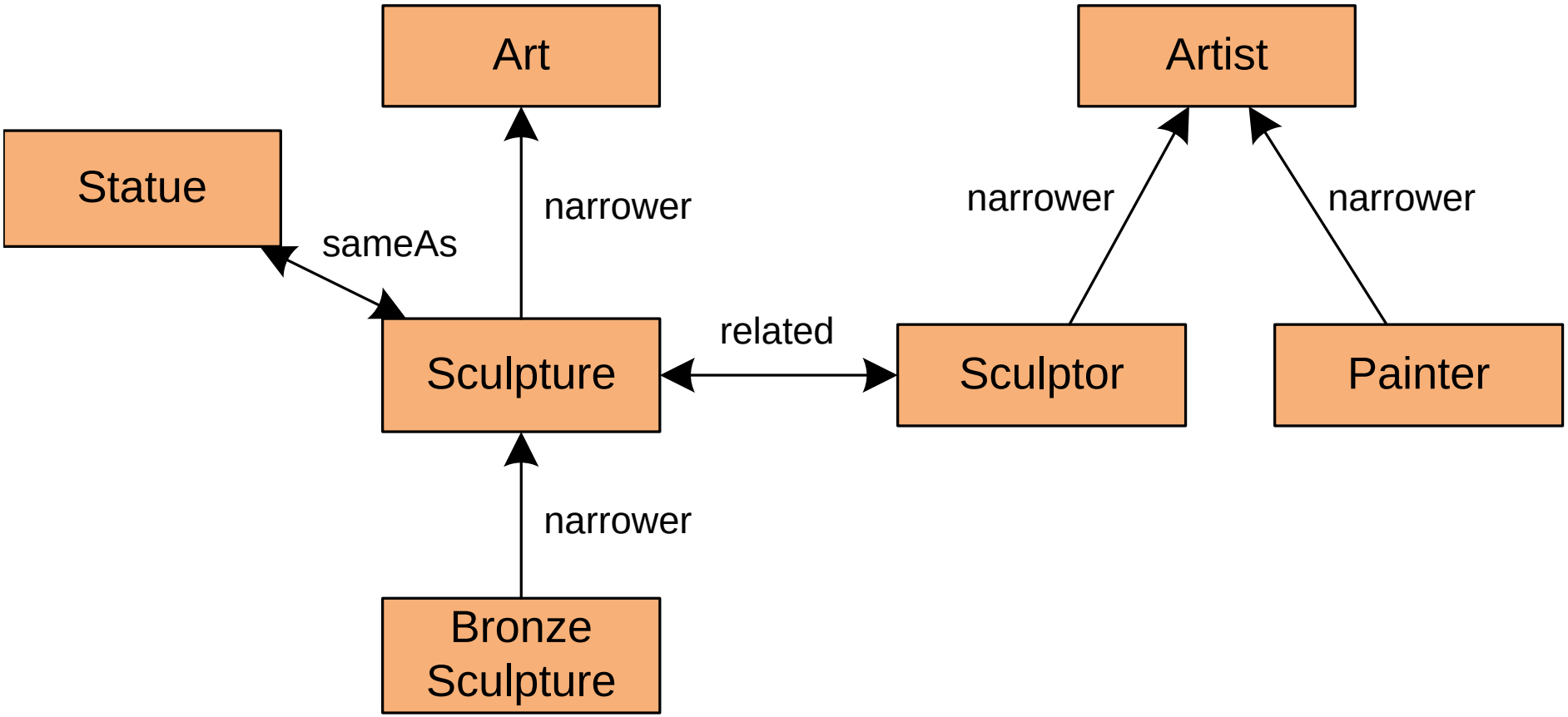
Example (National Authority Files)

- Library of Congress Subject Headings (LCSH) [LCS],
- German National Library (“Schlagwortnormdatei der Deutschen Nationalbibliothek”) [SWD] and
- Répertoire d’Autorité-Matière Encyclopédique et Alphabétique Unifié (RAMEAU) [RAM].

