Metadata Processing

Zuzana Nevěřilová

<xpopelk@fi.muni.cz>

17th March 2010

ℰ𝓊DML

The **EUROPEAN DIGITAL MATHEMATICS LIBRARY**

Contents

Metadata

For each unit:

- journal (e.g. Czechoslovak Mathematical Journal)

- volume (e.g. 55, 2005, 2)

- article (e.g. On convergence theory in fuzzy topological spaces and its applications)

- references (e.g. [1] R. G. Bartle: Nets and filters in topology, Amer. Math. Monthly 62 (1955), 551–557. 1955)

Metadata

For each unit:

- journal (e.g. Czechoslovak Mathematical Journal)

- volume (e.g. 55, 2005, 2)

- article (e.g. On convergence theory in fuzzy topological spaces and its applications)

- references (e.g. [1] R. G. Bartle: Nets and filters in topology, Amer. Math. Monthly 62 (1955), 551–557. 1955)

Metadata

For each unit:

- journal (e.g. Czechoslovak Mathematical Journal)

- volume (e.g. 55, 2005, 2)

- article (e.g. On convergence theory in fuzzy topological spaces and its applications)

- references (e.g. [1] R. G. Bartle: Nets and filters in topology, Amer. Math. Monthly 62 (1955), 551–557. 1955)
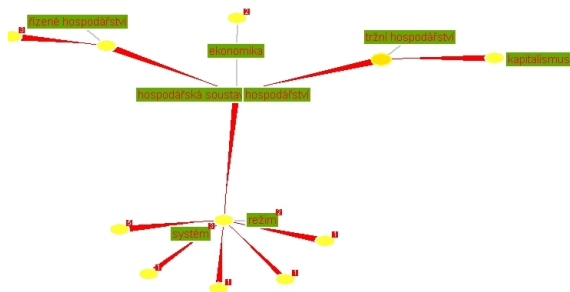
Metadata

For each unit:

- journal (e.g. Czechoslovak Mathematical Journal)

- volume (e.g. 55, 2005, 2)

- article (e.g. On convergence theory in fuzzy topological spaces and its applications)

- references (e.g. [1] R. G. Bartle: Nets and filters in topology, Amer. Math. Monthly 62 (1955), 551–557. 1955)

Visualization

"the visual representation of large-scale collections of non-numerical information"

can help understanding large amounts of information at once

generally useful for rather *sparse* and *complex* data sets

Visualization

"the visual representation of large-scale collections of non-numerical information"

can help understanding large amounts of information at once

generally useful for rather *sparse* and *complex* data sets

Visualization

"the visual representation of large-scale collections of non-numerical information"

can help understanding large amounts of information at once

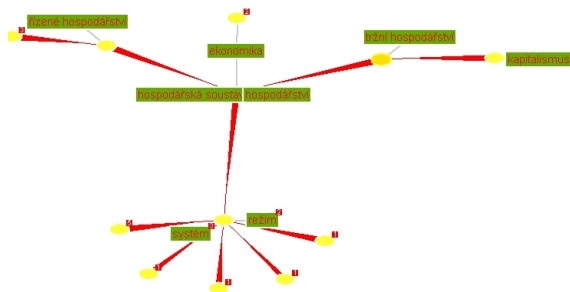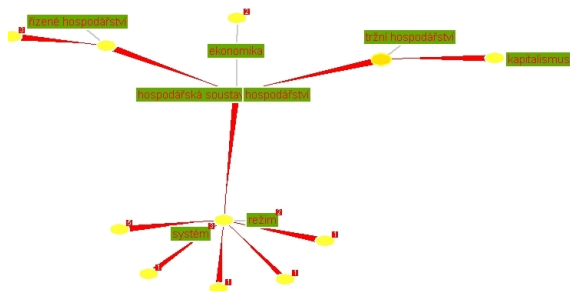generally useful for rather *sparse* and *complex* data sets

Visualization – contd.

The purpose $=$ another way to browse the library

- data: what metadata should be visualized?
  authors, articles, serials, years?, keywords?, references, MSC
  (classification), languages? no. of pages? no. of citations?

- graphics: what is important for users?
  what do users search on <dml.cz>?

