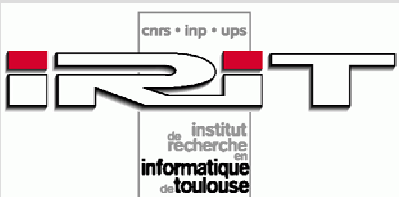


# Recherche et Exploration d'Information IRIT

- Josiane Mothe,  
Groupe SIG, Equipe RI  
Systèmes d'information généralisés  
Recherche d'information

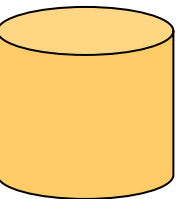
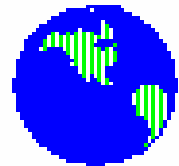
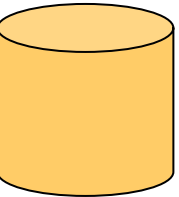


mothe@irit.fr

# Introduction

## Accès à l'information

### Sources



### Sources : hétérogènes

#### - Contenu

Contexte

Langue

#### - format

BD,

BD multidimen

Semi-structur

Multimedia

#### - stable ou flots

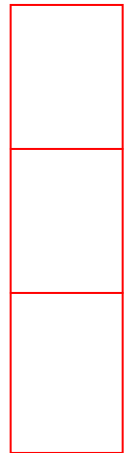
d'information

### Besoins utilisateur :

- courts ou à long terme

- information plus ou moins avancées

- résultats personnalisés

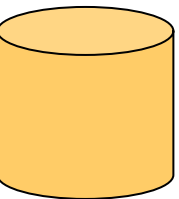
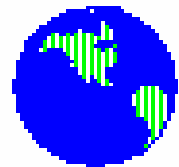
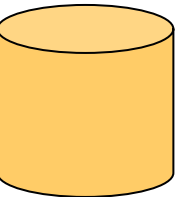


Profils

Besoins

# Introduction

Sources



Visualisation

Exploration

Indexation

## INDEXATION

Représentation des unités  
d'information

- audio
- video
- texte
- multi-media

Extraction d'éléments  
d'information ciblés

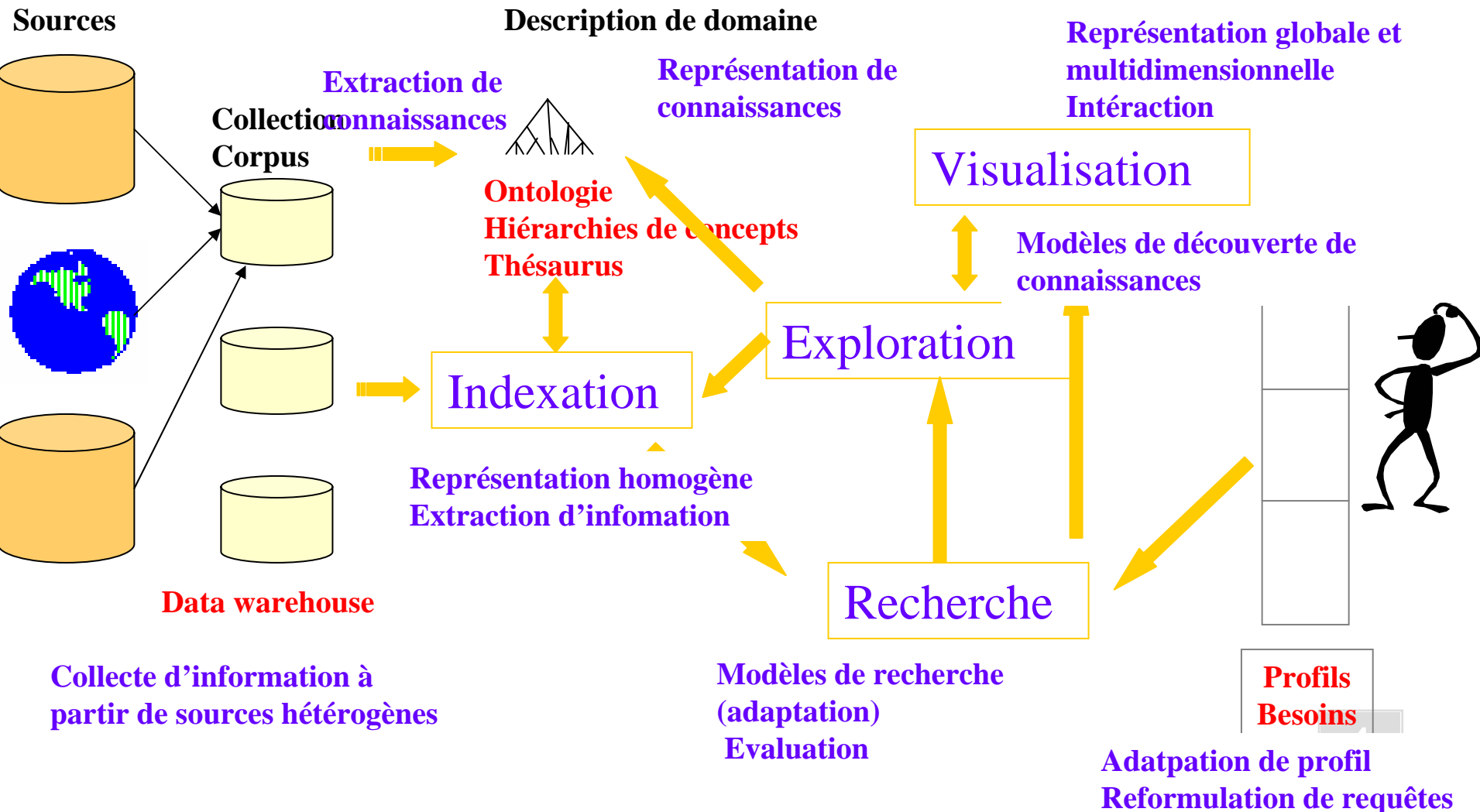
Recherche



Profils  
Besoins

3

# Introduction



## Quelques projets

- Mercure: Système de Recherche d'Information (réseau de neurones)
- IRAIA, eStage (5PCRD) : IRS, Domaines représentés via des hiérarchies de concepts (navigation)
- Easy-Dor : Interface de visualisation
- DocCube : Visualisation multi-dimensionnelle d'un ensemble de documents
- Tétralogie : DC à partir de textes