Transformation of an MAB entry into an Ontology



Excerpt of the generated Ontology



Overview

1 Motivation

2 Authority Control

3 Interoperability

4 Conclusion and Future Work

Concept for Interoperability

The relations between the concepts in the ontology are used to achieve interoperability.

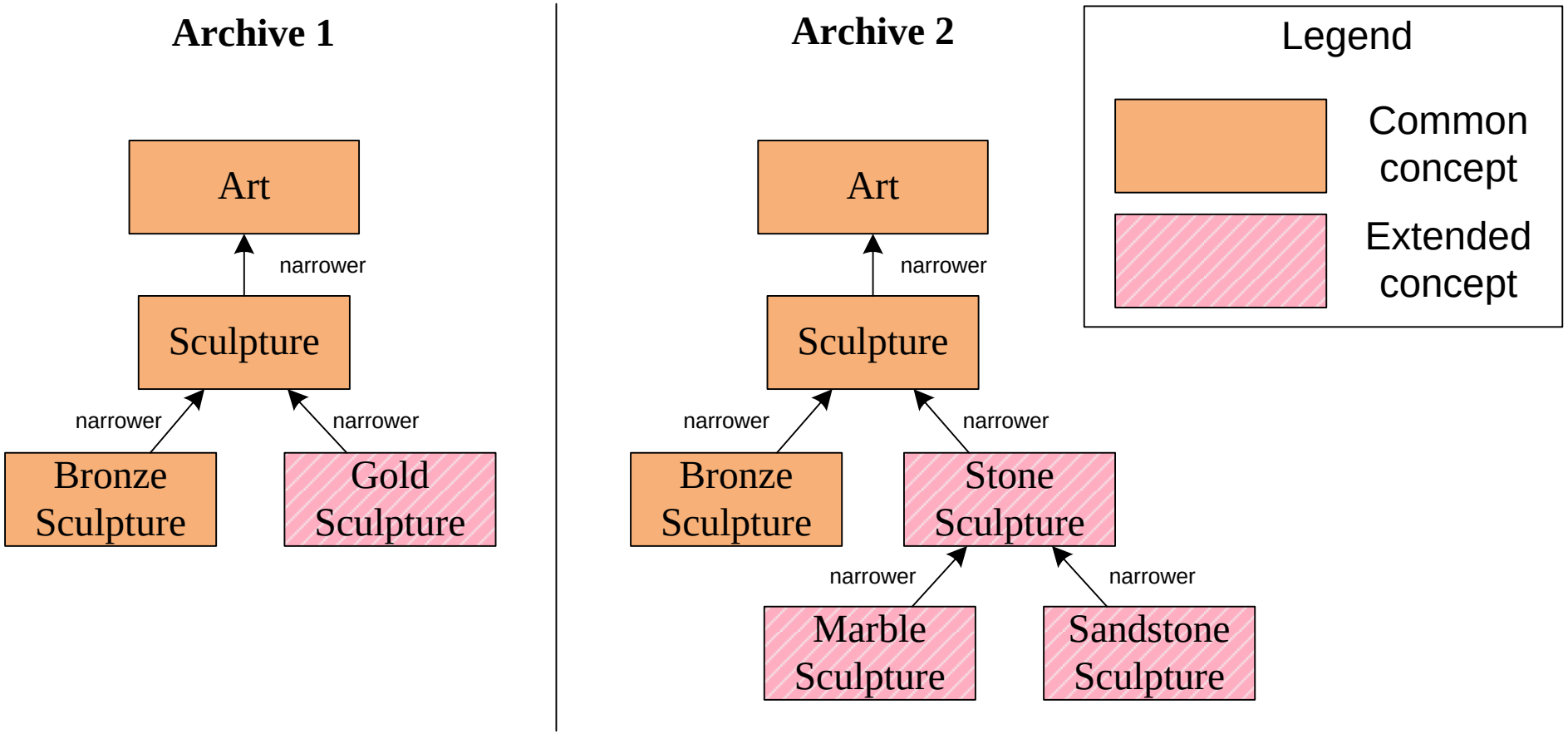
- *SameAs Relations* are used to transitively infer concepts that represent synonyms.
- *Related Relations* are used to infer concepts that are related directly. These relations are not transitively defined.
- *Narrower Relations* are used to infer concepts that represent concepts with a more specialized meaning.