- data conversion: Resource Description Framework (RDF)[1]

- server application providing the RDF data

- client side: Visual Browser [Nevěřilová, 2005]

---

[1] <http://www.w3.org/RDF/>

Visualization – contd.

The purpose = another way to browse the library

- data: what metadata should be visualized?
  authors, articles, serials, years?, keywords?, references, MSC
  (classification), languages? no. of pages? no. of citations?

- graphics: what is important for users?
  what do users search on <dml.cz>?

- data conversion: Resource Description Framework (RDF)[1]

- server application providing the RDF data

- client side: Visual Browser [Nevěřilová, 2005]

---

[1]<http://www.w3.org/RDF/>

Visualization – contd.

The purpose $=$ another way to browse the library

- data: what metadata should be visualized?
  authors, articles, serials, years?, keywords?, references, MSC
  (classification), languages? no. of pages? no. of citations?

- graphics: what is important for users?
  what do users search on <dml.cz>?

- data conversion: Resource Description Framework (RDF)[1]

- server application providing the RDF data

- client side: Visual Browser [Nevěřilová, 2005]

---

[1] <http://www.w3.org/RDF/>

Visualization – contd.

The purpose = another way to browse the library

- data: what metadata should be visualized?
  authors, articles, serials, years?, keywords?, references, MSC
  (classification), languages? no. of pages? no. of citations?

- graphics: what is important for users?
  what do users search on <dml.cz>?

- data conversion: Resource Description Framework (RDF)[1]

- server application providing the RDF data

- client side: Visual Browser [Nevěřilová, 2005]

---

[1] <http://www.w3.org/RDF/>

Visualization – contd.

The purpose $=$ another way to browse the library

- data: what metadata should be visualized?
  authors, articles, serials, years?, keywords?, references, MSC
  (classification), languages? no. of pages? no. of citations?

- graphics: what is important for users?
  what do users search on <dml.cz>?

- data conversion: Resource Description Framework (RDF)[1]

- server application providing the RDF data

- client side: Visual Browser [Nevěřilová, 2005]

---

[1] <http://www.w3.org/RDF/>

Building Semantic Network

metadata contains semantic information, but …

semantic network:

- relations of several types

  (nodes with one relation type usually form a tree, relations can be weighted)

- top ontology

Building Semantic Network

metadata contains semantic information, but …

semantic network:

- relations of several types

  (nodes with one relation type usually form *a tree*, relations can be weighted)

- top ontology

Building Semantic Network

metadata contains semantic information, but …

semantic network:

- relations of several types

  (nodes with one relation type usually form *a tree*, relations can be weighted)

- top ontology

Building Semantic Network

metadata contains semantic information, but …

semantic network:

- relations of several types

  (nodes with one relation type usually form *a tree*, relations can be weighted)

- top ontology

Keywords

```
fixed point
oscillation
regularity
natural operator
ideal
variety
stability
existence
compact
topological group
Orlicz spaces
```

how many are they?	2–5 per article
	but not all serials register keywords
how frequent are they?	$>$ 80 % hapax legomena
are the most frequent the less relevant?	maybe
part of speech	nouns or noun phrases
ambiguity	low

Keywords

```
fixed point
oscillation
regularity
natural operator
ideal
variety
stability
existence
compact
topological group
Orlicz spaces
```

how many are they?	2–5 per article
	but not all serials register keywords
how frequent are they?	$> 80 \%$ hapax legomena
are the most frequent the less relevant?	maybe
part of speech	nouns or noun phrases
ambiguity	low

Keywords

```
fixed point
oscillation
regularity
natural operator
ideal
variety
stability
existence
compact
topological group
Orlicz spaces
```

how many are they?	2–5 per article
	but not all serials register keywords
how frequent are they?	> 80 % hapax legomena
are the most frequent the less relevant?	maybe
part of speech	nouns or noun phrases
ambiguity	low

Keywords

```
fixed point
oscillation
regularity
natural operator
ideal
variety
stability
existence
compact
topological group
Orlicz spaces
```

how many are they?	2–5 per article
	but not all serials register keywords
how frequent are they?	$>$ 80 % hapax legomena
are the most frequent the less relevant?	maybe
part of speech	nouns or noun phrases
ambiguity	low

