Topic Maps

Metrics and Visualization

Benedicte.Le-Grand@lip6.fr
Laboratoire d'Informatique de Paris 6

KT'2001
Introduction

- Topic Maps: intended to structure large information pools
- Kind of semantic network above information resources
Information retrieval in topic maps

- Find answers to specific questions
  - Query languages

- Explore one topic in particular
  - List of topics

- What if there is no precise query?
  - Global understanding
  - Navigation
No precise subject of interest
Examples

- Global understanding
- Comparison between several topic maps about the same subject
- General knowledge
- Beginner user:
  - « What is this topic map about? »
  - « What are its main features? »
  - « Where should I start? »
Global understanding of the topic map

- Visualization
- Navigation
- But ...
This simple representation ...
… may become this one!
Our proposal

- Provide a map of the topic map
  - City metaphor
  - Different scales
  - Help user build his own cognitive map

- Representation as a virtual world
Definitions

- Object: topic or association

- Objects have characteristics called properties
  - What are these characteristics?

- Similarity:
  - Two objects are similar if they have a certain number of properties in common
  - How much is « a certain number »?
Algorithm

- Topic map
- Objects and properties
  - « regular » objects
  - Similarity matrix
- Objects coordinates
  - 2D Map
- Virtual World
- Topic map parser
- Classification algorithm
  - Topic map profile
- Multidimensional scaling
- Visualization tool
Example: music topic map

- Techquila’s topic map about music
- XML document, valid against xtm1.dtd
Topic Map Parser

Objects:
- Topics
- Associations

Properties: 2 scenarios
- Intuitive scenario
- Recursive scenario

KT2001
Intuitive scenario

- Object = element
- Properties = object’s attributes + its children’s attributes
- Only intrinsic characteristics
- The topic map is traversed once
Intuitive scenario for music.xtm

Object:
- t-the-clash

Properties:
- tt-band
- tt-music

Object:
- tt-band

Properties:
- t-the-clash
- http://www.topicmaps.org/xtm/1.0/psi-sort
Intuitive scenario for music.xtm

<association id="assoc6">
  <instanceOf>
    <topicRef xlink:href="at-recorded"/>
  </instanceOf>
  <member>
    <instanceOf>
      <topicRef xlink:href="tt-band"/>
    </instanceOf>
    <topicRef xlink:href="t-the-clash"/>
  </member>
  <member>
    <instanceOf>
      <topicRef xlink:href="tt-track"/>
    </instanceOf>
    <topicRef xlink:href="t-i-fought-the-law"/>
  </member>
</association>

Object:
assoc6

Properties:
at-recorded
tt-band
t-the-clash
tt-track
t-i-fought-the-law

KT'2001
Recursive scenario

- Elements’ properties become objects themselves
  - Superclasses know what their instances are
- Topics know what associations they are involved in
- Not only intrinsic characteristics
- 2 traversals of the topic map
Recursive scenario for music.xtm

Object: tt-band
Properties:
  tt-music
t-the-clash
assoc1
assoc2
assoc6

Object: t-the-clash
Properties:
  tt-band
http://www.topicmaps.org/xtm/1.0/psi-sort
assoc1
assoc2
assoc6
We need different scales

- There may be hundreds of topics and associations in a topic map
- Impossible to represent them all!
- Selection of topics and associations
  - Regular / singular objects
Topic map's profile

- What are the characteristics of this topic map?

- Is the topic map general or specific?
  - Are topics strongly related to each other?
  - Are these topics very similar?