Concept for Interoperability

The interoperability in an archive network is achieved by a combination of the following strategies:

- Each archive uses an authority-controlled ontology as a common knowledge base
- A local archive is only permitted to refine the concepts by defining narrower concepts
- Determine the similarity between the local concepts
- Infer related concepts in the archive network

Example of extension concepts

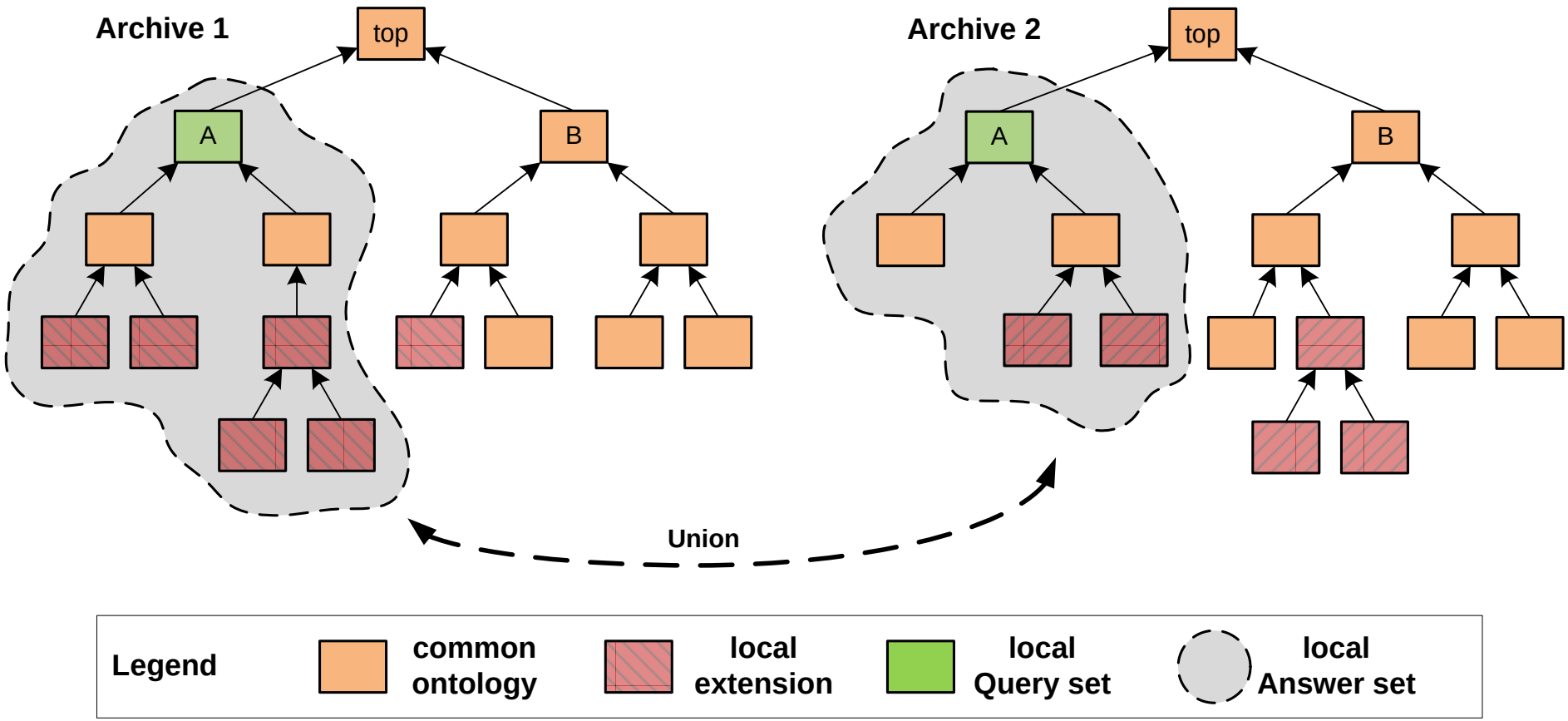


Using Authority Files for Semantical Interoperability

To answer a query that covers the entire archive network, the following cases must be considered:

1. If only concepts of the global base ontology occur in the query, then it can be answered directly
   - The archive network returns all information assigned to these common concepts.