- **Représentation de textes**
  - de chaque unité vs globalement
  - à partir d'un vocabulaire (catégorisation) vs des textes
- **Utilisation de vocabulaires (HC/ontologies)**
  - catégorisation / accès à l'information
  - visualisation globale / exploration
  - reformulation de requêtes
- **Aide à la construction de vocabulaires**

- **Unités: Représentation multi-facettes**
  - Principe [Ranganathan,30]
    - Objects/unités balisés (**attribut:valeurs**) ou facettes, ou dimensions
    - Facettes organisées hiérarchiquement
  - **Facettes : contenus et « meta »-informations**
    - contenu / meta-information : frontière mal définie
    - exemples
      - Mots-clés
      - auteurs, date, etc...
      - technique utilisée (papier scientifique, brevet)

- Unités: Représentation multi-facettes
  - 2 types de résultats
    - basé sur un vocabulaire contrôlé (// catégorisation)
    - basé sur un vocabulaire libre (// indexation)
  - Résultat : chaque unité est représentée par (attribut:valeurs)



## ■ Unités: Représentation multi-facettes

- Basée sur un vocabulaire libre :

- Techniques TAL
- Statistique

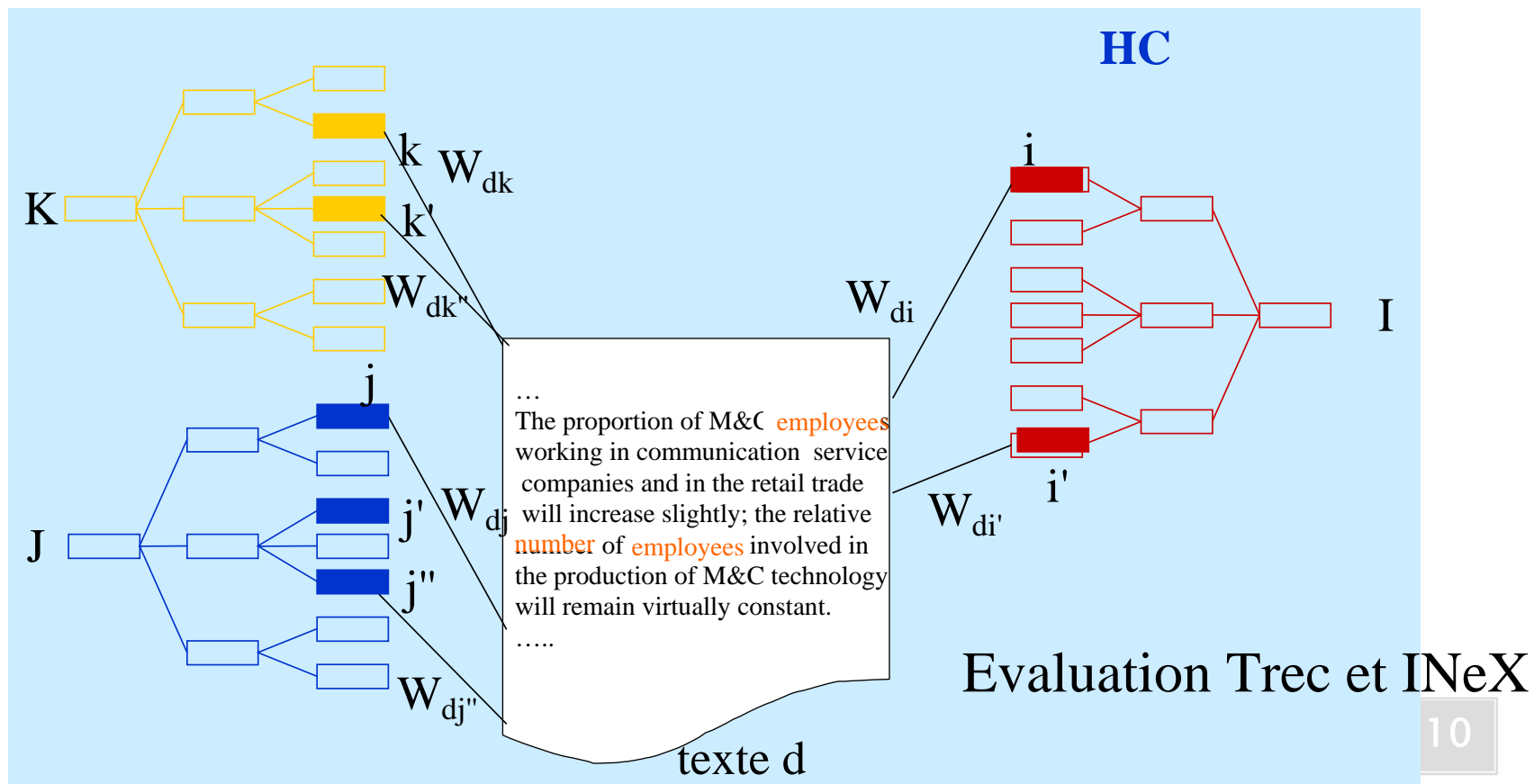
- Détection de termes (mots vides, radicalisation, groupes de mots)
- Synonymie, liens sémantiques connus
- Pondération (importance des termes)
- Dictionnaires (éventuellement)