- Profile = reference
Statistics computed on objects

- **Object**: topic or association

- **For each object**:
  - Does it have anything in common with other objects?
    - **Number of objects it shares properties with**
  - If yes, are they very close?
    - **Number of properties shared**
Conceptual Classification Algorithm

Objects and properties

« regular » objects

Classification algorithm
Topic map profile
Conceptual classification algorithm

- Galois classification algorithm
- Generation of a lattice of concepts
- Concepts: sets of objects which have properties in common
Classification algorithm

<table>
<thead>
<tr>
<th>R</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Objects statistics

Computation of statistics according to:

- The proportion of lattice concepts each object appears in
- The proportion of other objects it is grouped with
  - average and max values
- The proportion of properties it shares with other objects
  - average and max values
- The number of occurrences of each object in the topic map itself

KT2001
Objects statistics in music.xtm

Object tt-band Statistics

Occurrences in concepts : 6.3 %
Maximum number of related objects : 15.4 %
Average number of related objects : 15.4 %
Maximum number of common properties : 100.0 %
Average number of common properties : 50.0 %
Occurrences in the topic map : 1.7 %

Object t-the-clash Statistics

Occurrences in concepts : 12.5 %
Maximum number of related objects : 41.0 %
Average number of related objects : 17.9 %
Maximum number of common properties : 50.0 %
Average number of common properties : 33.3 %
Occurrences in the topic map : 6.9 %
Topic Map’s profile

- Computation of a weighted average
- Weights: number of occurrences in the topic map, divided by the total number of occurrences
music.xtm profile

Topic Maps Profile Statistics

Occurrences in concepts: 12.7%
Maximum number of related objects: 31.5%
Average number of related objects: 23.7%
Maximum number of common properties: 84.4%
Average number of common properties: 38.7%
Occurrences in the topic map: 2.6%
Selection of « regular » objects

- Comparison of objects statistics with the topic map’s profile values
- Computation of the standard deviation
- Selection condition:
  \[ O.stats + 1.5 \times \text{standard deviation} \geq \text{profile.stats} \]
- New list of object → new profile → new selection …
Music.xtm discarded objets

First iteration:

tt-musician  tt-country
 tt-city  t-disc-the-clash
t-disc-combat-rock  t-england
t-turkey  assoc1
Algorithm

- Input: any topic map
- Generation of a Galois lattice
- Profile computation
- Selection of "regular" objects
- Generation of a new lattice
- New profile computation
- ...
- No more discarded objects: algorithm stabilization
- Visualization as a virtual world
Stabilization of the algorithm

- When there are no more « singular » objects
- Set of objects which can be similar or not

- Computation of objects coordinates
- Multidimensional Scaling (MDS)

KT'2001
Results
Different levels of detail

- First view: only remaining objects

- More detailed view:
  - add objects discarded in the last (nth) iteration

- More detailed view:
  - add objects discarded in the (n-1)th iteration

- ...

KT'2001
Multidimensional Scaling

Position of topics

"Regular" Objects
Similarity Matrix

Objects coordinates
2D Map

Multidimensional Scaling
Similarity Matrix

- Notion of similarity between two objects

- Let two objects:
  - O1, set of properties P1
  - O2, set of properties P2

- Similarity (O1, O2) = \( \frac{\text{Card} (P1 \cap P2)}{\text{Card} (P1 \cup P2)} \)
Similarity example

- object : tt-artist with properties : [tt-person]
- object : tt-drummer with properties : [tt-musician]
- object : t-mick-jones with properties : [tt-person, tt-musician, http://www.topicmaps.org/xtm/1.0/psi-sort]

- Similarity between tt-artist and tt-drummer = 0.0 %
- Similarity between tt-artist and t-mick-jones = 33.3 %
- Similarity between tt-artist and t-nicky-headon = 33.3%
- Similarity between tt-drummer and t-mick-jones = 33.3%
- Similarity between tt-drummer and t-nicky-headon = 33.3%
- Similarity between t-mick-jones and t-nicky-headon = 100.0 %
**Multidimensional Scaling (MDS)**

- Multivariate data analysis
  - Position of topics on the map

- **Input**: similarity / distance matrix

- **Output**: 2D representation
Example : US Cities

- Input : distances between several US cities
- Output : 2D representation
- Distorsion : stress value
## US Cities: MDS Input