Using only concepts of the global base ontology

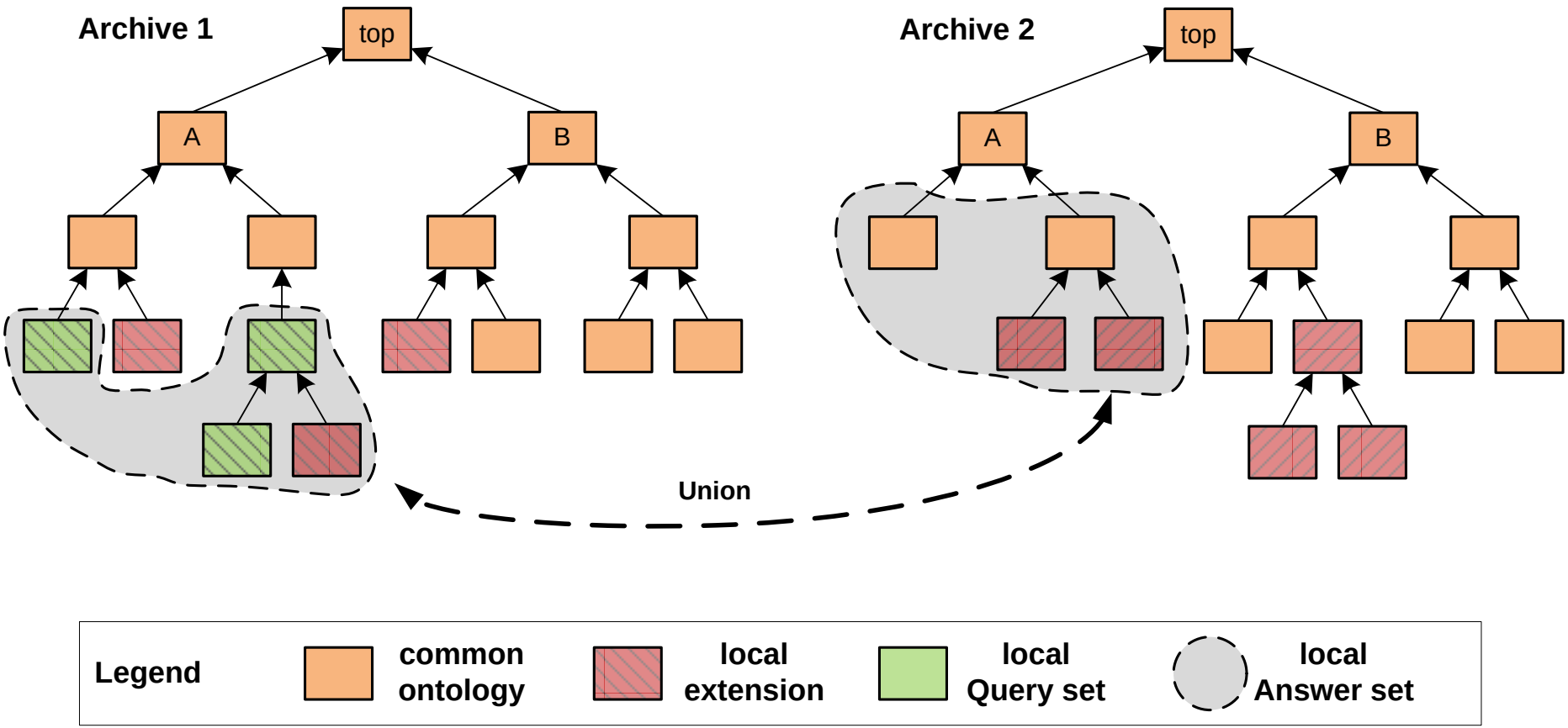


Using Authority Files for Semantical Interoperability

To answer a query that covers the entire archive network, the following cases must be considered:

1. If only concepts of the global base ontology occur in the query, then it can be answered directly
   - The archive network returns all information assigned to these common concepts.
2. If locally defined extensions occur in the query, then the generalization hierarchy is used to infer the closest common concepts in the common ontology.
   - The archive network returns all information assigned to this common concept or to locally defined narrower concepts of this concept.

