Topic Maps

Metrics and Visualization

Benedicte.Le-Grand@lip6.fr

Laboratoire d'Informatique de Paris 6

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

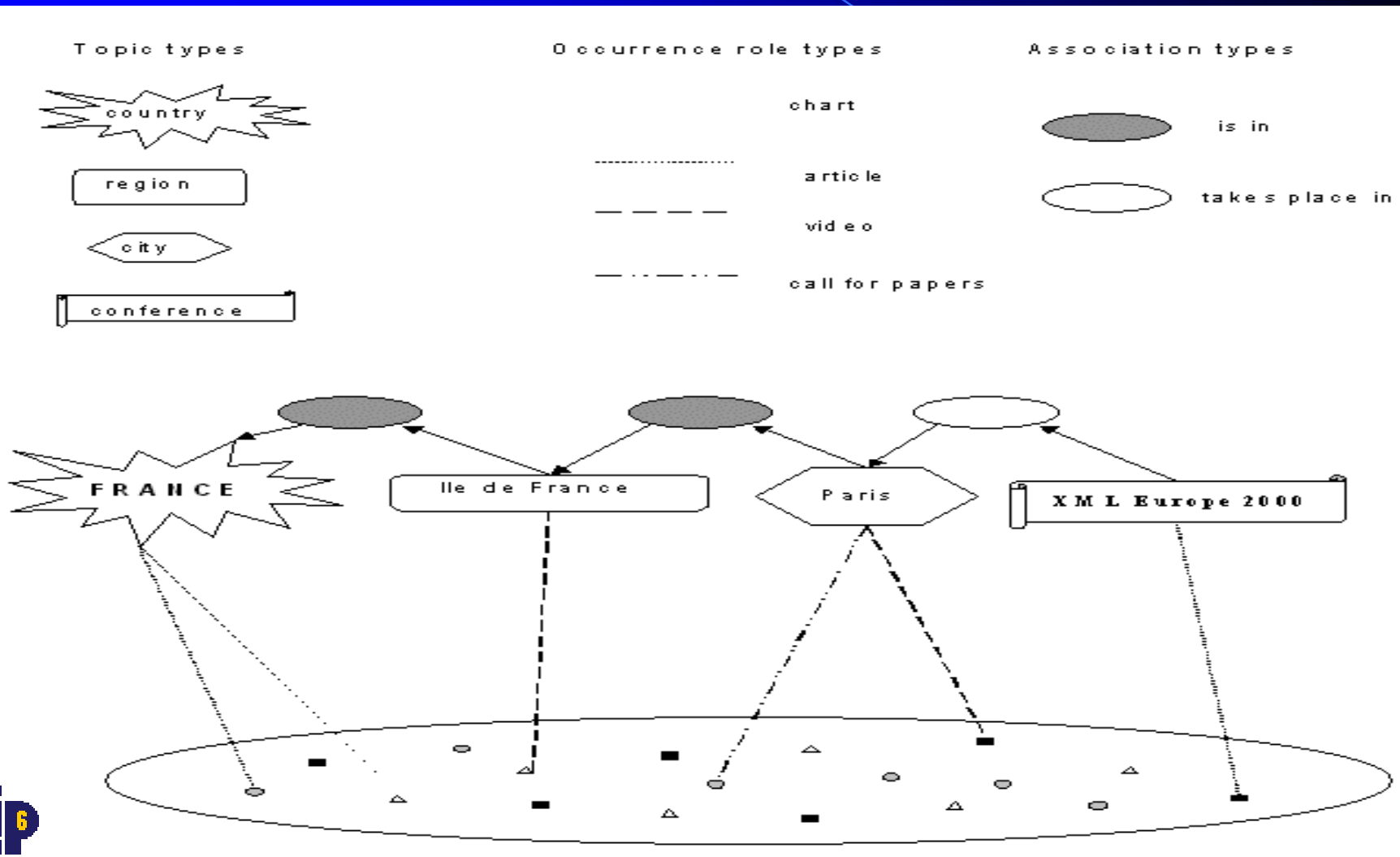
No precise subject of interest Examples

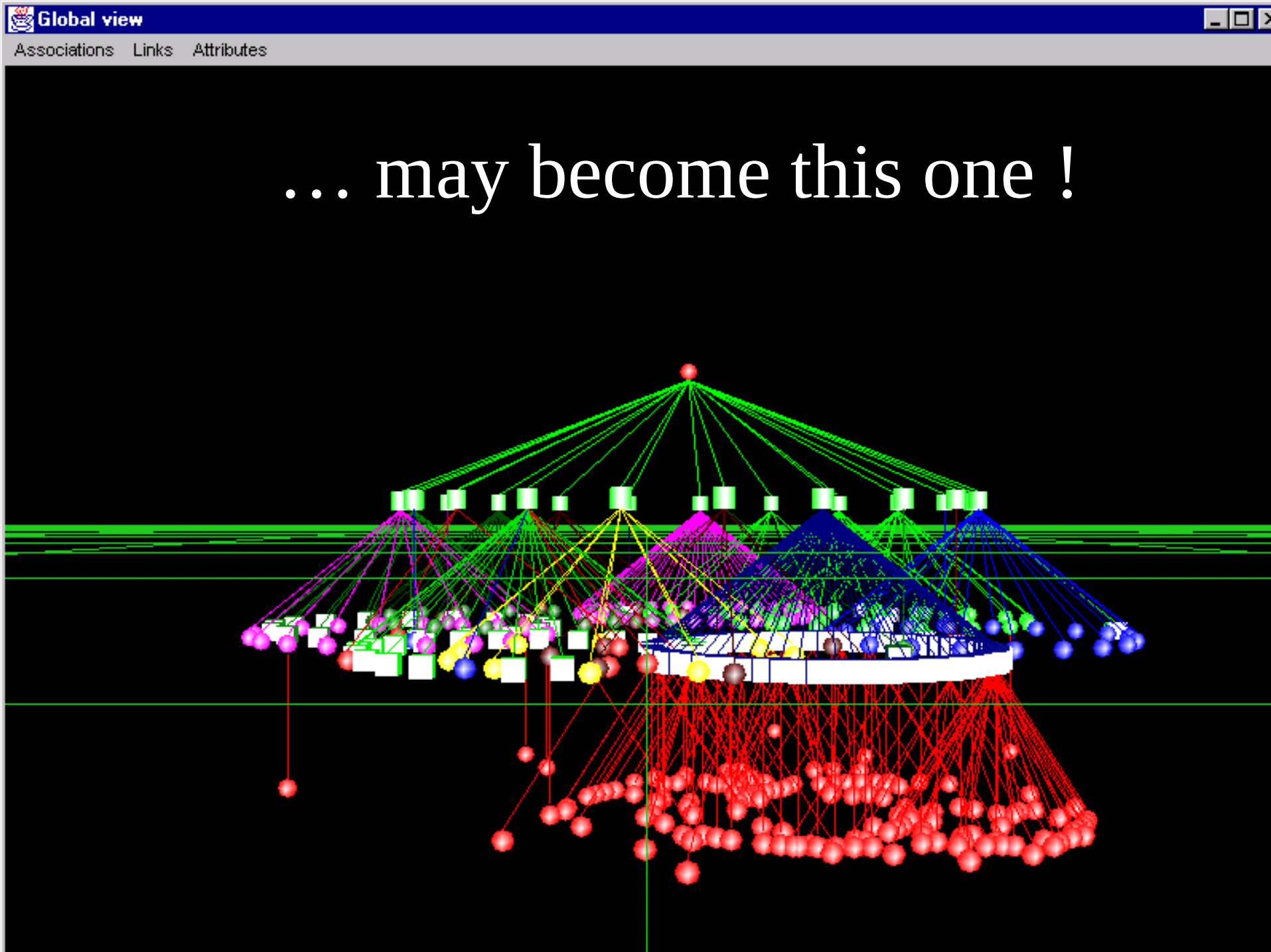
- Global understanding
- Comparison between several topic maps about the same subject
- General knowledge
- Beginner user :
  - « What is this topic map about ? »
  - « What are its main features ? »
  - « Where should I start ? »

Global understanding of the topic map

- Visualization
- Navigation
- But …

This simple representation …




Global view
Associations
Links
Attributes
… may become this one !