⇒ « bags of words » en RI traditionnelle

⇒ plus générique : ensemble de termes + sémantique (dimension, point de vue)

# Représentation

- Unités: Représentation multi-facettes
  - Vocabulaire contrôlé, catégorisation



# Représentation

- Unités
- Globale

## Représentation réduite

- Représentation globale de l'information (vs unités) pour extraire des relations entre éléments (KD)
- Point de départ : tables de contingence

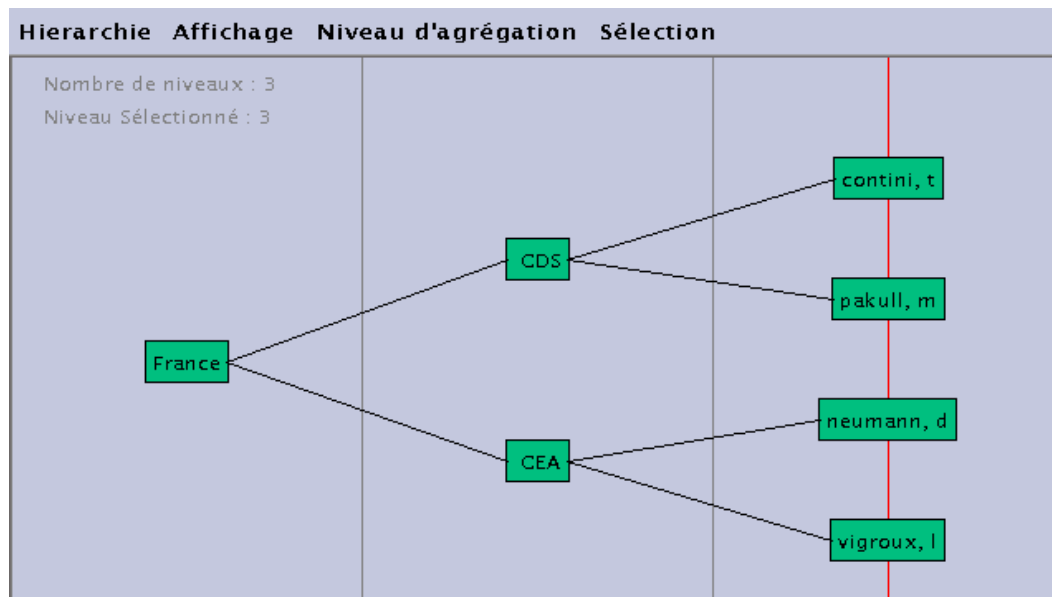
Dimensions

	<b>CDS</b>		<b>CEA</b>	
	<i>contini, t</i>	<i>pakull, m</i>	<i>neumann, d</i>	<i>vigroux, l</i>
<i>Variable Star</i>	2	6	1	6
<i>Peculiar Star</i>	3	9	1	4
<i>Cluster of Galaxy</i>	6	1	4	3
<i>Galaxy Nucleus</i>	12	7	6	14

# Représentation

## ■ Représentation réduite

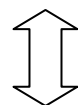
		<b>CDS</b>		<b>CEA</b>	
		<i>contini, t</i>	<i>pakull, m</i>	<i>neumann, d</i>	<i>vigroux, l</i>
<b>Star</b>	<i>Variable Star</i>	2	6	1	6
	<i>Peculiar Star</i>	3	9	1	4
<b>Galaxy</b>	<i>Cluster of Galaxy</i>	6	1	4	3
	<i>Galaxy Nucleus</i>	12	7	6	14



# Représentation

## ■ Représentation réduite

		<b>CDS</b>		<b>CEA</b>	
		<i>contini, t</i>	<i>pakull, m</i>	<i>neumann, d</i>	<i>vigroux, l</i>
<b>Star</b>	<i>Variable Star</i>	2	6	1	6
	<i>Peculiar Star</i>	3	9	1	4
<b>Galaxy</b>	<i>Cluster of Galaxy</i>	6	1	4	3
	<i>Galaxy Nucleus</i>	12	7	6	14



Changement de niveau d'agrégation

		<b>CDS</b>	<b>CEA</b>
<b>Star</b>	<i>Variable Star</i>	8	7
	<i>Peculiar Star</i>	12	5
<b>Galaxy</b>	<i>Cluster of Galaxy</i>	7	7
	<i>Galaxy Nucleus</i>	19	20

- Représentation de textes
  - de chaque unité vs globalement
  - à partir d'un vocabulaire (catégorisation) vs des textes
- Utilisation de vocabulaires (HC/ontologies)
  - catégorisation / accès à l'information
  - visualisation globale / exploration
  - reformulation de requêtes
- Aide à la construction de vocabulaires

# Utilisation : accès

## ■ Via HC

The screenshot displays the IRAIA Client software interface. The window title is "IRAIA Client". The interface includes a navigation bar with "Query by concepts", "Free text query", and "Query results" buttons. A language dropdown menu is set to "English". The main content area is divided into three columns: "Available Data Bases", "Industries", "Regions", and "Variables".

**Available Data Bases:**

- ifo-Konjunkturtest
- eurostat (LINDA)
- OECD (Main Economic Indicators)
- VGR (Konten)

**Industries:**

- Total industry
- Mining of coal and lignite; extraction of p...
- Extraction of crude petroleum and natur...
- Mining of uranium and thorium ores
- Mining of metal ores
- Other mining and quarrying
- Manufacture of food products and bever...
- Manufacture of tobacco products
- Manufacture of textiles
- Manufacture of wearing apparel; dressi...
- Tanning and dressing of leather; manu...
- Manufacture of wood and products of w...
- Manufacture of pulp, paper and paper p...
- Publishing, printing and reproduction o...
- Manufacture of coke, refined petroleum...
- Manufacture of chemicals and chemica...
- Manufacture of rubber and plastic produ...
- Manufacture of other non-metallic mine...