Using locally defined extensions

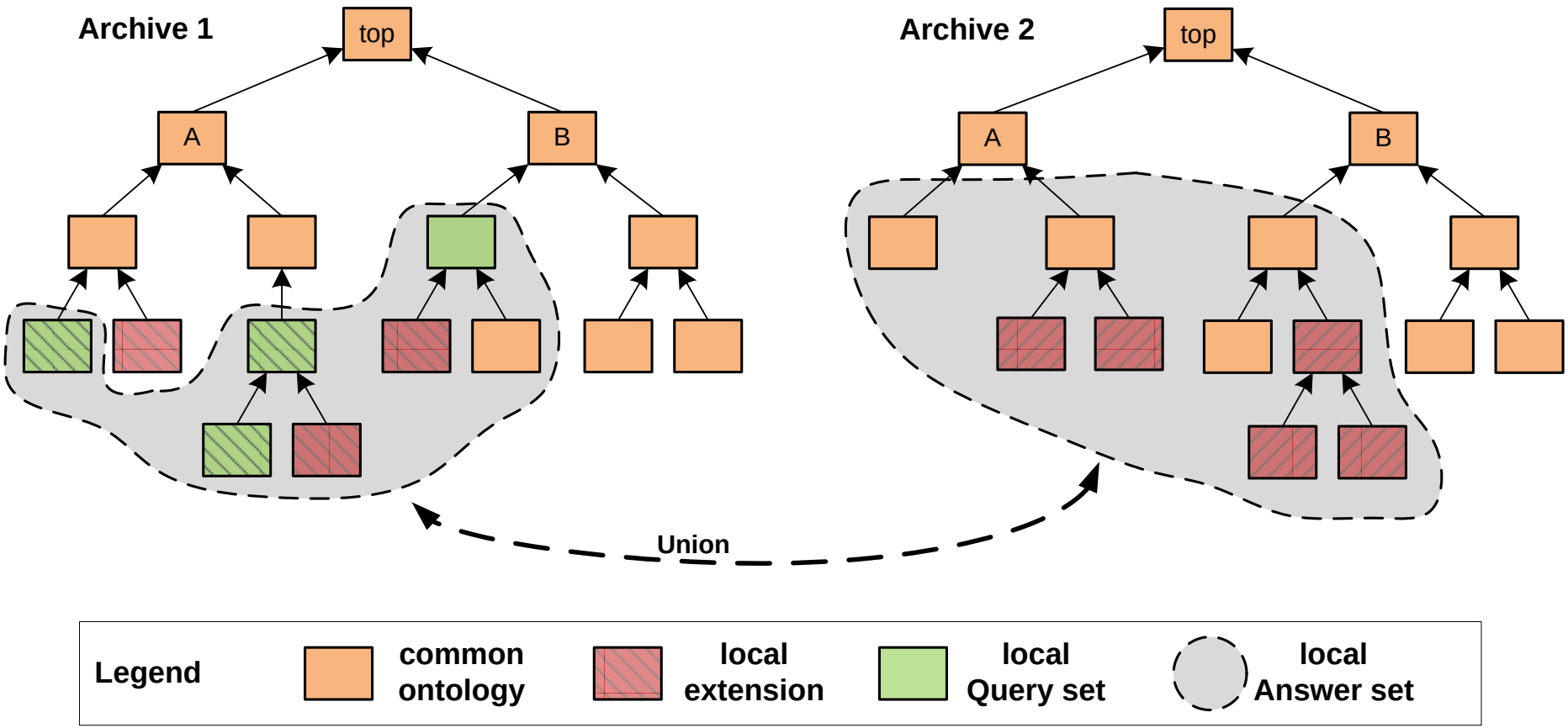


Using Authority Files for Semantical Interoperability

To answer a query that covers the entire archive network, the following cases must be considered:

1. If only concepts of the global base ontology occur in the query, then it can be answered directly
   - The archive network returns all information assigned to these common concepts.
2. If locally defined extensions occur in the query, then the generalization hierarchy is used to infer the closest common concepts in the common ontology.
   - The archive network returns all information assigned to this common concept or to locally defined narrower concepts of this concept.
3. If both, the concepts of the global base ontology and concepts of the locally defined extension, occur in the query
   - The archive network returns the union of all information assigned to the common concepts and to locally defined narrower concepts of these concepts.

Using global base ontology and local defined extension



Similarity between concepts

To determine the most relevant information, different similarity measures can be used.

Similarity Measures

- Edit distance by Levenshtein [Lev65]
- Information content [Res95]
- Semantic relations as in WordNet [LS08].

But these similarity measures are semi-automated and of limited use to determine semantical similarity.

Similarity between concepts

To determine the most relevant information, different similarity measures can be used.

Similarity Measures

- Edit distance by Levenshtein [Lev65]
- Information content [Res95]
- Semantic relations as in WordNet [LS08].

But these similarity measures are semi-automated and of limited use to determine semantical similarity.

Competency Questions [GF94]