Keywords

```
fixed point
oscillation
regularity
natural operator
ideal
variety
stability
existence
compact
topological group
Orlicz spaces
```

how many are they?	2–5 per article
	but not all serials register keywords
how frequent are they?	> 80 % hapax legomena
are the most frequent the less relevant?	maybe
part of speech	nouns or noun phrases
ambiguity	low

Keywords

```
fixed point
oscillation
regularity
natural operator
ideal
variety
stability
existence
compact
topological group
Orlicz spaces
```

how many are they?	2–5 per article
	but not all serials register keywords
how frequent are they?	$>$ 80 % hapax legomena
are the most frequent the less relevant?	maybe
part of speech	nouns or noun phrases
ambiguity	low

Math WordNet

- similar to LOIS (Lexical Ontologies for Legal Information Sharing) [Tiscornia, 2006]

- literals … keywords

- synsets … ?

- keywords in several languages … synset connection through WordNets' InterLingual Index [Vossen, 1998]

- relations … ?

Math WordNet

- similar to LOIS (Lexical Ontologies for Legal Information Sharing) [Tiscornia, 2006]

- literals … keywords

- synsets … ?

- keywords in several languages … synset connection through WordNets' InterLingual Index [Vossen, 1998]

- relations … ?

Math WordNet

- similar to LOIS (Lexical Ontologies for Legal Information Sharing) [Tiscornia, 2006]

- literals … keywords

- synsets … ?

- keywords in several languages … synset connection through WordNets' InterLingual Index [Vossen, 1998]

- relations … ?

Math WordNet

- similar to LOIS (Lexical Ontologies for Legal Information Sharing) [Tiscornia, 2006]

- literals … keywords

- synsets … ?

- keywords in several languages … synset connection through WordNets' InterLingual Index [Vossen, 1998]

- relations … ?

Math WordNet

- similar to LOIS (Lexical Ontologies for Legal Information Sharing) [Tiscornia, 2006]

- literals … keywords

- synsets … ?

- keywords in several languages … synset connection through WordNets' InterLingual Index [Vossen, 1998]

- relations … ?

Sparse Space of Relations

```
ANTONYM
NEAR_ANTONYM
NEAR_SYNONYM
HAS_HYPERONYM
HAS_HYPONYM
HAS_INSTANCE
HAS_HOLO_MEMBER
HAS_HOLO_PART
HAS_HOLO_PORTION
HAS_MERO_MEMBER
HAS_MERO_PART
HAS_MERO_PORTION
HAS_XPOS_HYPERONYM
FUZZYNYM
CAUSES
```

the only relation type we have from the classification is *domain*

can we use other math ontology such as EngMath [Gruber and Olsen, 1994]?

Sparse Space of Relations

```
ANTONYM
NEAR_ANTONYM
NEAR_SYNONYM
HAS_HYPERONYM
HAS_HYPONYM
HAS_INSTANCE
HAS_HOLO_MEMBER
HAS_HOLO_PART
HAS_HOLO_PORTION
HAS_MERO_MEMBER
HAS_MERO_PART
HAS_MERO_PORTION
HAS_XPOS_HYPERONYM
FUZZYNYM
CAUSES
```

the only relation type we have from the classification is *domain*

can we use other math ontology such as EngMath [Gruber and Olsen, 1994]?

Conclusion

studying metadata … preparing visualization … building Math WordNet

---

[2] <http://www.iter.org/sci/Pages/FusionFuels.aspx>

References

Gruber, T. R. and Olsen, G. R. (1994).
An ontology for engineering mathematics.
In *Fourth International Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann.

Nevěřilová, Z. (2005).
Visual Browser: A tool for visualising ontologies.
In *Proceedings of I-KNOW'05*, pages 453–461, Graz, Austria.
Know-Center in coop. with Graz Uni, Joanneum Research and Springer Pub. Co.

Tiscornia, D. (2004–2006).
Lexical ontologies for legal information sharing.

Vossen, P. (1998).
EuroWordNet – a multilingual database with lexical semantic networks.