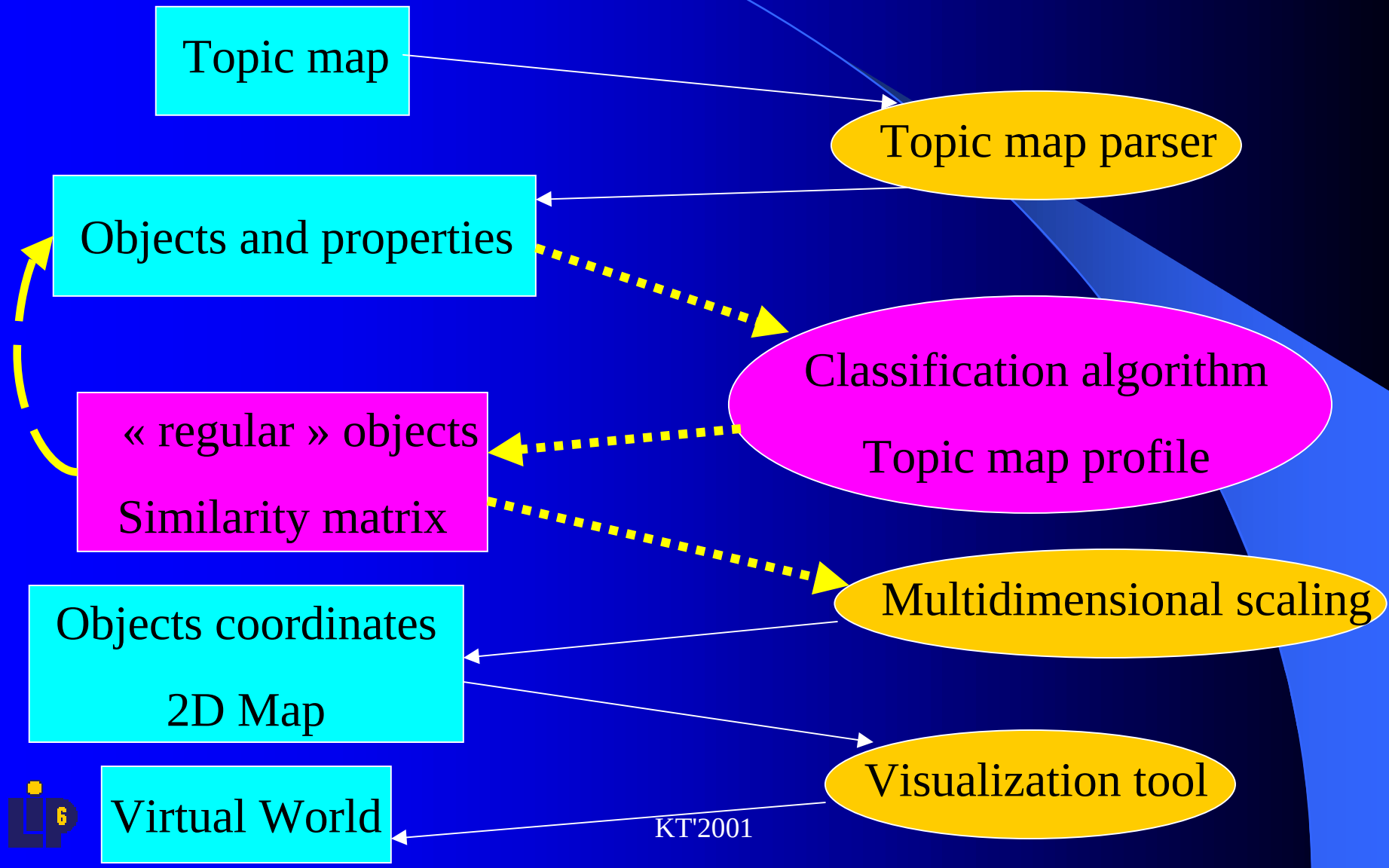
Our proposal

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

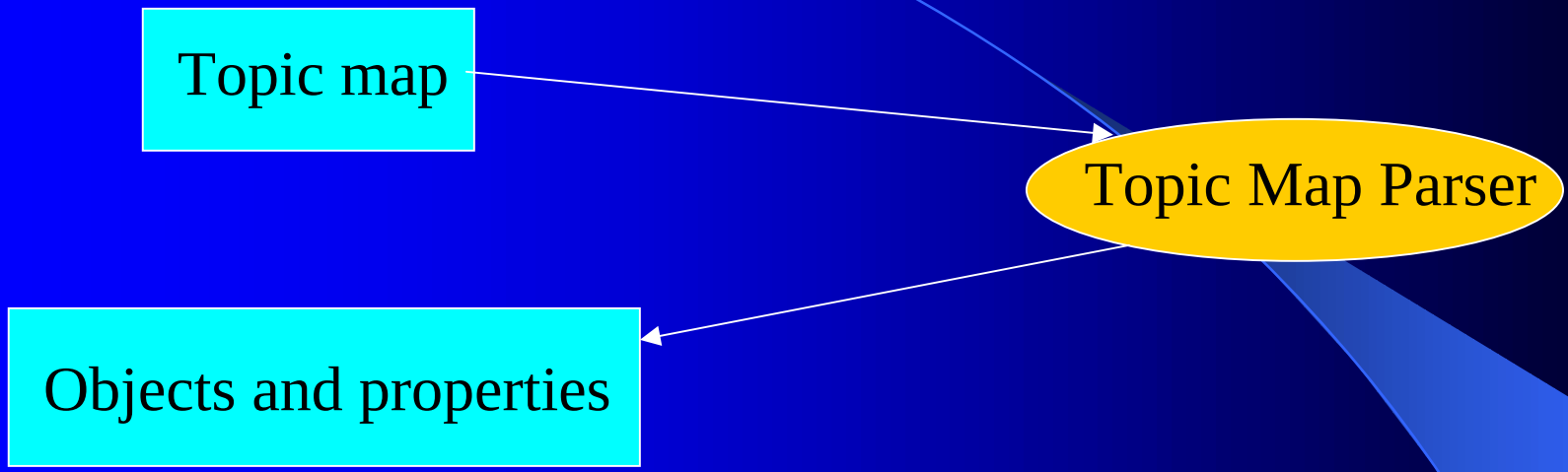
Algorithm


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

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

Intuitive scenario

- Object = element
- Properties = object's attributes
  + its children's attributes
- Only intrinsic caracteristics
- The topic map is traversed once

Intuitive scenario for music.xtm

```
<topic id="tt-band">
  <instanceOf>
    <topicRef xlink:href="tt-music"/>
  </instanceOf>
  <baseName>
    <baseNameString>Band</baseNameString>
  </baseName>
</topic>
<topic id="t-the-clash">
  <instanceOf>
    <topicRef xlink:href="tt-band"/>
  </instanceOf>
  <baseName>
    <baseNameString>The Clash</baseNameString>
    <variant>
      <parameters>
        <topicRef xlink:href="http://www.topicmaps.org/xtm/1.0/psi-sort"/>
      </parameters>
      <variantName>
        <resourceData>clash the</resourceData>
      </variantName>
    </variant>
    <variant>
      <parameters>
        <topicRef xlink:href="http://www.topicmaps.org/xtm/1.0/psi-sort"/>
      </parameters>
      <variantName>
        <resourceData>Clash, The</resourceData>
      </variantName>
    </variant>
  </baseName>
</topic>
```

Object :

tt-band

Properties :

tt-music

Object :

t-the-clash

Properties :

tt-band

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

Recursive scenario

- Elements' properties become objects themselves

  → Superclasses know what their instances are

- Topics know what associations they are involved in

- Not only intrinsic caracteristics

- 2 traversals of the topic map

Recursive scenario for music.xtm

```
<topic id="tt-band">
  <instanceOf>
    <topicRef xlink:href="tt-music"/>
  </instanceOf>
  <baseName>
    <baseNameString>Band</baseNameString>
  </baseName>
</topic>
<topic id="t-the-clash">
  <instanceOf>
    <topicRef xlink:href="tt-band"/>
  </instanceOf>
  <baseName>
    <baseNameString>The Clash</baseNameString>
    <variant>
      <parameters>
        <topicRef xlink:href="http://www.topicmaps.org/xtm/1.0/psi-sort"/>
      </parameters>
      <variantName>
        <resourceData>clash the</resourceData>
      </variantName>
    </variant>
    <variant>
      <parameters>
        <topicRef xlink:href="http://www.topicmaps.org/xtm/1.0/psi-sort"/>
      </parameters>
      <variantName>
        <resourceData>Clash, The</resourceData>
      </variantName>
    </variant>
  </baseName>
</topic>
```

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

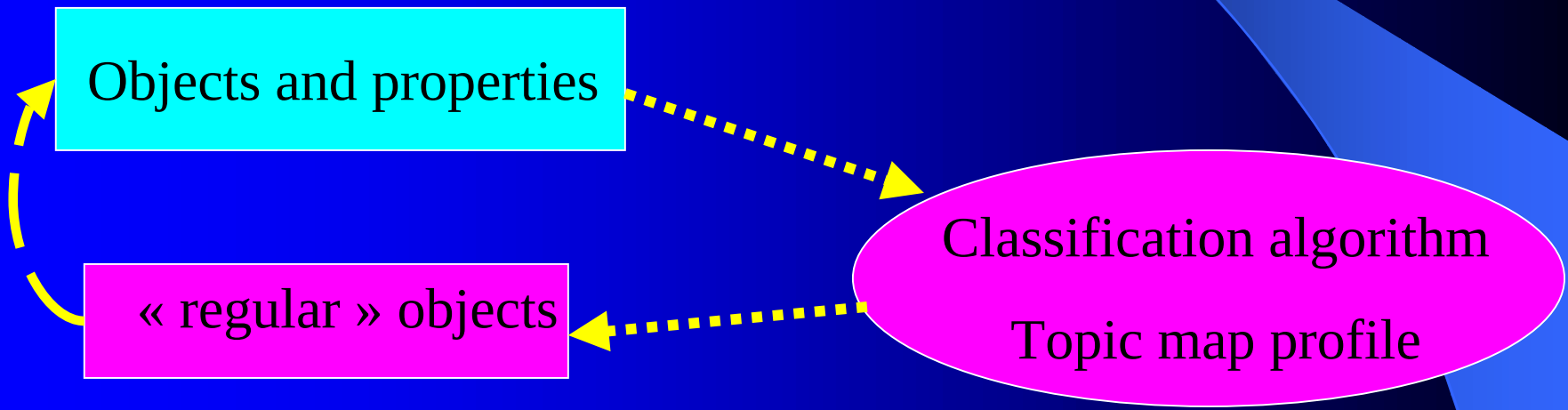
Topic map's profile

- What are the caracteristics 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

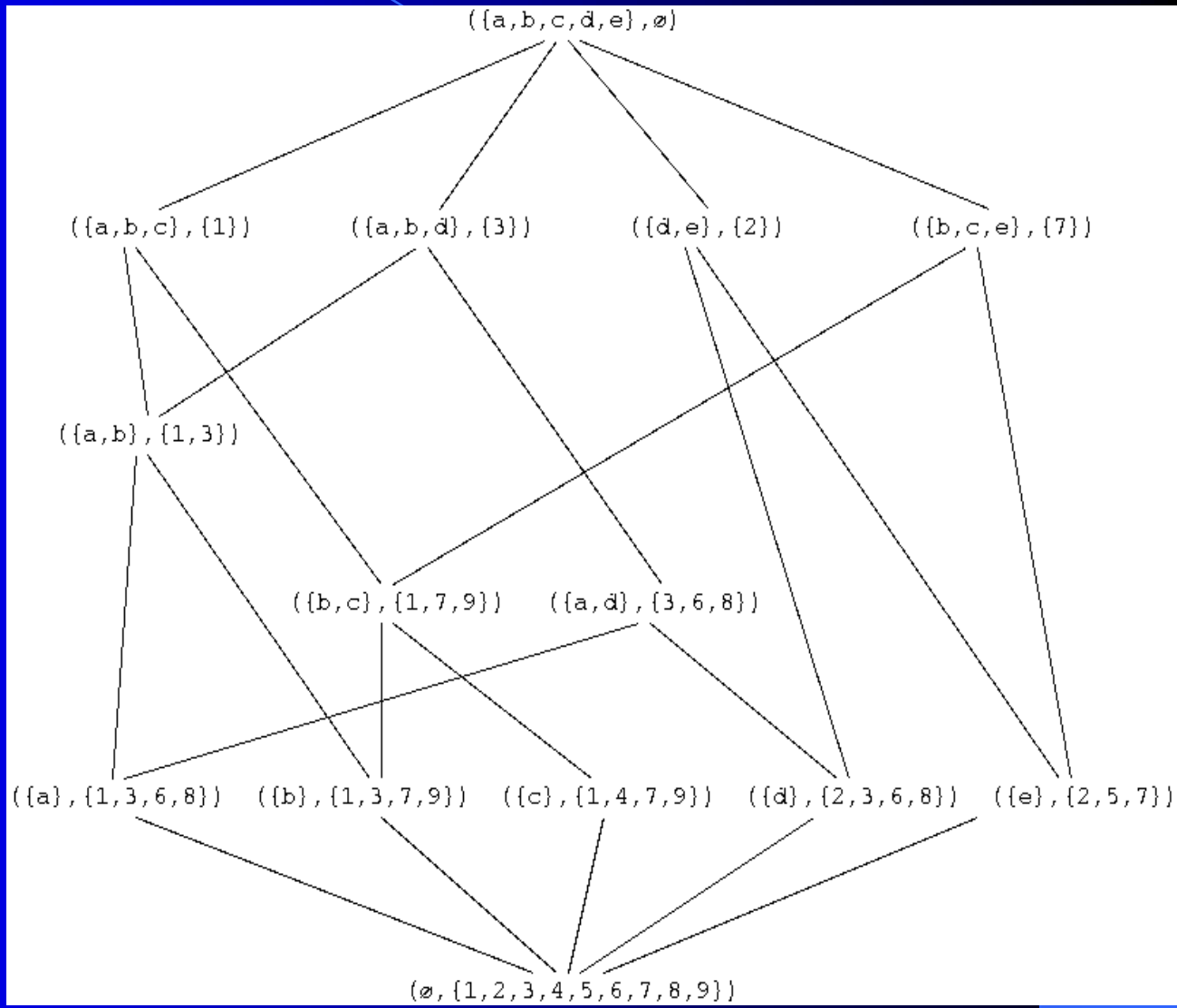


Conceptual classification algorithm

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

Classification algorithm

R	1	2	3	4	5	6	7	8	9
a	1	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	1	0	1
c	1	0	0	1	0	0	1	0	1
d	0	1	1	0	0	1	0	1	0
e	0	1	0	0	1	0	1	0	0



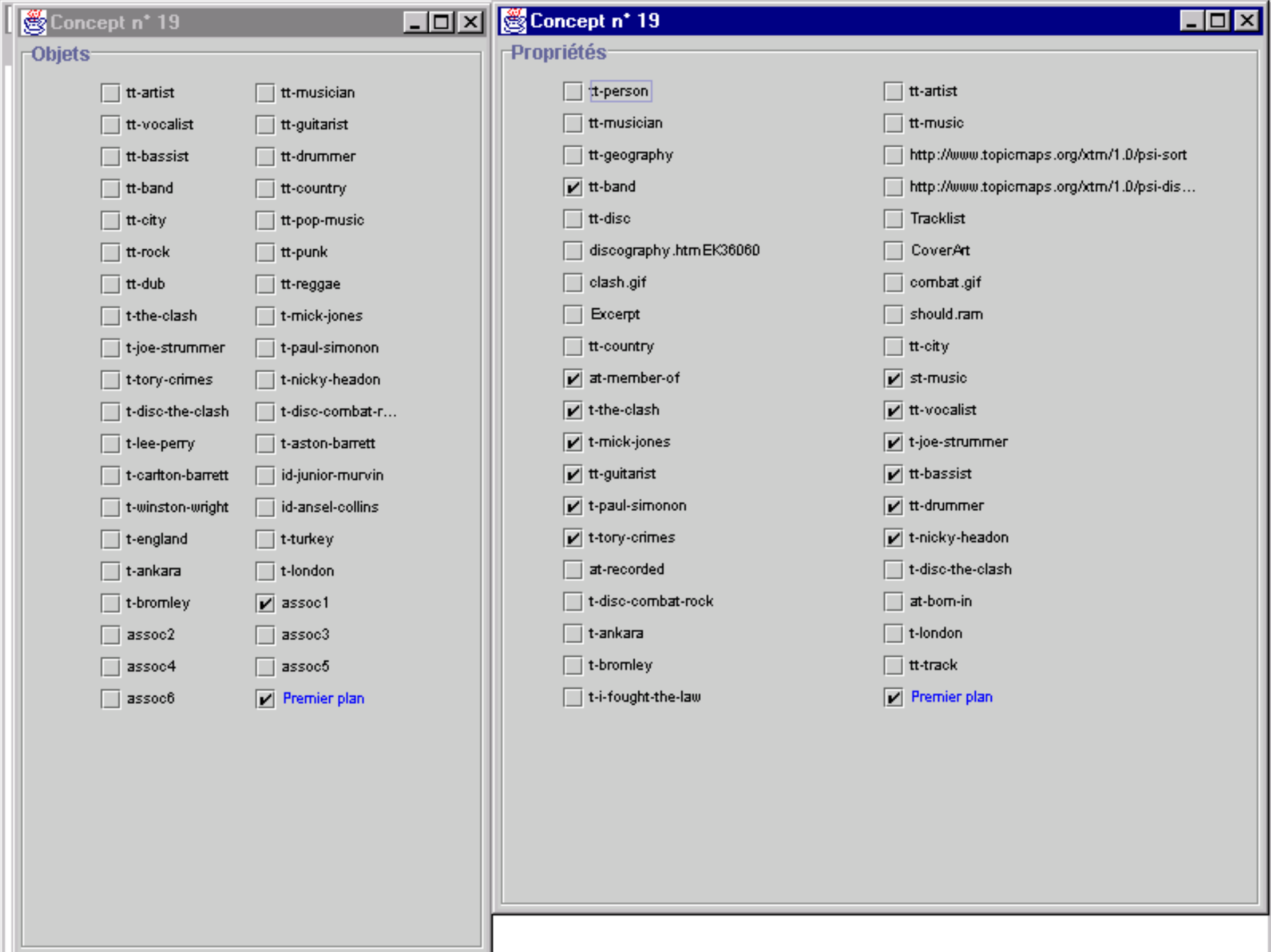
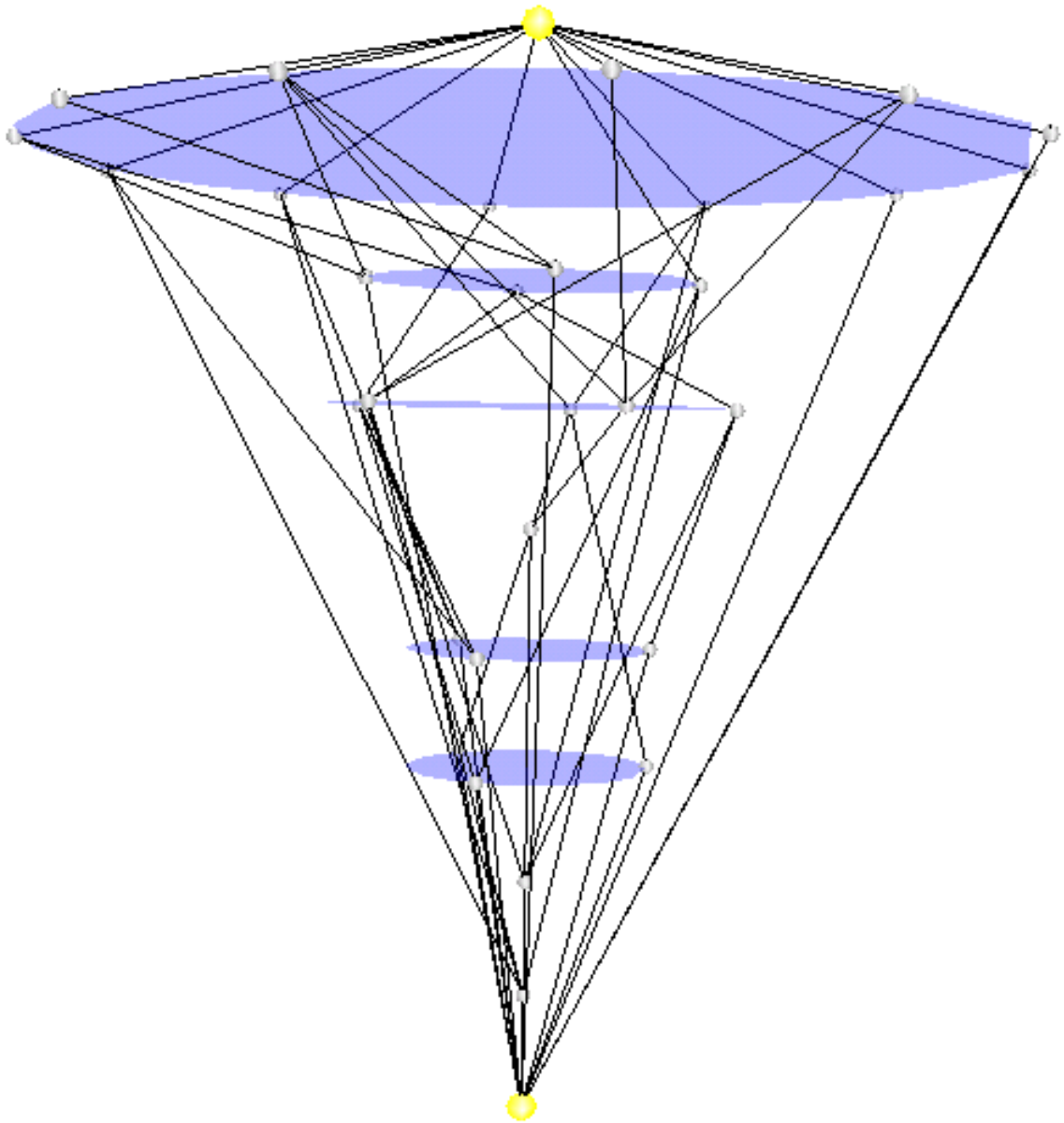

Concept n° 19
Objets
tt-artist
tt-musician
tt-vocalist
tt-guitarist
tt-bassist
tt-drummer
tt-band
tt-country
tt-city
tt-pop-music
tt-rock
tt-punk
tt-dub
tt-reggae
t-the-clash
t-mick-jones
t-joe-strummer
t-paul-simonon
t-tory-crimes
t-nicky-headon
t-disc-the-clash
t-disc-combat-r...
t-lee-perry
t-aston-barrett
t-carlton-barrett
id-junior-murvin
t-winston-wright
id-ansel-collins
t-england
t-turkey
t-ankara
t-london
t-bromley
assoc1
assoc2
assoc3
assoc4
assoc5
assoc6
Premier plan
Concept n° 19
Propriétés
t-person
tt-artist
tt-musician
tt-music
tt-geography
http://www.topicmaps.org/xtm/1.0/psi-sort
tt-band
http://www.topicmaps.org/xtm/1.0/psi-dis...
tt-disc
Tracklist
discography.htmEK36060
CoverArt
clash.gif
combat.gif
Excerpt
should.ram
tt-country
tt-city
at-member-of
st-music
t-the-clash
tt-vocalist
t-mick-jones
t-joe-strummer
tt-guitarist
tt-bassist
t-paul-simonon
tt-drummer
t-tory-crimes
t-nicky-headon
at-recorded
t-disc-the-clash
t-disc-combat-rock
at-born-in
t-ankara
t-london
t-bromley
tt-track
t-i-fought-the-law
Premier plan

Diagramme de Hasse n°0

Fichier Affichage Outils



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

Galois lattice

  - The number of occurrences of each object in the topic map itself

Topic map

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 * standard deviation ≥ profile.stats
- New list of object → new profile
  → new selection …

Music.xtm discarded objets

First iteration :

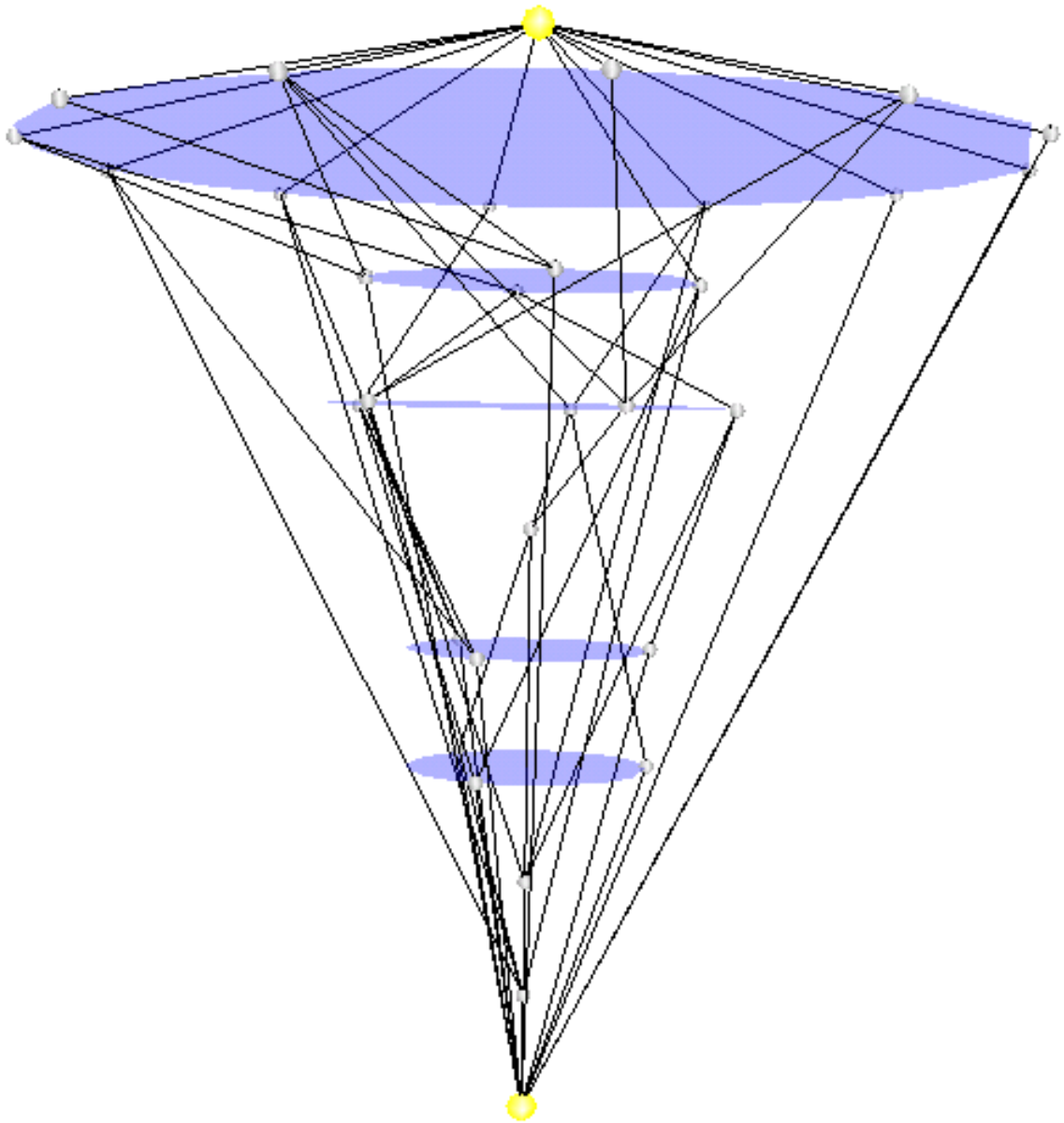
tt-musician
tt-city
t-disc-combat-rock
t-turkey
tt-country
t-disc-the-clash
t-england
assoc1

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

Diagramme de Hasse n°0

Fichier Affichage Outils



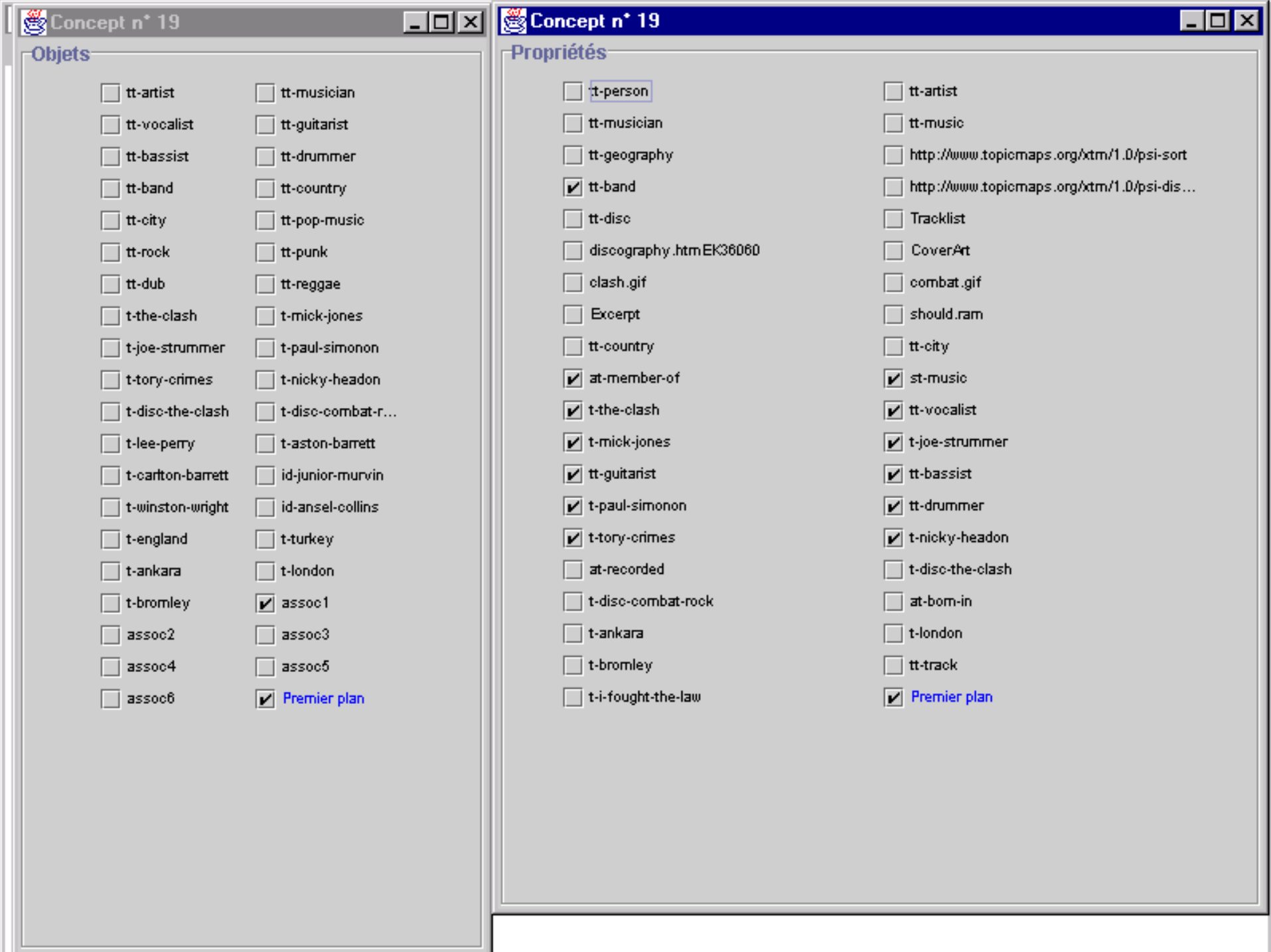
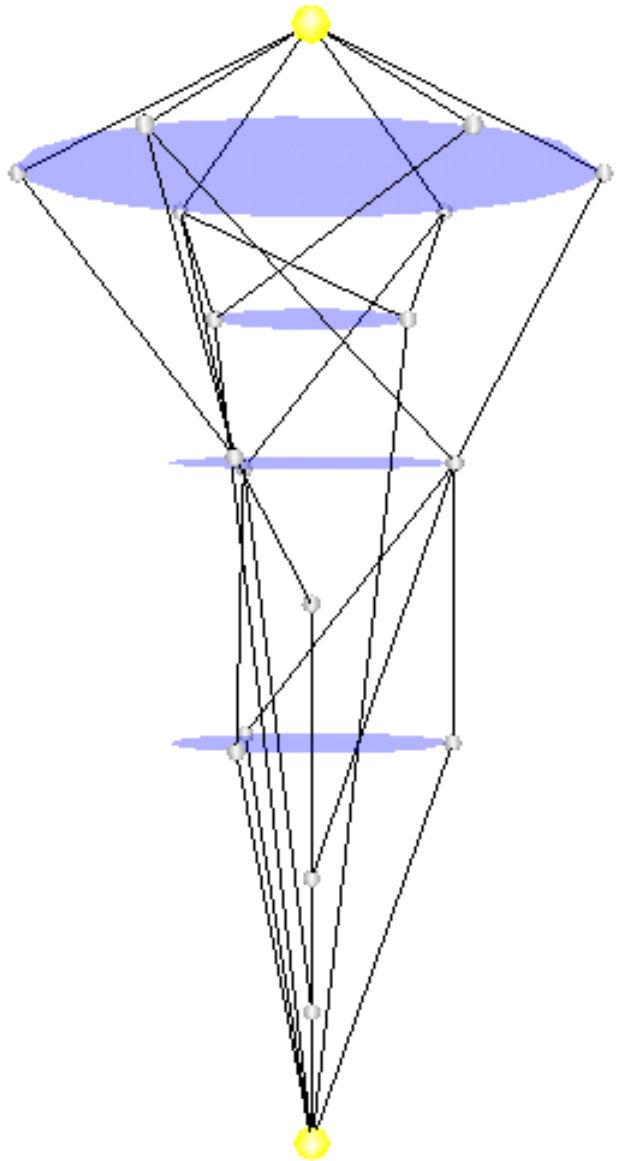

Concept n° 19
Objets
tt-artist
tt-musician
tt-vocalist
tt-guitarist
tt-bassist
tt-drummer
tt-band
tt-country
tt-city
tt-pop-music
tt-rock
tt-punk
tt-dub
tt-reggae
t-the-clash
t-mick-jones
t-joe-strummer
t-paul-simonon
t-tory-crimes
t-nicky-headon
t-disc-the-clash
t-disc-combat-r...
t-lee-perry
t-aston-barrett
t-carlton-barrett
id-junior-murvin
t-winston-wright
id-ansel-collins
t-england
t-turkey
t-ankara
t-london
t-bromley
assoc1
assoc2
assoc3
assoc4
assoc5
assoc6
Premier plan
Concept n° 19
Propriétés
t-person
tt-artist
tt-musician
tt-music
tt-geography
http://www.topicmaps.org/xtm/1.0/psi-sort
tt-band
http://www.topicmaps.org/xtm/1.0/psi-dis...
tt-disc
Tracklist
discography.htmEK36060
CoverArt
clash.gif
combat.gif
Excerpt
should.ram
tt-country
tt-city
at-member-of
st-music
t-the-clash
tt-vocalist
t-mick-jones
t-joe-strummer
tt-guitarist
tt-bassist
t-paul-simonon
tt-drummer
t-tory-crimes
t-nicky-headon
at-recorded
t-disc-the-clash
t-disc-combat-rock
at-born-in
t-ankara
t-london
t-bromley
tt-track
t-i-fought-the-law
Premier plan

Diagramme de Hasse n°0

Fichier Affichage Outils

Objets

- [ ] t-artist
- [ ] tt-vocalist
- [ ] tt-guitarist
- [ ] tt-bassist
- [ ] tt-drummer
- [ ] tt-band
- [ ] tt-pop-music
- [ ] tt-rock
- [ ] tt-punk
- [ ] tt-dub
- [ ] tt-reggae
- [ ] t-the-clash
- [ ] t-mick-jones
- [ ] t-joe-strummer
- [ ] t-paul-simonon
- [ ] t-tory-crimes
- [ ] t-nicky-headon
- [ ] t-lee-perry
- [ ] t-aston-barrett
- [ ] t-carlton-barrett
- [ ] id-junior-murvin
- [ ] t-winston-wright
- [ ] id-ansel-collins
- [ ] t-ankara
- [ ] t-london
- [ ] t-bromley
- [ ] assoc2
- [ ] assoc3
- [ ] assoc4
- [ ] assoc5
- [ ] assoc6
- [x] Premier plan



Propriétés

- [ ] tt-person
- [ ] tt-musician
- [ ] tt-music
- [ ] http://www.topicmaps.org/xtm/1.0/psi-sort
- [ ] tt-band
- [ ] http://www.topicmaps.org/xtm/1.0/psi-dis...
- [ ] tt-city
- [ ] at-recorded
- [ ] st-music
- [ ] t-the-clash
- [ ] tt-disc
- [ ] t-disc-the-clash
- [ ] t-disc-combat-rock
- [ ] at-born-in
- [ ] t-ankara
- [ ] t-joe-strummer
- [ ] t-london
- [ ] t-mick-jones
- [ ] t-paul-simonon
- [ ] t-bromley
- [ ] t-nicky-headon
- [ ] tt-track
- [ ] t-i-fought-the-law
- [x] Premier plan

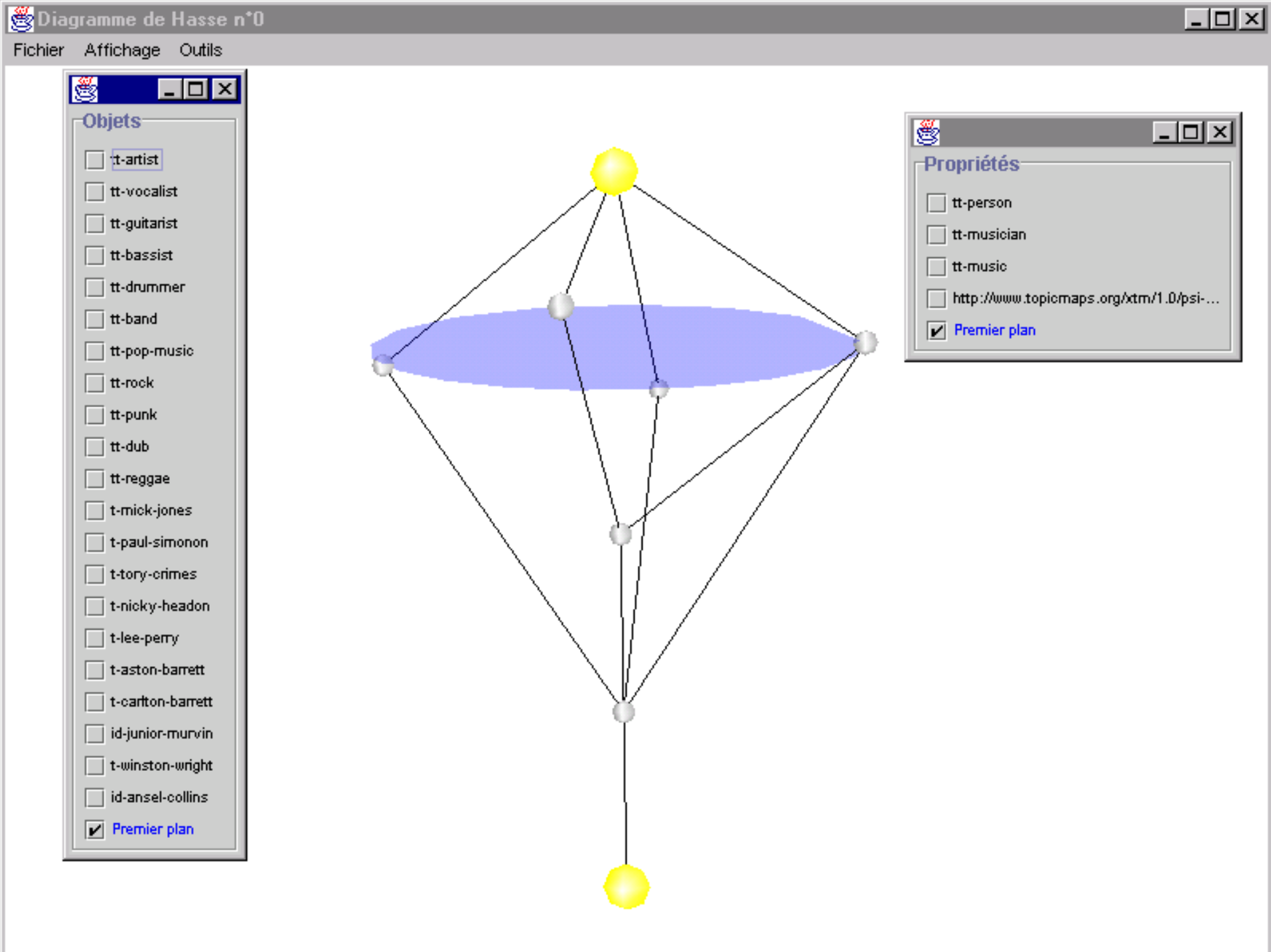


Diagramme de Hasse n°0
Fichier
Affichage
Outils
Objets
t-artist
tt-vocalist
tt-guitarist
tt-bassist
tt-drummer
tt-band
tt-pop-music
tt-rock
tt-punk
tt-dub
tt-reggae
t-mick-jones
t-paul-simonon
t-tory-crimes
t-nicky-headon
t-lee-perry
t-aston-barrett
t-carlton-barrett
id-junior-murvin
t-winston-wright
id-ansel-collins
Premier plan
Propriétés
tt-person
tt-musician
tt-music
http://www.topicmaps.org/xtm/1.0/psi-...
Premier plan

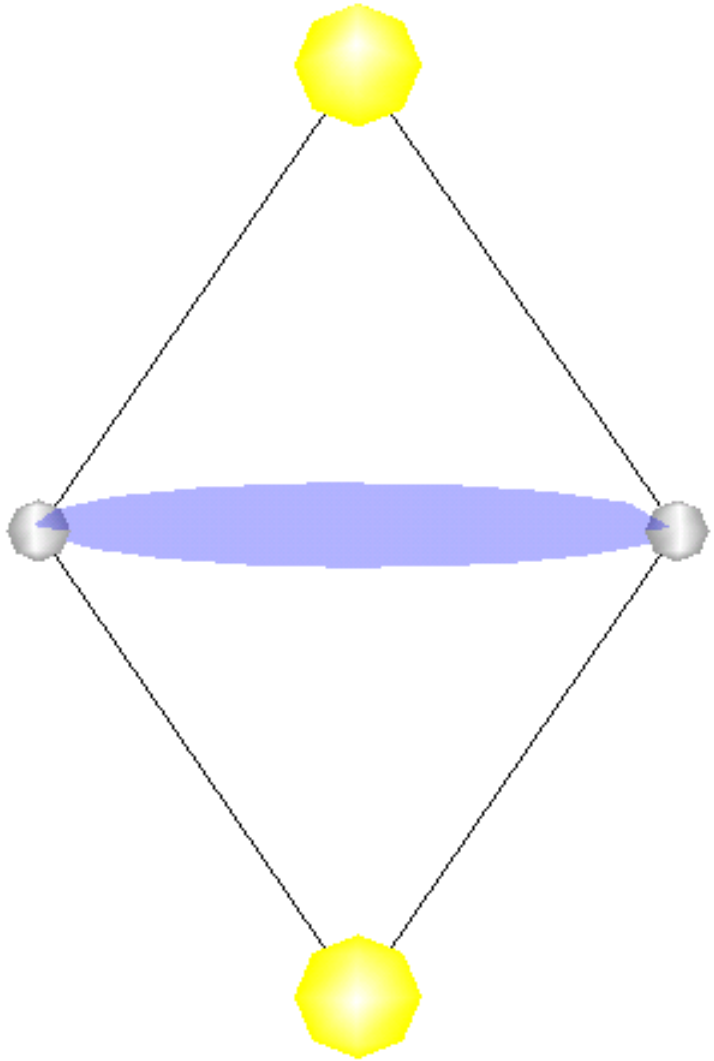
Diagramme de Hasse n°0

Fichier Affichage Outils

Co...

Objets

- tt-artist
- tt-vocalist
- tt-guitarist
- tt-bassist
- tt-drummer
- t-mick-jones
- t-paul-simonon
- t-tory-crimes
- t-nicky-headon
- t-lee-perry
- t-aston-barrett
- t-carlton-barrett
- id-junior-murvin
- t-winston-wright
- id-ansel-collins
- Premier plan



Concept n° 3

Propriétés

- tt-person
- tt-musician
- http://www.topicmaps.org/xtm/1.0/psi-...
- Premier plan

Stabilization of the algorithm

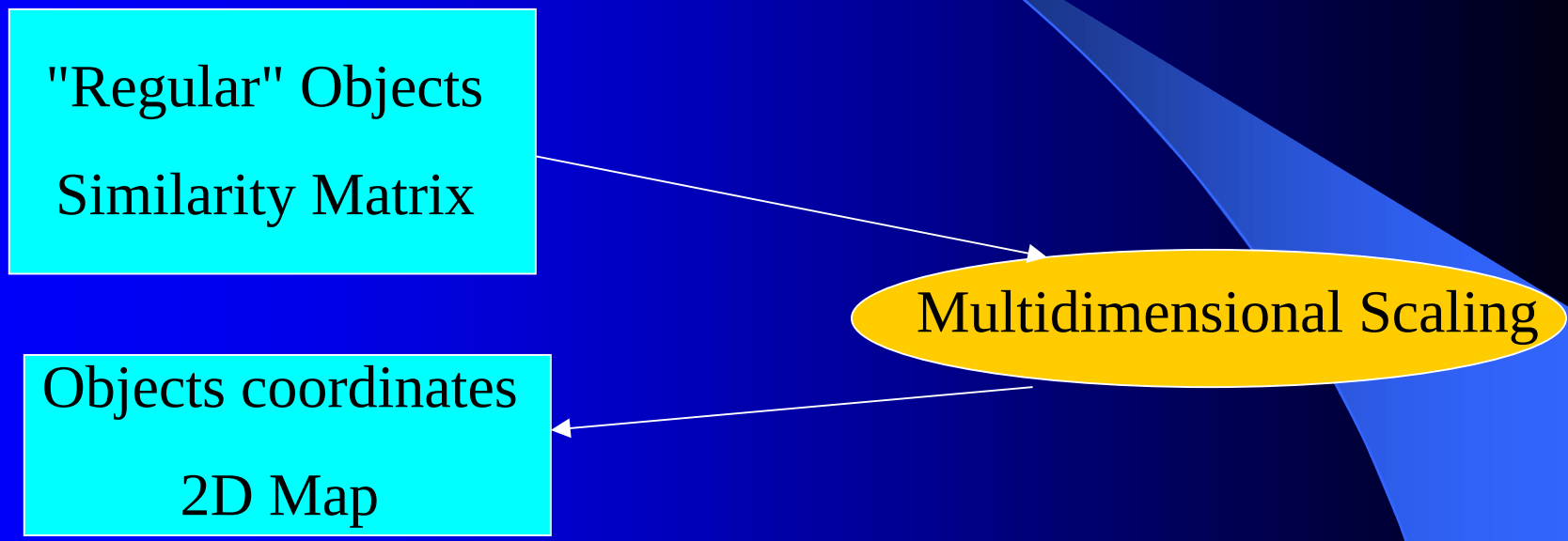
- When there are no more « singular » objects
- Set of objects which can be similar or not

⟹ Computation of objects coordinates
Multidimensional Scaling (MDS)

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
- ...

Multidimensional Scaling
Position of topics



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]
- object : t-nicky-headon 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

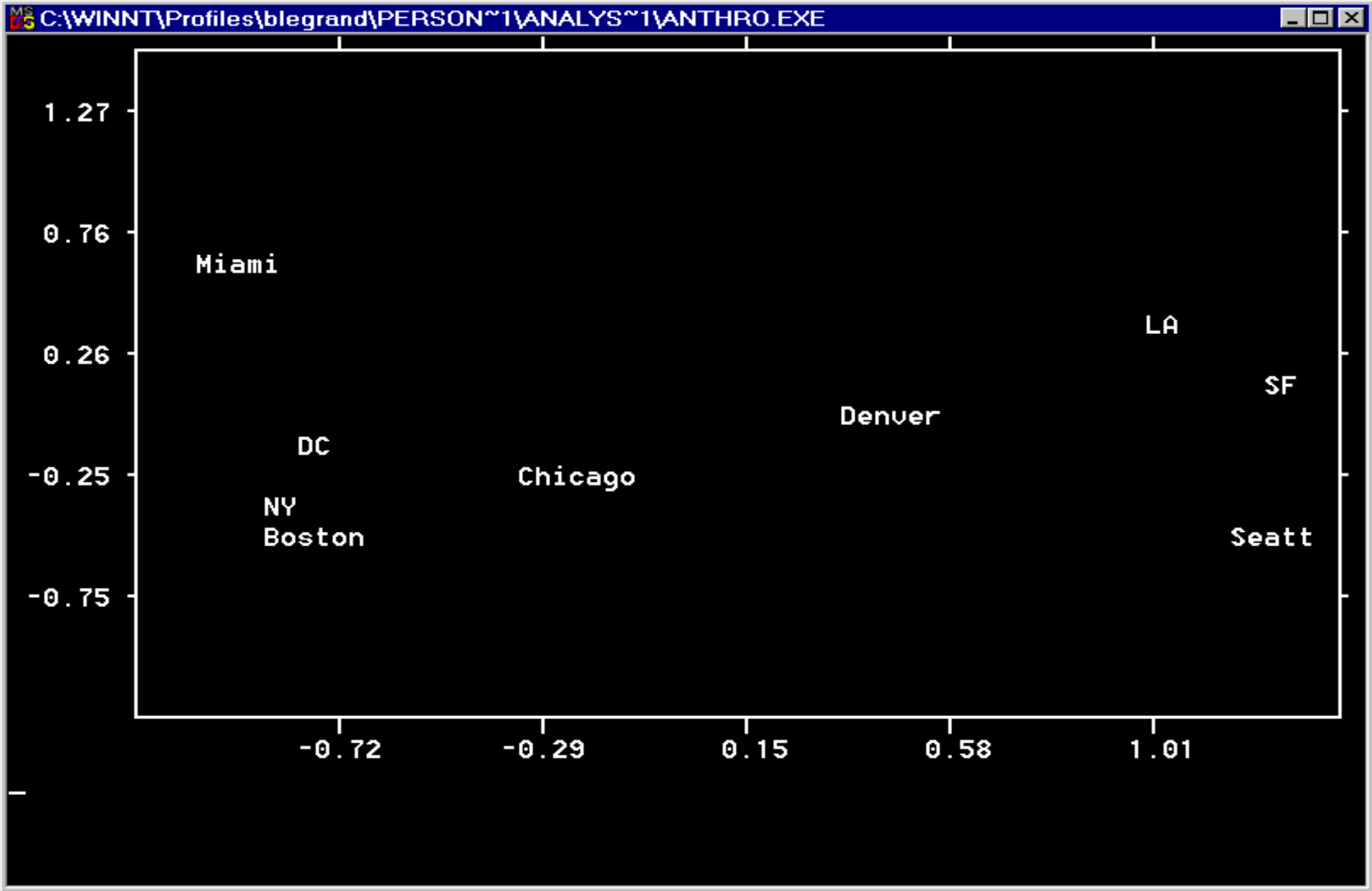
Example : US Cities

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

US Cities : MDS Input

	Boston	NY	DC	Miami	Chicago	Seattle	SF	LA	Denver
Boston	0	206	429	1504	963	2976	3095	2979	1949
NY	206	0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
Miami	1504	1308	1075	0	1329	3273	3053	2687	2037
Chicago	963	802	671	1329	0	2013	2142	2054	996
Seattle	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
Denver	1949	1771	1616	2037	996	1307	1235	1059	0

US Cities : MDS Output



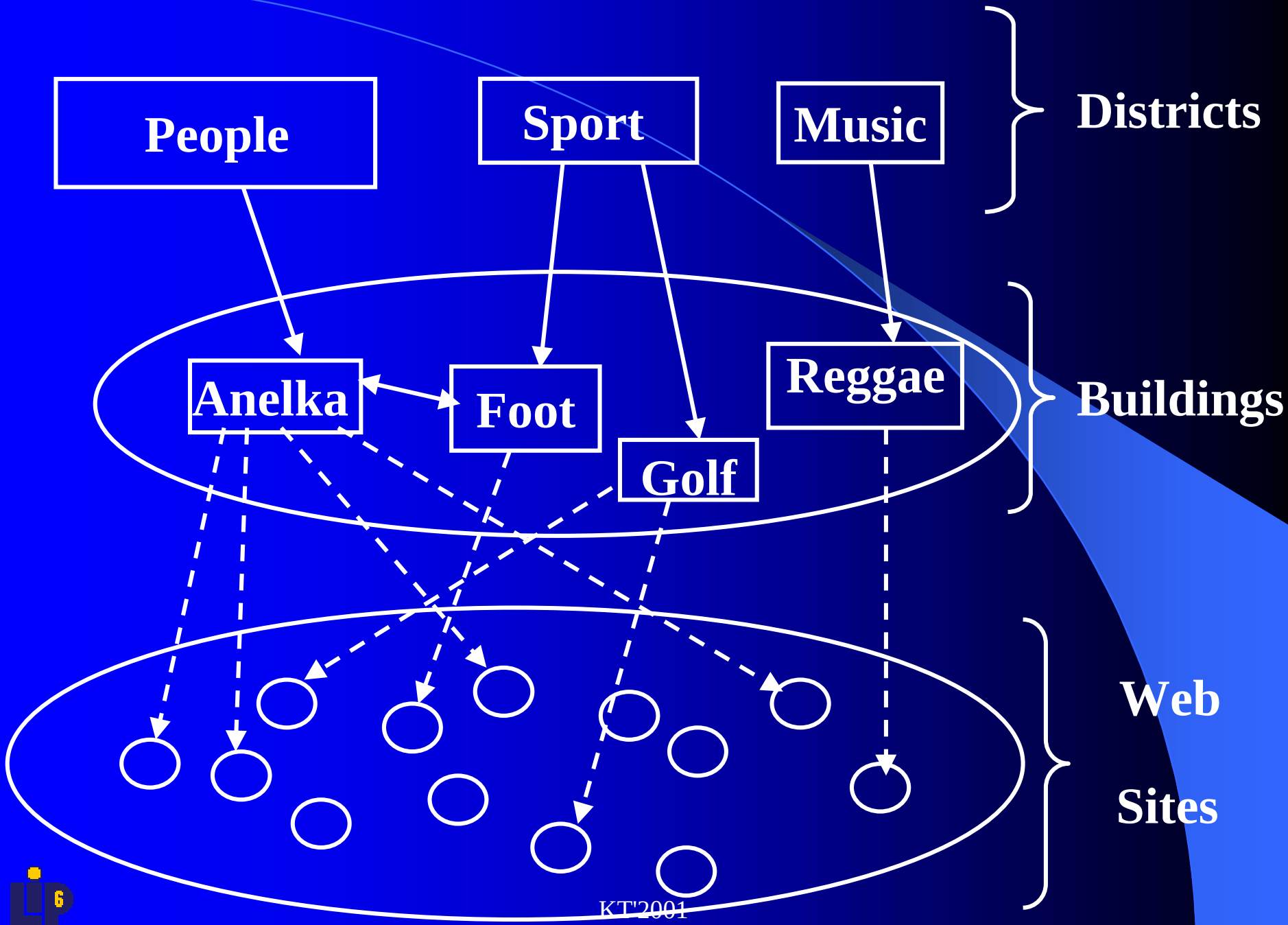
Topic Map Visualization : Virtual World

Objects coordinates

2D Map

Visualization tool

Virtual World


People
Sport
Music
Districts
Anelka
Foot
Golf
Reggae
Buildings
Web
Sites

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

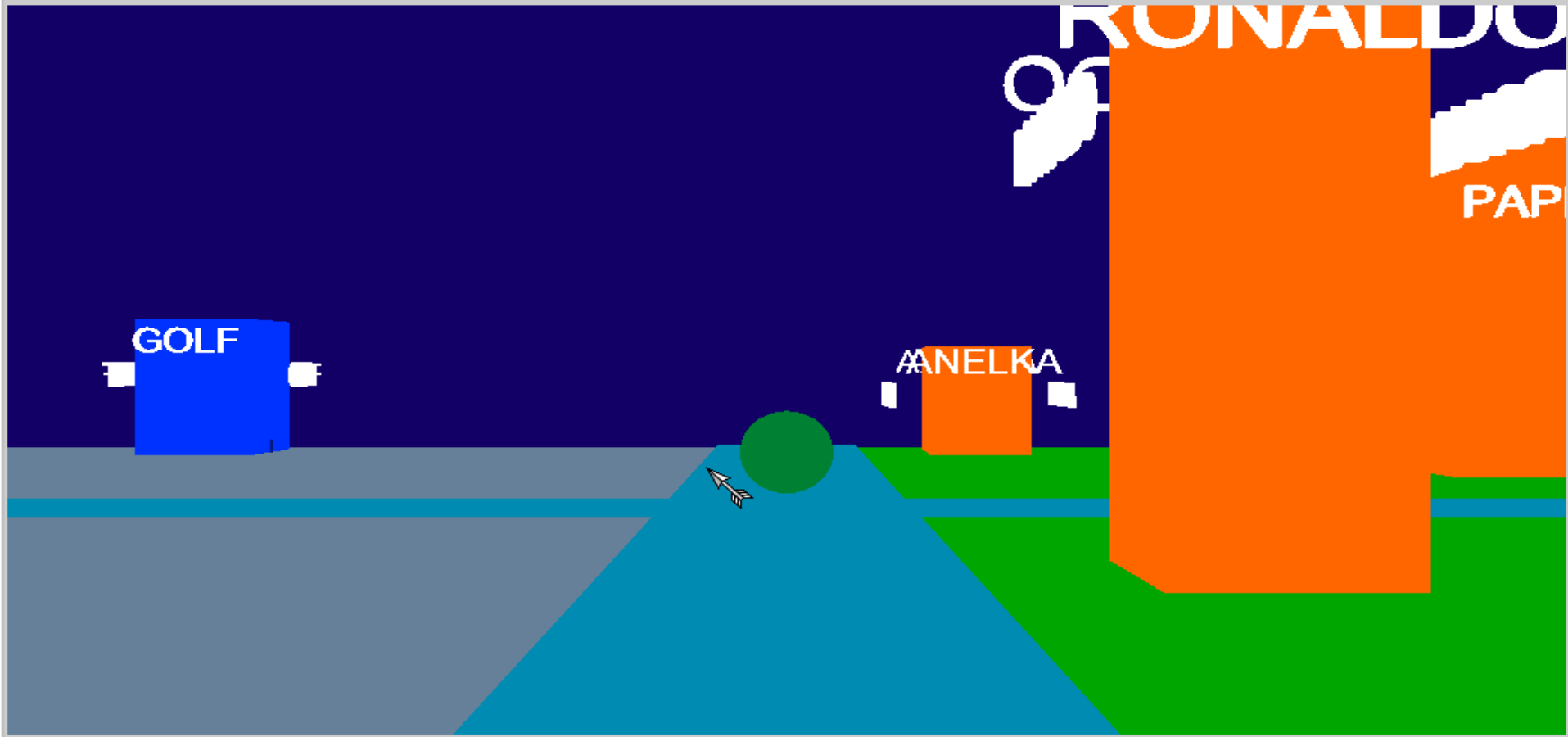
- Free navigation
  - Easy orientation with the map
  - Cognitive map
- Guided tours
  - Global tour - visit all clusters
  - Explore one district in particular

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

Le monde virtuel de la Recherche



Visite guidée terrestre

Visite guidée aérienne

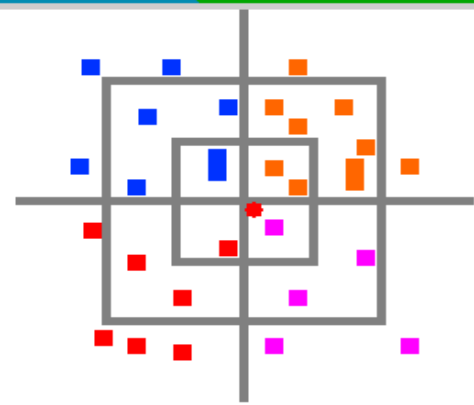
STOP

Contenu de l'immeuble :

Liste Contenu

Liens de l'immeuble : aucun immeuble sélectionné

Liste liens



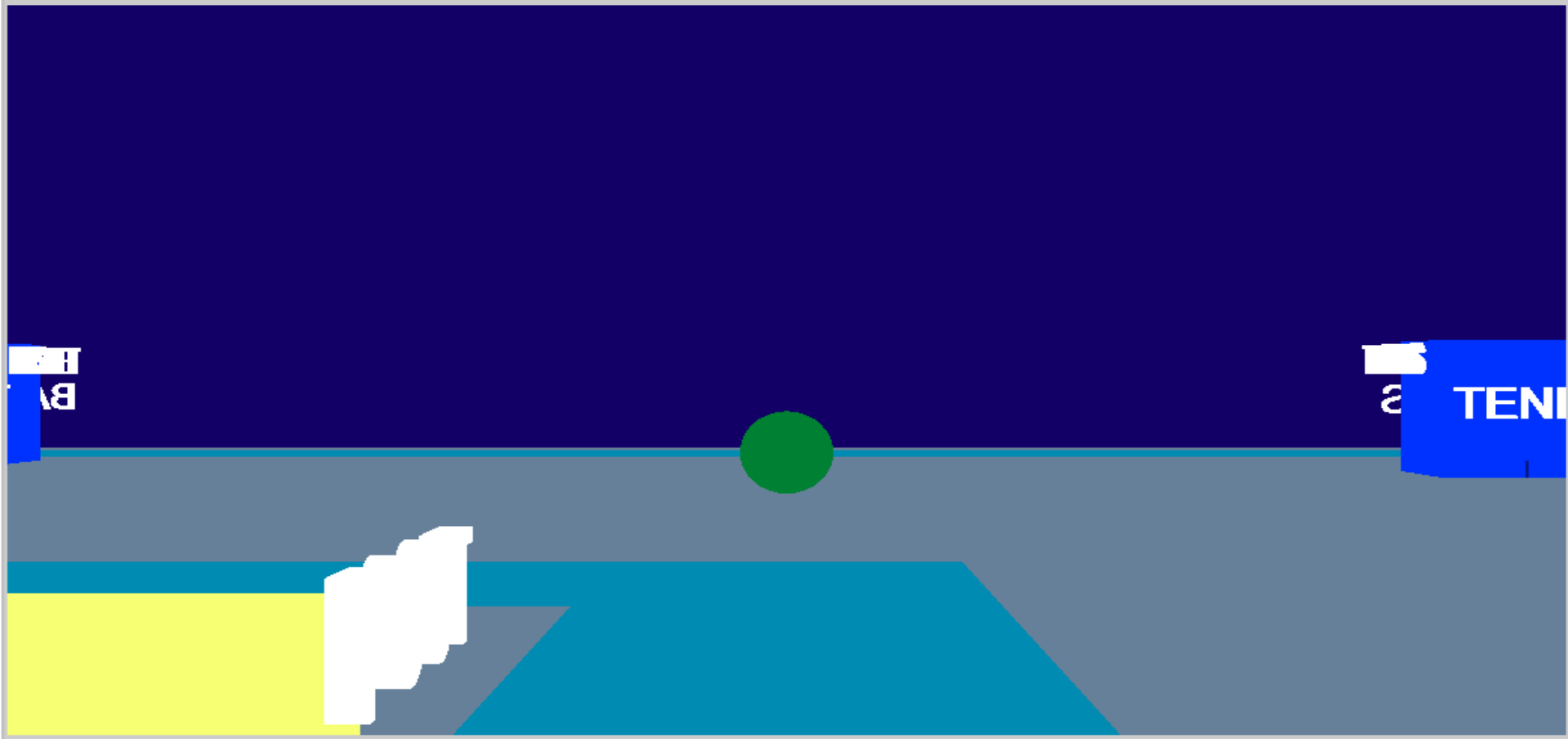
Aucun lien sélectionné

Demo : guided tours

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

Le monde virtuel

Le monde virtuel de la Recherche



Visite guidée terrestre

Visite guidée aérienne

STOP

Contenu de l'immeuble :

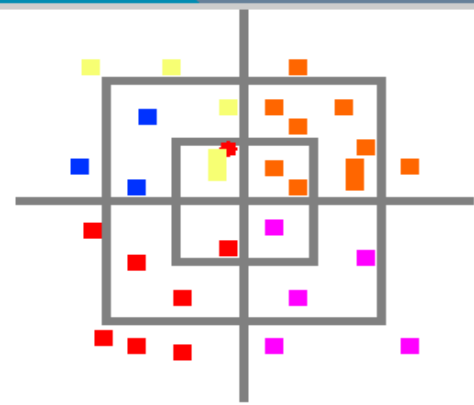
www.lisarazu.fr

Liens de l'immeuble : LISARAZU

FOOT
GOLF
VOLLEY
RUGBY

Go!

Aucun lien sélectionné



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 !