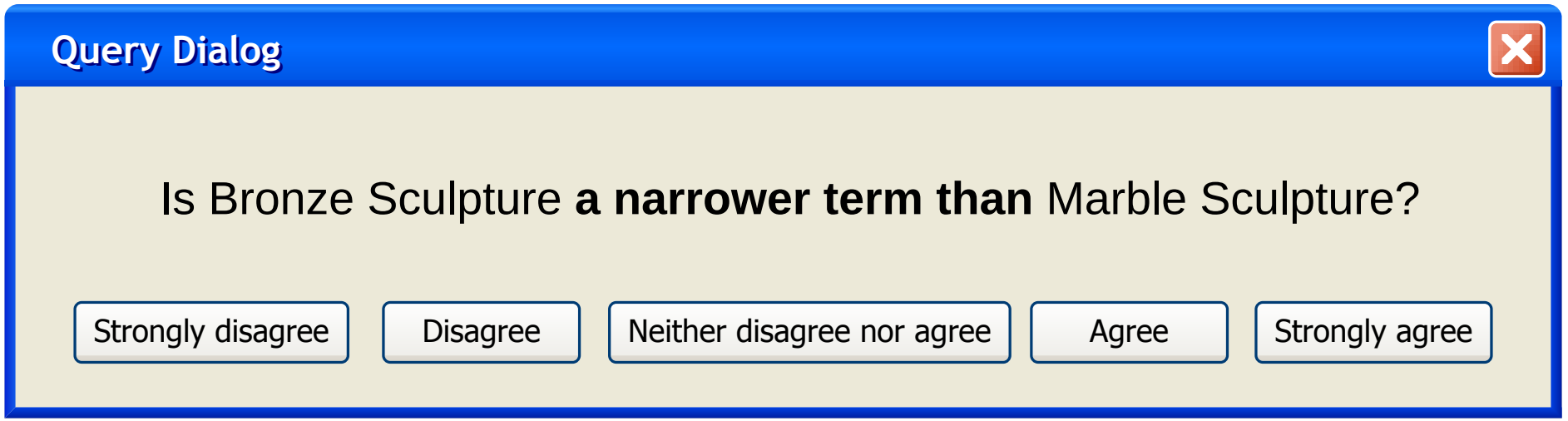
Example of Competency Questions

Is Bronze Sculpture **the same as** Marble Sculpture?
Is Bronze Sculpture **a narrower term than** Marble Sculpture?
Is Bronze Sculpture **a broader term than** Marble Sculpture?

Competency Questions [GF94]

Example of Competency Questions

Is Bronze Sculpture **the same as** Marble Sculpture?