<table>
<thead>
<tr>
<th></th>
<th>Boston</th>
<th>NY</th>
<th>DC</th>
<th>Miami</th>
<th>Chicago</th>
<th>Seattle</th>
<th>SF</th>
<th>LA</th>
<th>Denver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston</td>
<td>0</td>
<td>206</td>
<td>429</td>
<td>1504</td>
<td>963</td>
<td>2976</td>
<td>3095</td>
<td>2979</td>
<td>1949</td>
</tr>
<tr>
<td>NY</td>
<td>206</td>
<td>0</td>
<td>233</td>
<td>1308</td>
<td>802</td>
<td>2815</td>
<td>2934</td>
<td>2786</td>
<td>1771</td>
</tr>
<tr>
<td>DC</td>
<td>429</td>
<td>233</td>
<td>0</td>
<td>1075</td>
<td>671</td>
<td>2684</td>
<td>2799</td>
<td>2631</td>
<td>1616</td>
</tr>
<tr>
<td>Miami</td>
<td>1504</td>
<td>1308</td>
<td>1075</td>
<td>0</td>
<td>1329</td>
<td>3273</td>
<td>3053</td>
<td>2687</td>
<td>2037</td>
</tr>
<tr>
<td>Chicago</td>
<td>963</td>
<td>802</td>
<td>671</td>
<td>1329</td>
<td>0</td>
<td>2013</td>
<td>2142</td>
<td>2054</td>
<td>996</td>
</tr>
<tr>
<td>Seattle</td>
<td>2976</td>
<td>2815</td>
<td>2684</td>
<td>3273</td>
<td>2013</td>
<td>0</td>
<td>808</td>
<td>1131</td>
<td>1307</td>
</tr>
<tr>
<td>SF</td>
<td>3095</td>
<td>2934</td>
<td>2799</td>
<td>3053</td>
<td>2142</td>
<td>808</td>
<td>0</td>
<td>379</td>
<td>1235</td>
</tr>
<tr>
<td>LA</td>
<td>2979</td>
<td>2786</td>
<td>2631</td>
<td>2687</td>
<td>2054</td>
<td>1131</td>
<td>379</td>
<td>0</td>
<td>1059</td>
</tr>
<tr>
<td>Denver</td>
<td>1949</td>
<td>1771</td>
<td>1616</td>
<td>2037</td>
<td>996</td>
<td>1307</td>
<td>1235</td>
<td>1059</td>
<td>0</td>
</tr>
</tbody>
</table>
US Cities: MDS Output

- Miami
- DC
- NY
- Boston
- Chicago
- Denver
- LA
- SF
- Seattle
Topic Map Visualization: Virtual World

Objects coordinates

2D Map

Virtual World

Visualization tool

KT'2001
Buildings' Characteristics

- Number
- Name
- Visual characteristics
  - Color
  - Height, width, depth
- Coordinates
- Occurrences
- Associated buildings
Graphic User Interface

- 2D map
- Virtual world
- Selected building's characteristics
  - Occurrences
  - Associations
Navigation proposals

- **Free navigation**
  - Easy orientation with the map
  - Cognitive map

- **Guided tours**
  - Global tour - visit all clusters
  - Explore one district in particular
Demo : Free navigation

- Travel within the city with the keyboard

- Consistency 2D map / virtual world

- Building selection
  - Occurrences (web sites)
  - Associations : go to other buildings
    - manually
    - automatically (in the air) : take-off, landing
Le monde virtuel de la Recherche

Contenu de l'immeuble : 
Liens de l'immeuble : aucun immeuble sélectionné

Visite guidée terrestre
   Liste Contenu
   Liste liens

Visite guidée aérienne

Aucun lien sélectionné
Demo : guided tours

- On the ground
- In the air
- Stop the guided tour if needed
Conclusion

- Several levels of detail
  - Different scales according to user’s needs
- Topic map profile: reference
- Selection of topics and associations
- Representation of Topic Maps as Virtual Worlds
  - Cognitive map
  - Free navigation
  - Guided tours
Future Work

- Topic maps merging
  - Virtual worlds interoperability

- Populated topic maps
Thank you!