**Regions:**

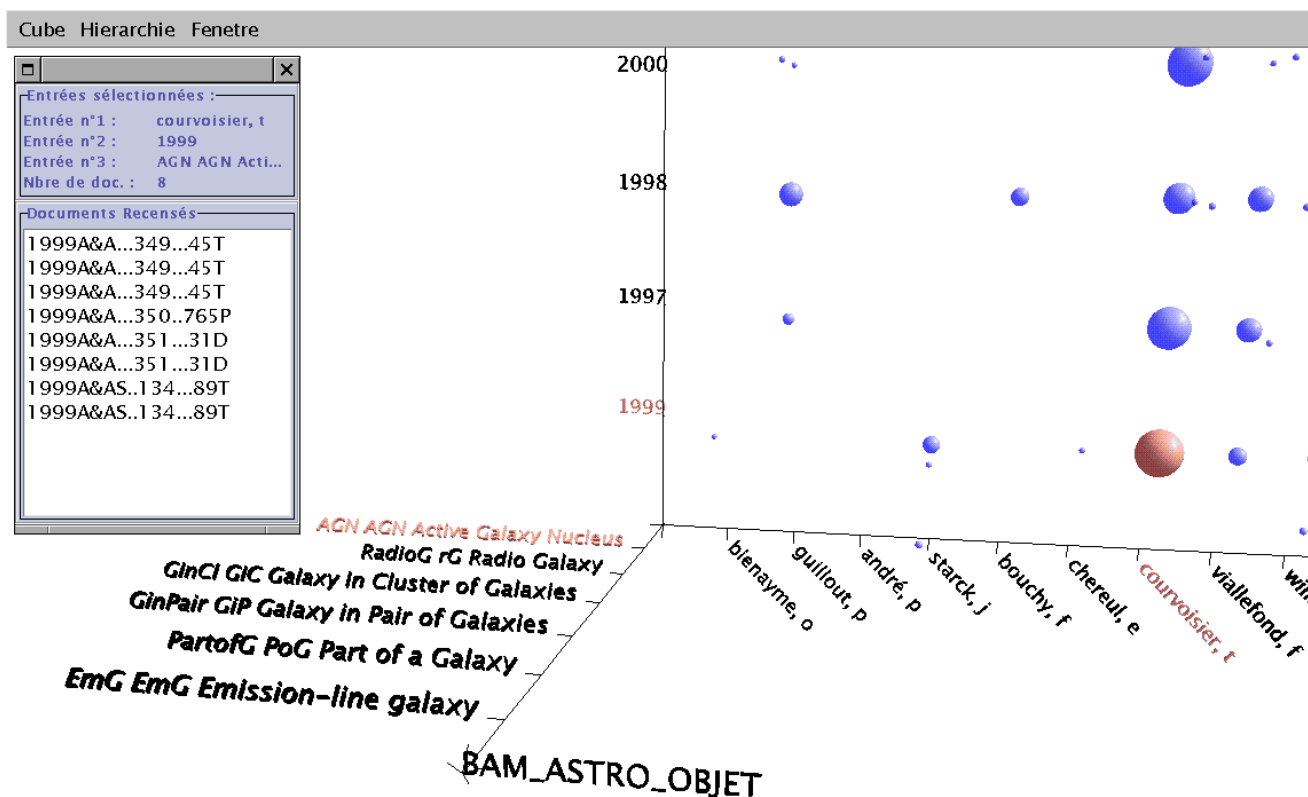
- EUR 15
- EUR 12
- Belgium
- Belgium and Luxembourg
- Denmark
- Germany
- Greece
- Spain
- France
- Ireland
- Italy
- Luxembourg
- Netherlands
- Portugal
- United Kingdom
- USA
- Japan
- Sweden
- Norway

**Variables:**

- Production
- Indices of turnover
- Indices of new orders
- Social indicators in industry
- Foreign trade indices
- Productivity indices
- Industrial producer price index

## ■ Vues globales (data cubes)

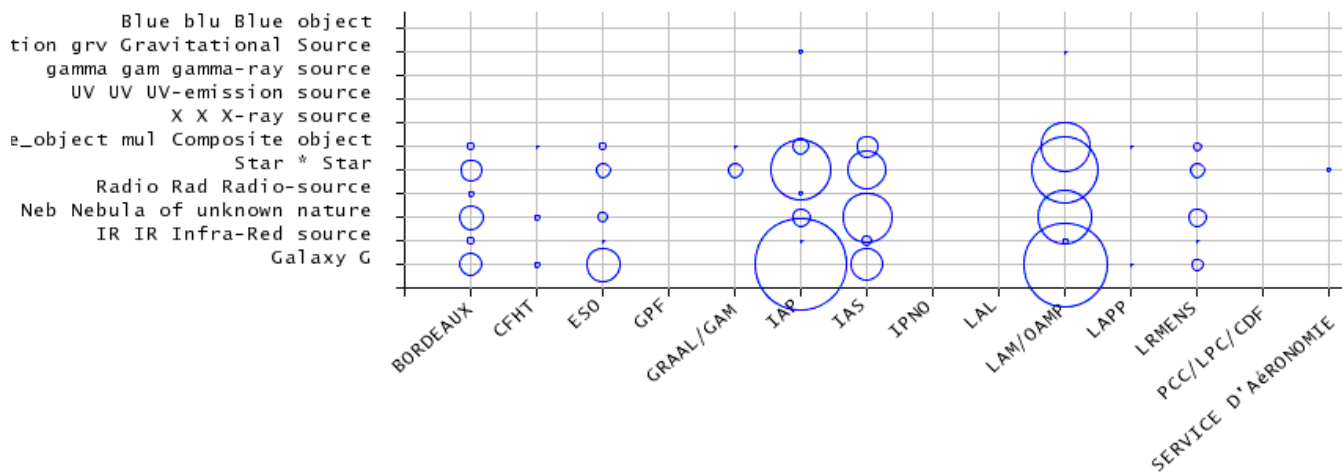
Systeme: DocCube





# Utilisation : accès

## ■ Vues globales (data cubes / slice)



Echelles

x2 1

x1 10

x1 50

/2 100

/10 100

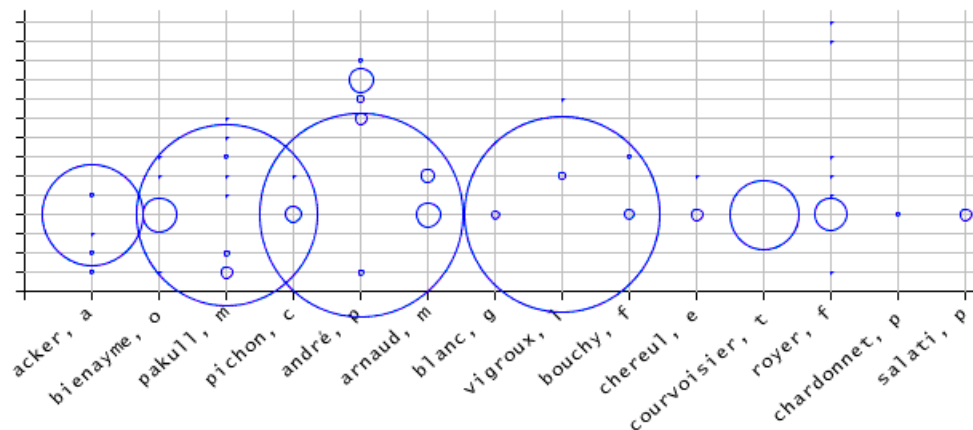
/100

Doit être supérieur à :

# Utilisation : accès

## ■ Vues globales (data cubes / slice)

ical Cepheid (delta Cep type)  
Cepheid variable Star  
sub-millimetric source  
Young Stellar Object  
nce Star (optically detected)  
Variable Star of Orion Type  
Be Star  
Spectroscopic binary  
Cluster of Galaxies  
riable Star of beta Cep type  
wn ? Object of unknown nature  
Post-AGB Star  
Ellipsoidal variable Star  
Double or multiple star



Echelles

x2

x1

/2

/10

/100

1

10

50

100

Doit être supérieur à :

## ■ Reformulation de requêtes

- Vocabulaire générique
  - choix des types de liens
  - évaluation WordNet (TREC)
- Vocabulaire spécifique
  - hiérarchies de concepts (liens père/fils et synonymie)
  - IRAIA

- **Représentation de textes**
  - de chaque unité vs globalement
  - à partir d'un vocabulaire (catégorisation) vs des textes
- Utilisation de vocabulaires (HC/ontologies)
  - catégorisation / accès à l'information
  - visualisation globale / exploration
  - reformulation de requêtes
- **Aide à la construction de vocabulaires**

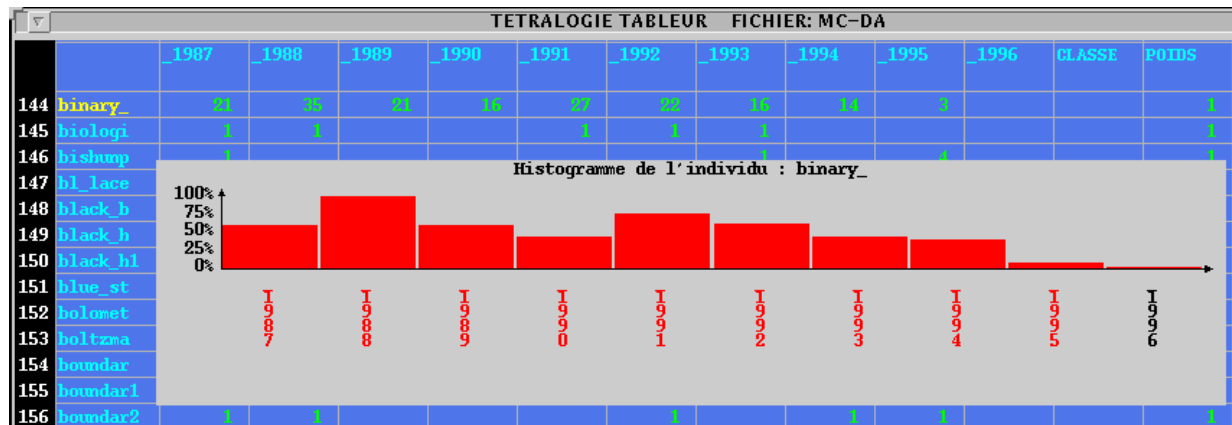
## ■ Découverte de connaissances

- Analyse de fréquence
- Classification
  - Supervisée (exple: typologie)
  - Non supervisée (exple: similarité)
- Dépendances
  - Entre valeurs de dimension
  - Evolution

## ■ Analyse de fréquence

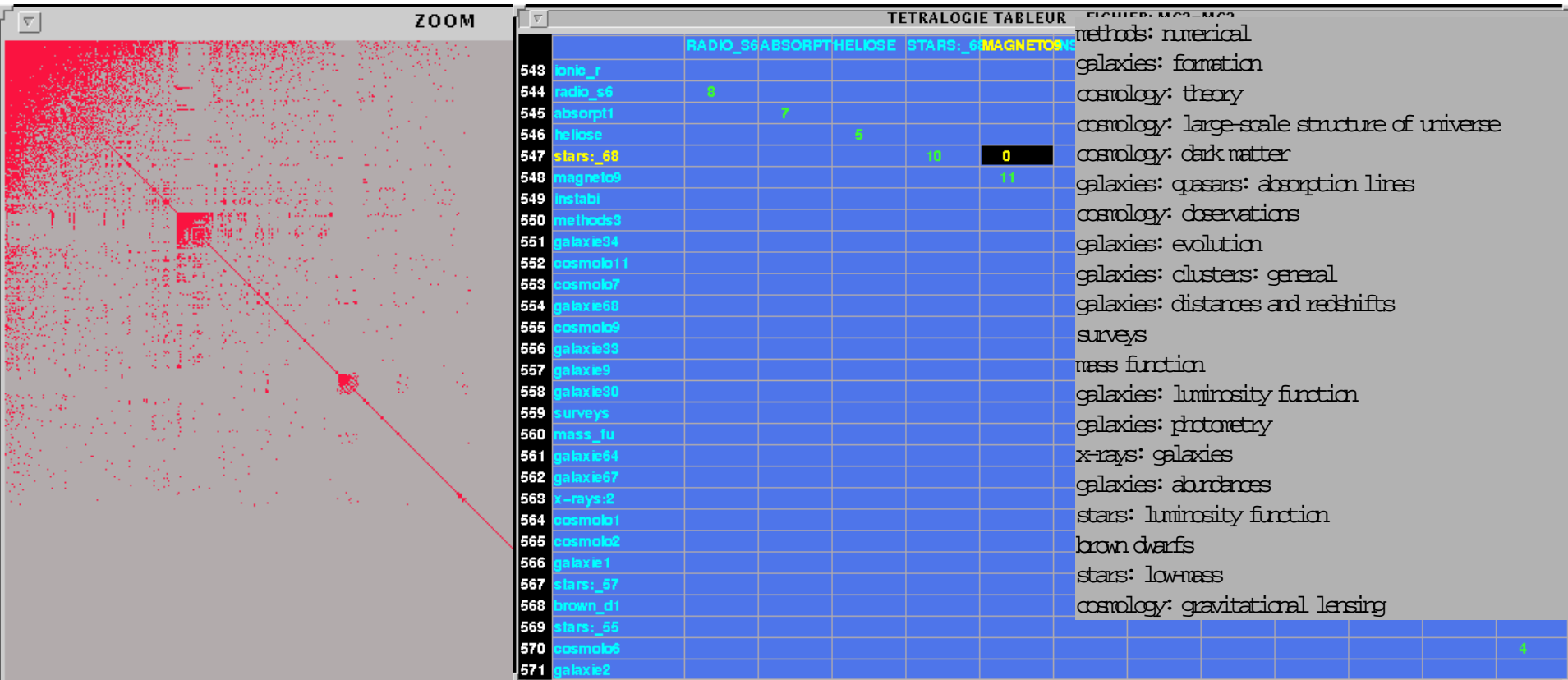
### ● Exemple

- terminologie importante et évolution
- en fonction des dimensions (instituts, personnes, etc...)



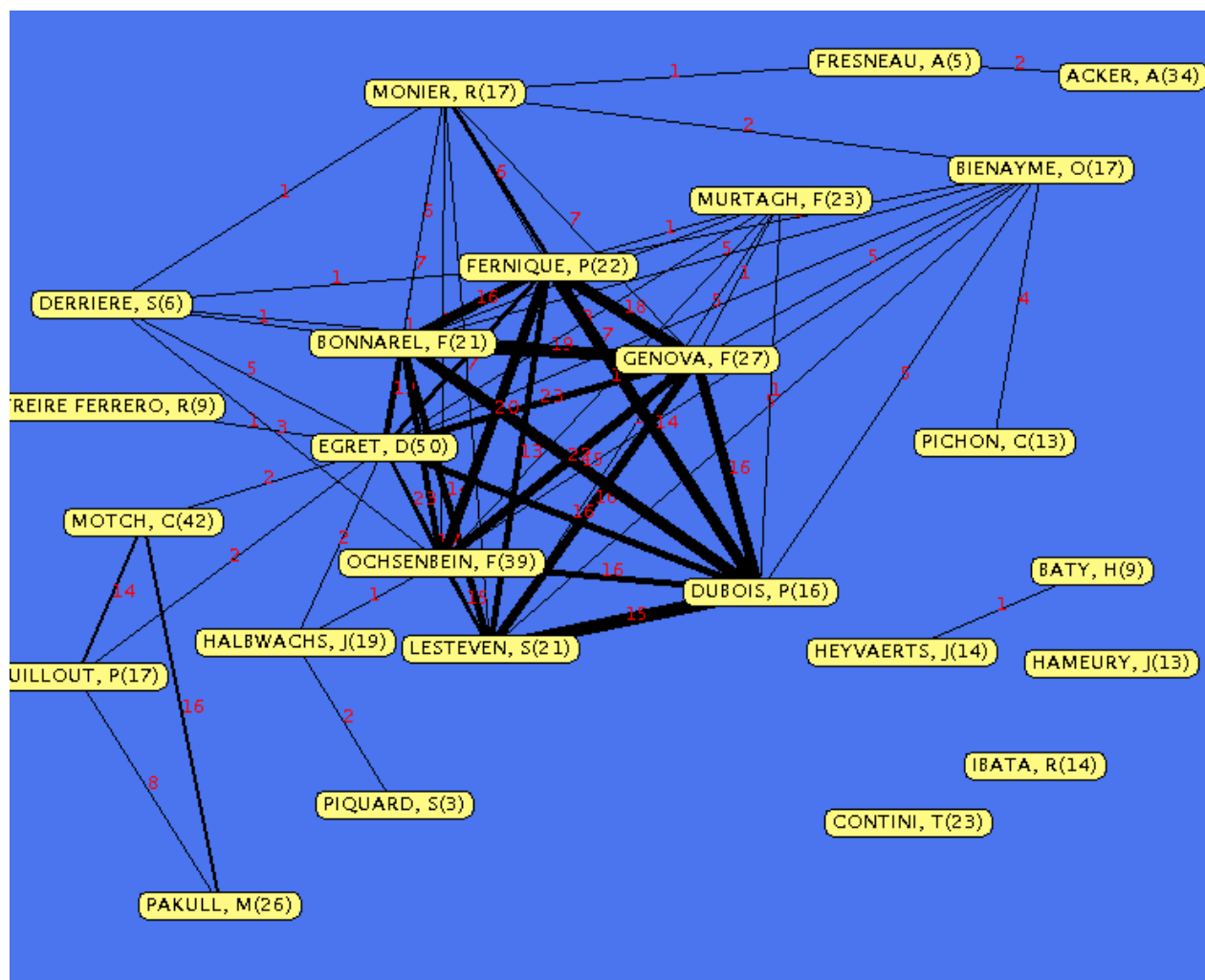


## ■ Classification

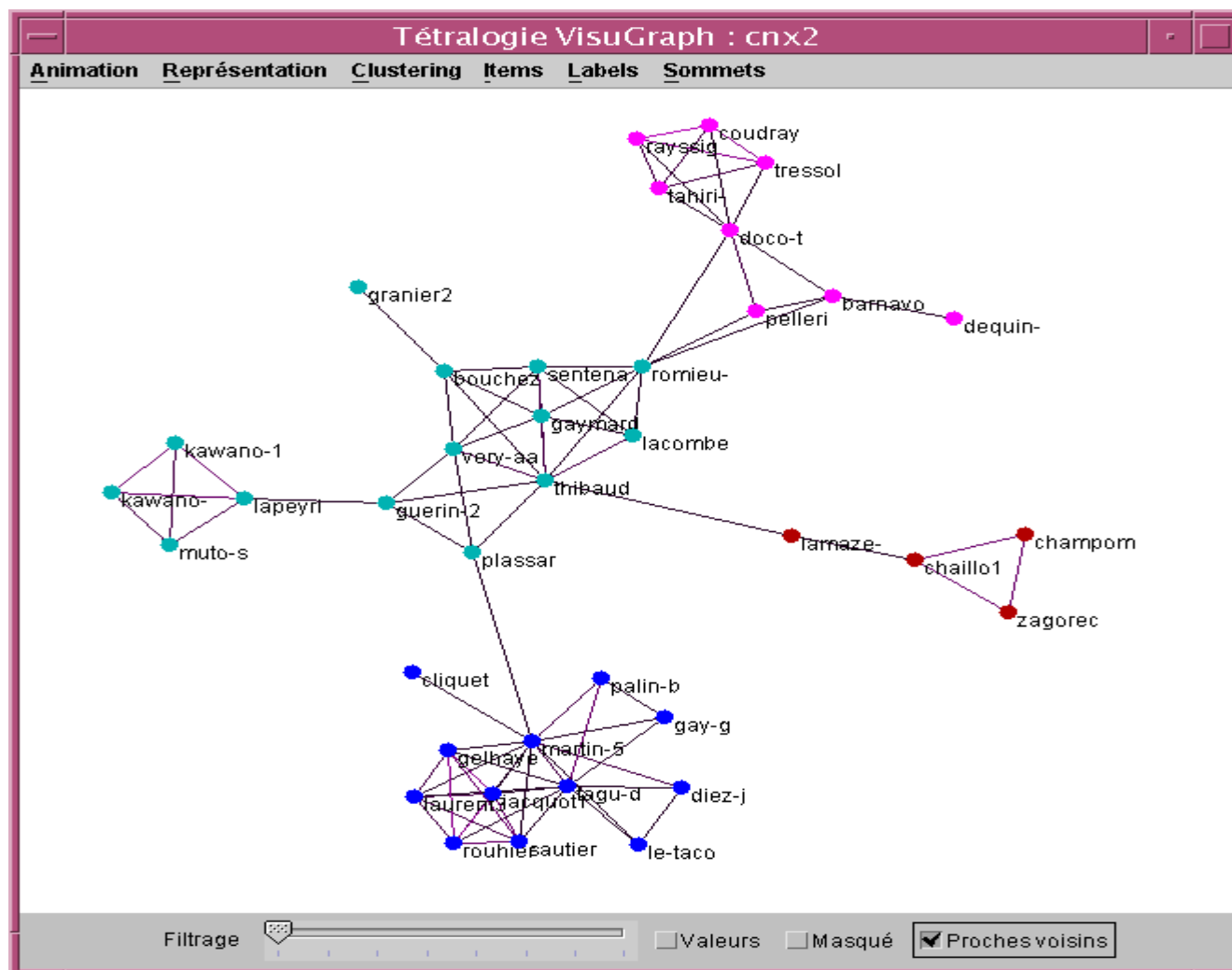




## ■ Lien entre information (Cooc)



## ■ Lien entre information (MCL- partitionnement)



## ■ Détection de corrélations

- Analyse factorielle des correspondances
- Avantages:
  - Donne plus d'importance aux spécificités
  - Représentation graphique utilisant les premiers axes
  - Différents types de croisement d'information

```

hubble-space-telescope
herbig-ae/be-star
monte-carlo-simulation
planetary-nebula
density-distribution
symbiotic-star
stars:early-type
statistical-model
space-telescope
schmidt-telescope
volume-density
late-type-giant
photometric-calibration
halo-population
stellar-sample
photographic-photometry
LATTANZI M.G.
----Emacs: temporaire2 (Fundament

```

```

emacs@pe200atlas.irit.fr
Buffers Files Tools Edit Search Help
colour-magnitude-diagram
star-cluster
proper-motion-star
rotation-curve
galactocentric-distance
ROESER S.
----Emacs: temporaire2

```

```

emacs@pe200atlas.irit.fr
Buffers Files Tools Edit Search
metallicity-gradient
symmetric-drift
density-law
kinematic-data
----Emacs: temporaire2

```

```

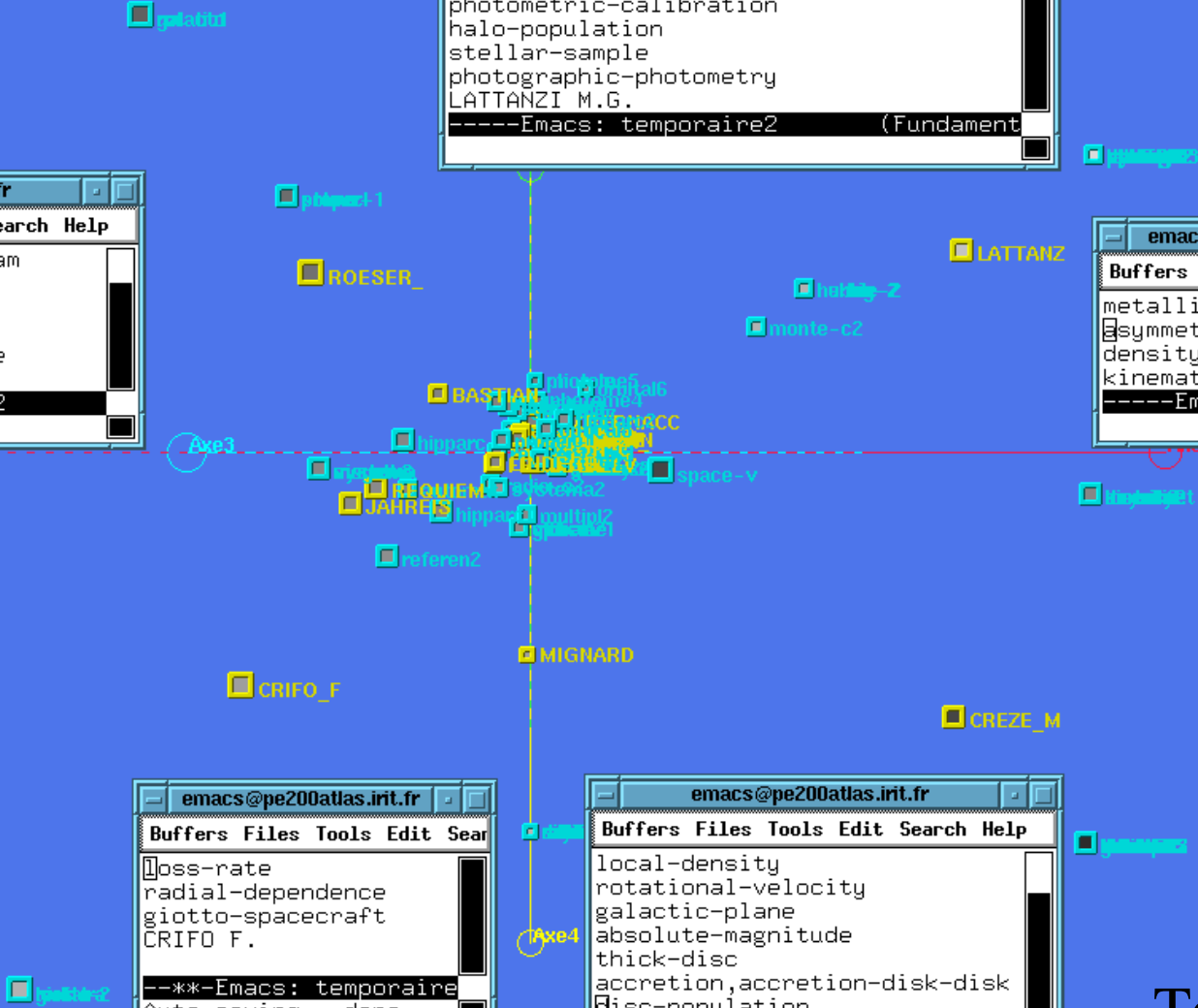
emacs@pe200atlas.irit.fr
Buffers Files Tools Edit Sear
loss-rate
radial-dependence
giotto-spacecraft
CRIFO F.
--*-Emacs: temporaire
Auto-saving...done

```

```

emacs@pe200atlas.irit.fr
Buffers Files Tools Edit Search Help
local-density
rotational-velocity
galactic-plane
absolute-magnitude
thick-disc
accretion, accretion-disk-disk
disc-population
solar-motion

```



# Conclusion

SRI- KD : solution intégrée

## Descriptions de domaine

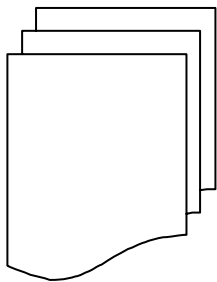


**Ontologie**  
**Hierarchies de concepts**  
**Thésaurus**

INDEXATION



AIDE à LA FORMULATION  
NAVIGATION



EXPLORATION