Is Bronze Sculpture **a narrower term than** Marble Sculpture?
Is Bronze Sculpture **a broader term than** Marble Sculpture?



Competency Questions [GF94]

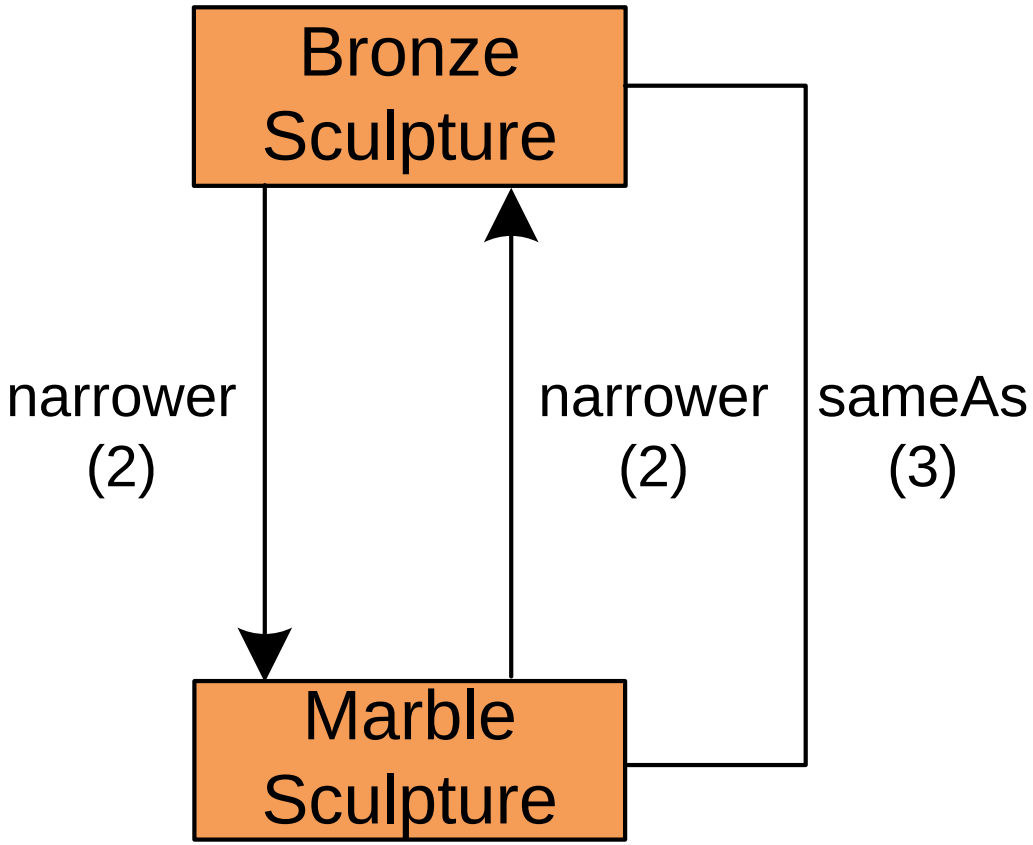
Example of Competency Questions

Is Bronze Sculpture **the same as** Marble Sculpture?
Is Bronze Sculpture **a narrower term than** Marble Sculpture?
Is Bronze Sculpture **a broader term than** Marble Sculpture?

Score for the answer options [Lik32]

Answer	**Score**
Strongly disagree	1
Disagree	2
Neither agree nor disagree	3
Agree	4
Strongly agree	5

Possible interpretation of the answers



Overview

1 Motivation

2 Authority Control

3 Interoperability

4 Conclusion and Future Work

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Conclusion

The presented approach can increase the interoperability of autonomous archives.

- Authority-controlled ontologies are distributed.
- Local archives are allowed to extend their local ontology.
- To answer queries the common ontology is applied.
- Relations between the concepts are inferred by reasoning.

Future Work

- Validate the transformation algorithm against the Library of Congress Subject Headings (LCSH) and the Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié (RAMEAU) [RAM].
- Evaluate similarity measures for Ontology Matching.
- Further research on the usability of the answers returned by the Competency Questions.
- Evaluate reasoning on ontologies with uncertain knowledge.

Conclusion

Thank you for your attention.

Questions?

Conclusion

Thank you for your attention.

Questions?


M. Gruninger and M. Fox.
The role of competency questions in enterprise engineering.
In *Workshop on Benchmarking- Theory and Practice.*, number WG5.7, Trondheim, Norway, 1994.


Library of Congress, Subject Headings (LCSH). available at authorities.loc.gov/.
last visited at 2009-11-19.


Vladimir I. Levenshtein.
Binary codes capable of correcting deletions, insertions, and reversals.
Soviet Physics Doklady, 4(163):845–848, 1965.


Rensis Likert.
A technique for the measurement of attitudes.
In *Archives of Psychology*, number 140, pages 1–55, 1932.


Feiyu Lin and Kurt Sandkuhl.
A survey of exploiting wordnet in ontology matching.
In Max Bramer, editor, *IFIP AI*, volume 276 of *IFIP*, pages 341–350. Springer, 2008.


Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU). available at rameau.bnf.fr/.
last visited at 2009-11-19.


Philip Resnik.
Using information content to evaluate semantic similarity in a taxonomy.
In *IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence*, pages 448–453, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.


Deutsche Nationalbibliothek: Schlagwortnormdatei (SWD). available at www.d-nb.de/standardisierung/normdateien/swd.htm.
last visited at 2009-11-19.

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Transformation algorithm

Data: Authority File with subjects

Result: Ontology with the concepts and their relations as OWL

Load the Authority File;

`Add Concepts to the Ontology;`

foreach *Authority Record* $\in$ *Authority File* **do**

  Load Concept represented by the Authority Record;

  if *Concept* $\notin$ *Ontology* **then**

    Add Concept as new class into the Ontology;

  end

end

`Add Relations to the Ontology;`

foreach *Authority Record* $\in$ *Authority File* **do**

  foreach *Relation* $\in$ *Authority Record* **do**

    Add Relation (narrower, sameAs, related) as a property relating the corresponding source and target classes of the Ontology;

  end

end

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Transformation algorithm

Data: Authority File with subjects

Result: Ontology with the concepts and their relations as OWL

Load the Authority File;

`Add Concepts to the Ontology;`

foreach *Authority Record* $\in$ *Authority File* **do**

  Load Concept represented by the Authority Record;

  if *Concept* $\notin$ *Ontology* **then**

    Add Concept as new class into the Ontology;

  end

end

`Add Relations to the Ontology;`

foreach *Authority Record* $\in$ *Authority File* **do**

  foreach *Relation* $\in$ *Authority Record* **do**

    Add Relation (narrower, sameAs, related) as a property relating the corresponding source and target classes of the Ontology;

  end

end

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Mapping MAB to SKOS

Mapping

001	Identifier	skos:Concept
070a	Institute	skos:inScheme
800s	name of the concept	skos:prefLabel
830s	the same concept	skos:exactMapping
850s	superordinate concept	skos:broader
860s	related concept	skos:related

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik