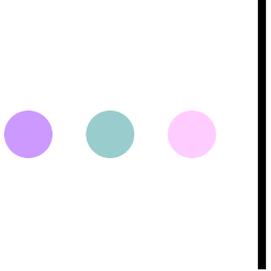




Intégration de données guidée par une ontologie *Application aux sciences du vivant*

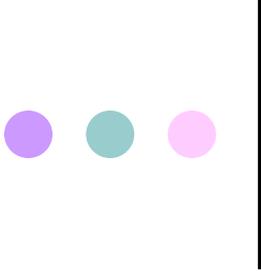
Juliette Dibie





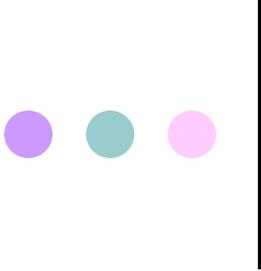
Introduction

- **Intégration de données et IC**
- Contexte
- Application à différents domaines



Intégration de données et IC

- L'ingénierie des connaissances : une thématique multidisciplinaire de l'Intelligence Artificielle
 - Acquisition et représentation de connaissances
 - Modélisation de connaissances
 - Intégration de données
 - Ontologie
 - ...

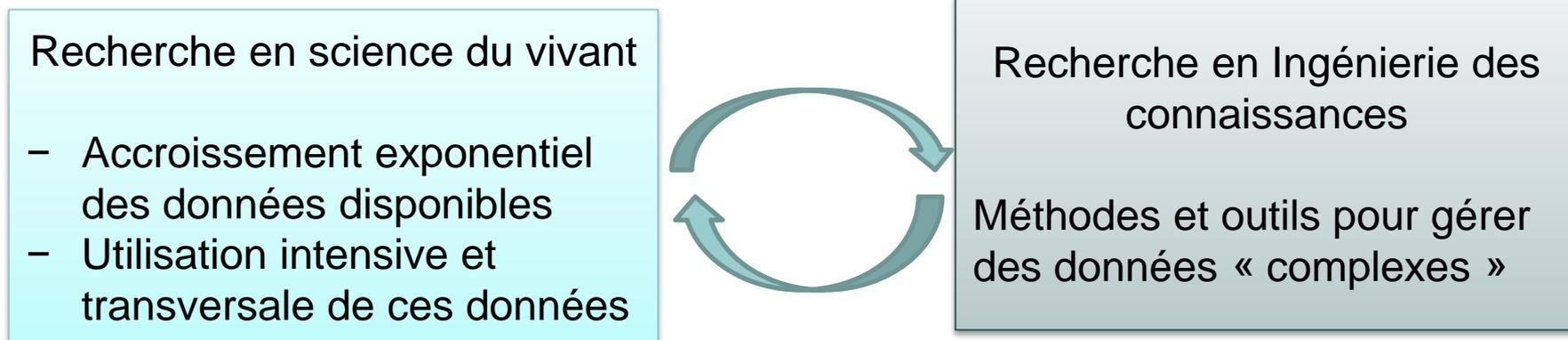


L'Ingénierie des connaissances à l'INRA

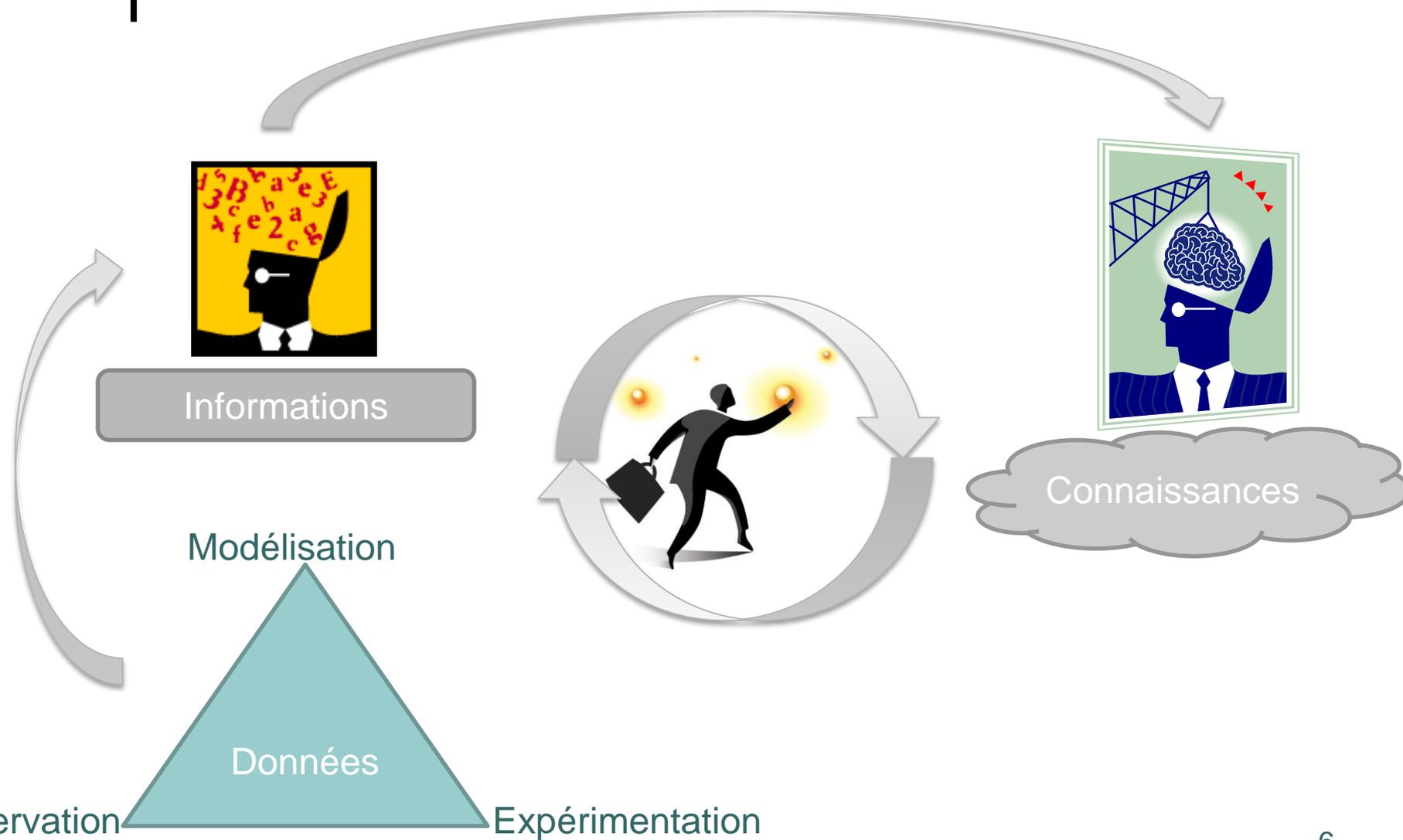
- Comment adapter et/ou développer des méthodes et des outils innovants de capitalisation et de modélisation de données et de connaissances afin de permettre leur exploitation (analyse, traitement, visualisation) et leur partage par des systèmes informatiques

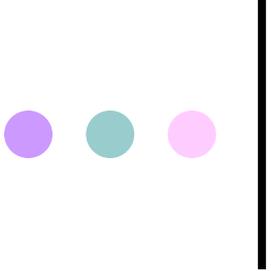
L'Ingénierie des Connaissances à l'INRA

- Les activités de recherche en Ingénierie des connaissances à l'INRA sont guidées par les besoins de recherche en sciences du vivant



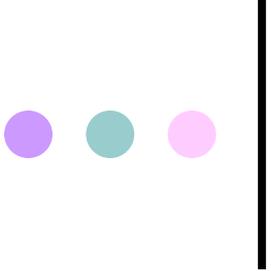
Qu'est-ce qu'une donnée ?





Introduction

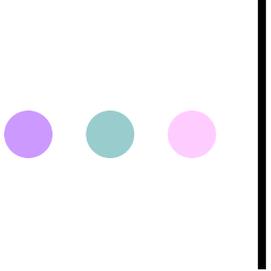
- Intégration de données et IC
- **Contexte**
- Application à différents domaines



Contexte

- Constat: dans de nombreux domaines d'application, les bases de connaissances locales sont **incomplètes**
 - Rareté relative de l'information (coût, confidentialité)
 - Nombre de données potentiellement intéressantes infini
- Idée: exploiter le Web qui est une gigantesque source d'information pour « **compléter** » ces bases
- Problématique: Comment **intégrer** des données pertinentes du Web dans les bases locales ?





Spécificités des données

- **Hétérogénéité**

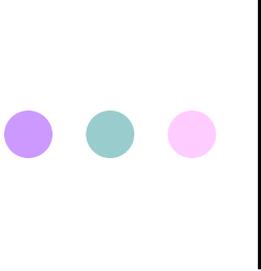
format, structure, vocabulaire

Exemple : paramètres étudiés

- **Imprécision**

notion imprécise, intervalle de valeurs, incertitude, similarités

- Complexité des processus biologiques
- Seuils de détection des capteurs

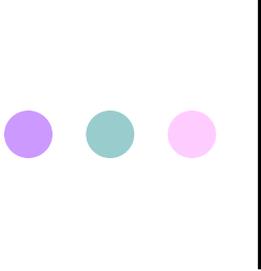


Hypothèses de travail

- Exploitation de la connaissance ontologique contenue dans les bases locales pour guider l'intégration des données du Web
- Exploitation des tableaux de données



**Intégration de données guidée
par une ontologie de domaine**



Notre travail de recherche

- Intégration de données guidée par une ontologie de domaine
 - Ontologie
 - Annotation sémantique
 - Interrogation flexible
 - Publication de données sur le Web



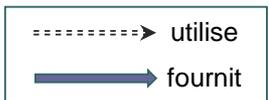
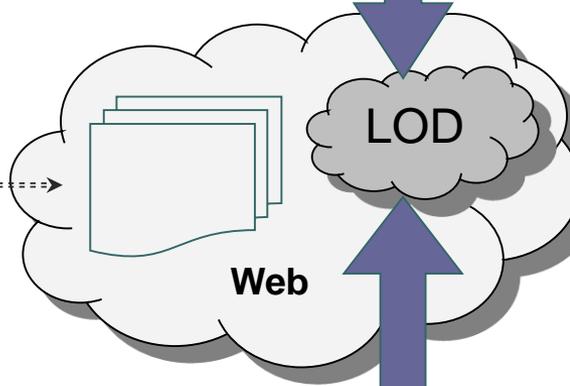
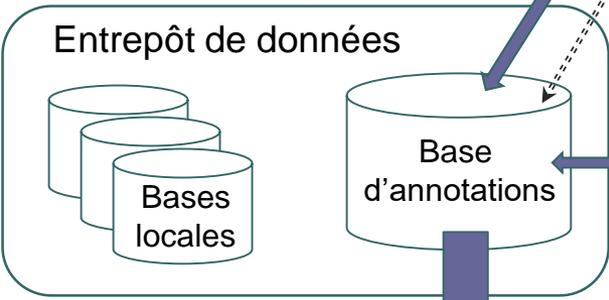
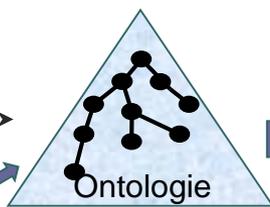
Le système ONDINE

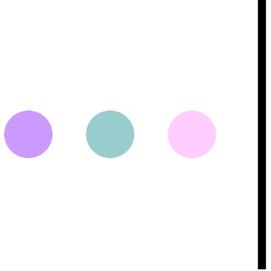
(Ontology based Data INtEgration)



requête

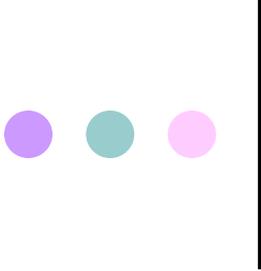
réponse





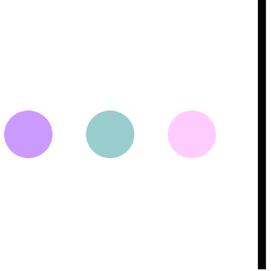
Introduction

- Intégration de données et IC
- Contexte
- **Application à différents domaines**



Travail de recherche appliqué

- Application au domaine du risque alimentaire
 - Analyse des risques microbiologiques dans les aliments
 - Analyse des risques chimiques dans les aliments
- Emballage alimentaire
- Bio raffinerie
- Blé dur



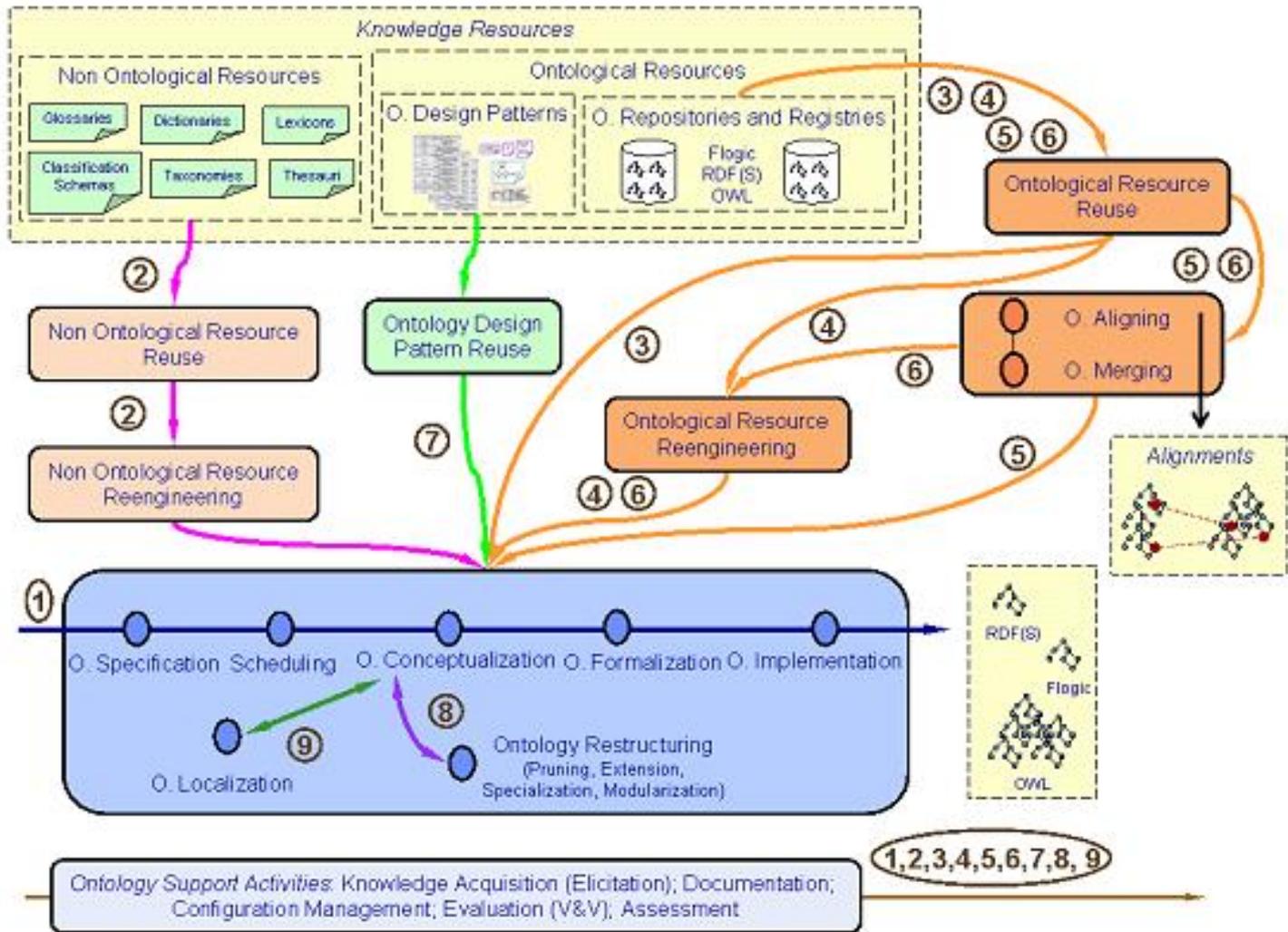
Sommaire

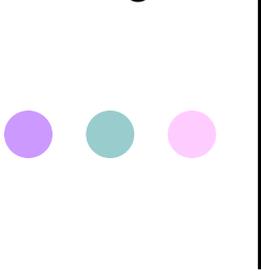
- **Ontologie: construction et modélisation**
- Annotation sémantique
- Interrogation flexible
- Evolution d'ontologie
- Alignements d'ontologies et LOD
- Conclusion et perspectives

Qu'est-ce qu'une ontologie?

- A quoi ça sert ?
 - Définir une « sémantique » d'un domaine du monde réel fondée sur un consensus entre les experts du domaine
 - Fournir une sémantique afin de :
 - Définir un vocabulaire commun, consensuel et partagé
 - Permettre son exploitation par une machine
 - Garantir une interopérabilité entre les données du domaine
- Qu'est-ce que c'est ?
 - Un ensemble de :
 - Concepts : objets, entités, attributs, états, événements, principes, idées, notions abstraites
 - Relations entre concepts
 - Règles et axiomes qui permettent le raisonnement

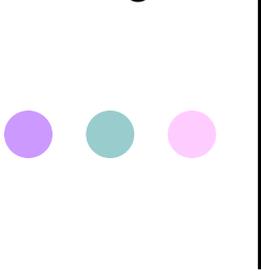
Construire une ontologie: la méthodologie NEON





Construire une ontologie

- Plusieurs étapes :
 1. Spécification de l'ontologie : quel domaine ? Pour qui ? Pour quoi faire ?
 2. Identification des éléments de l'ontologie et de leur définition
 3. Conceptualisation
 4. Intégration d'ontologies existantes
 5. Implémentation
 6. Evaluation
 7. Documentation



Identification des éléments de l'ontologie et de leur définition

- Techniques d'acquisition de connaissances
 - Réunions de brainstorming
 - Interviews d'experts
 - Analyses de documents
- Plusieurs réunions plénières afin de modéliser les connaissances avec les experts du domaine

Définition une RTO dédiée à la tâche d'annotation

- Besoin de définir une ressource termino-ontologique (RTO) dédiée à la tâche d'annotation sémantique de relations n-aire dans les tableaux de données



Exemple de tableau à annoter

Table 1: Permeabilities of MFC films and literature values for films of synthetic polymers and cellophane

Sample	Grammage (g/m ²)	Thickness (μm)	Air permeability (nm/Pa s)	Oxygen permeability in the material (ml m ⁻² day ⁻¹)
MFC film A	17 ±1	21 ±1	13 ± 2	17.0, 18.5
EVOH	–	25	–	3–5
Cellophane	–	21	–	3
Symbolic concept Packaging	Symbolic concept Ethylene Vinyl Alcohol	Quantity Thickness		Quantity O2_Permeability

Note: An arrow points from the 'Symbolic concept Packaging' cell to the 'EVOH' cell.

O2_Permeability relation

Un exemple de RTO



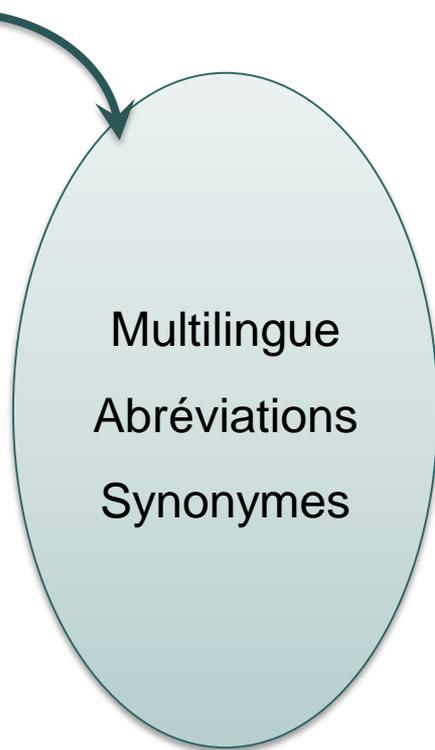
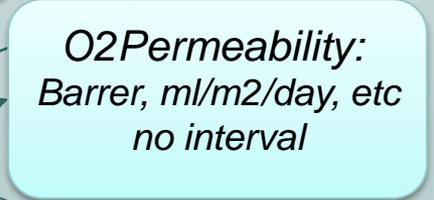
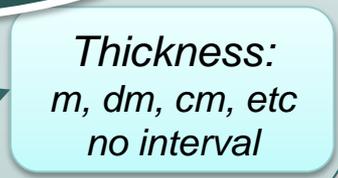
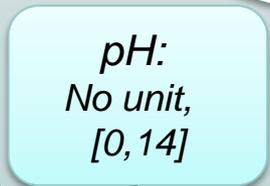
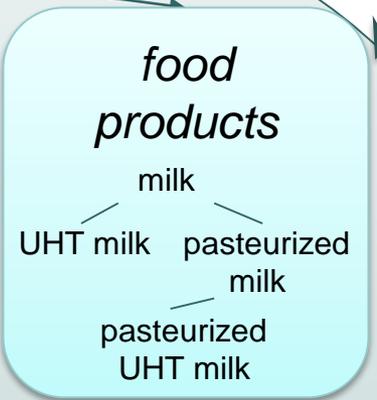
symbolic concept

Relation:

Signature:

- result concept
(range)

- Access concep
(domain)

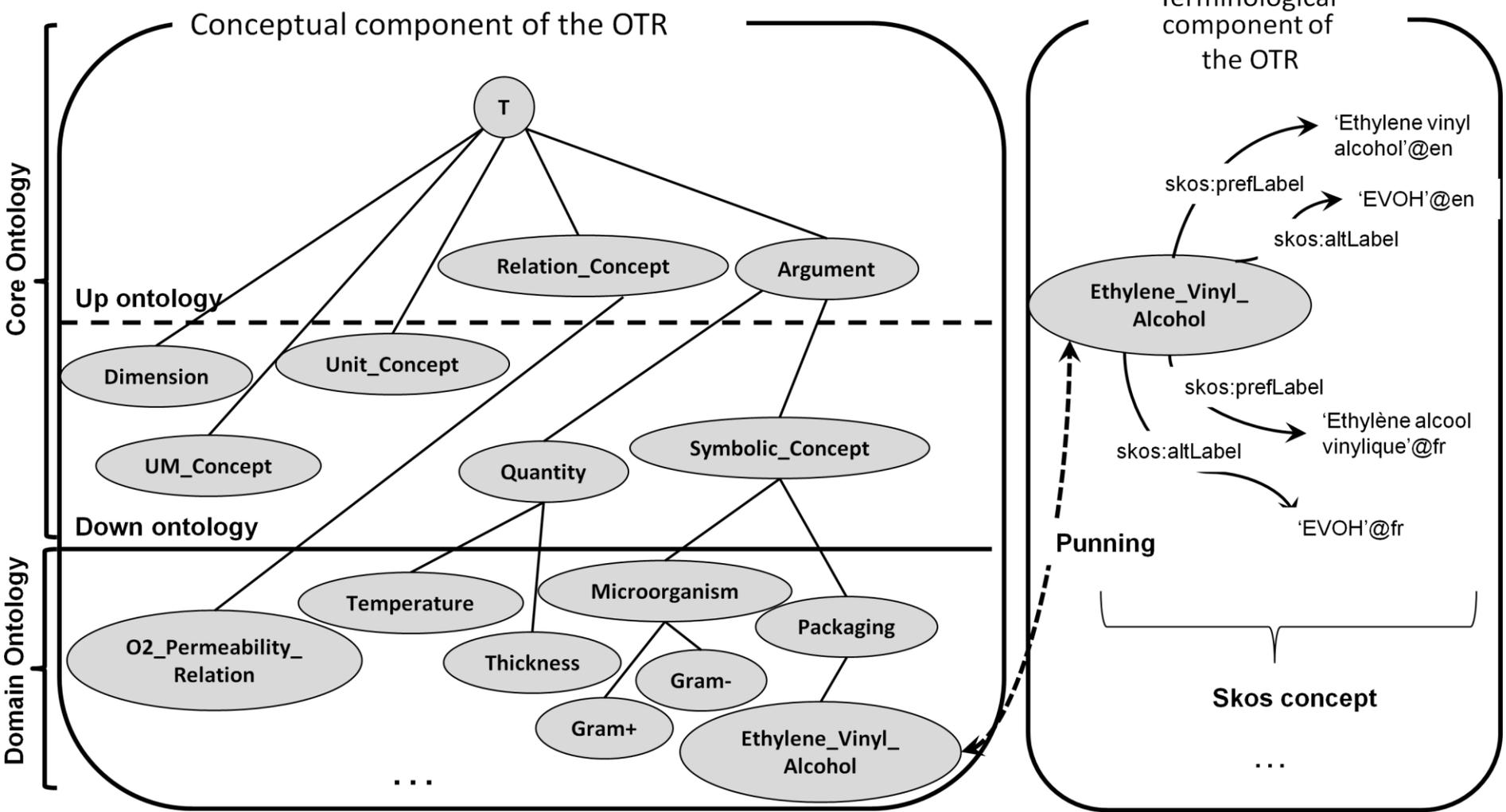


Ontologie

Terminologie

Quantity

Modélisation de la RTO



La composante conceptuelle de la RTO

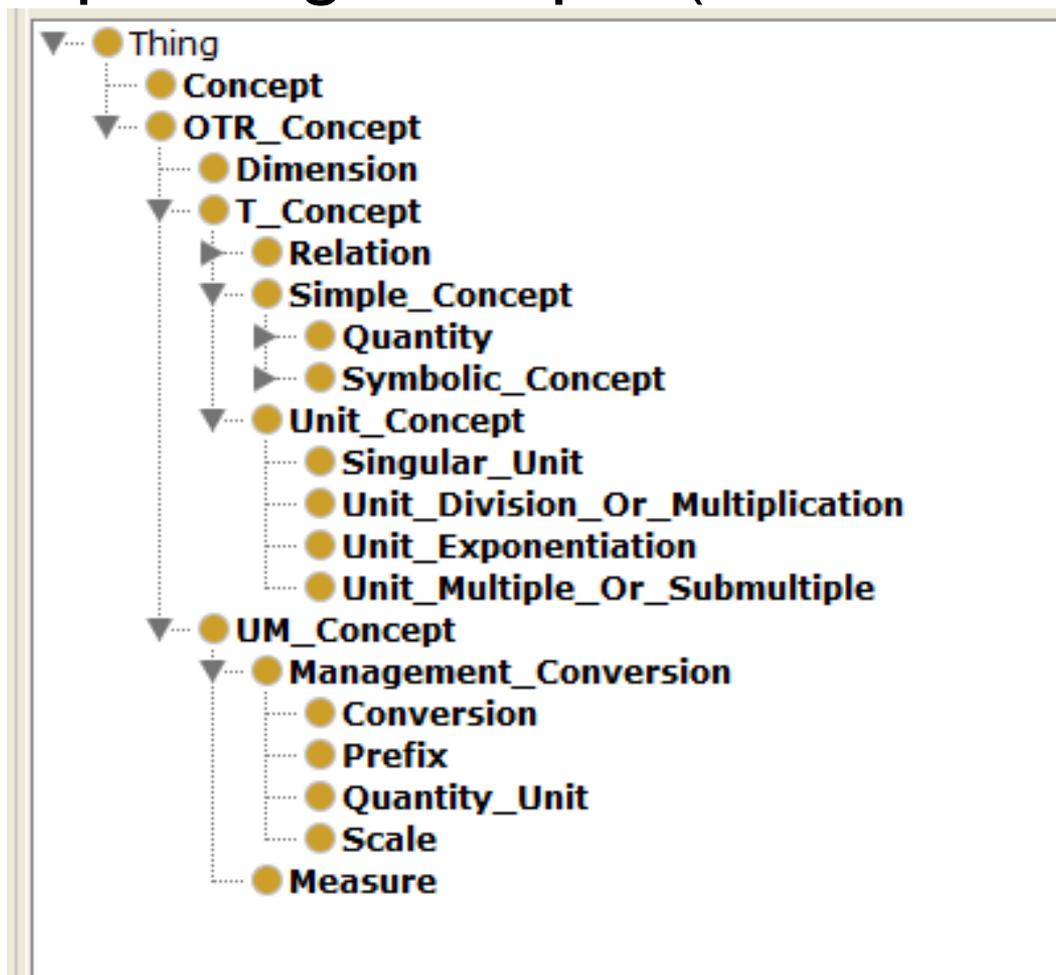
- La composante conceptuelle de la RTO est composée de deux parties :
- Une partie générique (Core ontology) :
 - Représente la structure de l'ontologie.
 - Dédiée à la tâche d'annotation des relations n-aires dans les tableaux du Web.
- Une partie spécifique (Domain ontology):
 - Elle dépend du domaine d'intérêt



RTO générique : elle permet l'instanciation des relations n-aires dans n'importe quelle application.

La composante conceptuelle de la RTO

- La partie générique (Core ontology) :

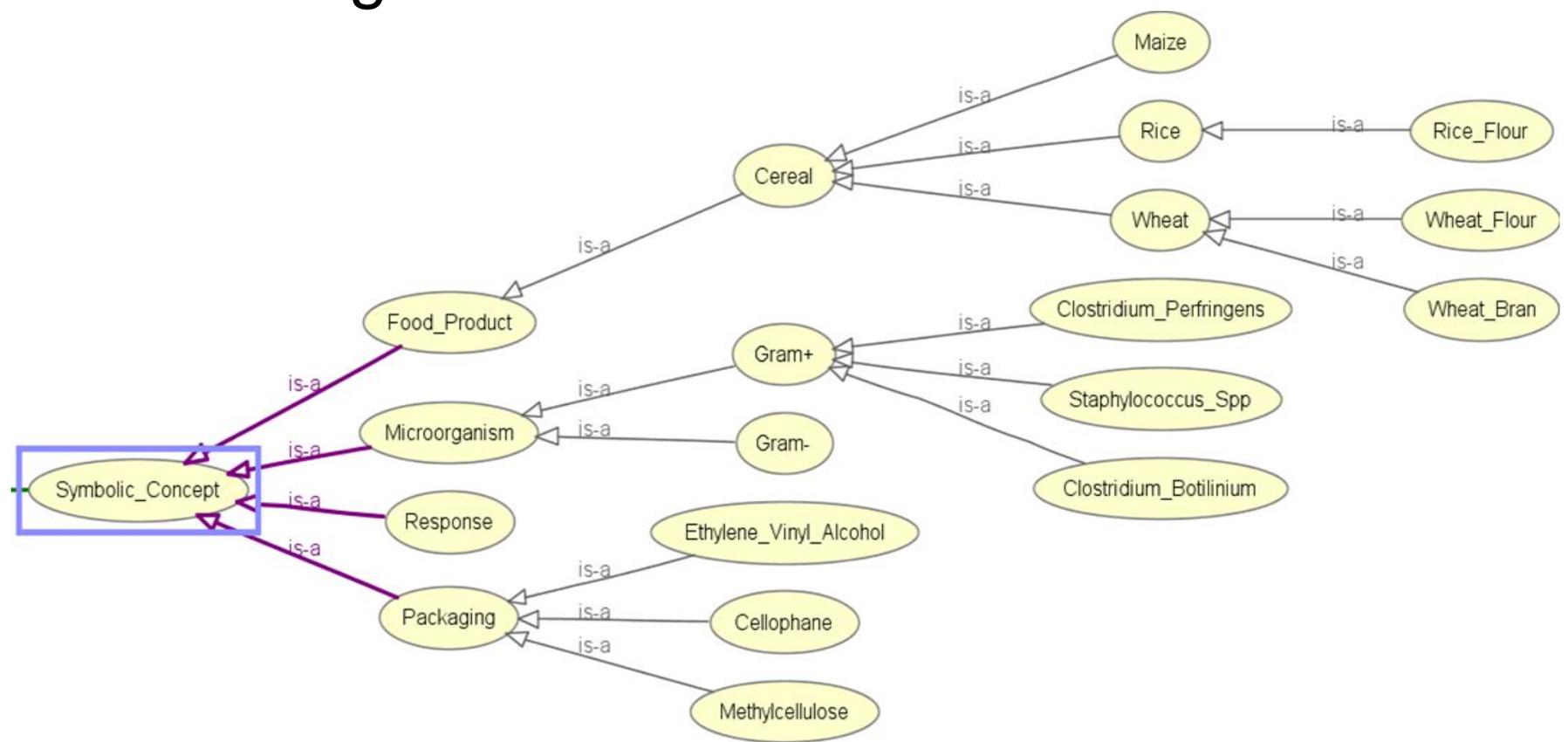


La composante conceptuelle de la RTO

- La partie spécifique permet de représenter les concepts spécifiques à un domaine d'application donné
- Elle est composée des sous-concepts des concepts génériques
 - Tous les concepts sont représentés par des classes OWL
 - Elle sont reliées par des relations de subsomption (subClassOf)
 - Elle sont mutuellement disjointes

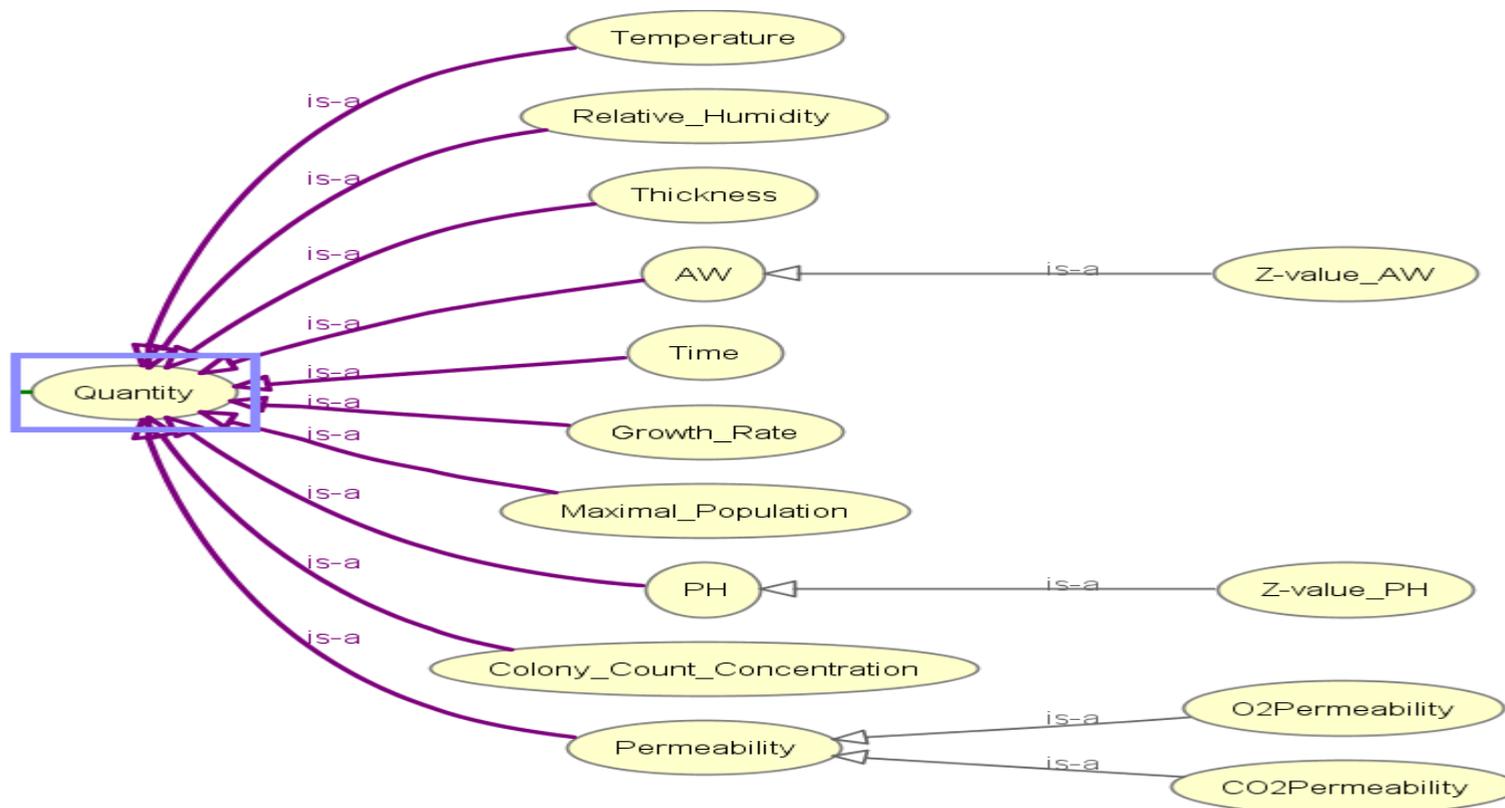
La composante conceptuelle de la RTO

- Les concepts symboliques sont organisées en hiérarchie



La composante conceptuelle de la RTO

- Une quantité est caractérisée par :
 - Un ensemble d'unités de mesure
 - Une dimension
 - Un intervalle de valeurs possible



La composante conceptuelle de la RTO

- Un exemple : la quantité `relative_humidity`

Description: `relative_humidity`

Equivalent To 

SubClass Of 

 **hasNumericalValue** *only* float[$\geq 0.0f$, $\leq 100.0f$]

 **hasUnitConcept** *only* ({Percent})

 **Quantity**

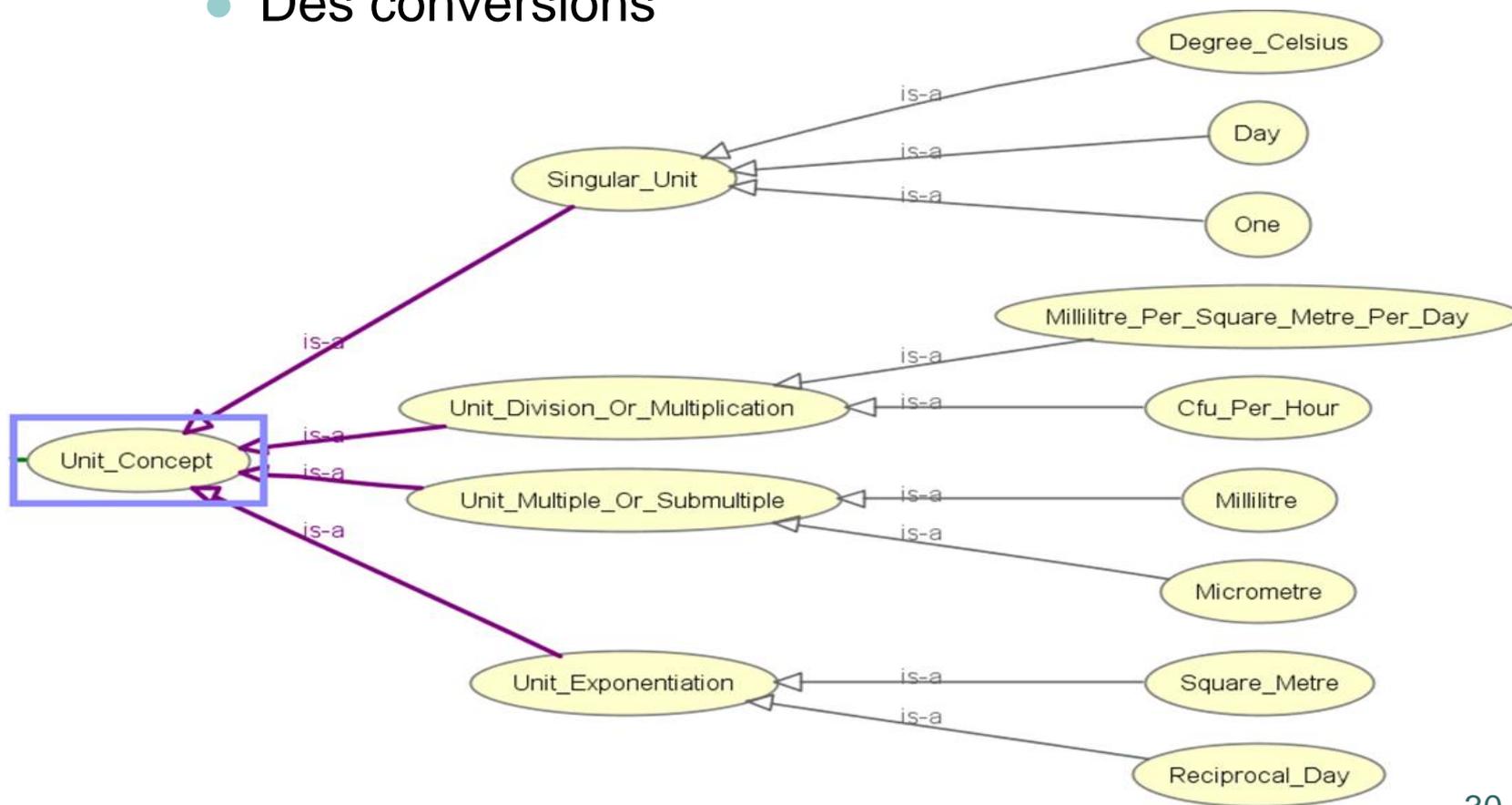
SubClass Of (Anonymous Ancestor)

 **hasDimension** *only* Dimension

 **hasUnitConcept** *min 1* Thing

La composante conceptuelle de la RTO

- Un concept unité est caractérisé par :
 - Une dimension
 - Des conversions



La composante conceptuelle de la RTO

- Un exemple : l'unité Degree_Fahrenheit

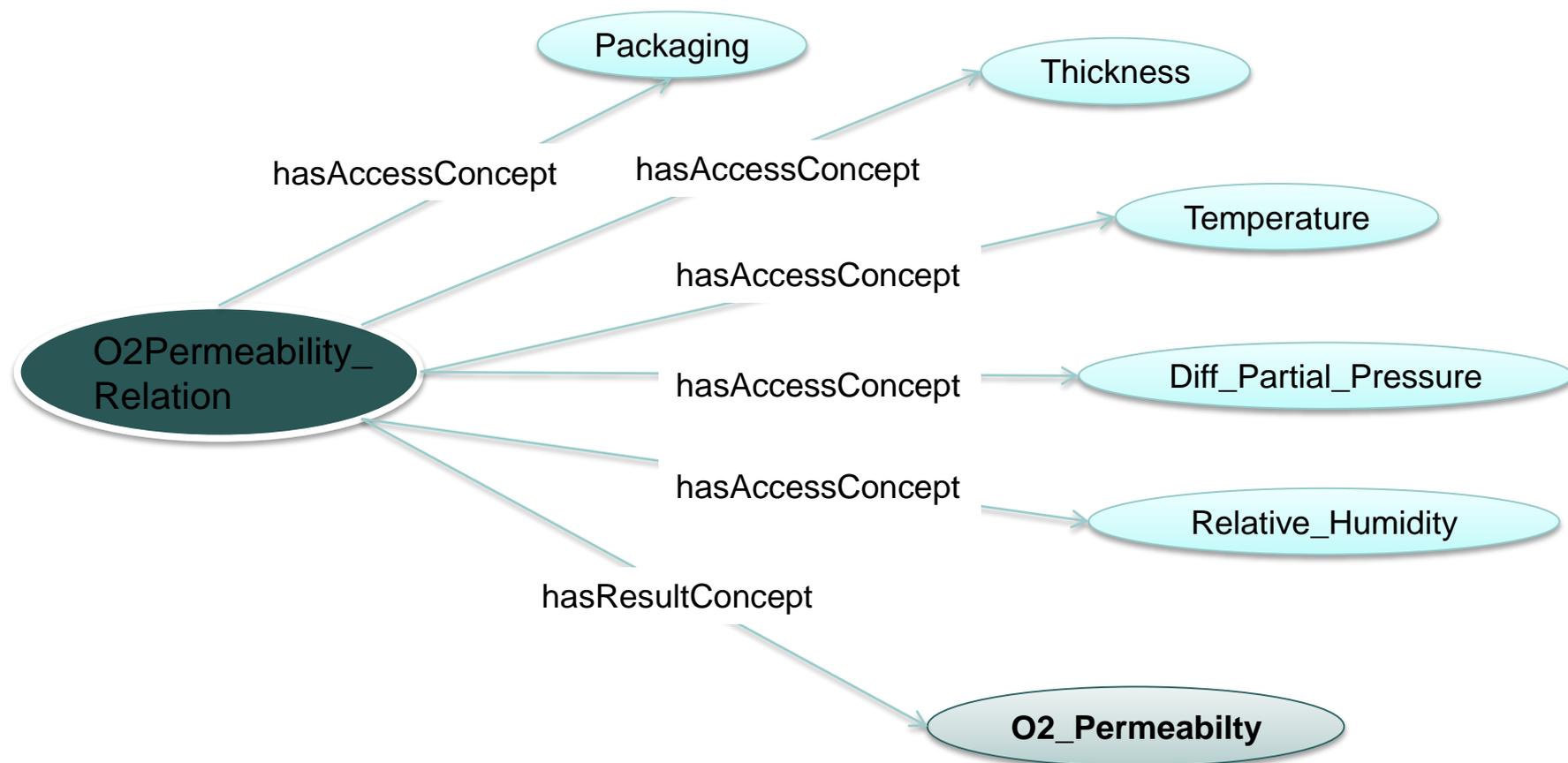
The image displays four panels from an ontology editor:

- Description: Degree_Fahrenheit**: Shows the class hierarchy with **Concept** and **Singular_Unit** as types.
- Property assertions: Degree_Fahrenheit**: Lists object properties including **hasDimension** (thermodynamic_temperature-dimensic) and **hasConversion** (FahrenheitToCelsius), which is circled in red. It also lists data properties for labels in English and French, and alternative labels like "oF" and "°F".
- Description: FahrenheitToCelsius**: Shows the class **Conversion** as a type.
- Property assertions: FahrenheitToCelsius**: Lists object properties **hasSourceUnit** (Degree_Fahrenheit), **hasScaleOrPrefix** (URI_0.556), and **hasTargetUnit** (Degree_Celsius), along with a data property **hasOffset** (-32.0f).

A red arrow points from the **hasConversion** property in the top-right panel to the **Description: FahrenheitToCelsius** panel, indicating the conceptual link between the two.

La composante conceptuelle de la RTO

- Une relation est caractérisée par :
 - sa signature





La composante conceptuelle de la RTO

- Un exemple de relation: O2_Permeability_relation

Description: o2_permeability_relation

Equivalent To

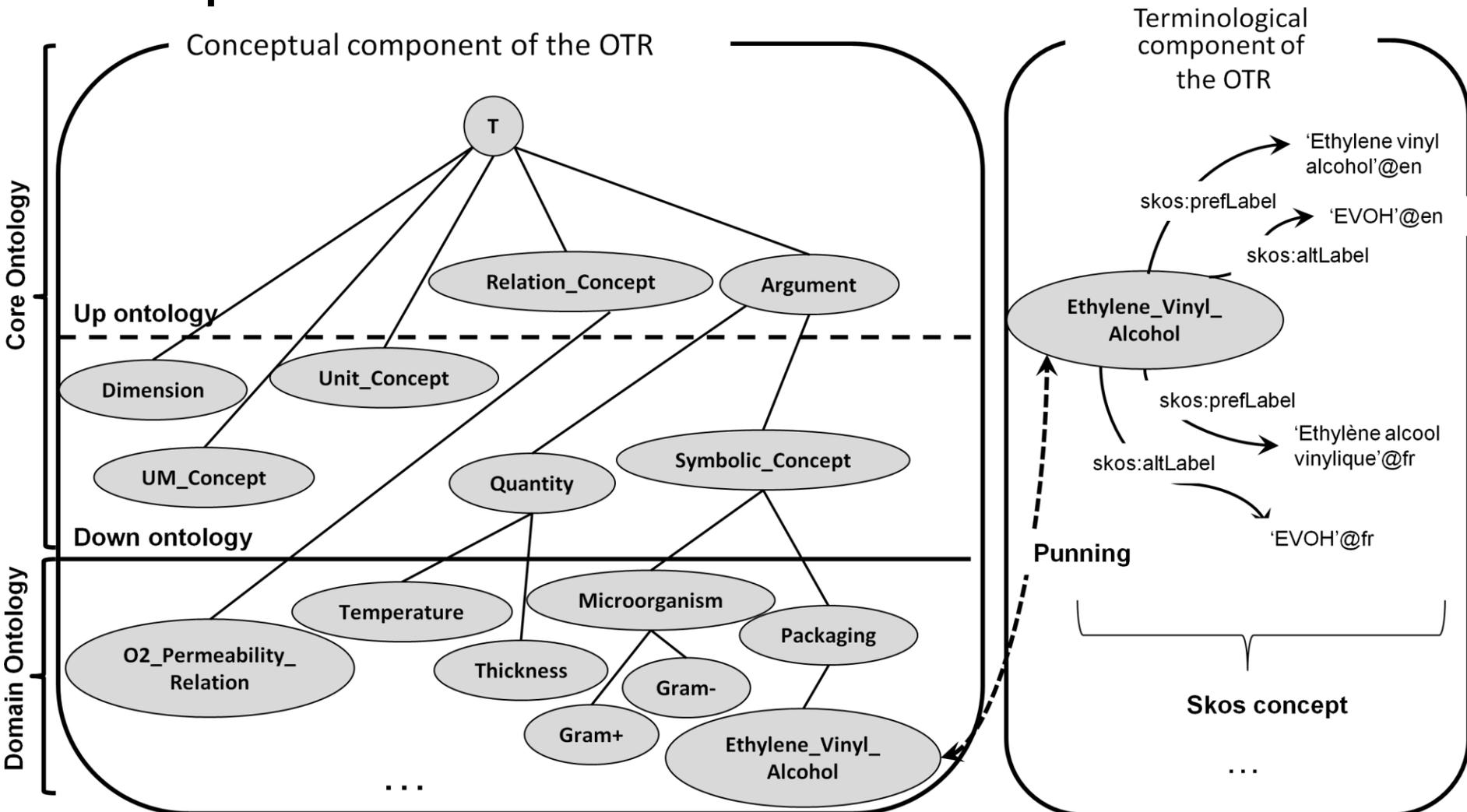
SubClass Of

- hasImportantAccessConcept **some** packaging
- hasImportantAccessConcept **some** partial_pressure_difference
- hasImportantAccessConcept **some** relative_humidity
- hasImportantAccessConcept **some** temperature
- hasImportantAccessConcept **some** thickness
- hasResultConcept **only** o2_permeability
- Relation

SubClass Of (Anonymous Ancestor)

- hasResultConcept **exactly** 1 Thing
- hasAccessConcept **min** 1 Thing

La composante terminologique de la RTO



La composante terminologique de la RTO

- Elle contient l'ensemble des termes associés aux concepts de la partie conceptuelle
- A chaque sous concept du concept générique «T_concept» est associé une partie terminologique :
 - Un seul label préféré par langue
 - Un ou plusieurs labels alternatifs par langue (synonymes et abréviations)



Les labels associés à un concept donné sont utilisés pour l'annotation sémantique des tableaux de données

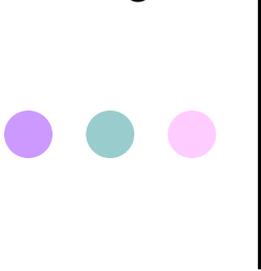
La composante terminologique de la RTO

- Utilisation des propriétés lexicales de skos (Simple Knowledge Organization Scheme) : prefLabel et altLabel.
- Utilisation de la technique du « Punning » (ou métamodélisation) possible en passant à OWL2-DL.

```
<owl:Class rdf:ID="Ethylene Vinyl Alcohol">
  <rdfs:subClassOf rdf:resource="#Packaging"/>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:prefLabel xml:lang="en">Ethylene vinyl alcohol</skos:prefLabel>
  <skos:altLabel xml:lang="en">EVOH</skos:altLabel>
  <skos:prefLabel xml:lang="fr">Ethylène alcool vinylique</skos:prefLabel>
  <skos:altLabel xml:lang="fr">EVOH</skos:altLabel>
</owl:Class>
```

Conclusion sur la modélisation de la RTO

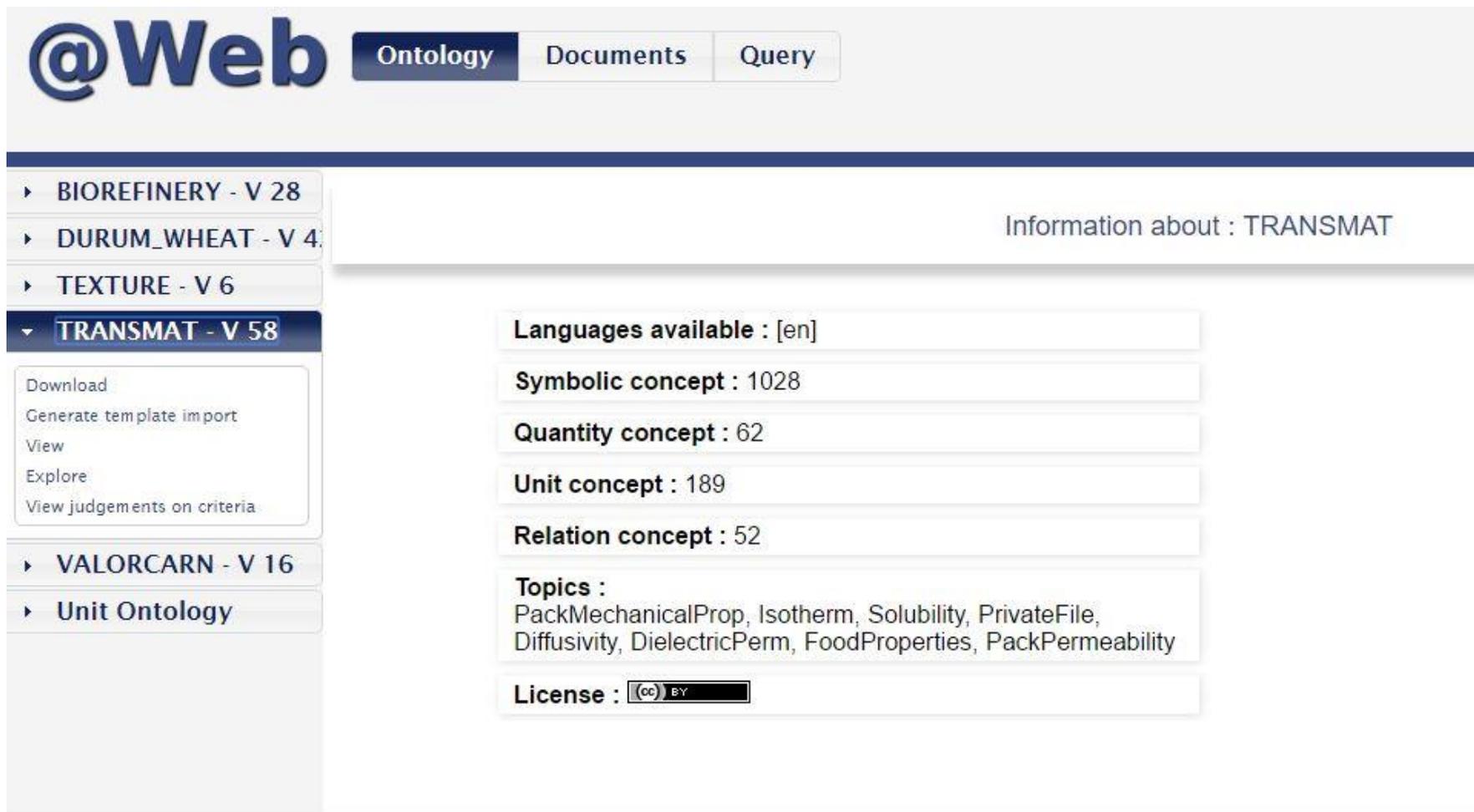
- Modélisation d'une RTO dédiée à la représentation de données expérimentales quantitatives :
 - Définition formelle de la RTO.
 - Définition formelle et représentation générique de relations n-aires.
 - Représentation des unités de mesures ainsi que leurs conversions.
 - Représentation des synonymes et abréviations dans plusieurs langues.
- Utilisation du modèle par le logiciel @Web
<http://pfl.grignon.inra.fr/atWeb/>



Conclusion sur la modélisation de la RTO

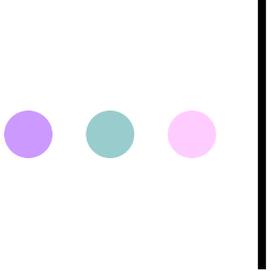
Domaine d'application	Concepts symboliques	Quantités	Relations
TRANSMAT http://agroportal.lirmm.fr/ontologies/TRANSMAT	1028	62	52
Biorefinery http://agroportal.lirmm.fr/ontologies/BIOREFINERY	188	54	32
Wheat_durum http://agroportal.lirmm.fr/ontologies/DURUM_WHEAT	104	8	4

Conclusion sur la modélisation de la RTO



The screenshot displays the @Web Ontology interface. The top navigation bar includes the @Web logo and three tabs: Ontology (selected), Documents, and Query. A left sidebar lists several ontologies: BIOREFINERY - V 28, DURUM_WHEAT - V 4, TEXTURE - V 6, TRANSMAT - V 58 (selected), VALORCARN - V 16, and Unit Ontology. Below the selected ontology, a menu offers actions: Download, Generate template import, View, Explore, and View judgements on criteria. The main content area is titled 'Information about : TRANSMAT' and lists the following statistics:

- Languages available : [en]
- Symbolic concept : 1028
- Quantity concept : 62
- Unit concept : 189
- Relation concept : 52
- Topics : PackMechanicalProp, Isotherm, Solubility, PrivateFile, Diffusivity, DielectricPerm, FoodProperties, PackPermeability
- License : 



Sommaire

- Ontologie: construction et modélisation
- **Annotation sémantique**
- Interrogation flexible
- Evolution d'ontologie
- Alignements d'ontologies et LOD
- Conclusion et perspectives

Le système ONDINE

(Ontology based Data INtEgration)



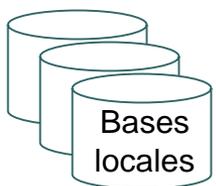
requête

réponse

Interrogation

Interrogation unifié et flexible des
différentes bases de l'entrepôt

Entrepôt de données



Base
d'annotations

Annotation

Annotation sémantique de tableaux
et de texte extraits de documents
scientifiques du Web pour enrichir
des bases locales

Ontologie

Publication

Evolution

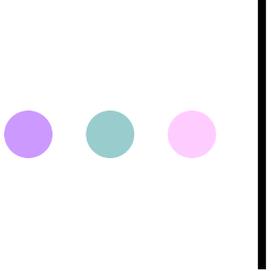
LOD

Web

Publication

-----> utilise

—————> fournit

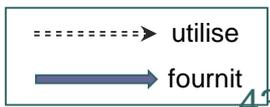
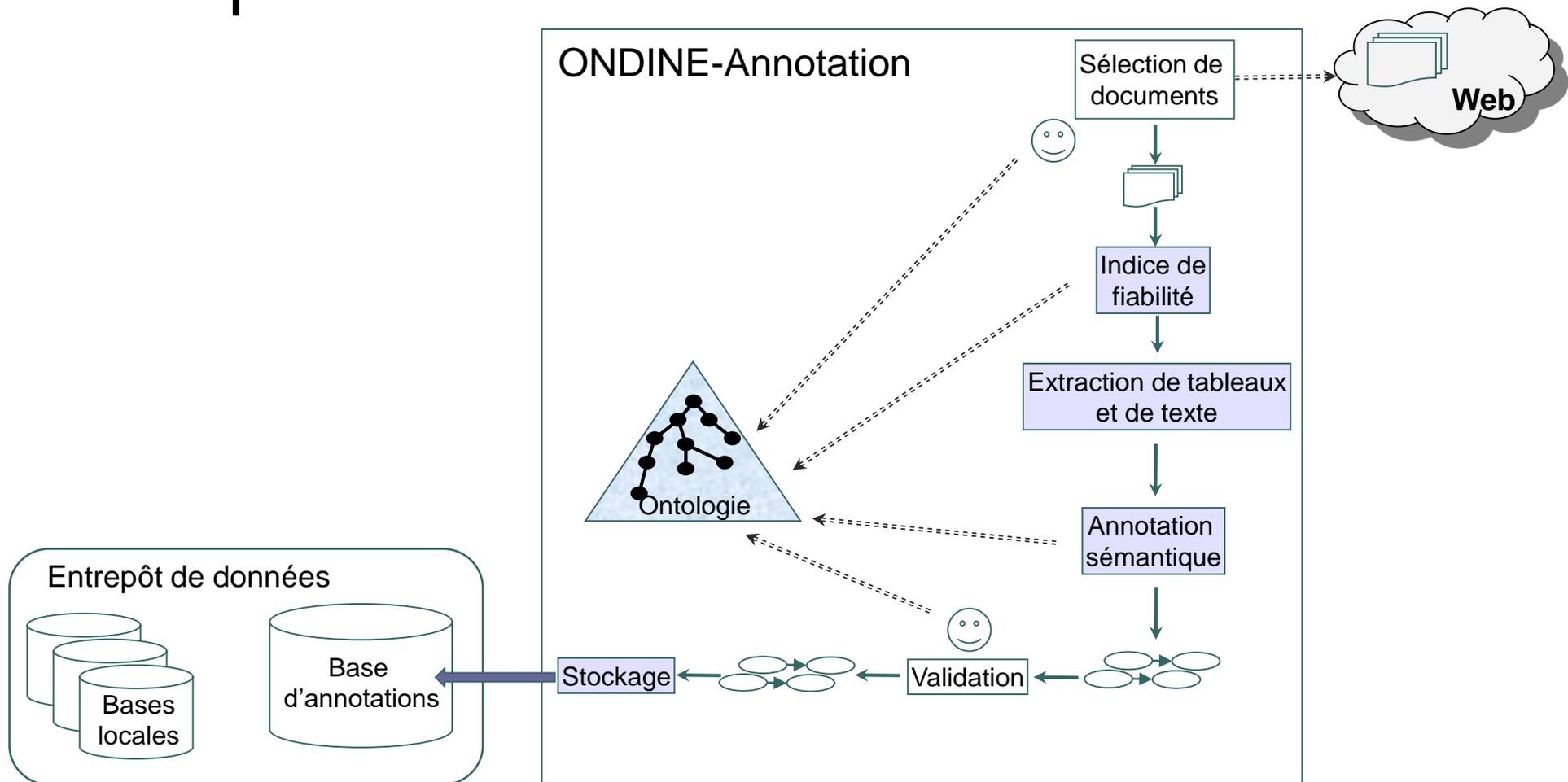


Sommaire

- Ontologie: construction et modélisation
- **Annotation sémantique**
 - **Annotation sémantique d'un tableau**
- Interrogation flexible
- Evolution d'ontologie
- Alignements d'ontologies et LOD
- Conclusion et perspectives

Le sous-système @Web

(Annotating Tables from the Web)





Annotation sémantique d'un tableau

- Thèse de Gaëlle Hignette soutenue en décembre 2007
- Annotation automatique guidée par une ontologie de domaine
- Approche agrégative des éléments du tableau (cellule, colonne, relation entre les colonnes)



Annotation sémantique d'un tableau

1. Distinction des colonnes numériques et symboliques
2. Annotation des colonnes symboliques
3. Annotation des colonnes numériques
4. Reconnaissance des relations



Cellules symboliques ou numériques

- Count the number of marks in the cell
 - numeric indicator: number in scientific notation or number followed by a unit
 - numeric mark : number or unit
 - symbolic mark : word which is nor a unit and nor an indicator of no result (nr, ...)
- Choose the type of the cell
 - numeric
 - More numeric marks than symbolic marks
 - else symbolic
 - Unknown if same number of numeric and symbolic



Cellules symboliques ou numériques

Products	Samples tested	Positive for Campylobacter (%)	Year	Reference
Chilled chicken <i>symb.</i>	22 <i>num.</i>	68.2 <i>num.</i>	1984 <i>num.</i>	Gill and Harris, 1984 <i>symb.</i>
Chicken products <i>symb.</i>	1320 (approx – not specified) <i>unknown</i>	0.07 <i>num.</i>	1992/1994 <i>num.</i>	Campbell and Gilbert, 1995 <i>symb.</i>

symb.

num.

num.

num.

symb.

Annotation sémantique d'un tableau

1. Distinction des colonnes numériques et symboliques
2. Annotation des colonnes symboliques
3. Annotation des colonnes numériques
4. Reconnaissance des relations



Etape 2: Annotation des colonnes symboliques (1)

- Etude du **contenu** de la colonne

Food	Contaminant	Contamination level	Stn
Rice bran <i>Food product</i>	Ochratoxin A <i>Contaminant</i>	1.65-1.95 ng/g	10 000
Chili powder <i>Food product</i>	Ochratoxin A <i>Contaminant</i>	2.34-4.91 ng/g	10 000
Grape raisins <i>Food product</i>	Ochratoxin A <i>Contaminant</i>	0.93-1.20 ng/g	5 000
<i>Food product</i>	<i>Contaminant</i>		



Etape 2: Annotation des colonnes symboliques (2)

- Etude du **titre** de la colonne

Food {(Food product,0.98), (Contaminant,0)}	Contaminant {(Food product,0), (Contaminant,1)}	Contamination level	Stn
Rice bran	Ochratoxin A	1.65-1.95 ng/g	10 000
Chili powder	Ochratoxin A	2.34-4.91 ng/g	10 000
Grape raisins	Ochratoxin A	0.93-1.20 ng/g	5 000
<i>Food product</i>	<i>Contaminant</i>		
<i>Food product</i>	<i>Contaminant</i>		



Annotation sémantique d'un tableau

1. Distinction des colonnes numériques et symboliques
2. Annotation des colonnes symboliques
3. Annotation des colonnes numériques
4. Reconnaissance des relations



Etape 3: Annotation des colonnes numériques (1)

o Etude du contenu de la colonne

Food	Contaminant	Contamination level	Stn
Rice bran	Ochratoxin A	1.65-1.95 ng/g	10 000
Chili powder	Ochratoxin A	2.34-4.91 ng/g	10 000
Grape raisins	Ochratoxin A	0.93-1.20 ng/g	5 000
		<i>Contamination level</i> <i>Loq</i> <i>Lod</i>	<i>Samples total number</i> <i>Samples positive</i>

Nom	Unités	intervalles
Contamination level	ng/g, µg/g, µg/kg, µg/l	[0,10]
Loq	ng/g, µg/g, µg/kg, µg/l	
Lod	ng/g, µg/g, µg/kg, µg/l	
Samples total number	no unit	[0-100 000]
Samples positive	no unit	[0-100 000]

Types
Numériques
de
l'ontologie



Etape 3: Annotation des colonnes numériques (2)

- Etude du **titre** de la colonne

Food	Contaminant	Contamination level {(Contamination level,1), (Loq,0), (Lod,0)}	Stn {{(Samples total number,0), (Samples positive,0)}
Rice bran	Ochratoxin A	1.65-1.95 ng/g	10 000
Chili powder	Ochratoxin A	2.34-4.91 ng/g	10 000
Grape raisins	Ochratoxin A	0.93-1.20 ng/g	5 000
		<i>Contamination level</i> <i>Loq</i> <i>Lod</i>	<i>Samples total number</i> <i>Samples positive</i>

Annotation sémantique d'un tableau

1. Distinction des colonnes numériques et symboliques
2. Annotation des colonnes symboliques
3. Annotation des colonnes numériques
4. **Reconnaissance des relations**
 - Identification des relations
 - Instanciation des relations

Etape 4: Reconnaissance des relations (1)

○ Identification des relations

Food	Contaminant	Contamination level	Stn
Rice bran	Ochratoxin A	1.65-1.95 ng/g	10 000
Chili powder	Ochratoxin A	2.34-4.91 ng/g	10 000
Grape raisins	Ochratoxin A	0.93-1.20 ng/g	5 000
<i>Food product</i>	<i>Contaminant</i>	<i>Contamination level</i>	<i>Attribute</i>

Nom	Type résultat	Types d'accès
Contamination range	Contamination level	Food product, Contaminant, Samples total number
Loq Relation	Loq	Food product, Contaminant, Samples total number
Lod Relation	Lod	Food product, Contaminant, Samples total number
Number Samples positive	Samples positive	Food product, Contaminant, Samples total number

Relations de l'ontologie

Etape 4: Reconnaissance des relations (1)

○ Identification des relations

Food	Contaminant	Contamination level	Stn
Rice bran	Ochratoxin A	1.65-1.95 ng/g	10 000
Chili powder	Ochratoxin A	2.34-4.91 ng/g	10 000
Grape raisins	Ochratoxin A	0.93-1.20 ng/g	5 000
<i>Food product</i>	<i>Contaminant</i>	<i>Contamination level</i>	<i>Attribute</i>

Nom	Type résultat	Types d'accès
Contamination range	Contamination level	Food product, Contaminant, Samples total number

Pas de titre

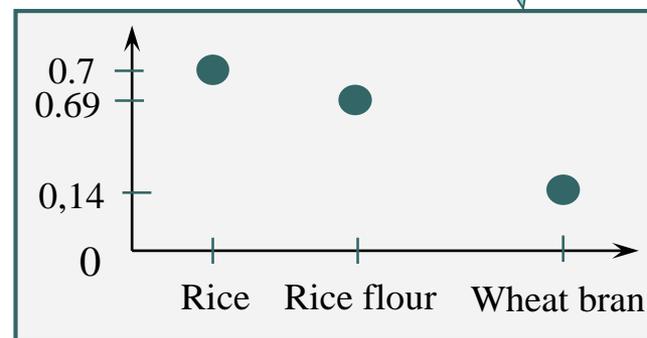
degré de pertinence(Contamination Range)=3/4

Etape 4: Reconnaissance des relations (2)

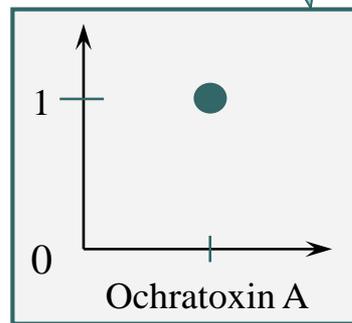
○ Instanciation des relations

Food	Contaminant	Contamination level	Stn
Rice bran	Ochratoxin A	1.65-1.95 ng/g	10 000
Chili powder	Ochratoxin A	2.34-4.91 ng/g	10 000
Grape raisins	Ochratoxin A	0.93-1.20 ng/g	5 000

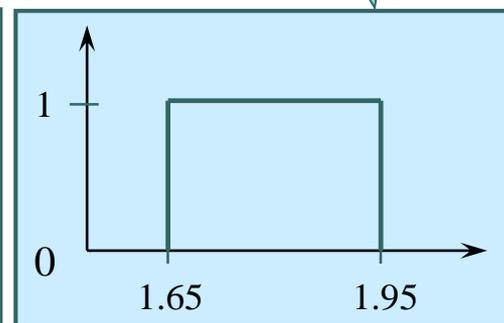
Food products *Contaminant* *Contamination level* *Attribute*



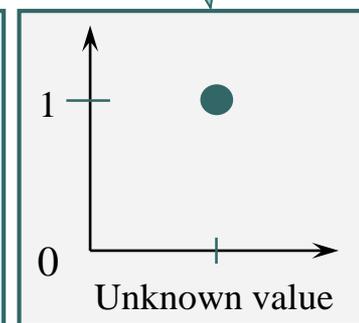
Similarité



Similarité



Imprécision



Similarité

Choix du modèle des sous-ensembles flous

- Pour représenter les données imprécises
 - Expérimentations répétées
 - Seuils de détection des capteurs
 - Connaissances exprimées de manière floue

« *Decimal reduction time of *Listeria monocytogenes* in skimmed milk at 63.3°C is between [25.8, 28.2] (in seconds)* »

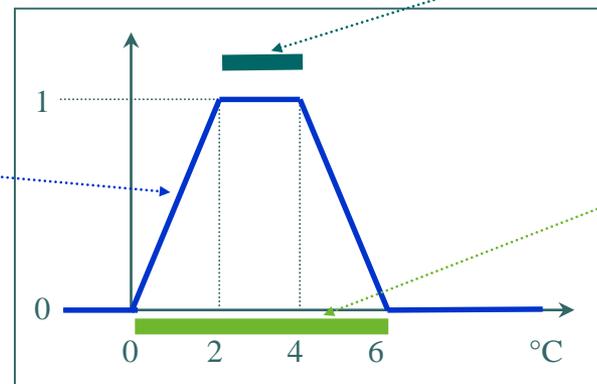
« *The bacterium *Escherichia Coli* is suspected to be responsible, but the bacterium *Shigella* is not excluded* »

- Pour représenter les données annotées

Choix du modèle des sous-ensembles flous

- Assouplissement de la notion de sous-ensemble d'un ensemble (classique)
- Degré d'appartenance d'un élément au sous-ensemble
 - classique : {vrai, faux}
 - flou : $[0, 1]$

fonction d'appartenance



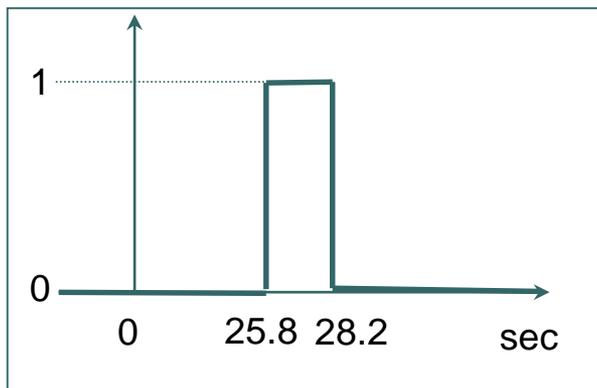
noyau

support

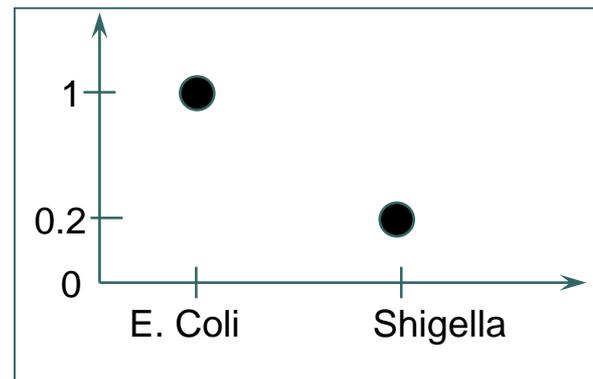
Température Réfrigération

Choix du modèle des sous-ensembles flous

- Une distribution de possibilités représentée, sous la forme d'un sous-ensemble flou, un ensemble priorisé de valeurs vraisemblables mutuellement exclusives (Zadeh 1978).



DecimalReductionTime
Continuous Fuzzy Set



ResponsibleBacterium
Discrete Fuzzy Set



Choix du modèle des sous-ensembles flous

- Expressivité (valeur précise, intervalle, valeur floue, ...)
- Homogénéité données/requêtes (nécessaire pour l'interrogation)
- Étudié dans des travaux antérieurs

Conclusion sur l'annotation sémantique

- Intégration de tableaux du Web : solution à l'incomplétude des bases locales
- Annotation des tableaux : solution à l'hétérogénéité des données locales et des données du Web
- Prise en compte des différentes sémantiques du flou dans l'annotation :
 - données imprécises,
 - similarités entre termes du Web et termes de l'ontologie,
 - degrés de pertinence associés aux relations identifiées pour annoter un tableau



Conclusion sur l'annotation sémantique

- Utilisation des standards du Web sémantique pour représenter les tableaux (en XML) et leurs annotations (en RDF)
- Implémentation : le système @Web
- Expérimentations dans plusieurs domaines différents



Références bibliographiques

- Patrice Buche, Juliette Dibie-Barthelemy, Liliana Ibanescu and Lydie Soler (2013). Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource. TKDE - IEEE Transactions on Knowledge and Data Engineering, vol. 25(4), IEEE Computer Society, pp. 805-819
- 8. G. Hignette, P. Buche, O. Couvert, J. Dibie-Barthélemy, D. Doussot, O. Haemmerlé, E. Mettler, L. Soler (2008). Semantic annotation of Web data applied to risk in food. International Journal of Food Microbiology, vol. 128(1), Elsevier, pp. 174-180.
- P. Buche, J. Dibie-Barthélemy, O. Haemmerlé and G. Hignette (2006). Fuzzy semantic tagging and flexible querying of XML documents extracted from the Web. Journal of Intelligent Information Systems 26(1), 2006, Kluwer, pp 25-40.

Logiciel @Web

Sélection des concepts relations de la RTO TRANSMAT pertinents pour annoter le tableau d'origine figurant dans la partie supérieure de la fenêtre

Manual Annotation of Table 2. Central Composite Design Arrangement and Responses

Original table

variable	levels	responses
T(°C)	RH (%)	CO ₂
9	14.6	258
39	14.6	314
9	85.3	1147
39	85.3	2235
3	50	317
45	50	1026
24	0	86

Annotated table

- Destruction kinetics
- Growth kinetics
- MIC relation
- Max population relation
- O₂ Permeability_relation
- Product properties AW
- Product properties PH

O₂ Permeability_relation ▾
optional | important | result

name	Thickness	Temperature	Relative_Humidity	Packaging	Partial pressure	CO ₂ Permeability	O ₂ Permeability
CO ₂ Permeability_Relation	/	/	/	/	/		
O ₂ Permeability relation	/	/	/	/	/		

Ok

Cancel



Logiciel @Web

Le tableau annoté avec la RTO TRANSMAT

Manual Annotation of Table 2. Central Composite Design Arrangement and Responses

Original table

variable	levels	responses		
T(°C)	RH (%)	CO ₂ permeability (amols ⁻¹ m ⁻¹ Pa ⁻¹)	O ₂ permeability (amols ⁻¹ m ⁻¹ Pa ⁻¹)	select-ivity
9	14.6	258	111	2.3
39	14.6	314	131	2.4
9	85.3	11475	1011	11.3
39	85.3	22353	863	25.9
3	50	317	181	1.7
45	50	1026	233	4.4
24	0	88	77	1.1

Annotated table

n°	Temperature Unit : °C	Thickness Unit : µm	Partial pressure Unit : Pa	CO ₂ Permeability Unit : mol/m/s/Pa	Relative_Humidity Unit : %	Packaging	O ₂ Permeability Unit : mol/m/s/Pa
1	9	[7.700e+1 ; 8.300e+1]		2.580e-16	14.6	Proteins	1.110e-16
2	39	[7.700e+1 ; 8.300e+1]		3.140e-16	14.6	Proteins	1.310e-16
3	9	[7.700e+1 ; 8.300e+1]		1.148e-14	85.3	Proteins	1.011e-15
4	39	[7.700e+1 ; 8.300e+1]		2.235e-14	85.3	Proteins	8.630e-16
5	3	[7.700e+1 ; 8.300e+1]		3.170e-16	50	Proteins	1.810e-16
6	45	[7.700e+1 ; 8.300e+1]		1.026e-15	50	Proteins	2.330e-16
7	24	[7.700e+1 ; 8.300e+1]		8.800e-17	0	Proteins	7.700e-17

Logiciel @Web

Annotation de la première ligne du tableau
d'origine avec la RTO TRANSMAT



Manual Annotation of Table 2 Central Composite Design Arrangement a

Original value :

Type :

Concept :

- Packaging
 - Multilayers Structures
 - Biopolymers
 - Polysaccharides
 - Proteins**
 - Biopolyesters
 - Rubbers
 - Thermoplastic Alloys
 - Thermoplastic Elastomers
 - Thermoplastic
 - Thermosets

Selected Concept :

PrefLabel : AltLabel :

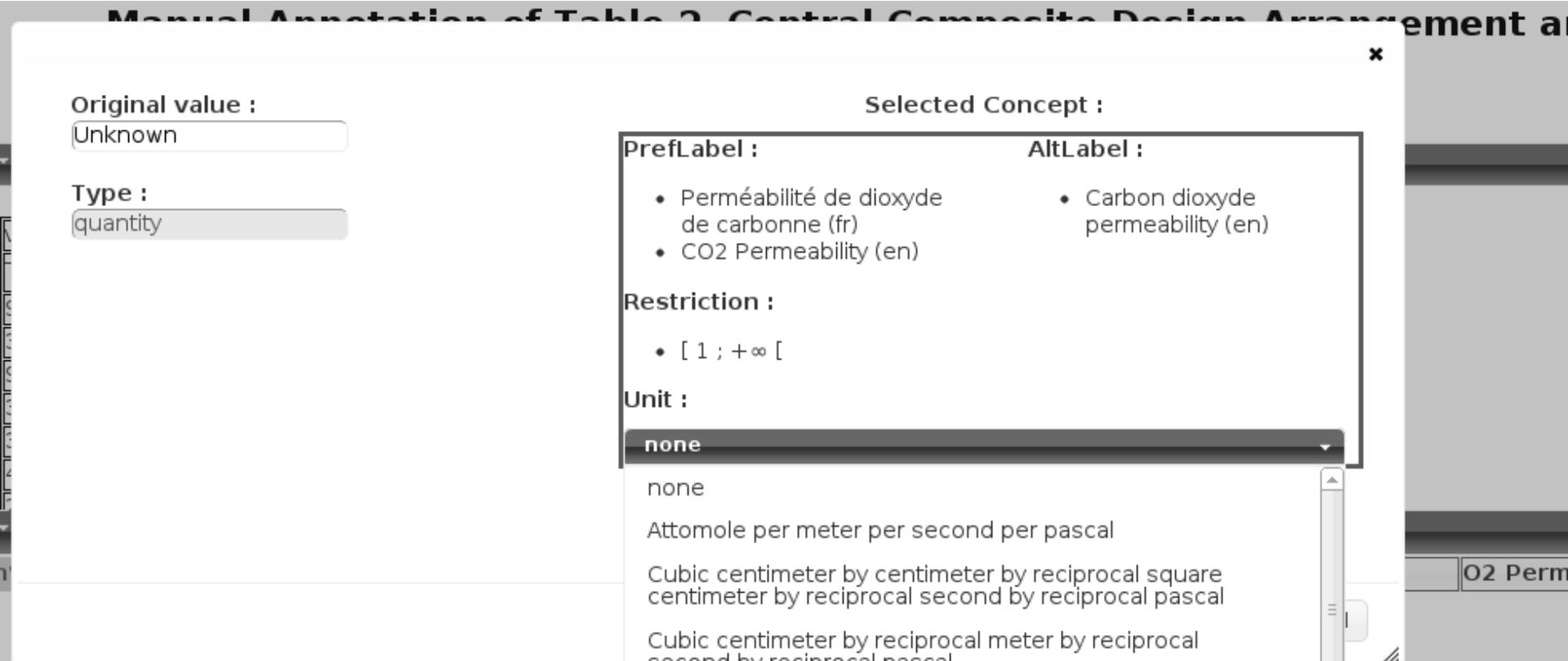
- Proteins (en)

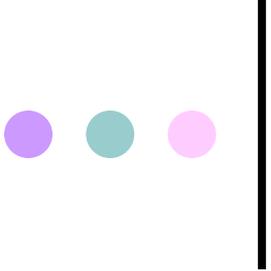
Save Cancel

39	[7.700e+1 ; 8.300e+1]	3.140e-16	14.6	Proteins	1.310e-
9	[7.700e+1 ; 8.300e+1]	1.148e-14	85.3	Proteins	1.011e-
20	[7.700e+1 ; 8.300e+1]	2.225e-14	85.3	Proteins	8.620e-

Logiciel @Web

Sélection dans la RTO TRANSMAT d'une unité de mesure pour l'associer à la quantité CO2Permeability





Sommaire

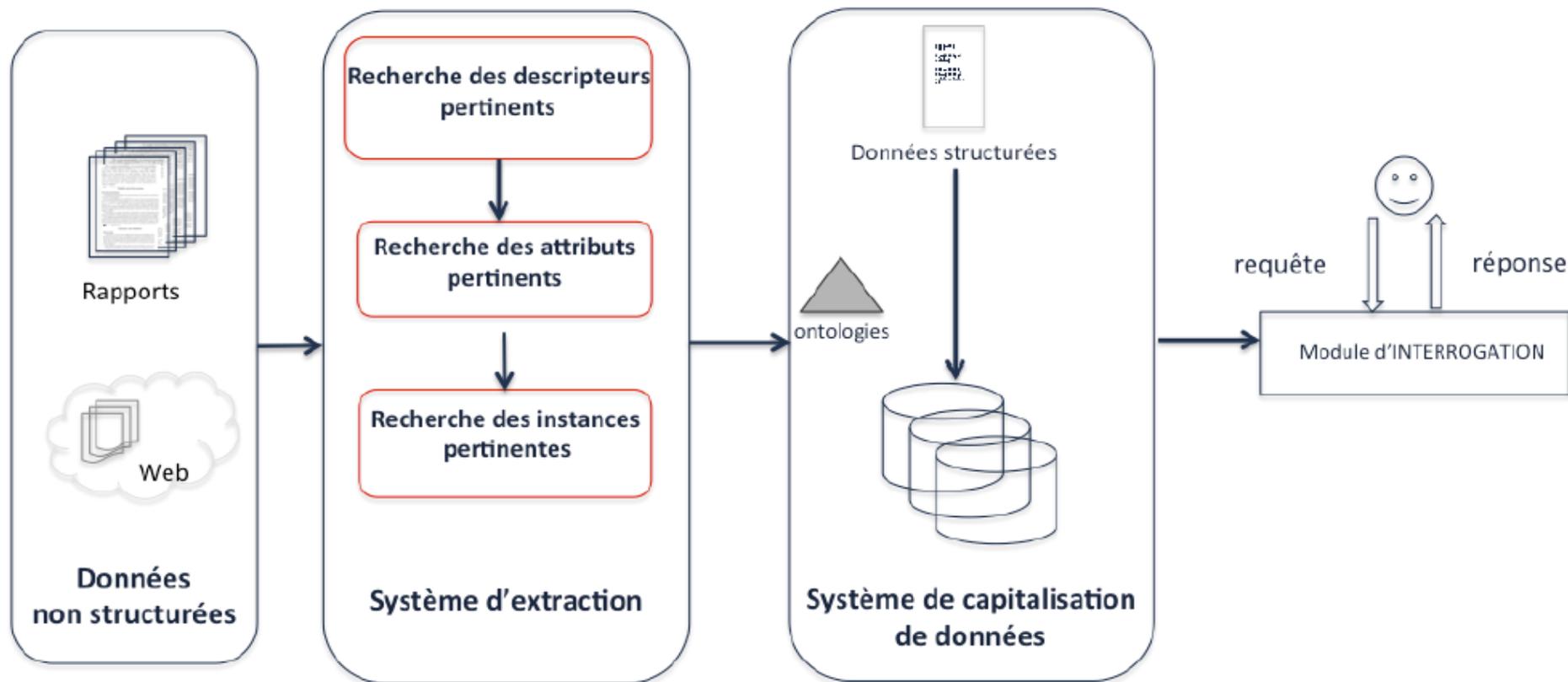
- Ontologie: construction et modélisation
- **Annotation sémantique**
 - **Annotation sémantique d'un texte**
- Interrogation flexible
- Evolution d'ontologie
- Alignements d'ontologies et LOD
- Conclusion et perspectives



Extraction d'arguments de relations n-aires dans les textes guidée par une RTO de domaine

- Thèse de Lilia Berrahou soutenue en septembre 2015
- Exemple : Eight apple wedges were packaged into polypropylene trays and wrap-sealed using a **64 μm** thickness **polypropylene** film with a permeability to oxygen of **110 $\text{cm}^3\text{m}^{-2}\text{bar}^{-1}\text{day}^{-1}$** at **23°C** and **0% RH**.

Contexte : Extraction d'information à partir de texte



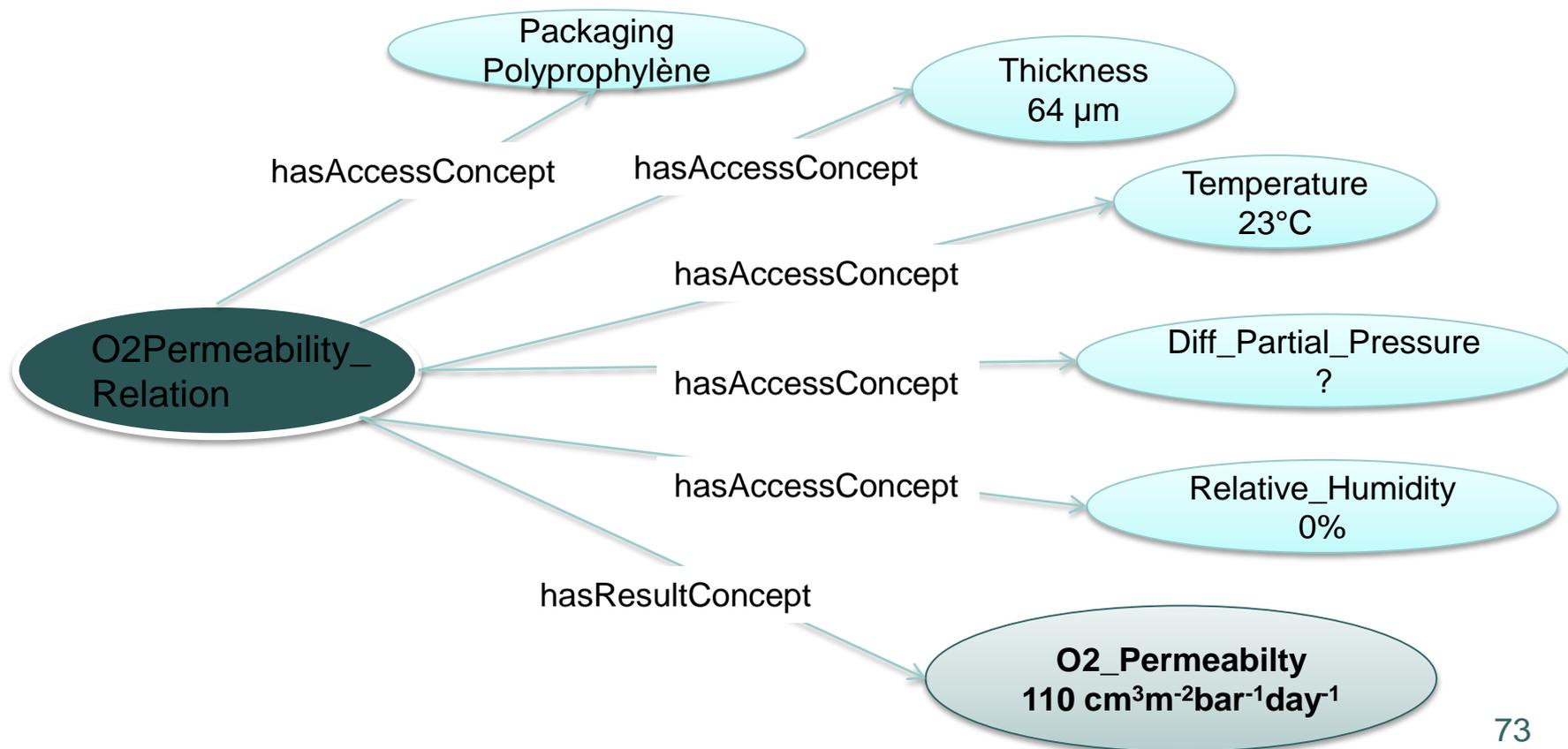


Modélisation d'un résultat expérimental en relation n-aire

- Résultat expérimental
 - Un objet étudié
 - Ensemble de paramètres de contrôle
 - Ensemble d'attributs (valeurs numériques et unités de mesure)

Modélisation d'un résultat expérimental en relation n-aire

Exemple : Eight apple wedges were packaged into polypropylene trays and wrap-sealed using a **64 μm** thickness **polypropylene** film with a permeability to oxygen of **110 $\text{cm}^3\text{m}^{-2}\text{bar}^{-1}\text{day}^{-1}$** at **23°C** and **0% RH**.



Les verrous : extractions des instances

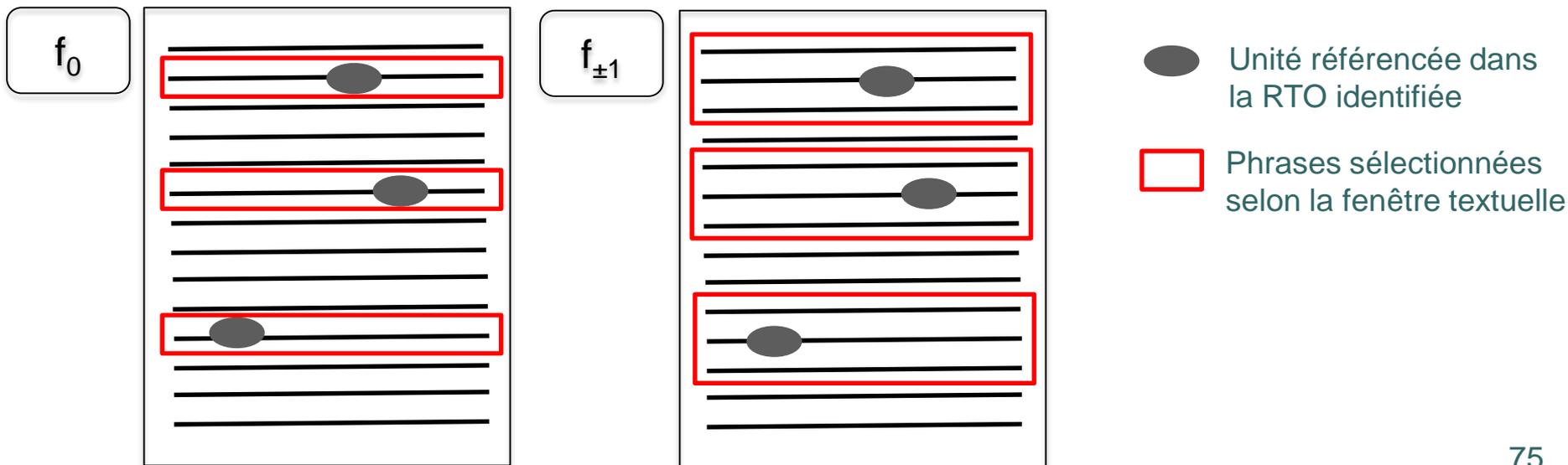
- Les arguments de l'instance sont dispersés dans le texte sur plusieurs phrases et plusieurs paragraphes
- Expression implicite de certains arguments (surtout quantitatifs)
- Expression partielle de l'instance
- Trouver l'instance d'intérêt

Exemple : E. coli O157 :H7 had been previously spread with 0.1 mL of inoculum containing 10^5 CFU/mL of tested bacterium. Plates were incubated at 37°C for 48 h.



Hypothèse de travail

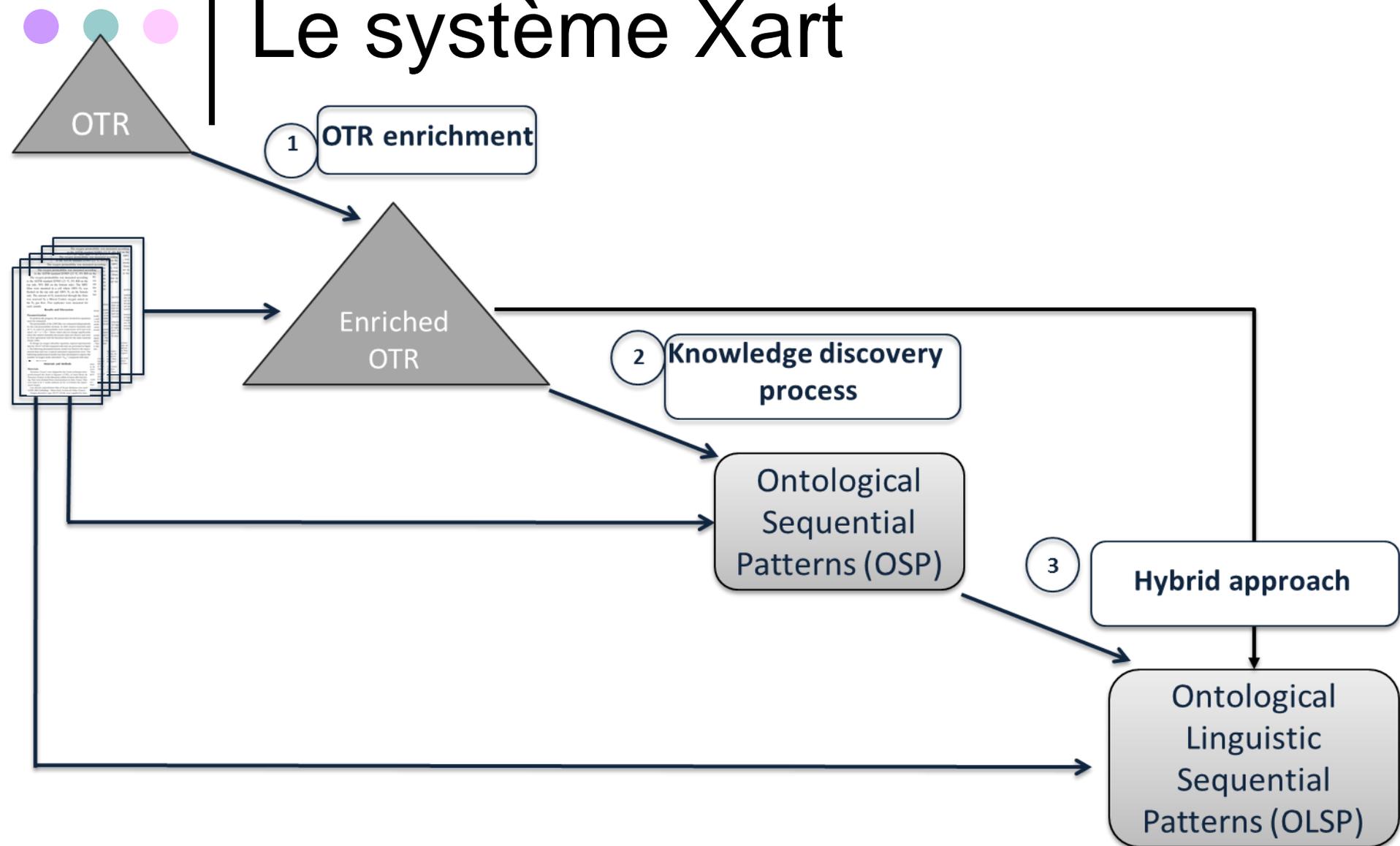
- Unité de mesure de la RTO: contexte favorable à la découverte d'arguments de la relation n-aire
- Unités textuelles pertinentes :
 - La phrase pivot
 - La fenêtre textuelle



Les verrous: extractions des unités de mesure

- Fortes variations typographiques des unités de mesure dans les textes
- Quelques exemples de variants d'unités de la perméabilité
 - $\text{cm}^3.\mu\text{m}/\text{cm}^2.\text{d.kPa}$
 - $\text{cm}^3.\mu\text{m}/\text{m}^2.\text{d.kPa}$: conversion de cm^2 en m^2
 - $\text{cm}^3.\text{um}/\text{m}^2.\text{d.kPa}$: changement de caractère μm en um
 - $\text{cm}^3 \mu\text{m m}^{-2}.\text{d}^{-1} \text{kPa}^{-1}$: aucun séparateur caractéristique
- Verrous
 - les localiser automatiquement
 - les identifier

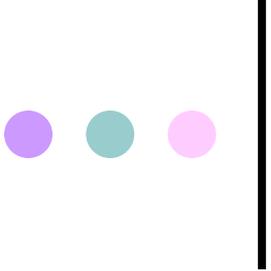
Le système Xart





Références bibliographiques

- Soumia Lilia Berrahoua, Patrice Buche, Juliette Dibie and Mathieu Roche (2017). Xart: Discovery of correlated arguments of n-ary relations in text. *Expert Systems with Applications*, 73, pp. 115-124. DOI: 10.1016/j.eswa.2016.12.028
- Soumia Lilia Berrahou, Patrice Buche, Juliette Dibie, Mathieu Roche (2016). Découverte et extraction d'arguments de relations n-aires corrélés dans les textes. *Revue des Nouvelles Technologies de l'Information*, Hermann, 2016, Fouille de Données Complexes vol. RNTI-E-31, pp. 37-56
- Soumia Lilia Berrahou, Ludovic Lebras, Patrice Buche, Juliette Dibie-Barthélemy, Mathieu Roche (2014). Enrichissement d'une RTO par l'ajout de termes spécialisés. In W 2014, *Revue des Nouvelles Technologies de l'Information*, vol. RNTI-W-2, pp.59-74



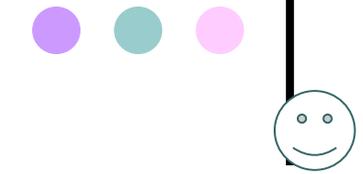
Sommaire

- Ontologie: construction et modélisation
- Annotation sémantique
- **Interrogation flexible**
- Evolution d'ontologie
- Alignements d'ontologies et LOD
- Conclusion et perspectives

Interrogation flexible

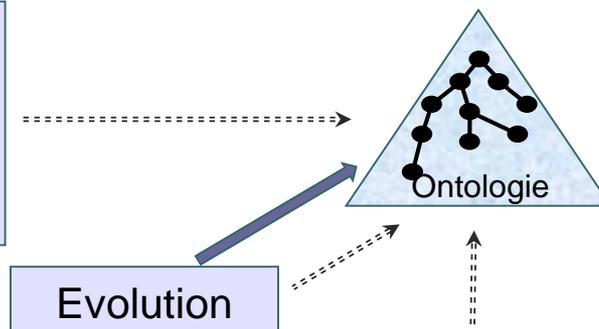
Le système ONDINE

(Ontology based Data INtEgration)



requête
réponse

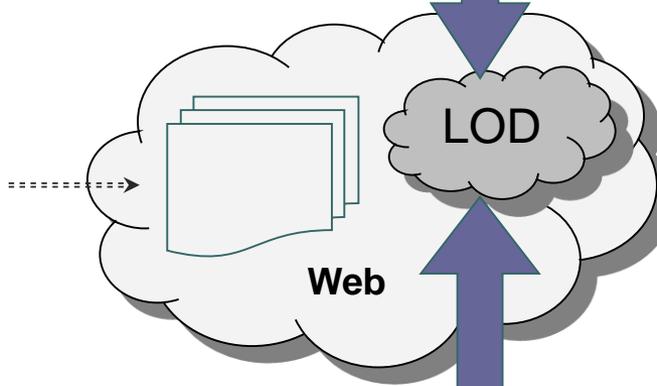
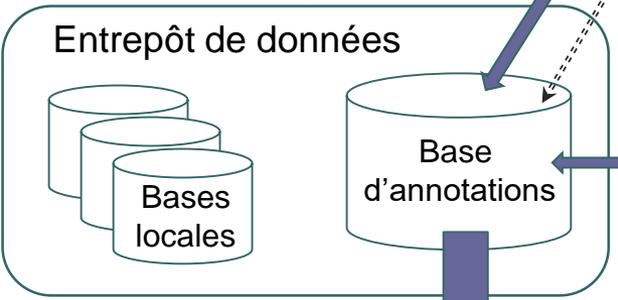
Interrogation
Interrogation unifié et flexible des différentes bases de l'entrepôt



Publication

Evolution

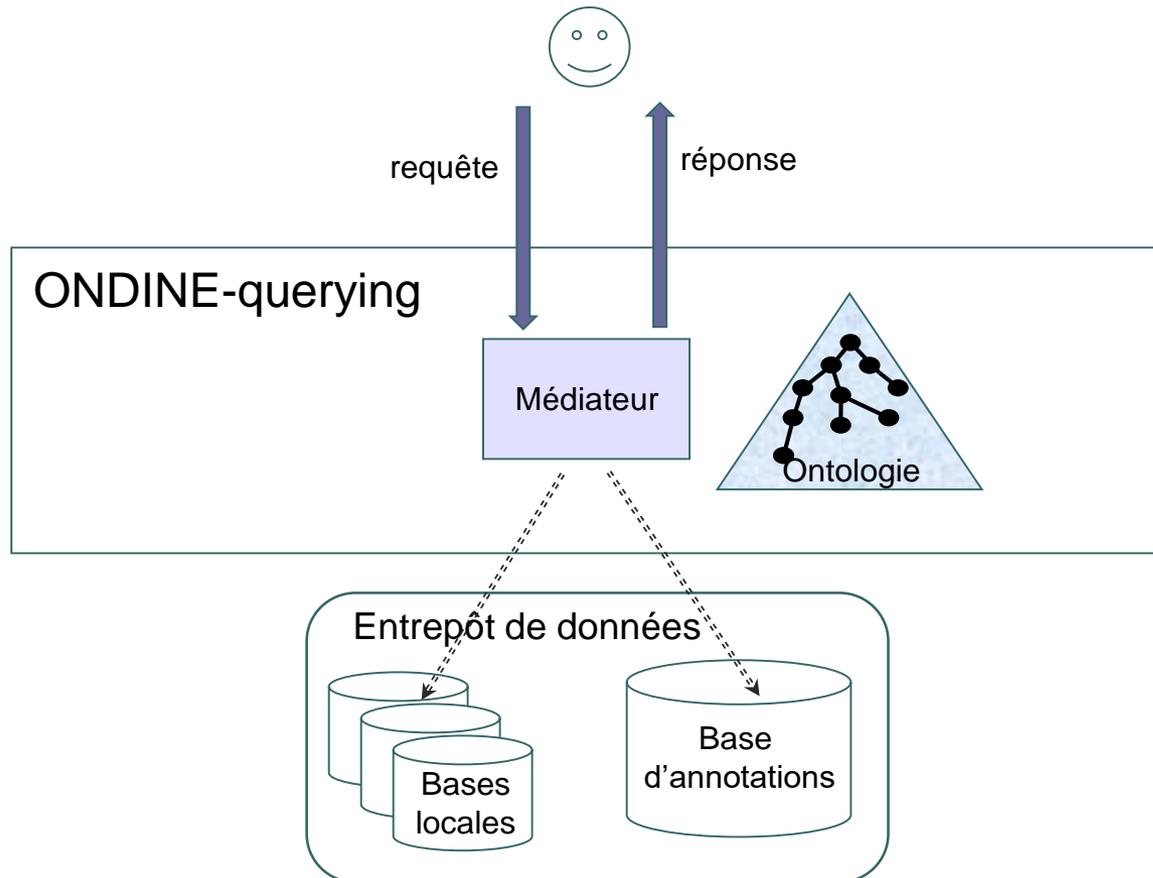
Annotation
Annotation sémantique de tableaux et de texte extraits de documents scientifiques du Web pour enrichir des bases locales



-----> utilise
—————> fournit

Publication

Le sous-système d'interrogation flexible





Interrogation flexible guidée par une ontologie

- Besoin de prendre en compte les différentes sémantiques du flou dans l'interrogation (données imprécises et données annotées)
- Solution à l'hétérogénéité des données (médiateur)
- Solution à l'incomplétude des bases locales...

Trois façons de traiter l'incomplétude

- Complétion de la base
 - découverte de données sur le Web
- Requêtes flexible
 - expressions de préférences dans les requêtes de l'utilisateur
 - Priorisation des valeurs recherchées
- Généralisation automatique des requêtes



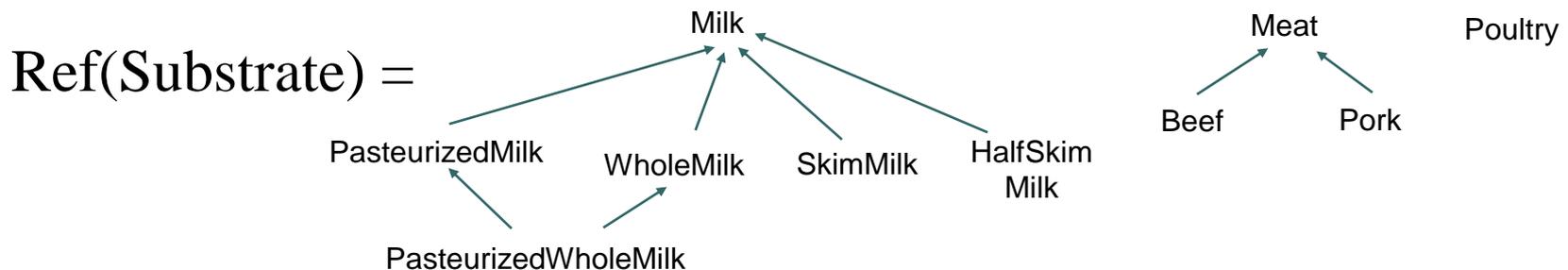
ONDINE querying: le langage MIEL

A, ensemble fini d'attributs

$\forall a \in A, \text{Type}(a) \in \{\text{numérique, symbolique, hiérarchisé}\}$

Domaine de référence d'un attribut

- si $Type(a) = \text{numérique}$, $Ref(a)$ est un sous-ensemble de \mathbb{R}
 - $Ref(pH) = [0, 14]$
- si $Type(a) = \text{symbolique}$, $Ref(a)$ est un ensemble de constantes symboliques
 - $Ref(\text{Author}) = \{S.Ajjarapu, C.P.Rivituso, M.Zwietering\}$
- si $Type(a) = \text{hiérarchisé}$, $Ref(a)$ est un ensemble de constantes symboliques muni d'un ordre partiel



MIEL - Syntaxe d'une requête

Attributs de projection

$\{a_1, \dots, a_p \mid$

$\text{Vue}(a_1, \dots, a_n) \wedge$

$(a_{p+1} = v_{p+1}) \wedge \dots \wedge (a_m = v_m) \}_{1 \leq p \leq m \leq n}$

Vue dans laquelle la requête est posée

Critères de sélection

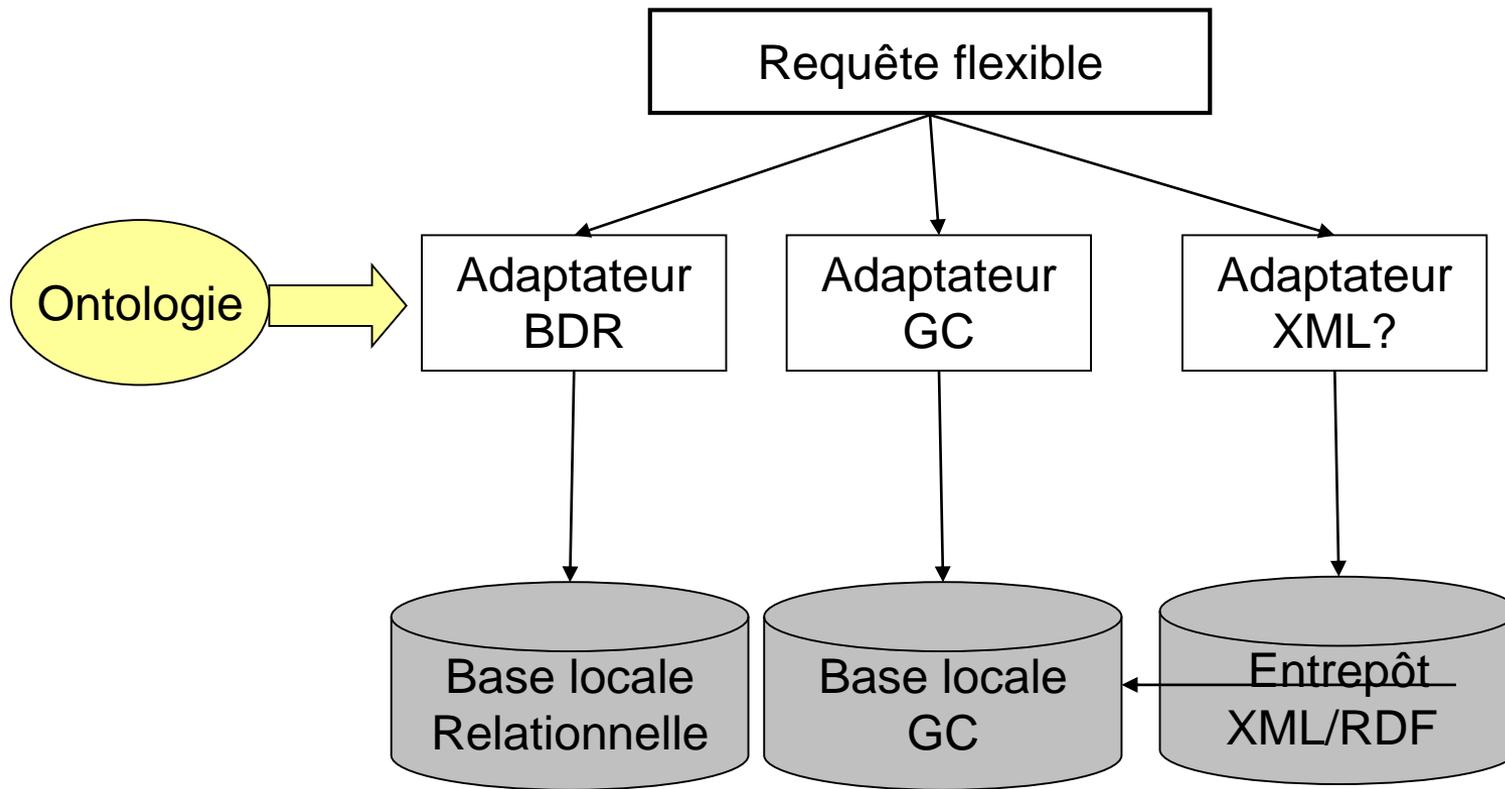
MIEL – Forme d'une réponse

- Réponse à une requête dans MIEL

$\{a_1, \dots, a_p \mid \text{Vue}(a_1, \dots, a_n) \wedge (a_{p+1} = v_{p+1}) \wedge \dots \wedge (a_m = v_m)\}_{1 \leq p \leq m \leq n}$

tuple $\{\tau[a_1, \dots, a_p]\}$

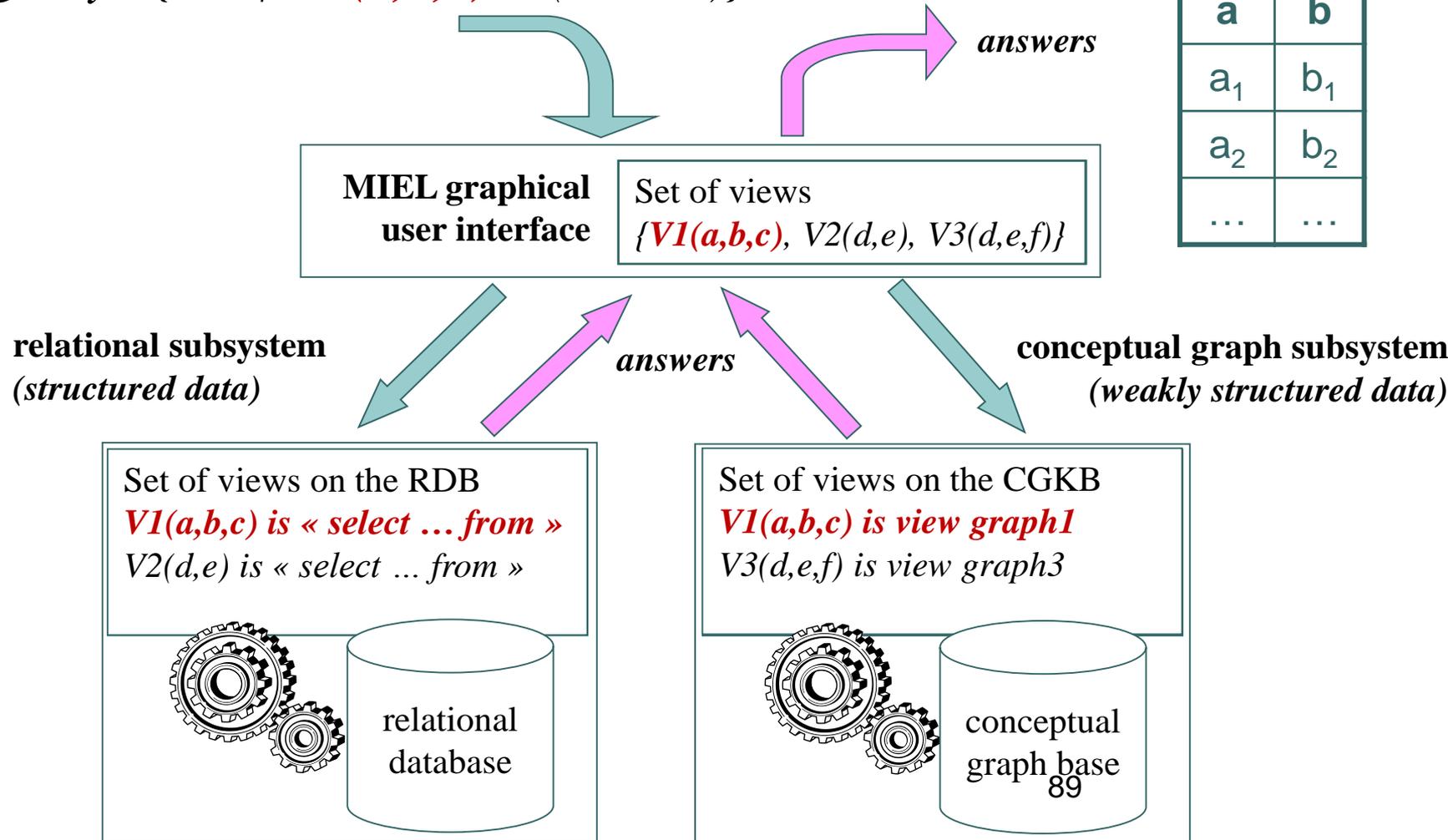
Interrogation flexible



Exécution d'une requête MIEL

Query : $\{a,b \mid V1(a,b,c) \wedge (c = 37)\}$

a	b
a ₁	b ₁
a ₂	b ₂
...	...





Interrogation flexible

- Interrogation flexible via un adaptateur GC
 - **Le modèle des graphes conceptuels**
 - Le sous-système GC
 - Représentation du flou dans le modèle des GC
 - Extension de l'adaptateur GC
- Interrogation flexible d'une BDR
- Interrogation flexible via un adaptateur SPARQL

Le modèle des graphes conceptuels

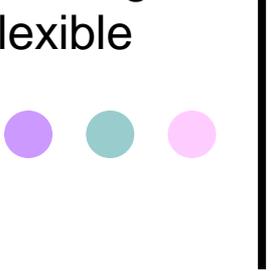
- Conceptual Structures : Information Processing in Mind and Machine, (Sowa, 1984)
- Objectif : « allier la puissance de la logique à l'expressivité de la langue naturelle »

Pourquoi utilisons-nous ce modèle

- Lisibilité pour des non-informaticiens
- Noyau théorique robuste
 - fondé sur la théorie des graphes
 - lien étroit avec la logique des prédicats
- Faiblement structuré par nature
- Distinction claire terminologie / assertions
- Implémentations existantes
 - CoGITo / CoGITanT/ Corese / CoGui
- Formalisation utilisée (Mugnier et Chein, 1996)

Le modèle des graphes conceptuels

- Deux ensembles distincts de connaissances
 - le support, ou connaissances terminologiques (ou ontologie)
 - vocabulaire
 - contraintes d'utilisation du vocabulaire
 - les graphes, ou connaissances assertionnelles
 - faits proprement dits

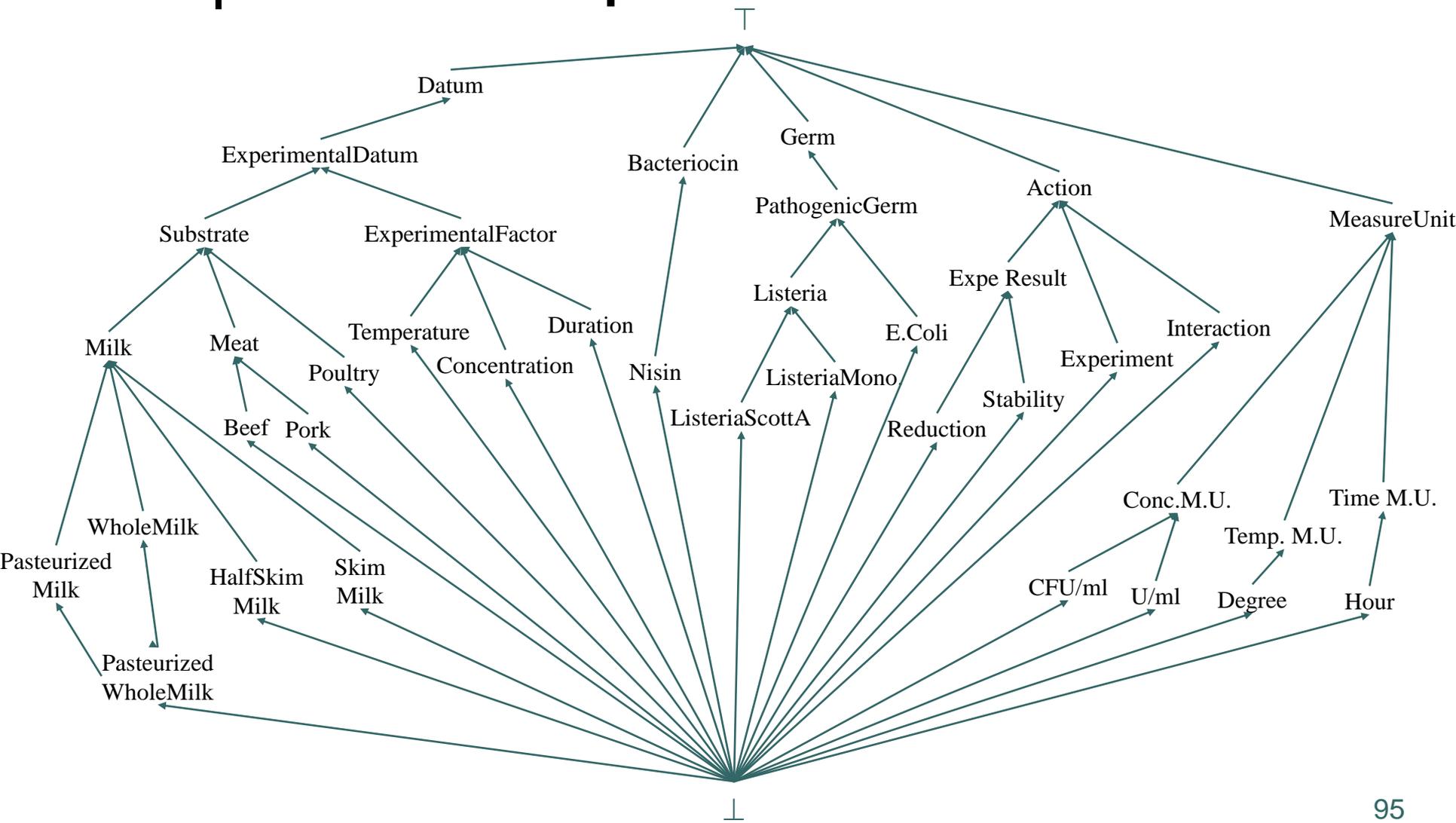


GC – Le support

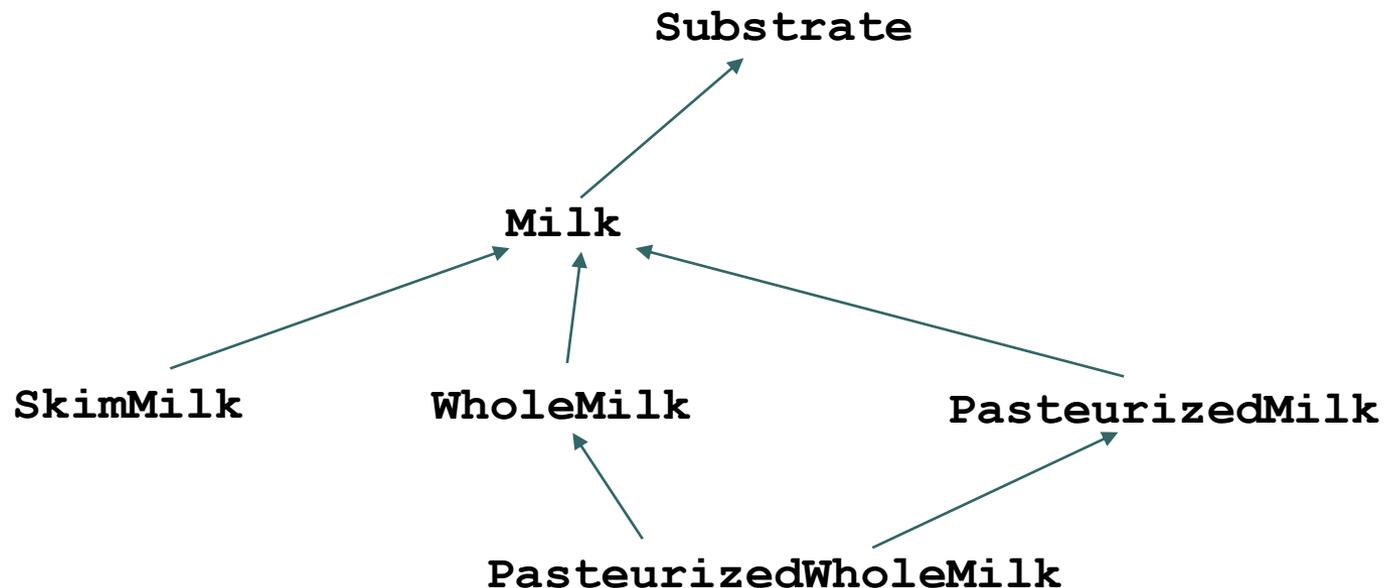
- Un support S est un quintuplet $(T_c, T_r, M, \tau, \sigma)$
- T_c est l'ensemble des types de concepts
 - description des entités, attributs, états, événements...
 - partiellement ordonné par la relation « sorte-de »



Tc – L'ensemble des types de concepts



Interprétation logique de Tc



$\forall x \text{ (SkimMilk}(x) \rightarrow \text{Milk}(x))$

Tr – L'ensemble des types de relations

- Description des liens entre concepts
- Partiellement ordonnés
- Relations « traditionnellement »
pourvues d'une sémantique pauvre
 - Agt (Action, Universal)
 - Obj (Action, Universal)
 - Res (Action, ExperimentalResult)
 - Subst (Action, Substrate)

M – L'ensemble des marqueurs

- Les individus (ou instances) d'un concept donné
 - « C.P. Rivituso » (instance du type de concept « Author »)
 - « Journal of Food Protection » (instance de « Journal »)
- Le marqueur générique permet de désigner un indéterminé
 - noté * ou omis
 - une Expérience, un Substrat...

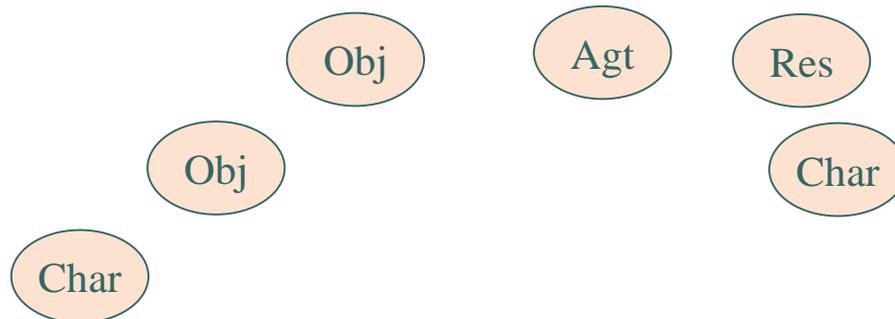
Les graphes conceptuels (1)

- Un graphe conceptuel $G = (R, C, U, lab)$ est :
 - un multi-graphe
 - plusieurs arêtes peuvent lier 2 mêmes sommets
 - biparti
 - les sommets sont partitionnés en 2 classes (R et C)
 - tout sommet d'une classe n'est relié qu'à des sommets de l'autre classe
 - non nécessairement connexe

connexe : à partir d'un sommet, on peut atteindre tous les autres sommets du graphe en parcourant les arêtes

Les graphes conceptuels (2)

- R est l'ensemble des sommets relations
 - représentés par des ovales
 - étiquetés par un type de relation $\in TR$



Les graphes conceptuels (3)

- C est l'ensemble des sommets concepts
 - représentés par des rectangles
 - étiquetés par un couple (type de concept \in TC, marqueur $\in M \cup \{*\}$)

ListeriaScottA : *

Nisin : *

Reduction : *

Experiment : E1

Interaction : I1

SkimMilk : *

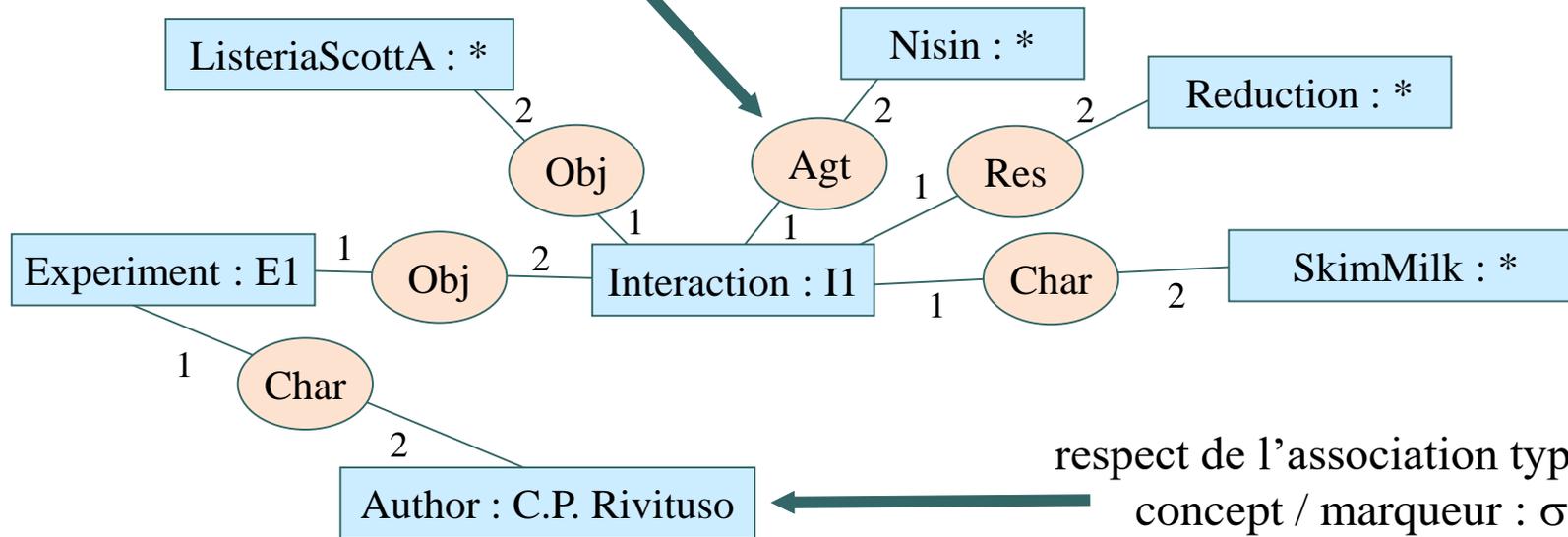
Author : C.P. Rivituso

Les graphes conceptuels (4)

- U est l'ensemble des arêtes
 - les arêtes adjacentes à un sommet relation r sont totalement ordonnées, et numérotées de 1 à degré(r)

respect de la signature des relations

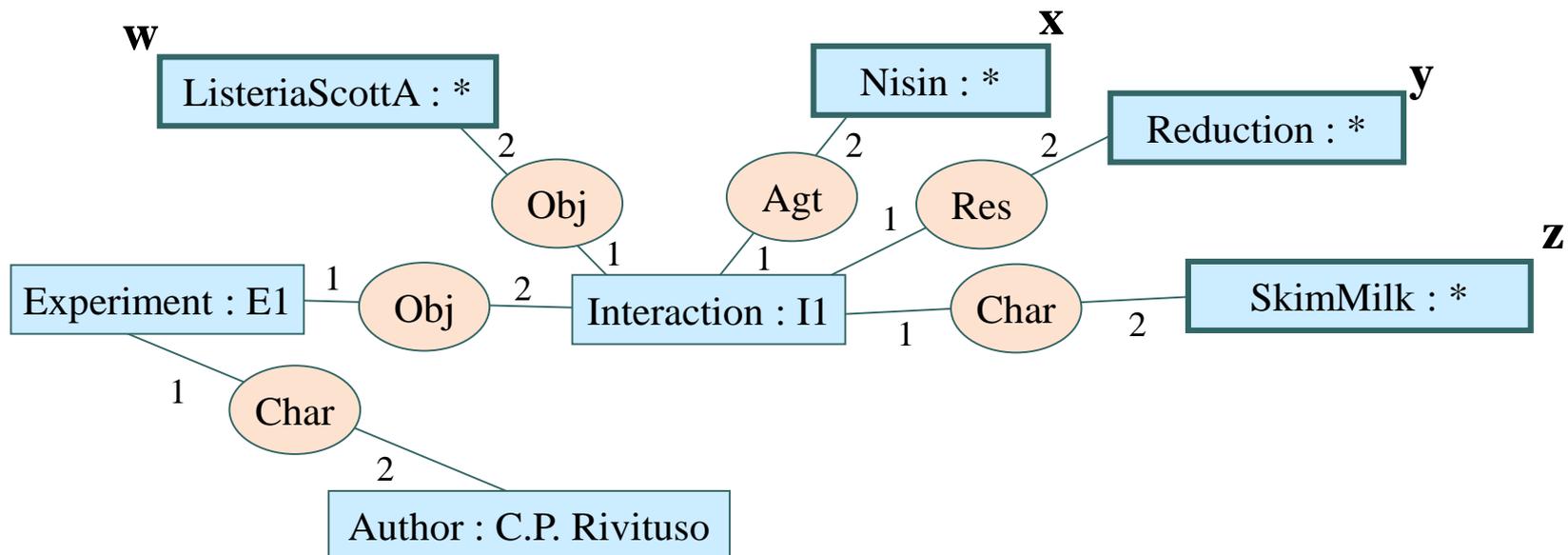
Agt (Action, Universel) : τ



respect de l'association type de concept / marqueur : σ

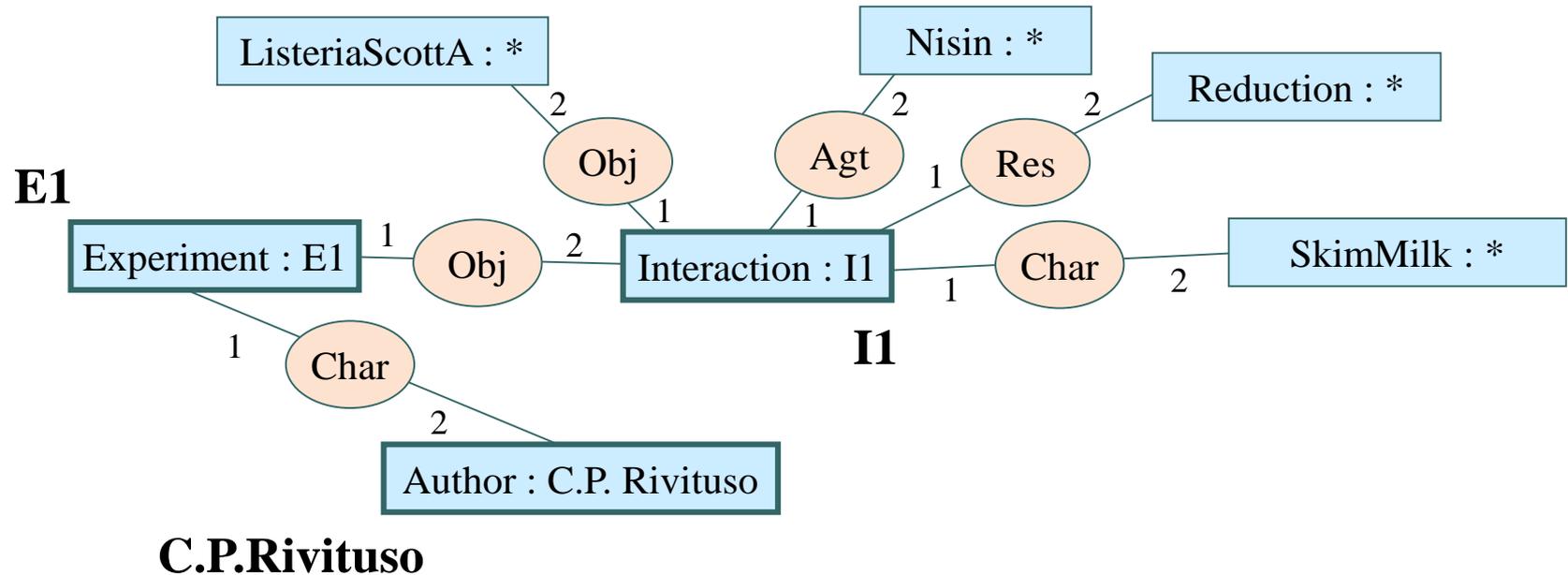
Interprétation logique d'un GC (1)

- Chaque sommet concept générique de G est mis en bijection avec une variable distincte de $\Phi(G)$



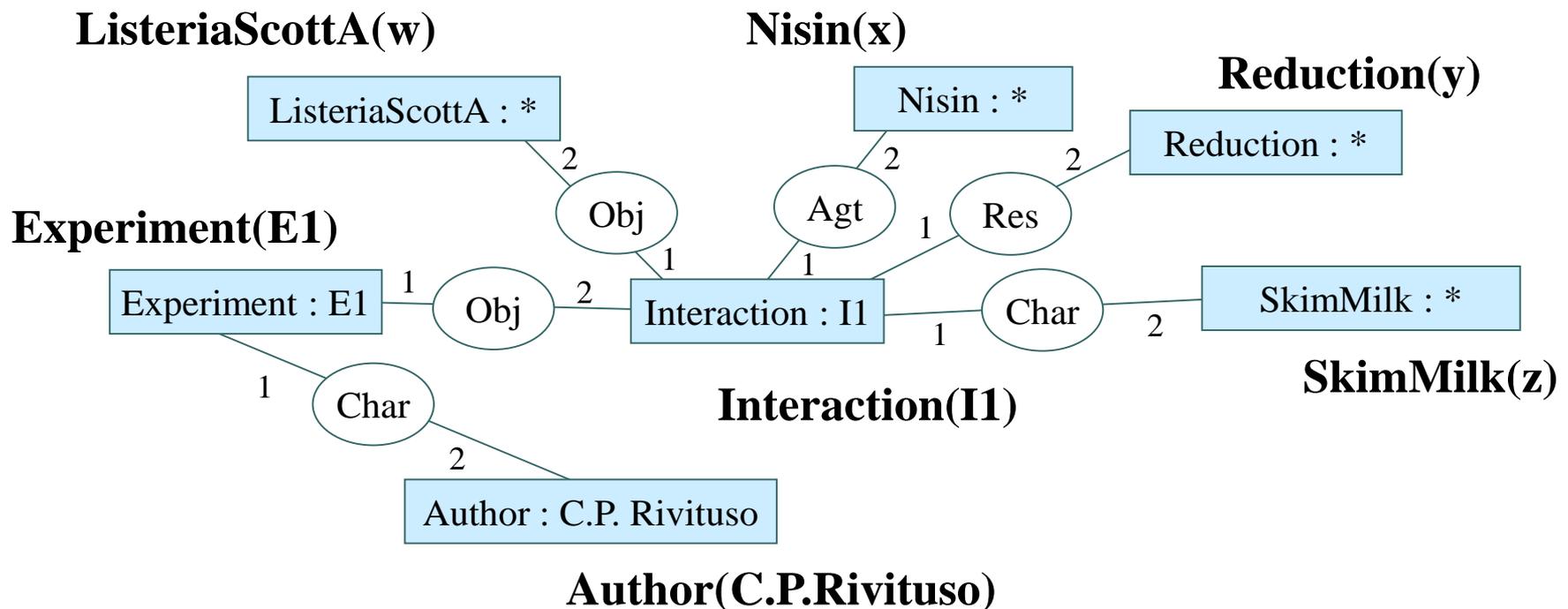
Interprétation logique d'un GC (2)

- Chaque marqueur individuel de G est mis en bijection avec une constante de $\Phi(G)$



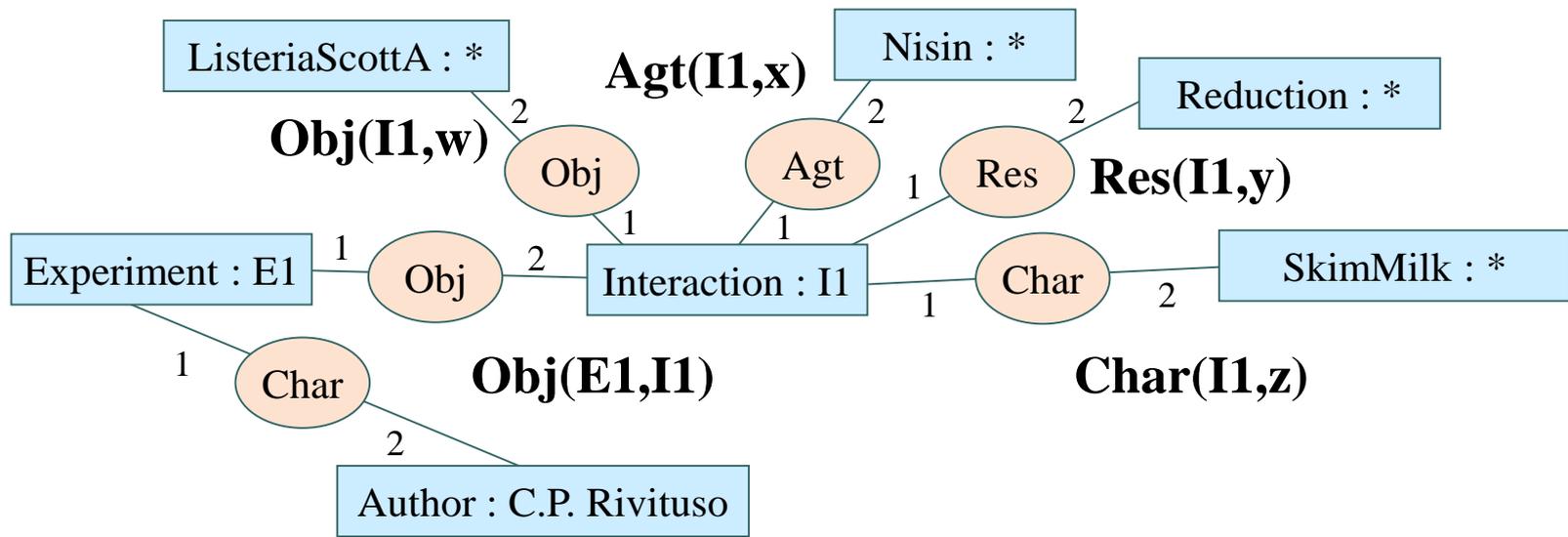
Interprétation logique d'un GC (3)

- Chaque sommet concept de G est représenté par un prédicat unaire de $\Phi(G)$, nommé par le type de concept



Interprétation logique d'un GC (4)

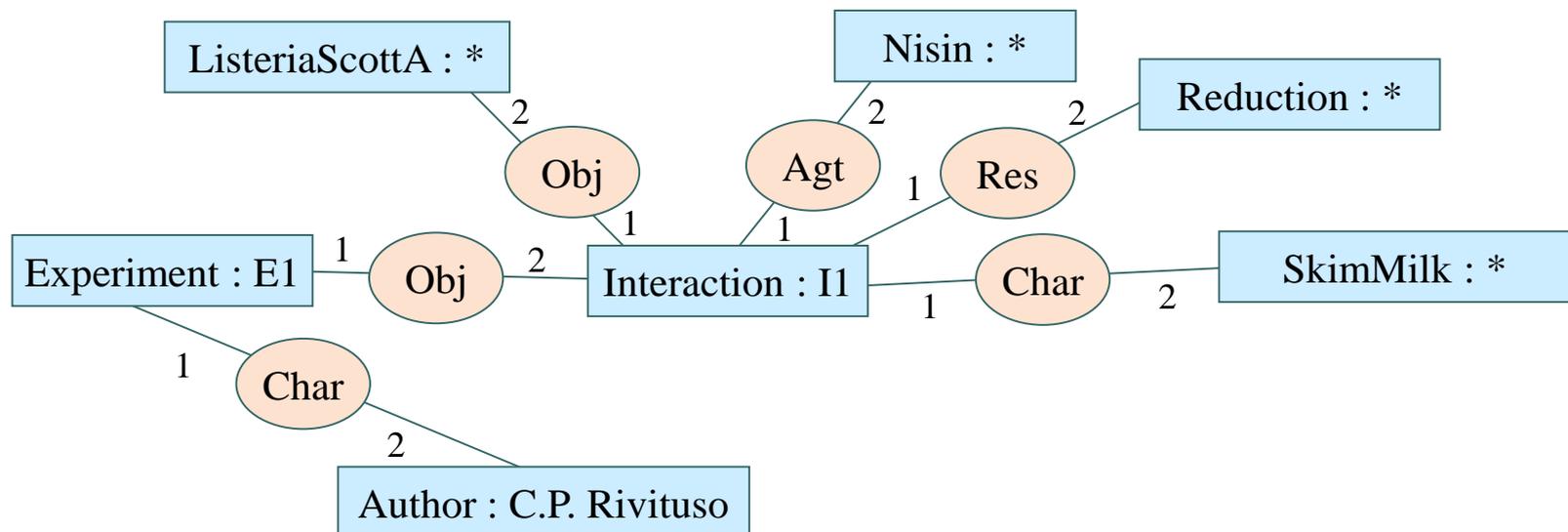
- Chaque sommet relation r de G est représenté par un prédicat de $\Phi(G)$ d'arité $\text{degré}(r)$, nommé par le type de relation



Char(E1,C.P.Rivituso)

Interprétation logique d'un GC (5)

- $\Phi(G)$ est la fermeture existentielle de la conjonction des prédicats associés à chaque sommet du graphe

$$\exists w \exists x \exists y \exists z (\text{Experiment}(E1) \wedge \text{Interaction}(I1) \wedge \text{Author}(\text{C.P.Rivituso}) \wedge \text{ListeriaScottA}(w) \wedge \text{Nisin}(x) \wedge \text{Reduction}(y) \wedge \text{SkimMilk}(z) \wedge \text{Obj}(E1,I1) \wedge \text{Obj}(I1,w) \wedge \text{Agt}(I1,x) \wedge \text{Res}(I1,y) \wedge \text{Char}(I1,z) \wedge \text{Char}(E1,\text{C.P.Rivituso}))$$


Spécialisation / projection (1)

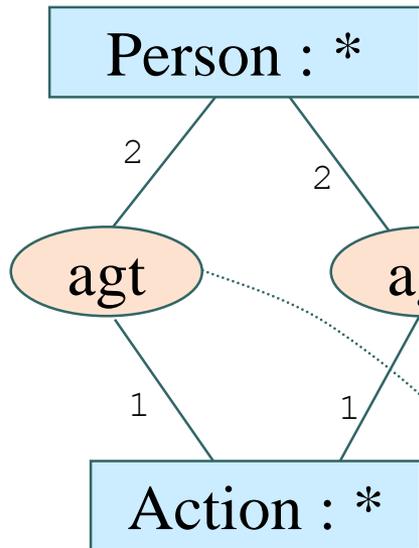
- Un graphe $G1$ est dit “spécialisation” d’un graphe $G2$ (noté $G1 \leq G2$) si $G1$ contient une information plus spécifique que celle de $G2$
- Comment déterminer que $G1$ est “plus spécifique” que $G2$?
- L’opération de projection

Interrogation
flexible

Spécialisation / projection

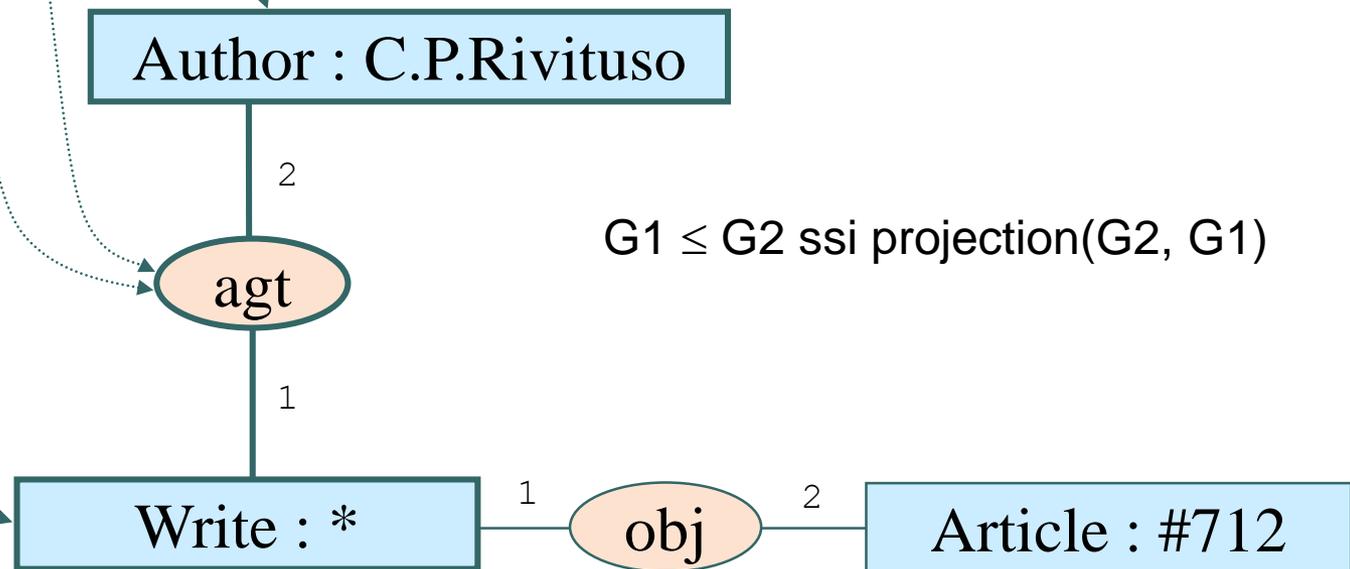
(2)

G2



- opération de morphisme de graphe (appariement)
- restrictions d'étiquettes de sommets possibles
- conservation des arêtes et de leur numérotation
- non nécessairement injective

G1



Spécialisation / projection (3)

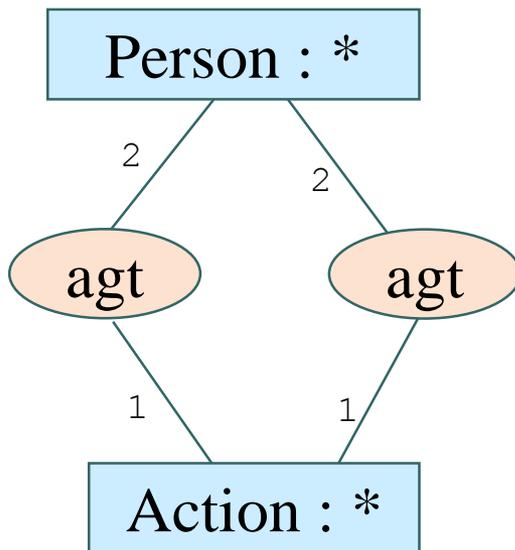
- La projection est un problème NP-complet dans le cas général (algorithme de backtrack)
 - explosion combinatoire si grand nombre de sommets « banalisés » (molécules de chimie par exemple)
 - peu réaliste dans le cas général (les étiquettes des sommets ont un fort pouvoir discriminant)
- Certains cas particuliers sont polynomiaux
 - projection d'un arbre dans un graphe

Spécialisation / projection (4)

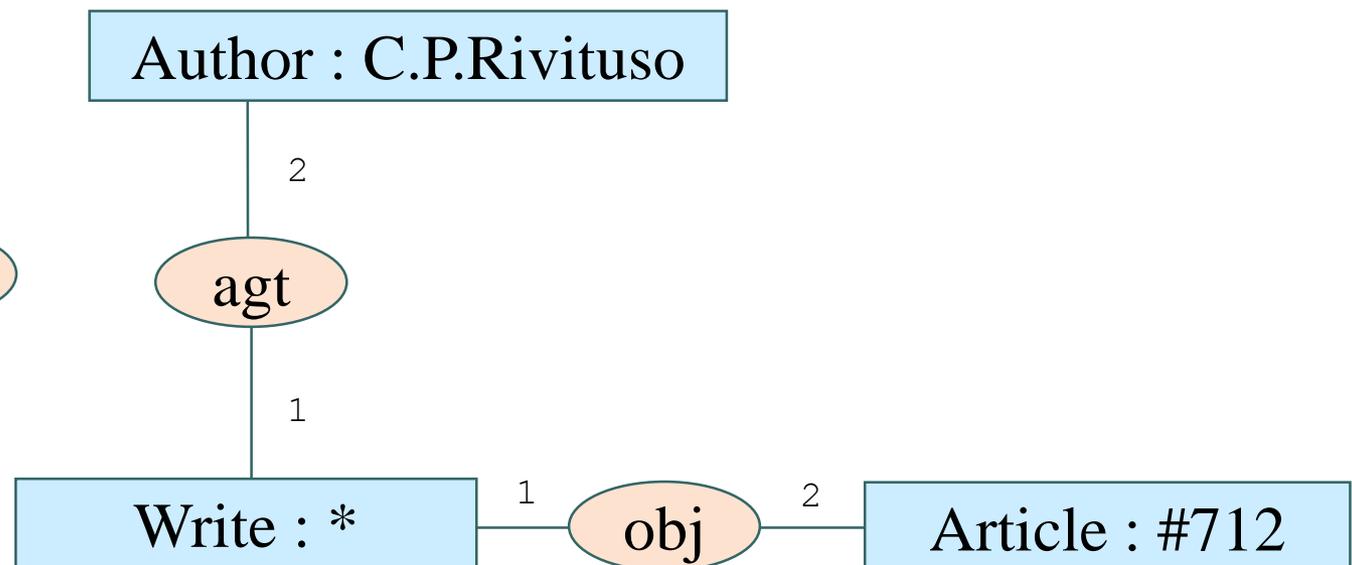
- Lien avec la déduction logique

$$G2 \leq G1 \text{ ssi } \Phi(G2) \rightarrow \Phi(G1)$$

G1

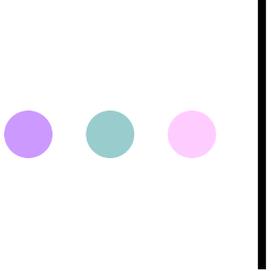


G2



CoGITo / CoGITaNT/ CoGui (LIRMM)

- Bibliothèque de classes C++
- Haute teneur en pointeurs
- Interface de saisie de graphes CoGui
<http://www.lirmm.fr/cogui/>

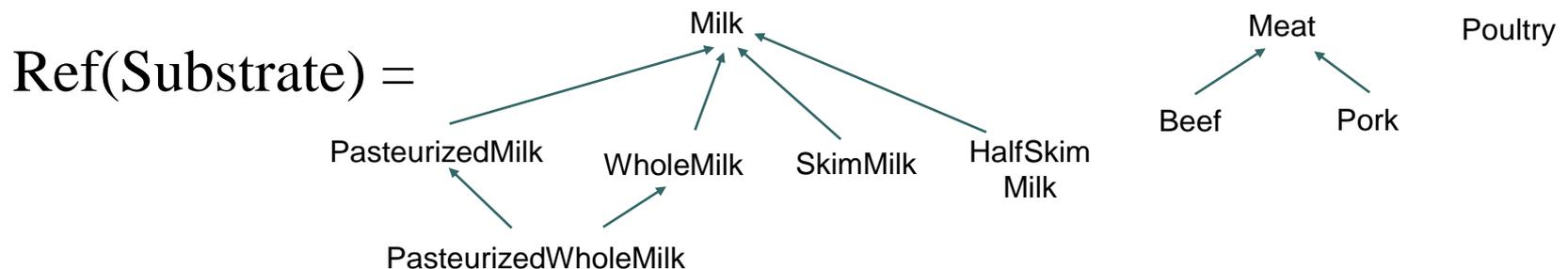


Interrogation flexible

- Interrogation flexible via un adaptateur GC
 - Le modèle des graphes conceptuels
 - **Le sous-système GC**
 - Représentation du flou dans le modèle des GC
 - Extension de l'adaptateur GC
- Interrogation flexible d'une BDR
- Interrogation flexible via un adaptateur SPARQL

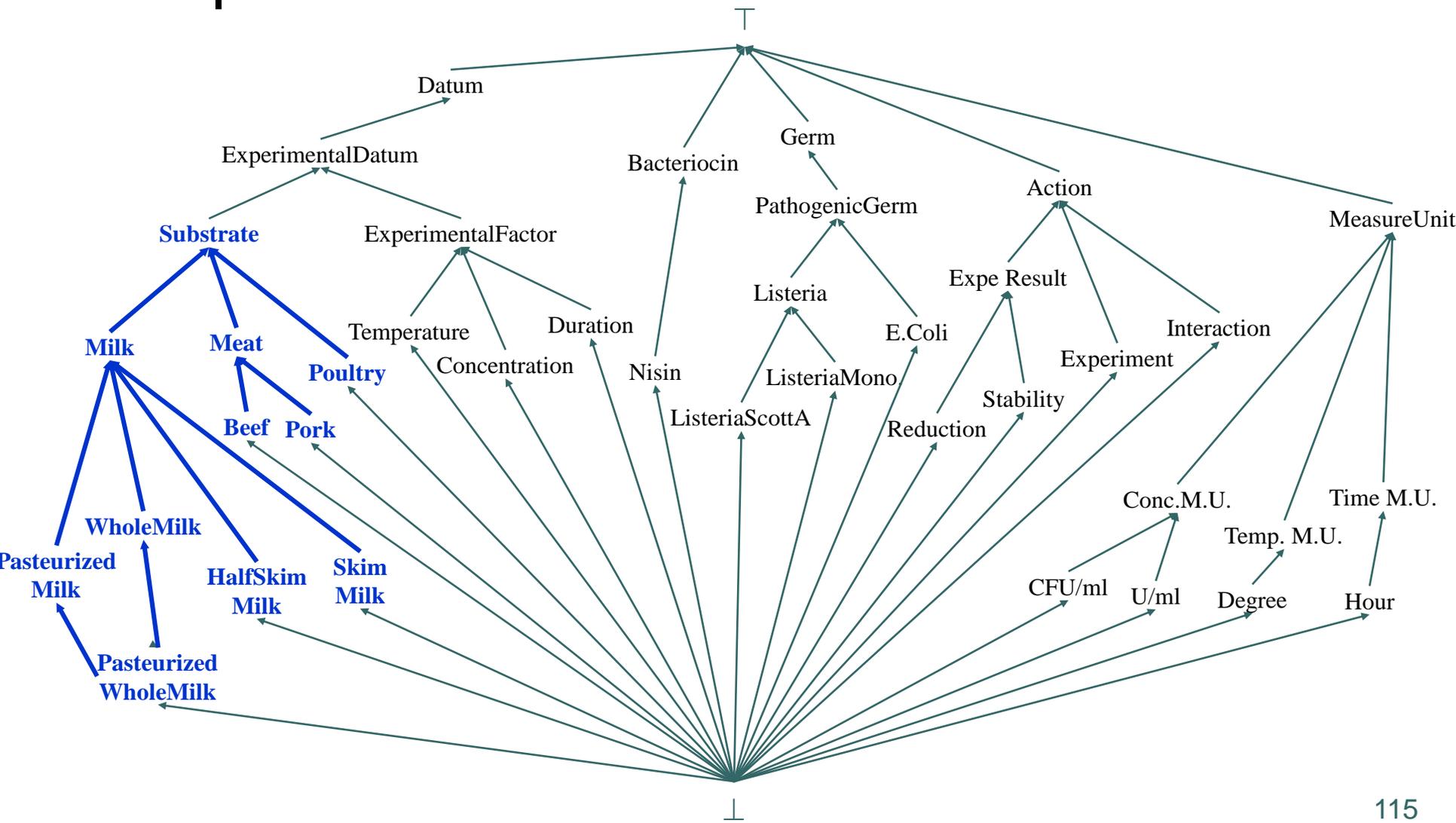
MIEL – Le sous-système GC

- Représentation de l'ontologie à l'aide du modèle des graphes conceptuels
 - les numériques et symboliques apparaissent en tant que marqueurs
 - les attributs de type hiérarchisé apparaissent ainsi que leurs valeurs en tant que types de concepts



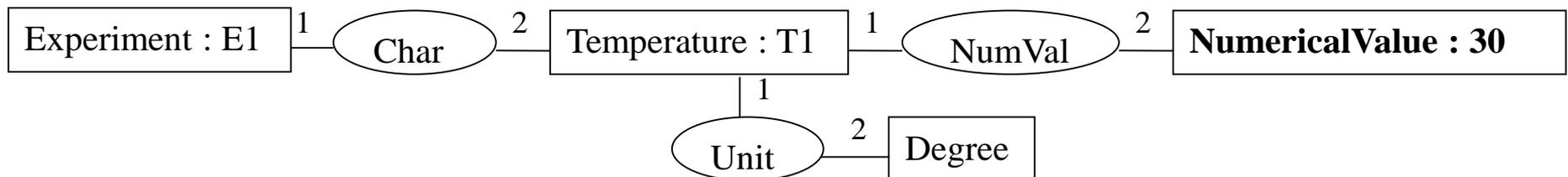
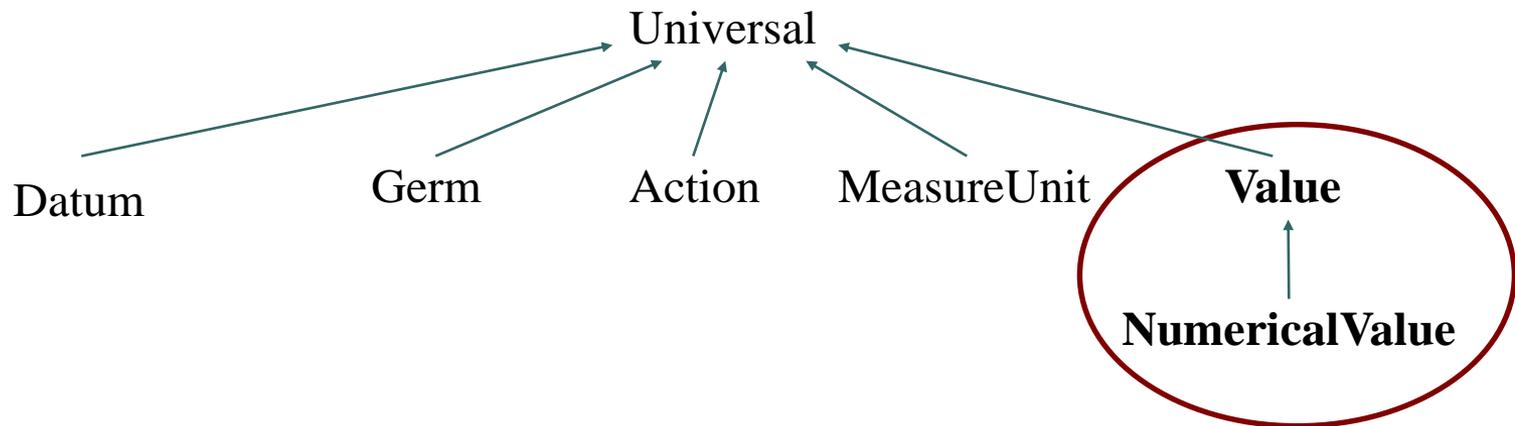


MIEL – Le sous-système GC



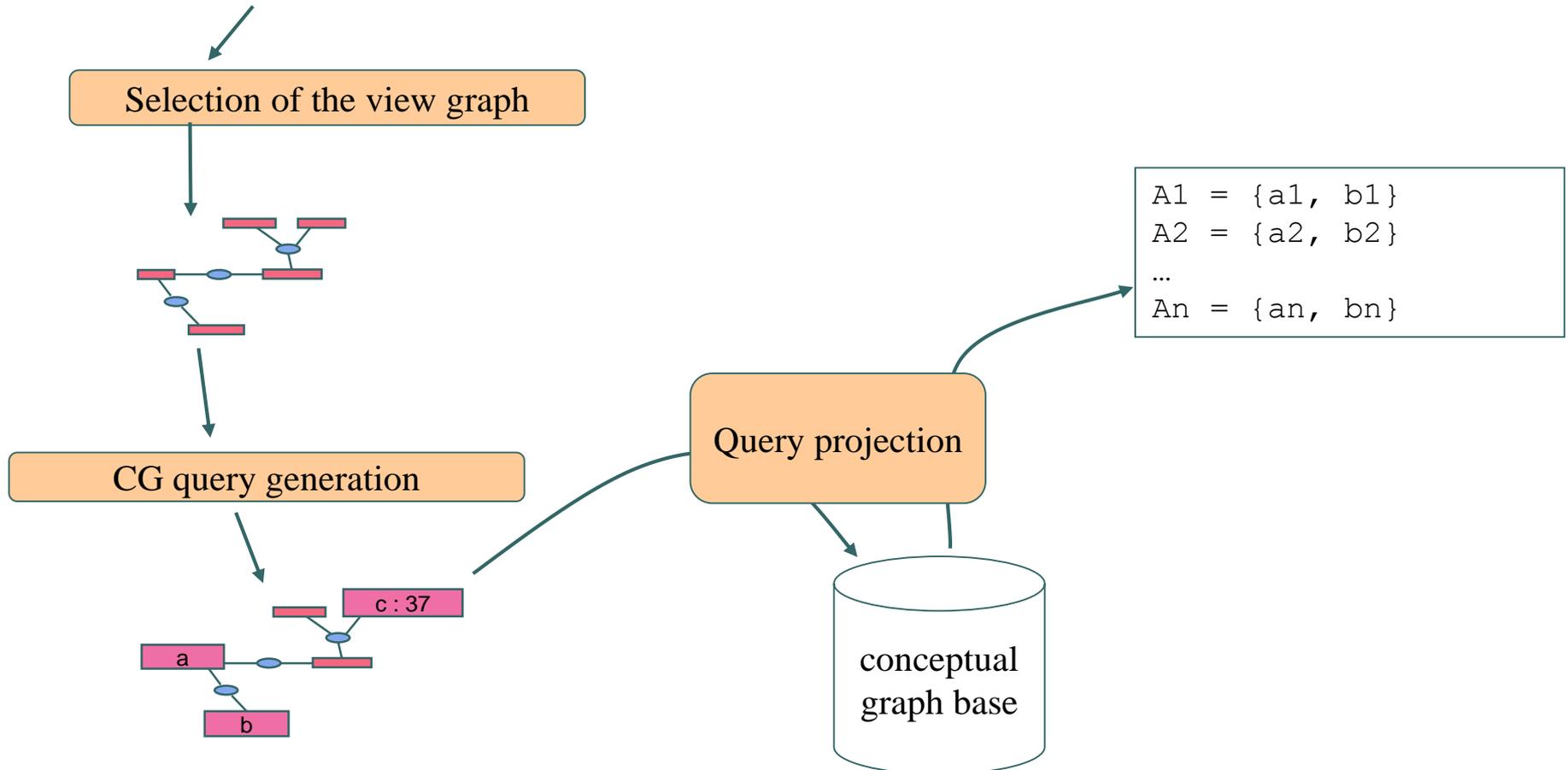
MIEL – Le sous-système GC

- Représentation du numérique dans les GC



Exécution d'une requête GC

Query : $\{a, b \mid \mathbf{V1(a,b,c)} \wedge (c = 37)\}$



Exécution d'une requête GC

Vue :

Bacteriocin interaction

Attributs de projection :

Pathogenic Germ, Expe. Result

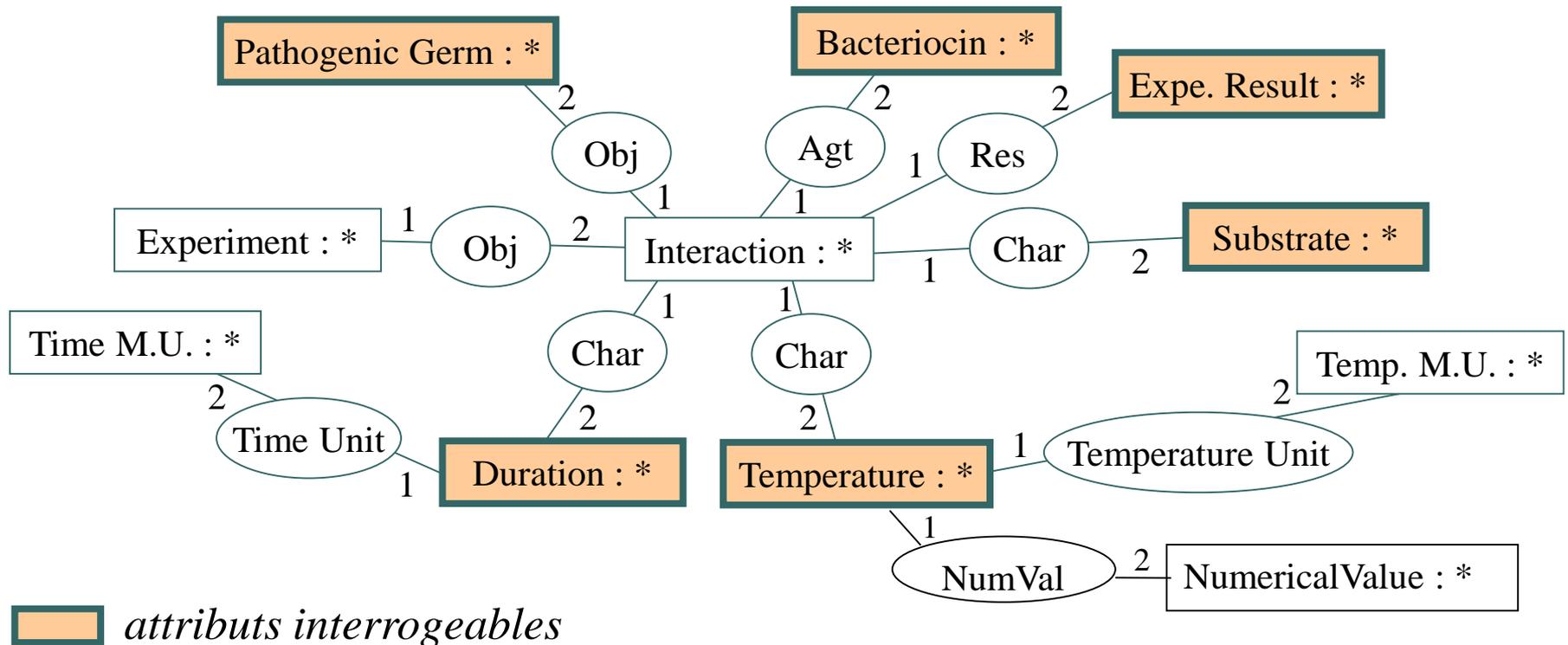
Critères de sélection :

(Temperature = 37)

(Bacteriocin = Nisin)

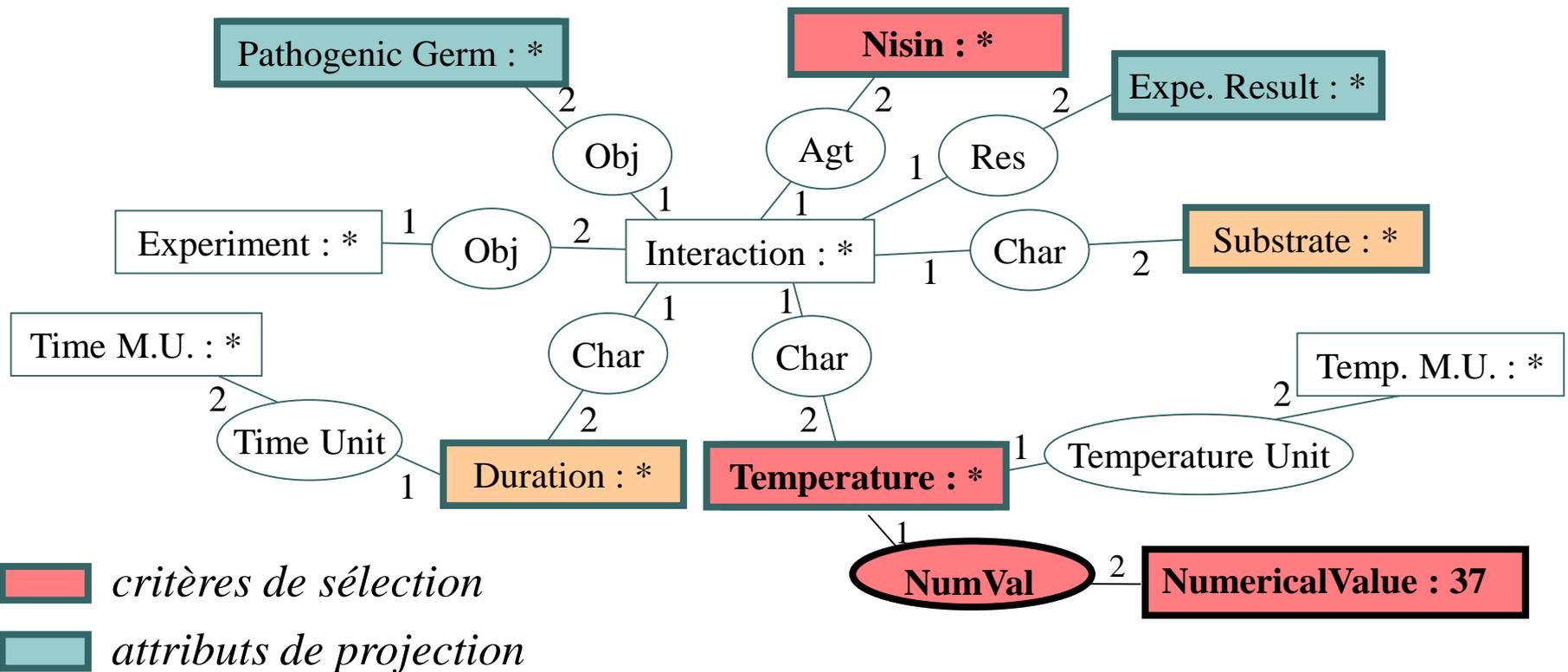
Exécution d'une requête GC

Graphe vue correspondant à la vue **Bacteriocin interaction**



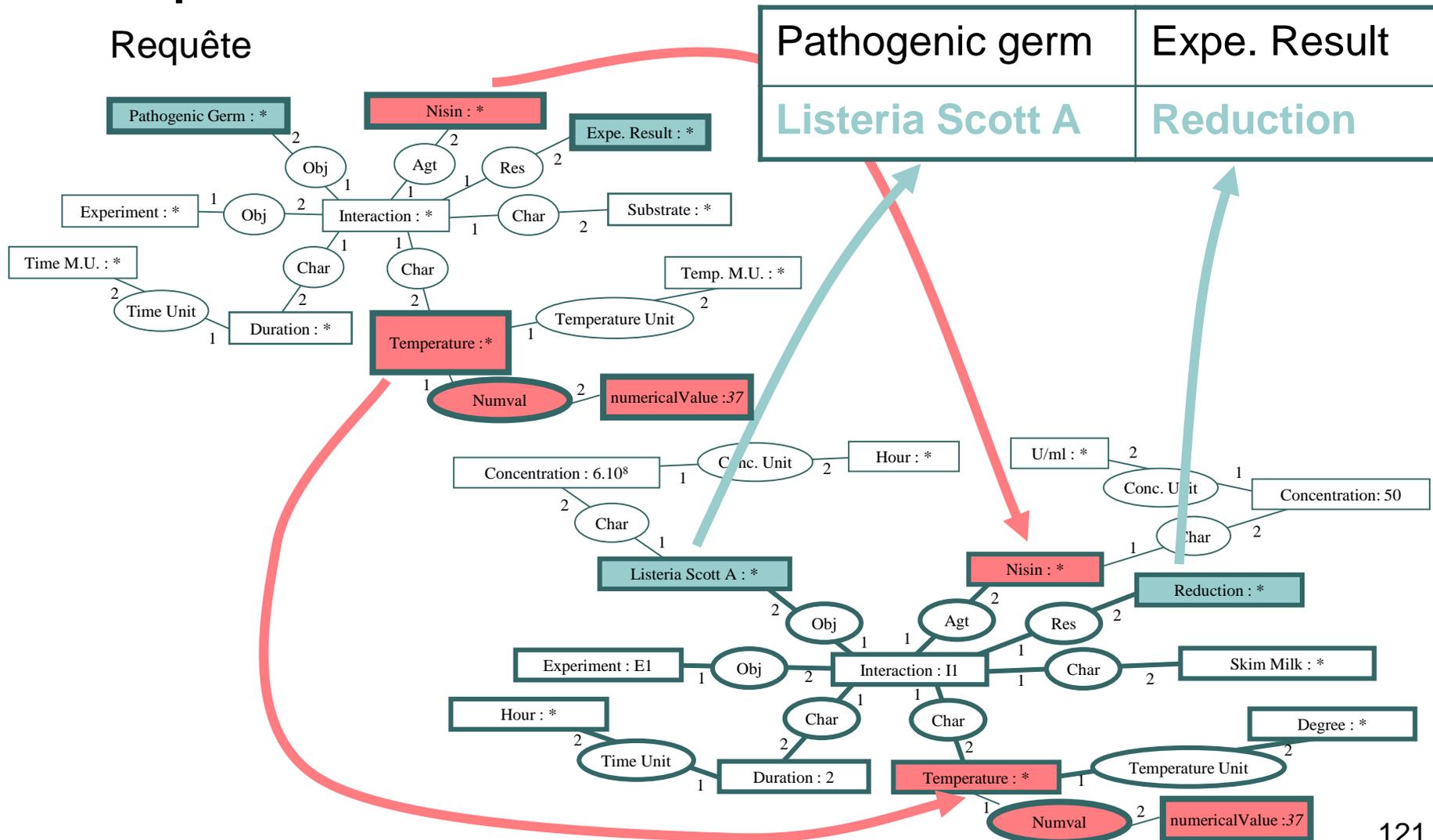
Exécution d'une requête GC

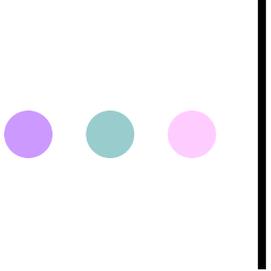
Le graphe requête : instantiation partielle du graphe vue



Exécution d'une requête GC

Requête



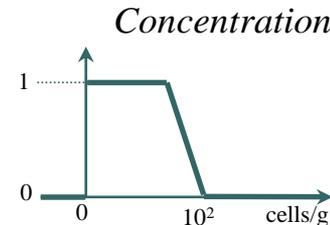
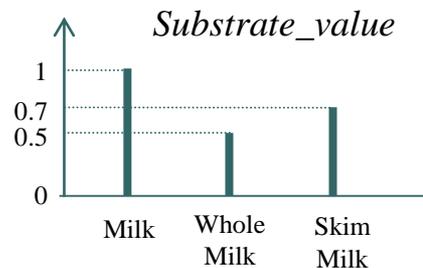
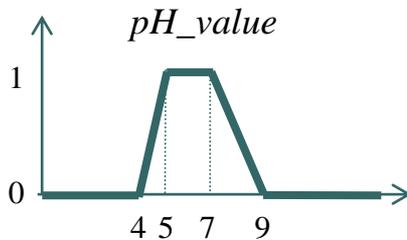


Interrogation flexible

- Interrogation flexible via un adaptateur GC
 - Le modèle des graphes conceptuels
 - Le sous-système GC
 - **Représentation du flou dans le modèle des GC (Thèse Rallou Thomopoulos, déc 2013)**
 - Extension de l'adaptateur GC
- Interrogation flexible d'une BDR
- Interrogation flexible via un adaptateur SPARQL

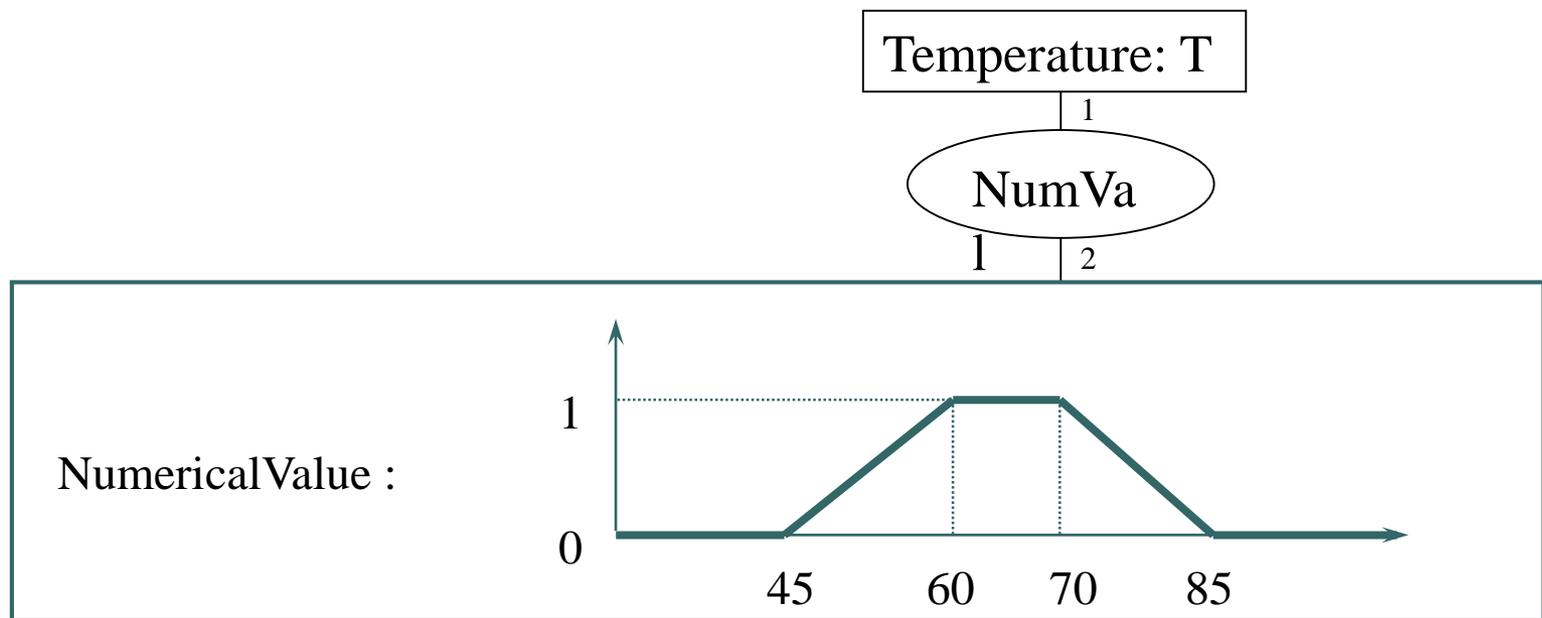
Représentation du flou dans le modèle des GC

- Utilisation des sous-ensembles flous
 - représentation de données imprécises et des données annotées à l'aide de distributions de possibilités
- Valeurs possibles d'un attribut a
 - Les sous-ensembles flous sont définis sur l'ensemble des valeurs possibles de l'attribut a noté $Ref(a)$



Graphes conceptuels flous

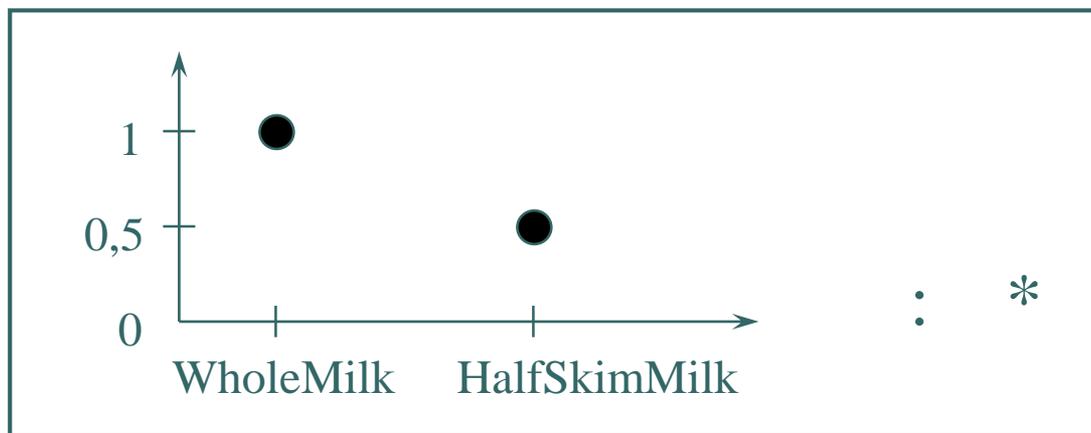
- Exemple 1 : une température floue



- Un **concept avec marqueur flou** est un concept dont le marqueur est un sous-ensemble flou défini sur l'ensemble des marqueurs individuels du support

Graphes conceptuels flous

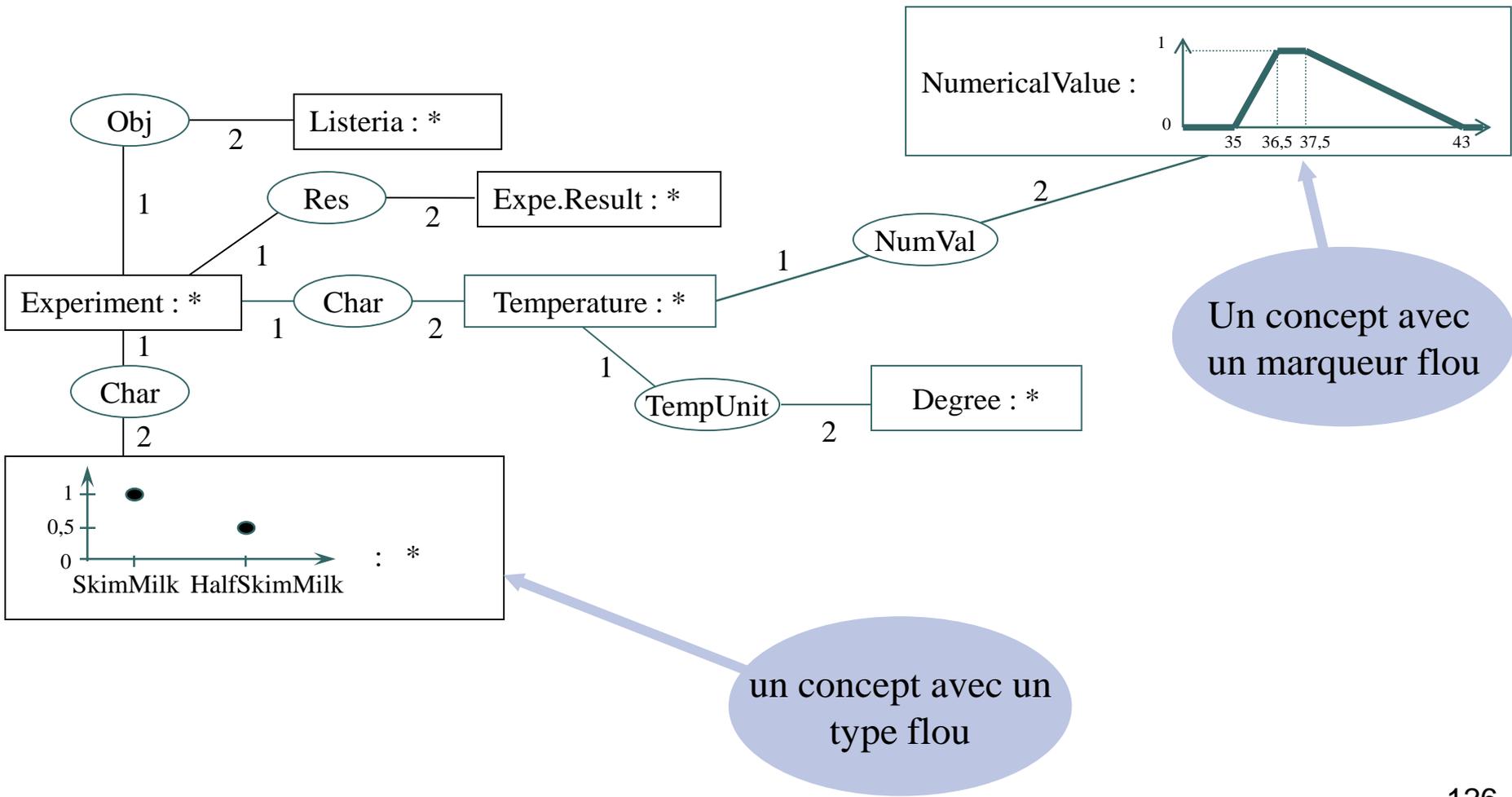
- Exemple 2 : un substrat flou

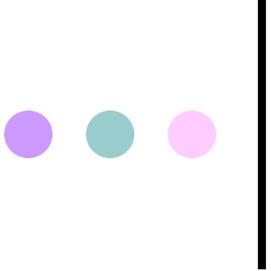


- Un **concept avec type flou** est un concept dont le type est un sous-ensemble flou défini sur un ensemble de types de concepts du support. Son marqueur est générique.

Interrogation flexible

Graphes conceptuels flous





Interrogation flexible

- Interrogation flexible via un adaptateur GC
 - Le modèle des graphes conceptuels
 - Le sous-système GC
 - Représentation du flou dans le modèle des GC
 - **Extension de l'adaptateur GC (Thèse Rallou Thomopoulos, déc 2013)**
- Interrogation flexible d'une BDR
- Interrogation flexible via un adaptateur SPARQL

Interrogation
flexible

Critères de sélection flous (1)

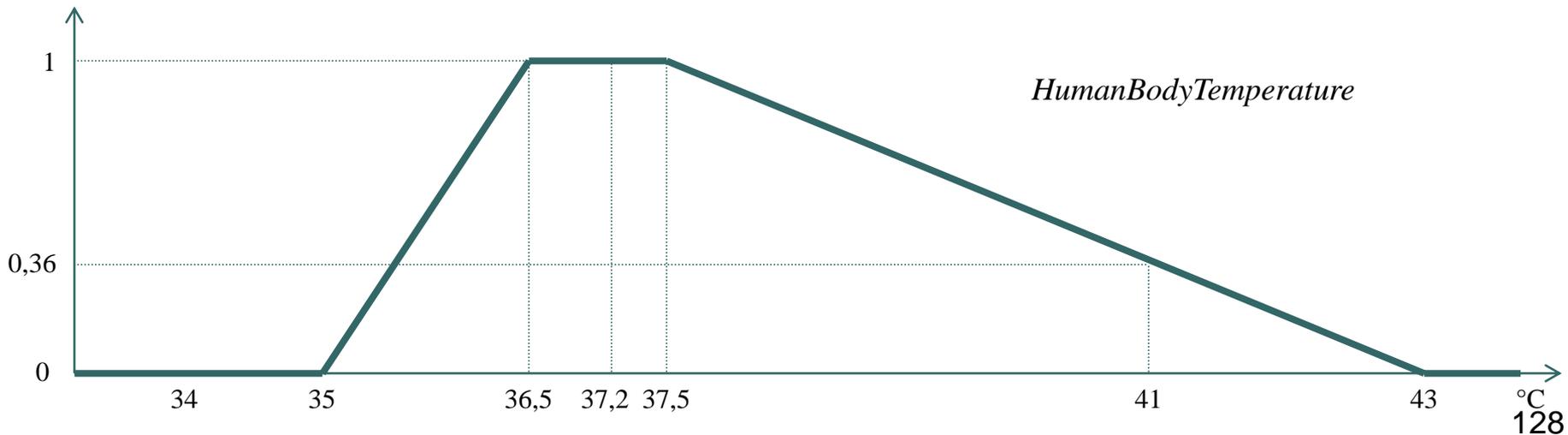
○ Expressions de préférences

(attr \approx val)

attribute name

fuzzy set

(Temperature \approx HumanBodyTemperature)

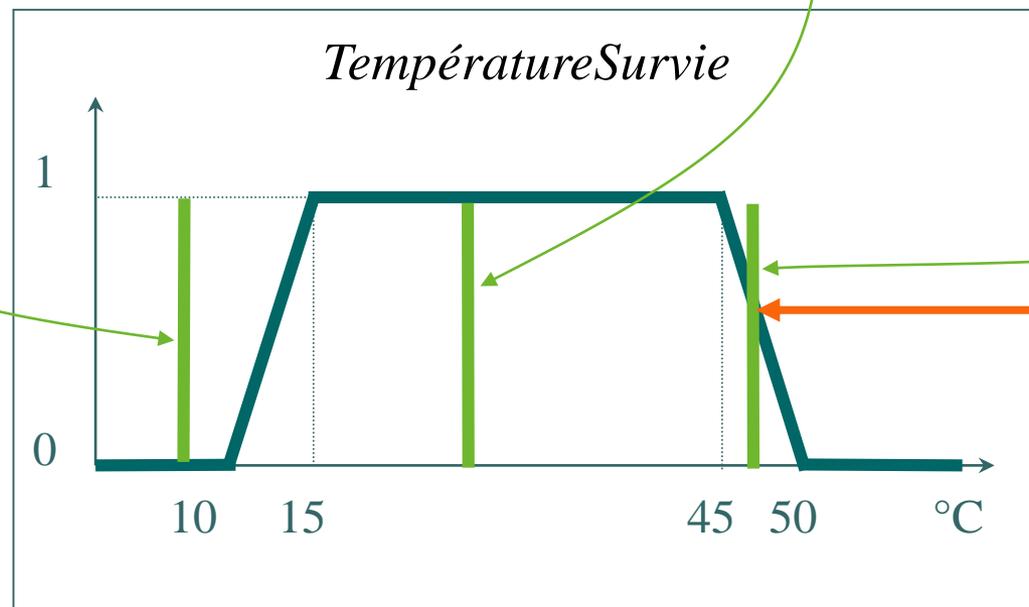


Critères de sélection flous (2)

- Appariement d'un critère de sélection flou à une valeur précise

*ne répond pas
à la requête*

répond à la requête



0,57

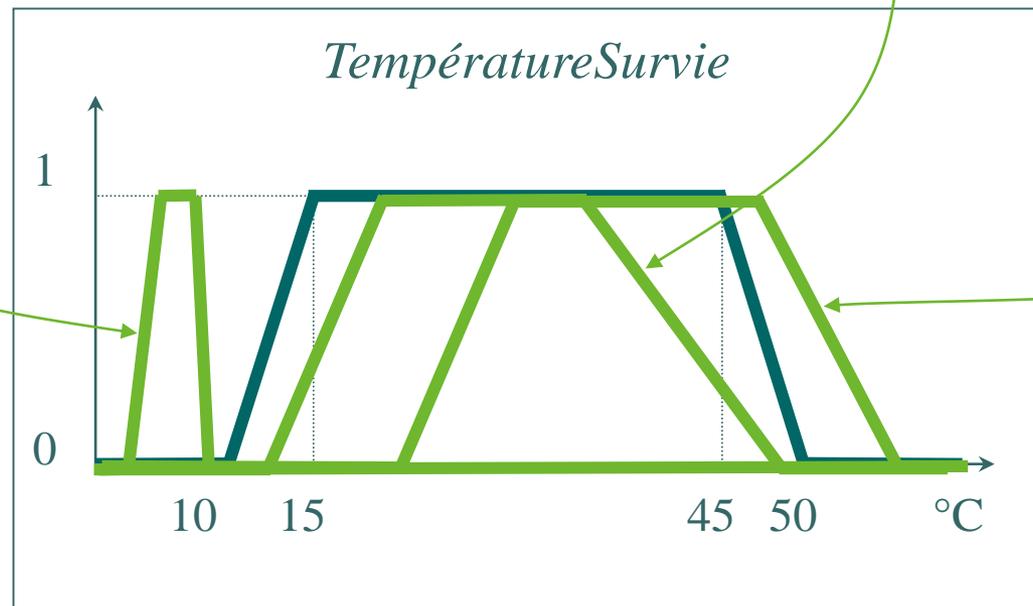
???

Critères de sélection flous (3)

- Appariement d'un critère de sélection flou à une valeur floue

*ne répond pas
à la requête*

répond à la requête



Degrés d'adéquation d'une donnée floue à un critère flou

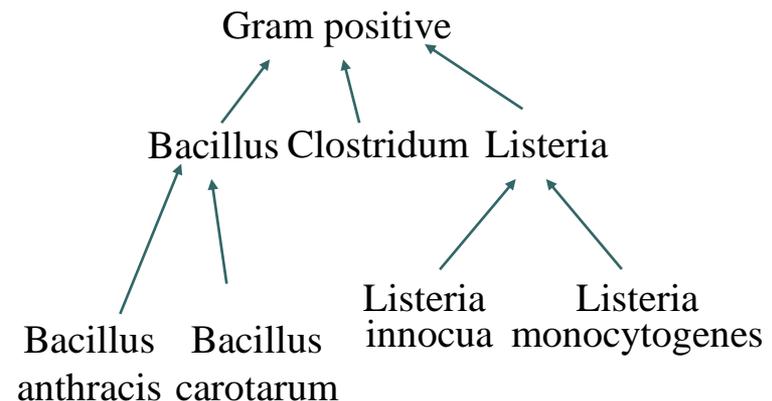
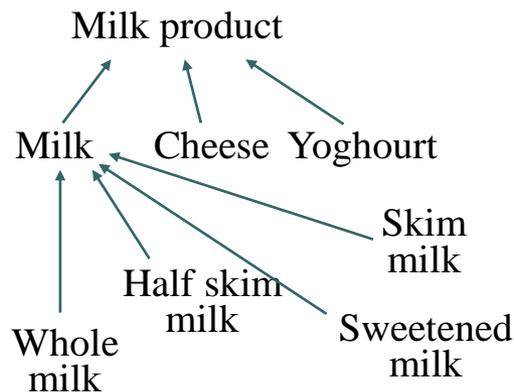


<u>Degrés d'adéquation</u>	Degré de possibilité [Zadeh 78]	Degré de nécessité [Dubois&Prade 88]
<u>Formule</u>	$\Pi(Q;D) = \sup (\min (\mu_Q; \pi_D))$	$N(Q;D) = 1 - \Pi(\bar{Q};D)$ $= \inf (\max (\mu_Q; 1-\pi_D))$
<u>Exemple</u>		
<u>Signification</u>	mesure d'intersection	mesure d'inclusion

- Les sous-ensembles flous à comparer doivent être définis sur le même domaine

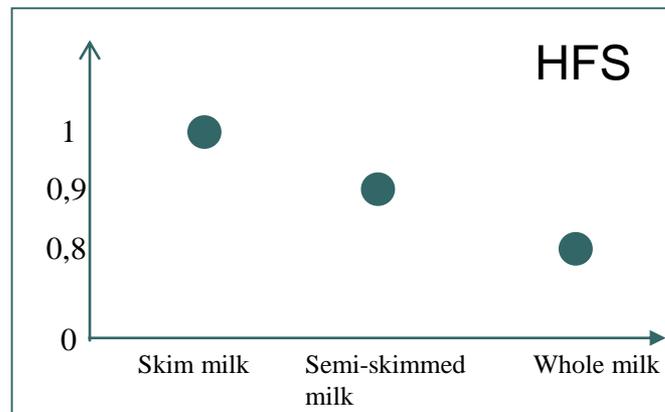
Ensembles flous définis sur un domaine de valeurs hiérarchisé

- Dans nos applications biologiques, les domaines de valeurs sont organisés en taxonomie:
 - Une première relation d'ordre représentée par la relation « sorte de » est définie entre les valeurs
 - Exemples des taxonomies « food products » et « microorganisms ».



Ensembles flous définis sur un domaine de valeurs hiérarchisé

- Les critères de sélection flous associent des préférences aux valeurs recherchées
 - Une relation d'ordre est définie sur le domaine de valeurs.



Ensembles flous définis sur un domaine de valeurs hiérarchisé

- Une relation d'ordre pour définir des préférences sur le domaine de valeurs
- Une deuxième relation d'ordre représentée par la relation « sorte de » est également définie.
- L'hypothèse classique d'indépendance entre les degrés pour les domaines de valeurs « plats » n'est plus vraie.

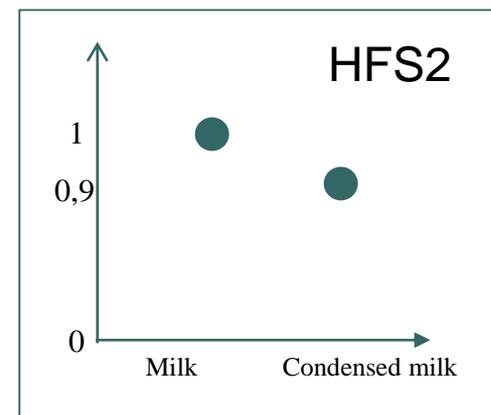
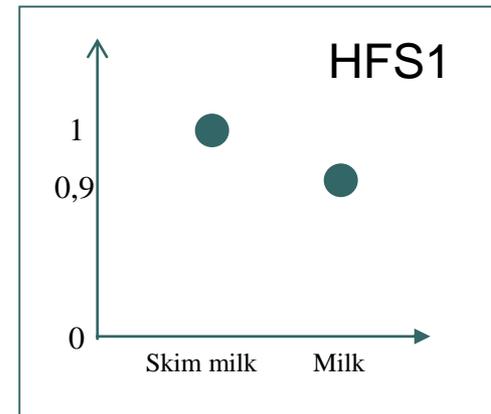
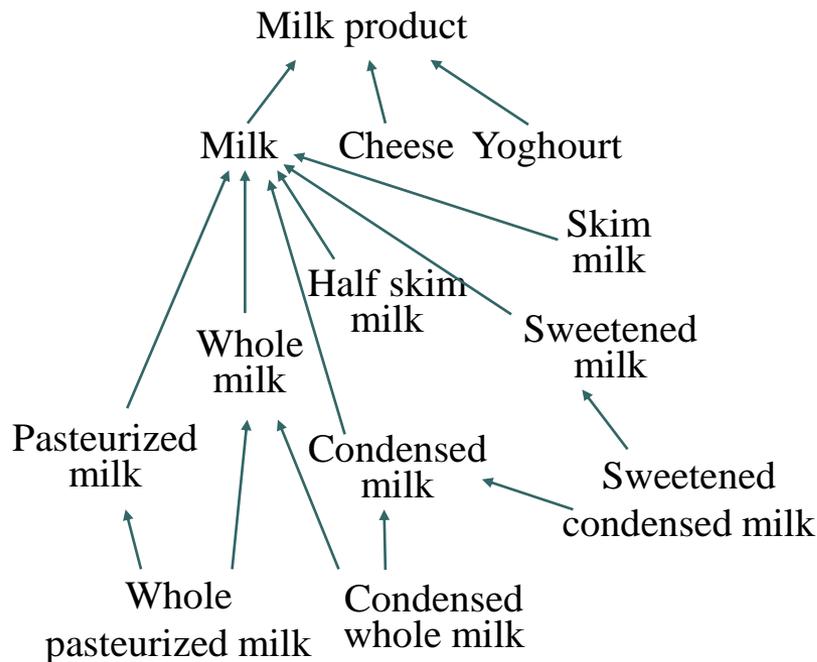
Les deux relations d'ordre doivent être mises en adéquation.

Sous-ensembles flous définis sur un domaine hiérarchisé

- Quelle signification quand deux termes comparables (au sens de la relation « sorte de ») ont des degrés de préférence différents ?
- Peut-on utiliser la taxonomie pour élargir les requêtes afin d'obtenir plus de réponses pertinentes ?

Notion de sous-ensemble flou hiérarchique (HFS)

- Un HFS est un ensemble flou défini sur un **sous-ensemble** de sa hiérarchie de référence.



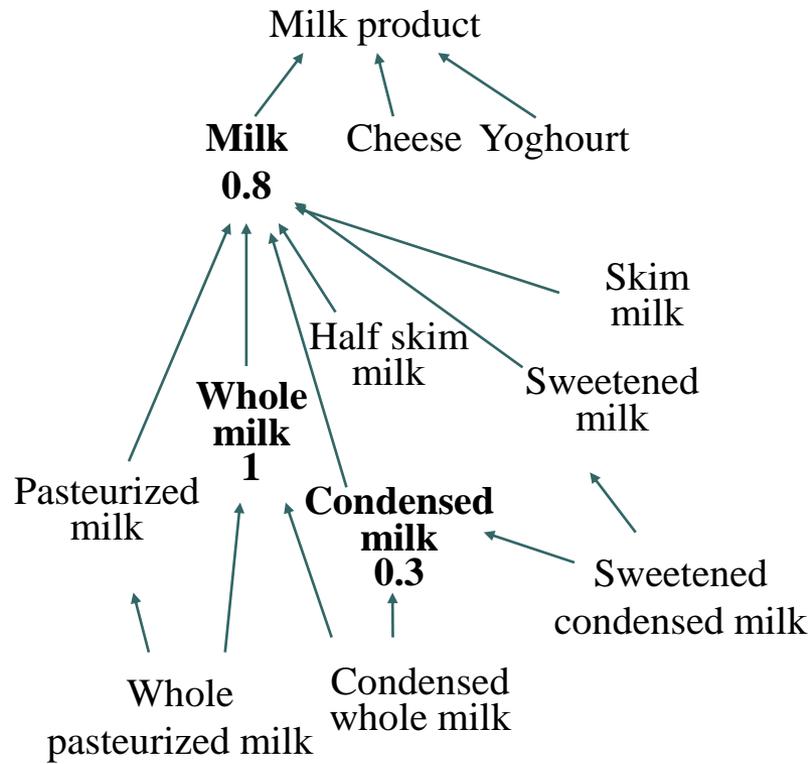
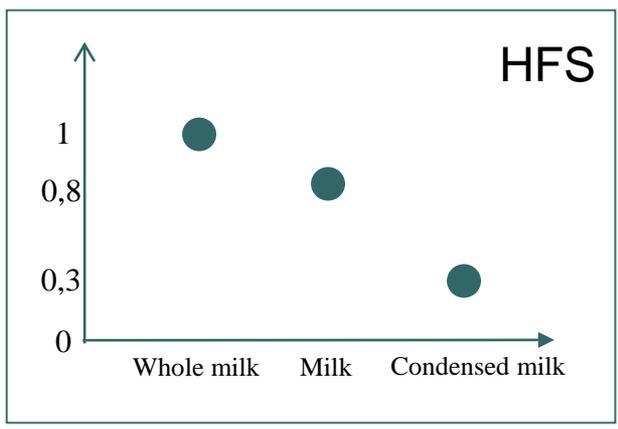
Notion de sous-ensemble flou hiérarchique (HFS) (2)

- Deux constatations
 - Sémantique: les utilisateurs considèrent que les termes plus spécifiques que ceux du HFS doivent également être recherchés
 - Opérationnelle: les opérateurs classiques de comparaison de la logique floue ne peuvent pas être utilisés directement sur les HFS



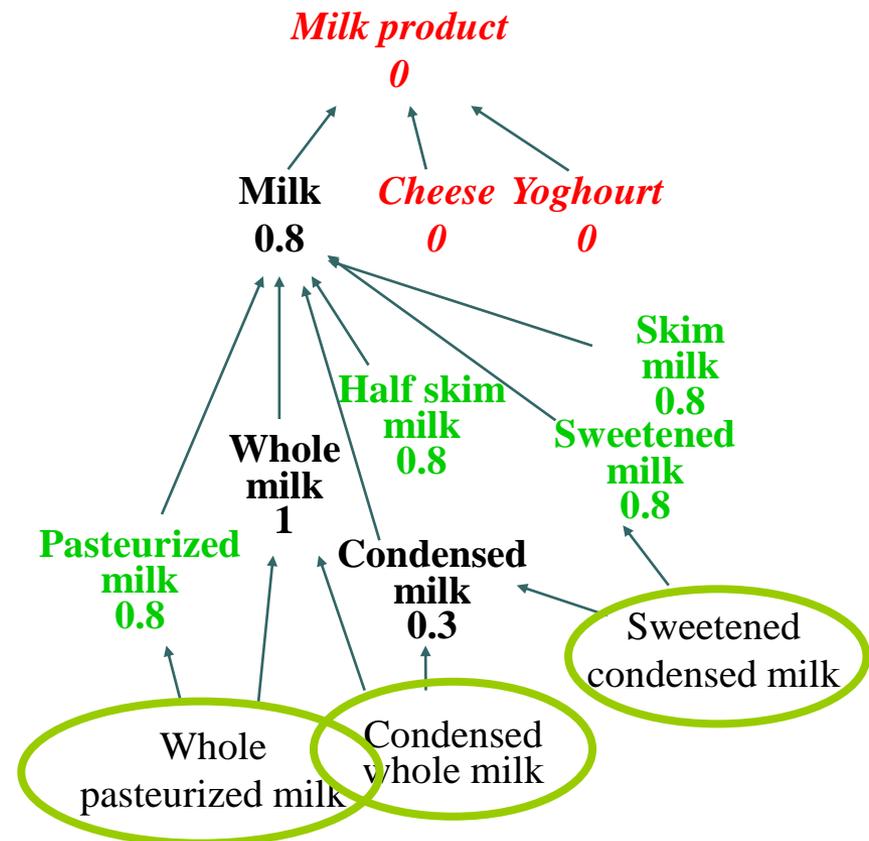
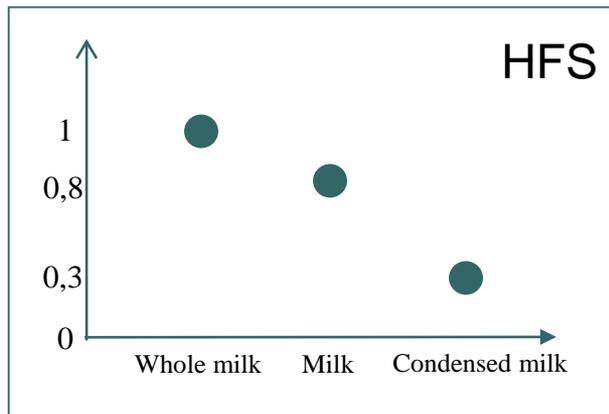
Fermeture d'ensemble flou hiérarchique

- La fermeture d'un HFS est un HFS défini sur l'intégralité de sa hiérarchie de référence en utilisant la relation « sorte de ».



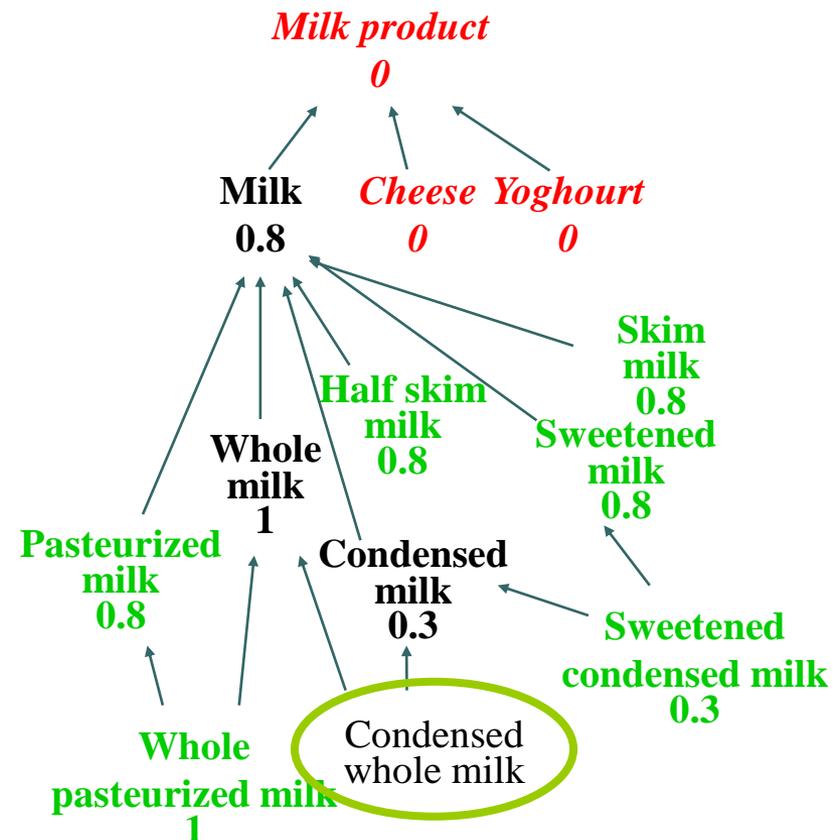
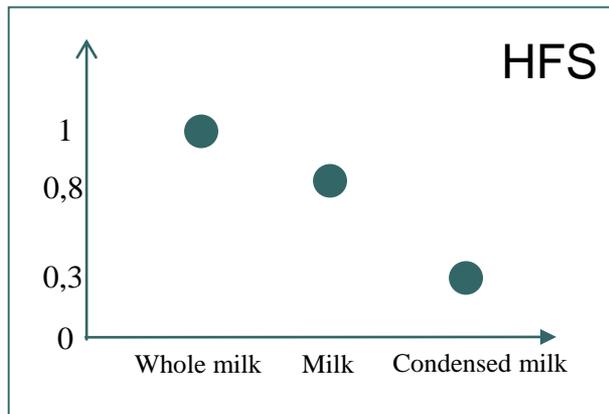
Fermeture d'ensemble flou hiérarchique

- La fermeture d'un HFS est un HFS défini sur l'intégralité de sa hiérarchie de référence en utilisant la relation « sorte de ».



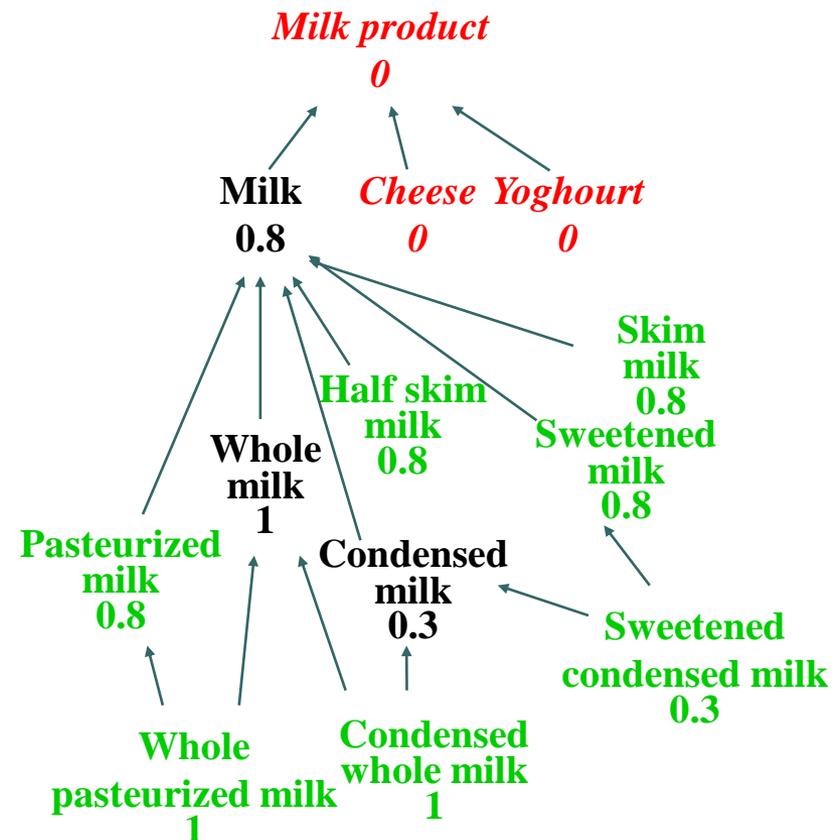
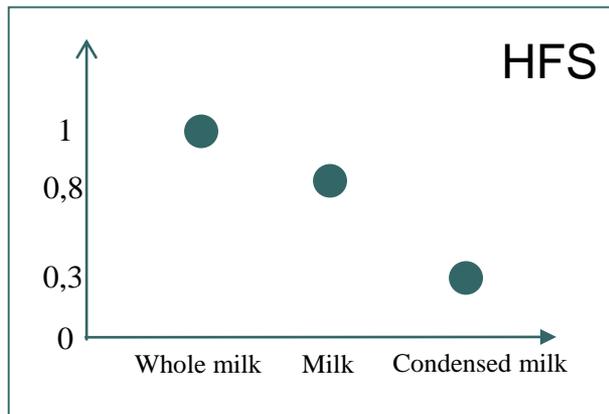
Fermeture d'ensemble flou hiérarchique

- La fermeture d'un HFS est un HFS défini sur l'intégralité de sa hiérarchie de référence en utilisant la relation « sorte de ».



Fermeture d'ensemble flou hiérarchique

- La fermeture d'un HFS est un HFS défini sur l'intégralité de sa hiérarchie de référence en utilisant la relation « sorte de ».



Interrogation flexible



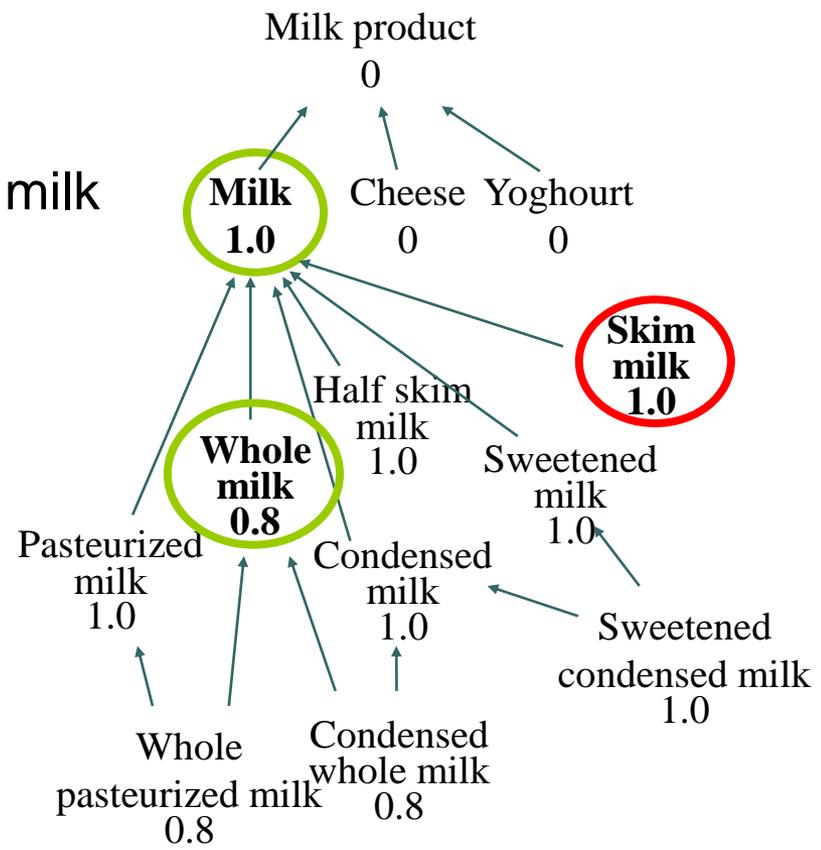
Notion d'ensemble flou hiérarchique minimal

- Deux HFS différents peuvent avoir même fermeture.

HFS1= 1.0/Milk + 0.8/Whole milk

HFS2= 1.0/Milk + 0.8/Whole milk + 1.0/Skim milk

Dans HFS2, Skim milk est **déductible**.



Notion d'ensemble flou hiérarchique minimal

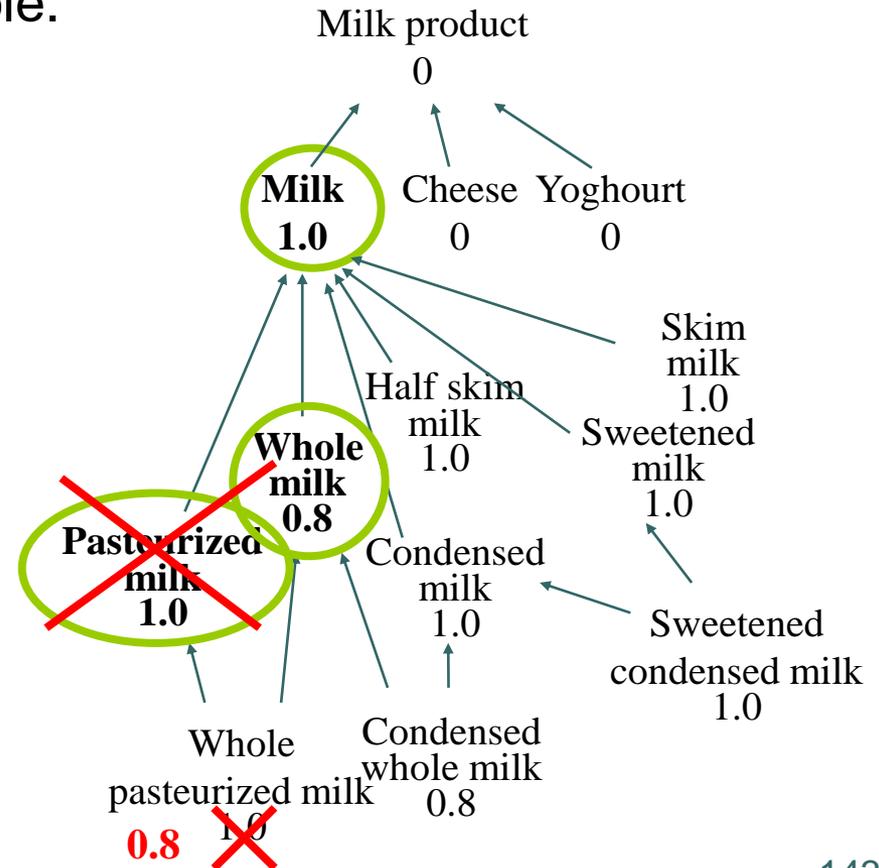
- Dans une classe d'équivalence donnée (pour une fermeture F donnée), un HFS est dit minimal si sa fermeture est F et si aucun des termes de son domaine de valeurs n'est déductible.

HFS1 = 1.0/Milk + 0.8/Whole milk + 1.0/Pasteurized milk

1.0/Pasteurized milk est déductible de 1.0/Milk, mais on ne peut pas le supprimer !

HFS2 = 1.0/Milk + 0.8/Whole milk + 1.0/Whole pasteurized milk

Le HFS minimal (HFS2) est **unique**



Notion d'ensemble flou hiérarchique minimal

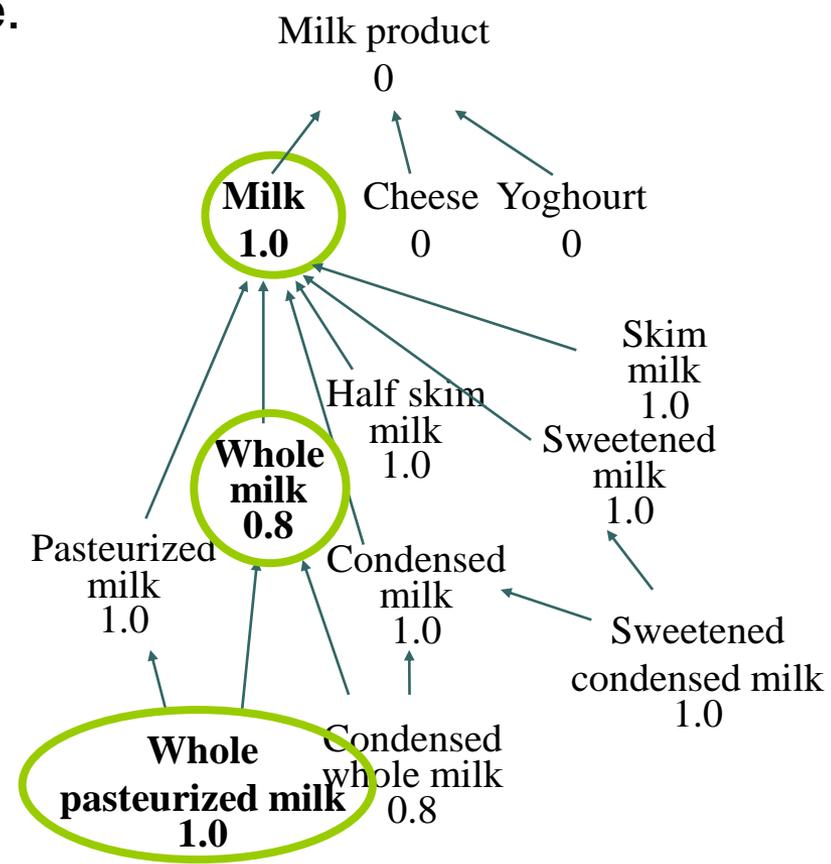
- Dans une classe d'équivalence donnée (pour une fermeture F donnée), un HFS est dit minimal si sa fermeture est F et si aucun des termes de son domaine de valeurs n'est déductible.

HFS1 = 1.0/Milk + 0.8/Whole milk + 1.0/Pasteurized milk

1.0/Pasteurized milk est déductible de 1.0/Milk, mais on ne peut pas le supprimer !

HFS2 = 1.0/Milk + 0.8/Whole milk + 1.0/Whole pasteurized milk

Le HFS minimal (HFS2) est **unique**

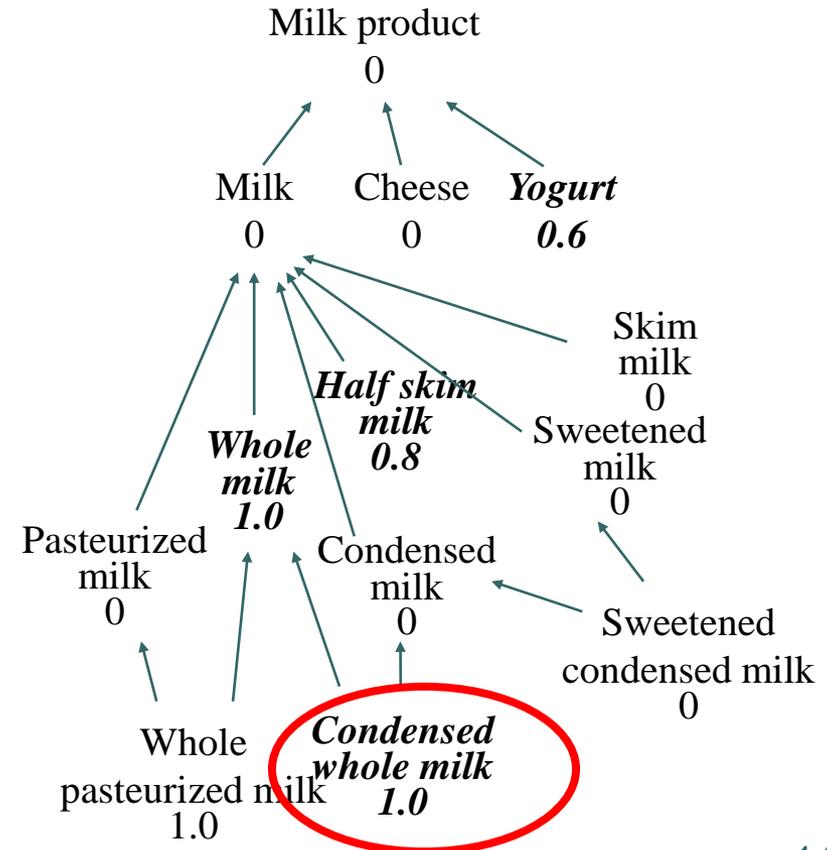
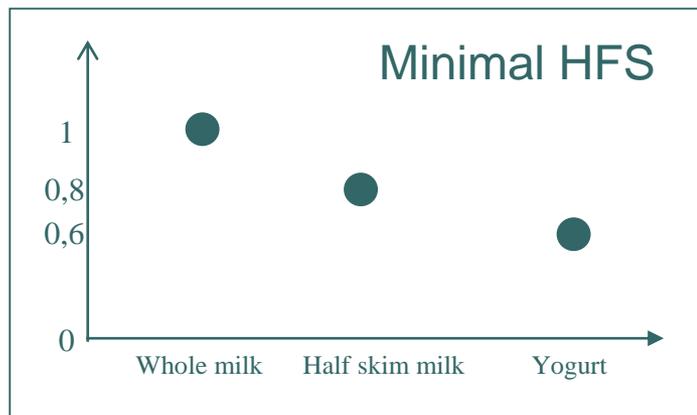
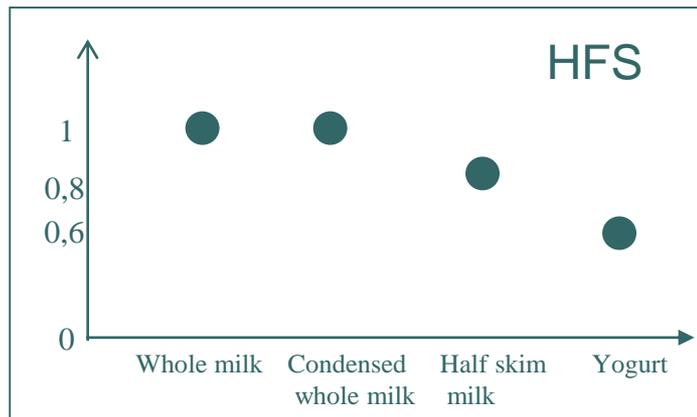


Généralisation d'un ensemble flou hiérarchique

- Utiliser un ensemble flou hiérarchique comme critère de sélection ne garantit pas que l'utilisateur récupérera suffisamment de données.
- Idée: Généralisation du HFS en utilisant la hiérarchie de référence pour obtenir plus de réponses pertinentes.

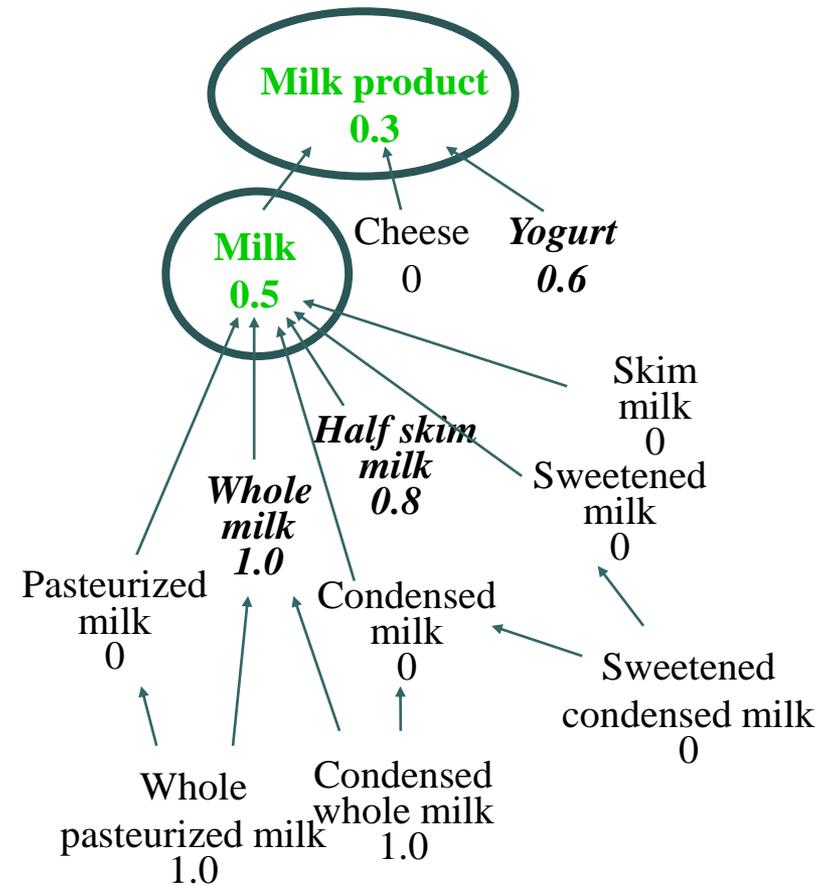
Généralisation d'un ensemble flou hiérarchique

Etape 1: La généralisation est effectuée sur l'HFS minimal pour garantir la même généralisation pour tous les HFS d'une même classe d'équivalence.



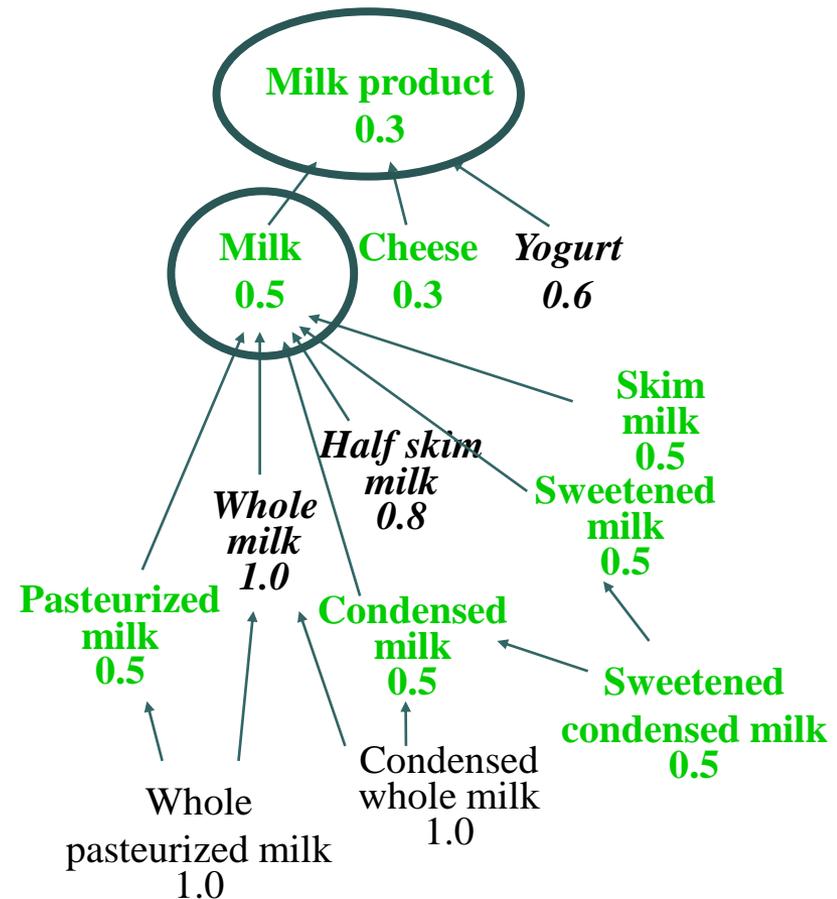
Généralisation d'un ensemble flou hiérarchique

- Etape 2: La généralisation est composée d'un ensemble de généralisation élémentaires



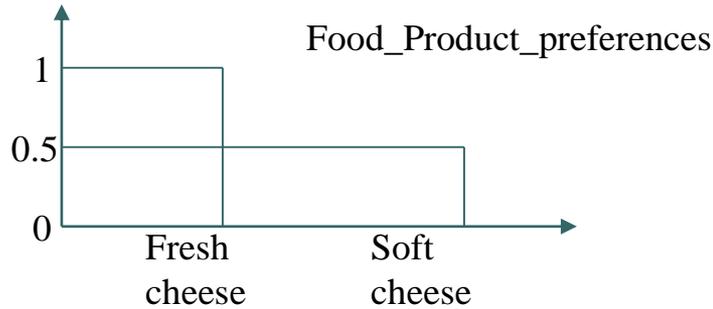
Généralisation d'un ensemble flou hiérarchique

- Etape 2: La généralisation est composée d'un ensemble de généralisation élémentaires
- Etape 3: La fermeture du HFS généralisé est calculée avant d'interroger la base.

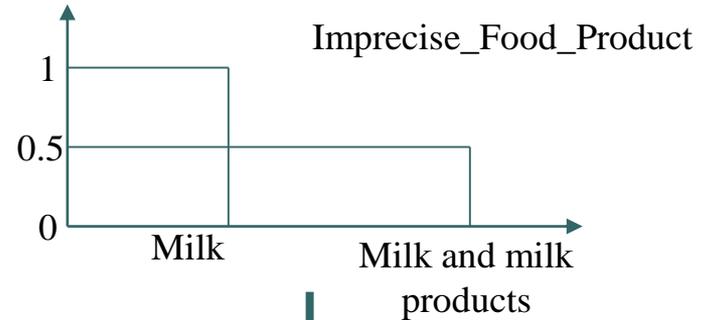
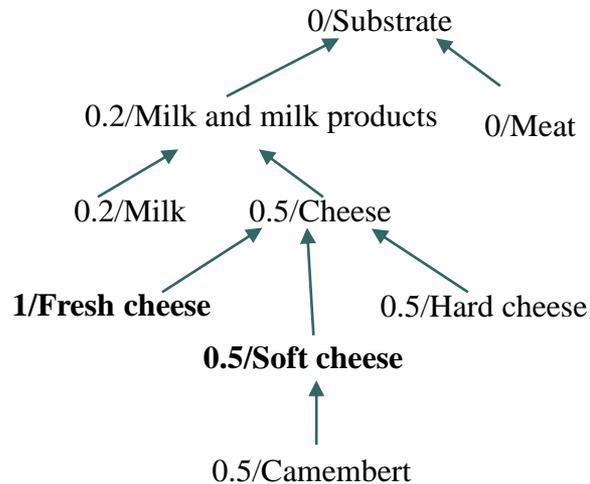


Interrogation flexible

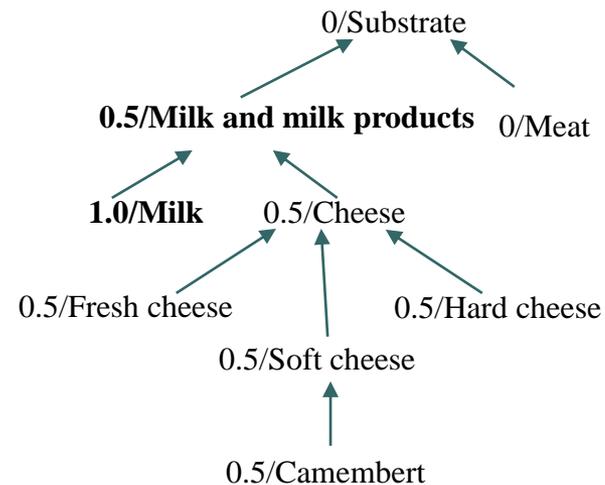
Comparaison des préférences de l'utilisateur aux données imprécises: rendre les domaines comparables



fermeture + généralisation



fermeture



Satisfaction d'un critère de sélection flou dans MIEL

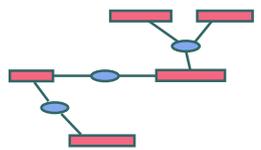
- Satisfaction d'un critère de sélection ($attr \approx val$)
 - degré de possibilité $\pi(val, \tau[attr])$
 - degré de nécessité $n(val, \tau[attr])$
- Adéquation d'un tuple à une requête ayant p critères de sélection
 - $\Pi = \min(\pi_1, \dots, \pi_p), N = \min(n_1, \dots, n_p)$



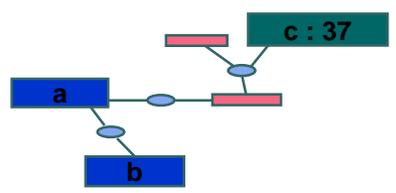
Exécution d'une requête GC (1)

Q1 : {a,b | **V1(a,b,c)** ∧ (c ≈ HumanBodyTemperature)}

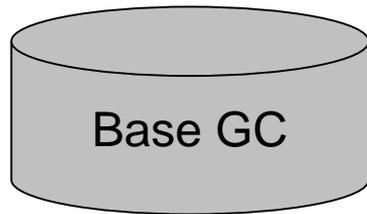
Sélection du (des) graphe(s) vue(s)



Génération de la (des) requête(s) GC



Projection approximative



A1 = {a1, b1, degrés}
A2 = {a2, b2, degrés}
...
An = {an, bn, degrés}

Prise en compte des données imprécises, des similarités, des degrés de pertinence

Exécution d'une requête GC (2)

- Requête:
« *Existe-t-il un niveau de contamination du riz et éventuellement des céréales par l'ochratoxin A compris entre 1.5 et 2 ?* »
- Sélection d'une vue : **Contamination Range**

Nom	Type résultat	Types d'accès
Contamination range	Contamination level	Food product, Contaminant, Samples total number
Loq Relation	Loq	Food product, Contaminant, Samples total number
Lod Relation	Lod	Food product, Contaminant, Samples total number
Number Samples positive	Samples positive	Food product, Contaminant, Samples total number

- Choix des attributs de projection
Contamination level, Food Product, Contaminant

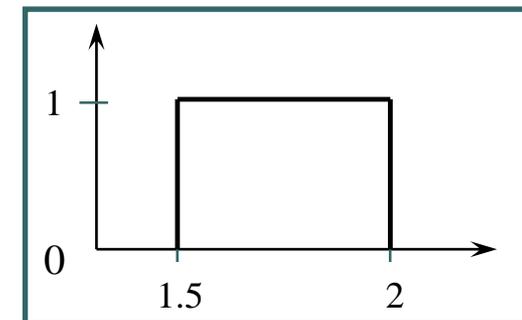
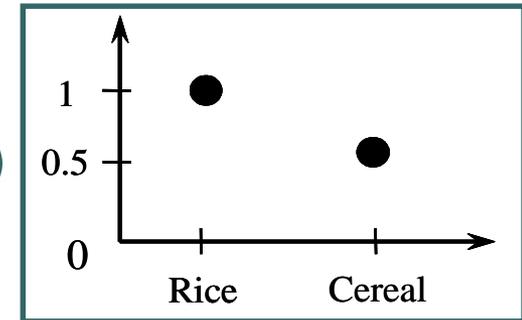
Exécution d'une requête GC (3)

- Requête:
« *Existe-t-il un niveau de contamination du riz et éventuellement des céréales par l'ochratoxin A compris entre 1.5 et 2 ?* »
- Définition des critères de sélection

$$\text{Food Product} = (1/\text{Rice} + 0.5/\text{Cereal})$$

$$\text{Contaminant} = (1/\text{Ochratoxin A})$$

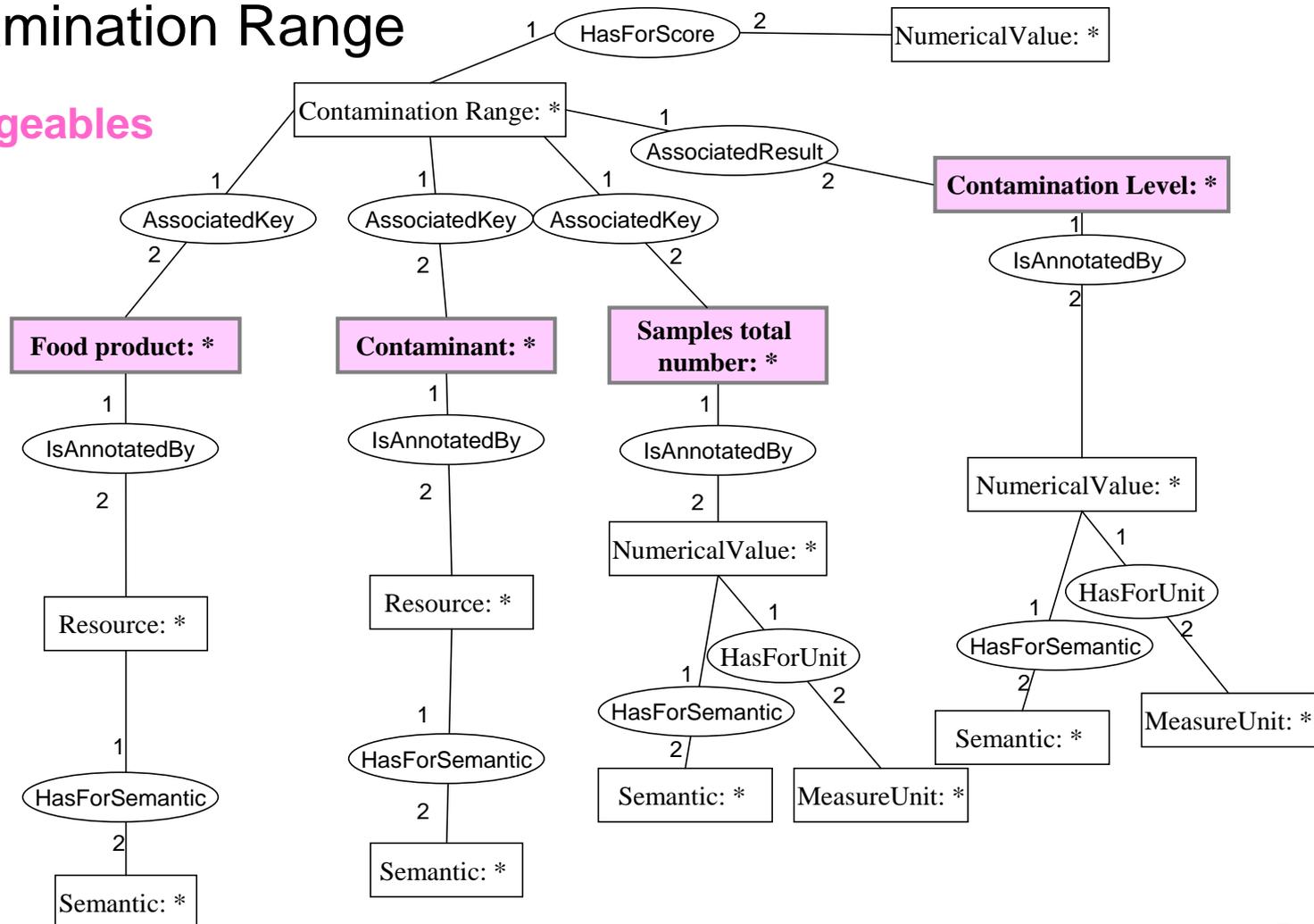
$$\text{Contamination level} = [1.5, 2]$$



Exécution d'une requête GC (4)

Vue Contamination Range

Attributs interrogeables

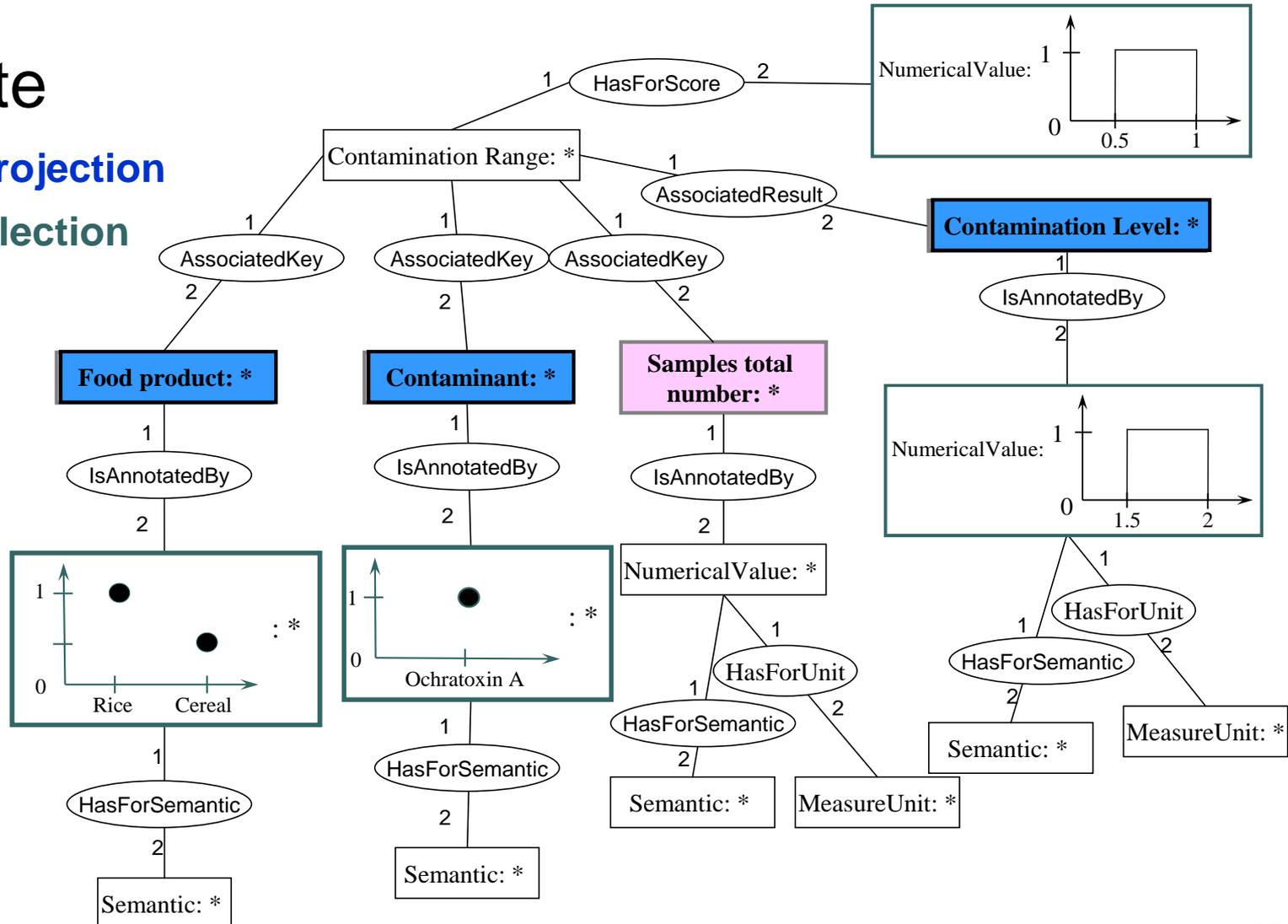


Exécution d'une requête GC (5)

Requête

Attributs de projection

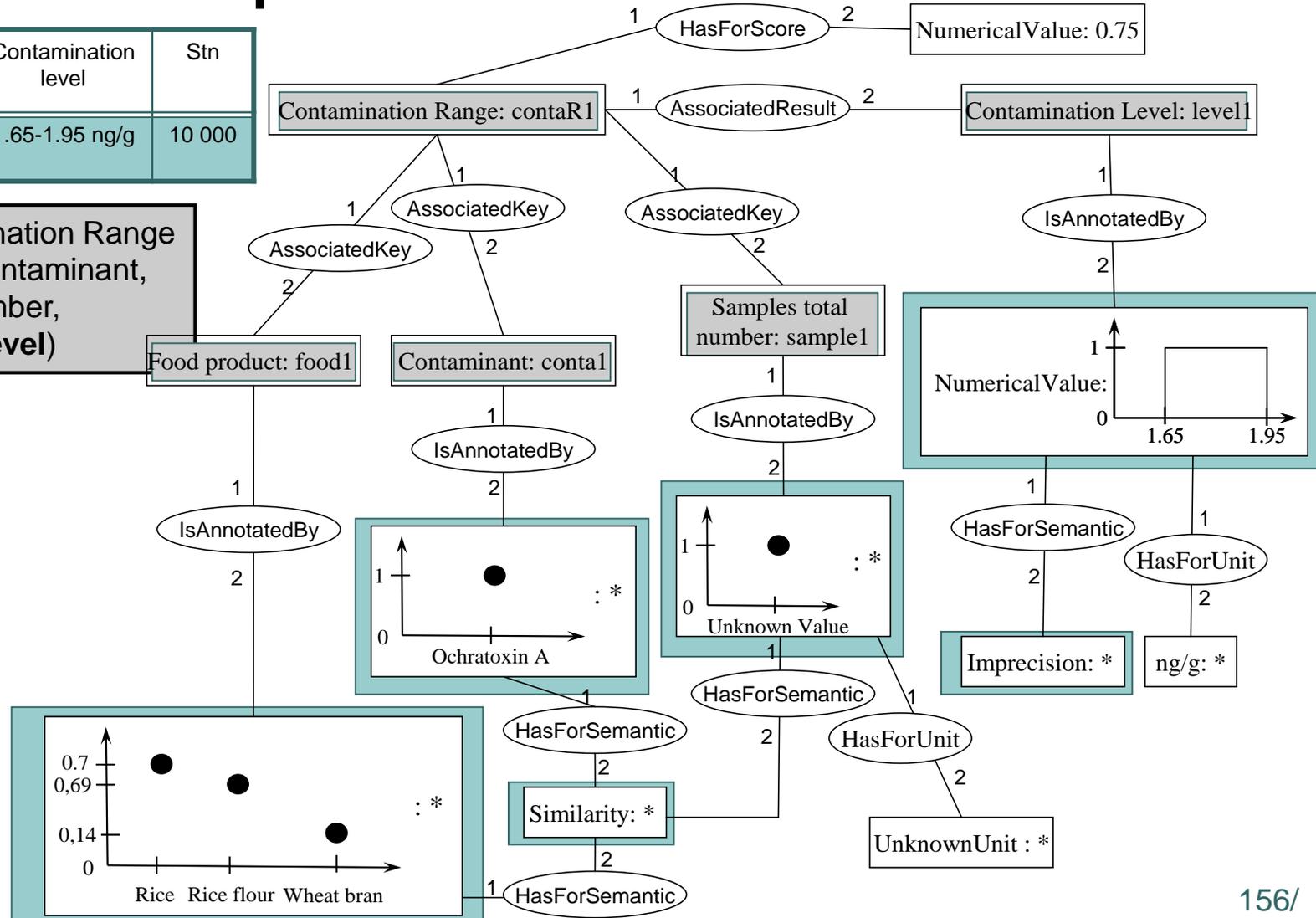
Critères de sélection



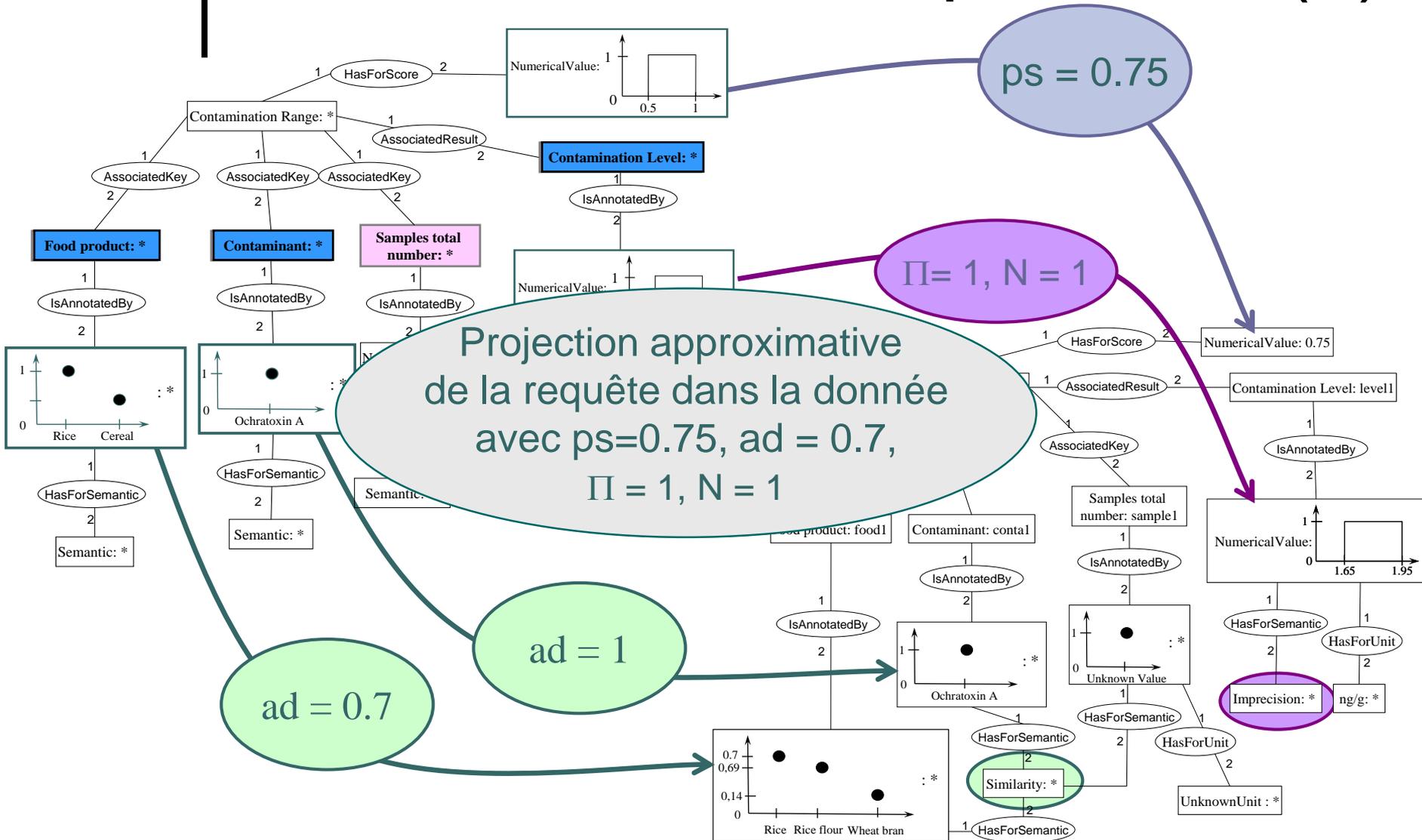
Un exemple de graphe conceptuel flu

Food	Contaminant	Contamination level	Stn
Rice bran	Ochratoxin A	1.65-1.95 ng/g	10 000

Relation Contamination Range
(Food product, Contaminant, Samples total number, Contamination level)

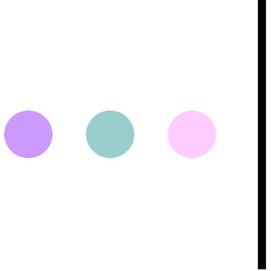


Exécution d'une requête GC (6)



Conclusion sur l'interrogation flexible via un adaptateur GC

- Expression de préférences dans les critères de sélection : solution à l'incomplétude des bases locales et des données du Web
- Extension de l'adaptateur GC : solution à l'hétérogénéité des données locales et des données du Web
- Prise en compte des différentes sémantiques du flou dans l'interrogation :
 - degré d'adéquation pour les données imprécises,
 - degré de possibilité et de nécessité pour les similarités,
 - score de pertinence de l'annotation par des relations de l'ontologie,
 - ordre total sur les réponses



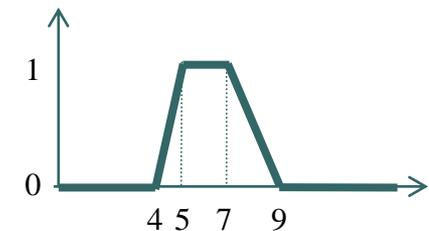
Interrogation flexible

- Interrogation flexible via un adaptateur GC
 - Le modèle des graphes conceptuels
 - Le sous-système GC
 - Représentation du flou dans le modèle des GC
 - Extension de l'adaptateur GC
- **Interrogation flexible d'une BDR**
- Interrogation flexible via un adaptateur SPARQL

Le sous-système relationnel

Valeurs numériques floues

Expeld	Substrate	Fuzzy pH
27	Pork	221
39	SkimMilk	223

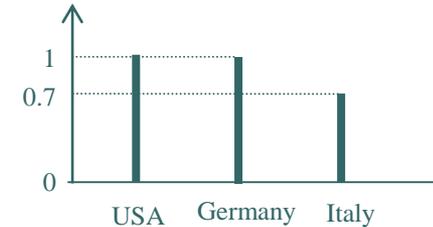


FuzzySetId	MinSupp	MinKer	MaxKer	MaxSupp
221	4	5	7	9
222	6	6	6	6
223	9	10	12	12.5

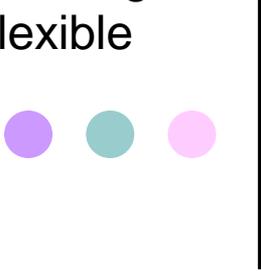
Le sous-système relationnel

- Valeurs symboliques ou hiérarchiques floues

Substrate	FuzzyOriginId
Pork	101



FuzzyOriginId	Value	Degree
101	USA	1.0
101	Germany	1.0
101	Italy	0.7



Le sous-système relationnel

- 2 tables permettent de stocker le domaine de référence pour les attributs à domaine hiérarchisé

*Ref*_{Substrate}

SubstrateName
Milk
WholeMilk
PasteurizedMilk
PasteurizedWholeMilk

*Hier*_{Substrate}

SubstrateName1	SubstrateName2
Milk	WholeMilk
Milk	PasteurizedMilk
PasteurizedMilk	PasteurizedWholeMilk
WholeMilk	PasteurizedWholeMilk

Exécution d'une requête BDR

$$Q1 : \{a,b \mid \mathbf{VI}(a,b,c) \wedge (c \approx \text{HumanBodyTemperature})\}$$

Sélection de la vue correspondant à la requête

Base relationnelle

$$\begin{aligned} A1 &= \{w1, \dots, wn, \delta1\} \\ A2 &= \{x1, \dots, xn, \delta2\} \\ A3 &= \{y1, \dots, yn, \delta3\} \\ A4 &= \{z1, \dots, zn, \delta4\} \end{aligned}$$

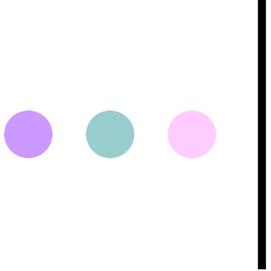
$$\mathbf{VI}(a,b,c) = \ll \text{from } R1, R2, R3 \text{ where } (R1.d = R2.d) \text{ and } (R2.e = R3.f) \gg$$

Exécution de la requête

« Défuzzification » de la requête

Construction de la requête SQL floue

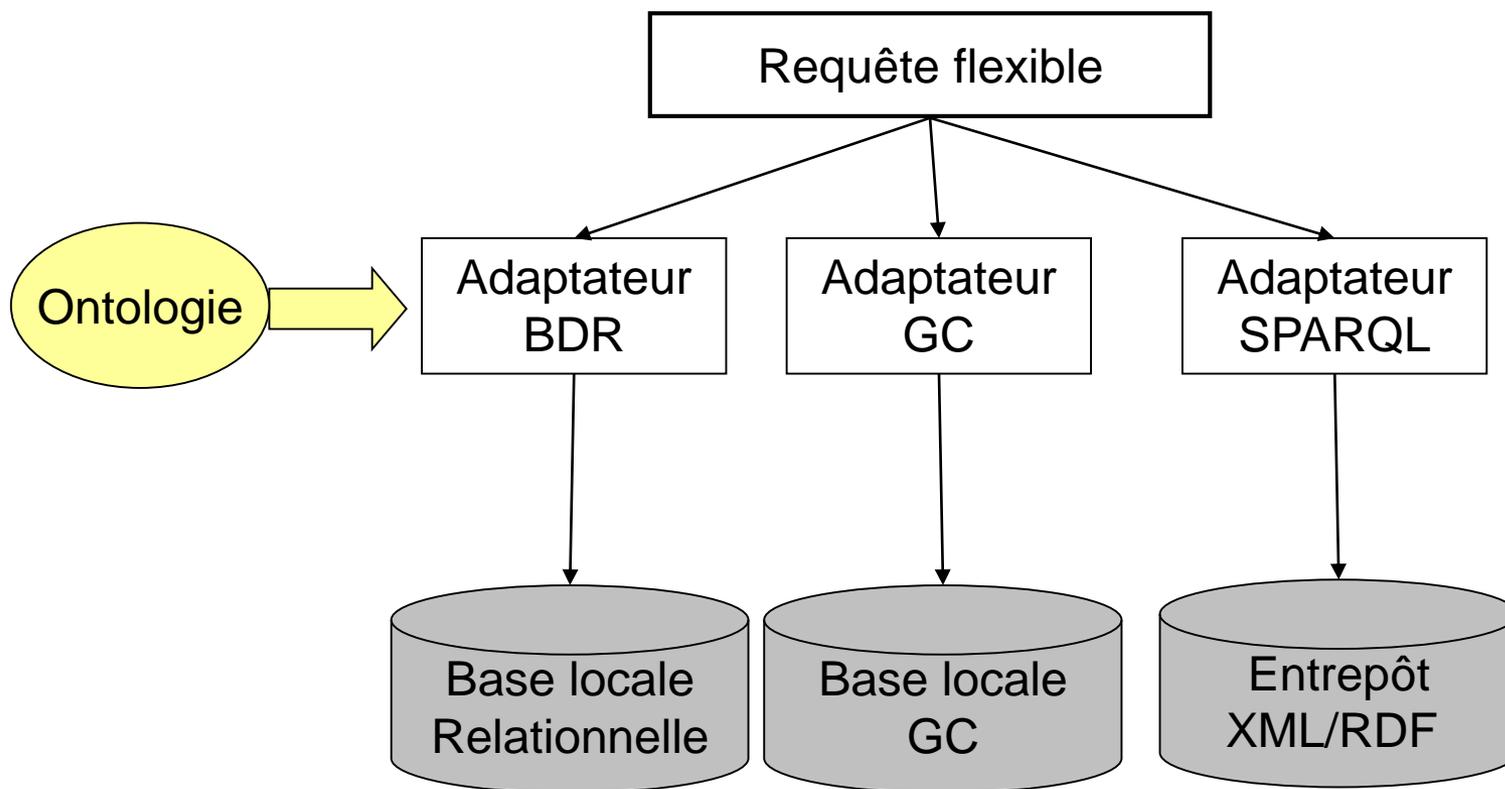
$$Q1' = \ll \text{select } a,b \text{ from } R1, R2, R3 \text{ where } (R1.d = R2.d) \text{ and } (R2.e = R3.f) \text{ and } (c \approx \text{HumanBodyTemperature}) \gg$$



Interrogation flexible

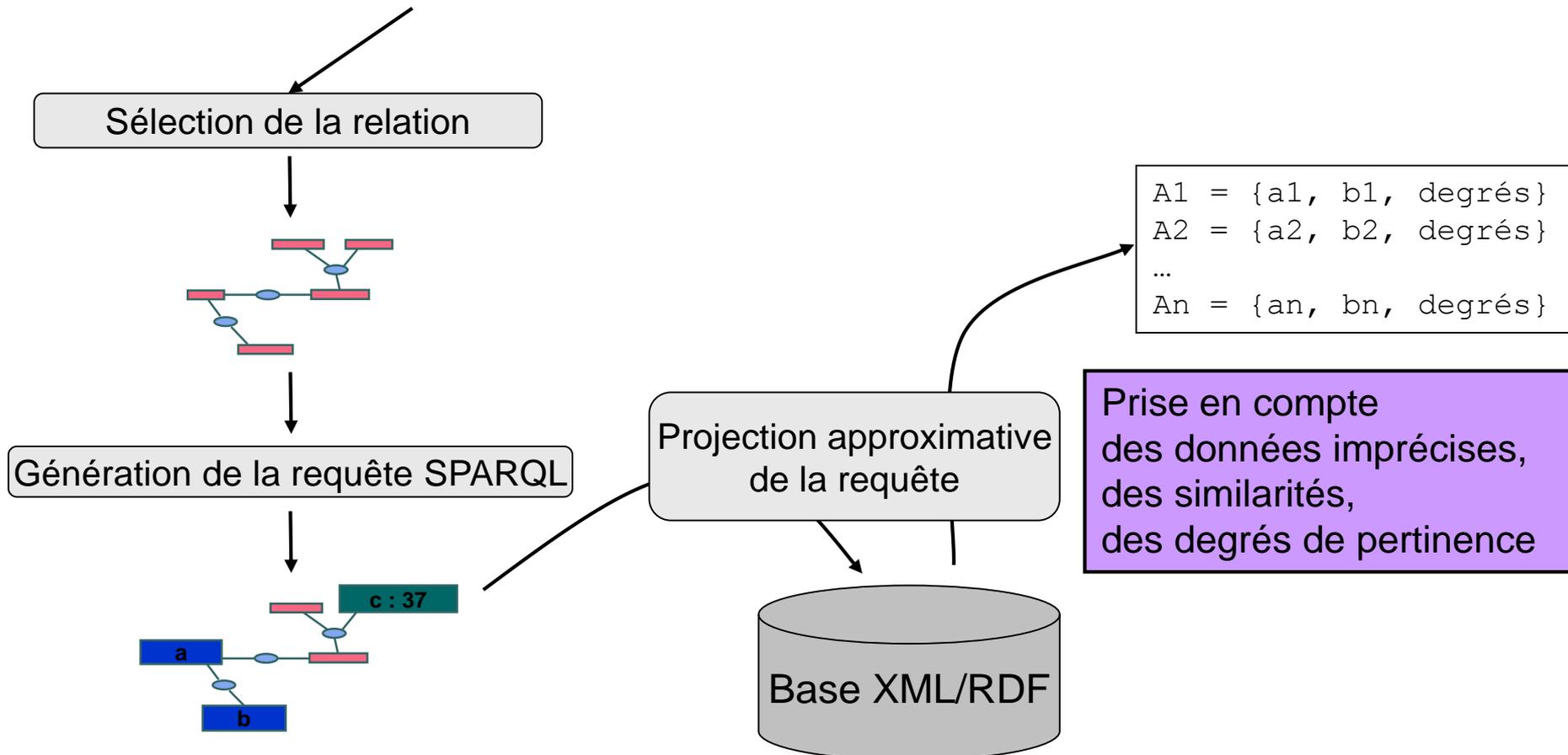
- Interrogation flexible via un adaptateur GC
 - Le modèle des graphes conceptuels
 - Le sous-système GC
 - Représentation du flou dans le modèle des GC
 - Extension de l'adaptateur GC
- Interrogation flexible d'une BDR
- **Interrogation flexible via un adaptateur SPARQL**

Interrogation flexible en utilisant SPARQL



Exécution d'une requête SPARQL

Requête : $\{a, b \mid R1(a, b, c) \wedge (c = 37)\}$



Interrogation
flexible

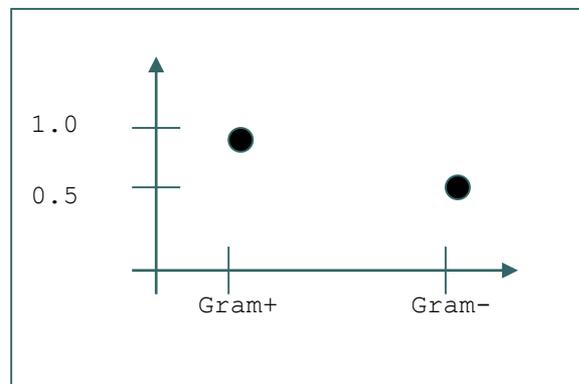
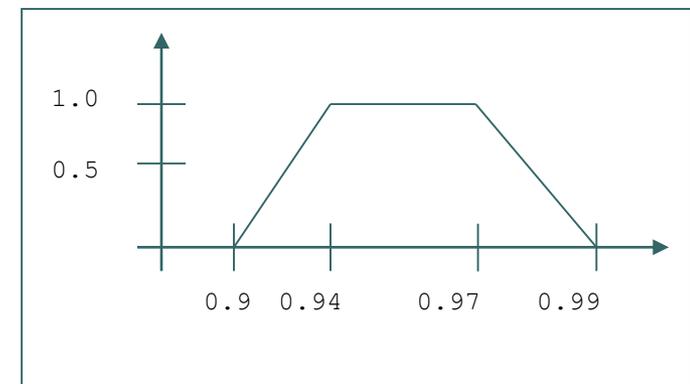
Un exemple de requête

Attributs de projection

Relation d'intérêt

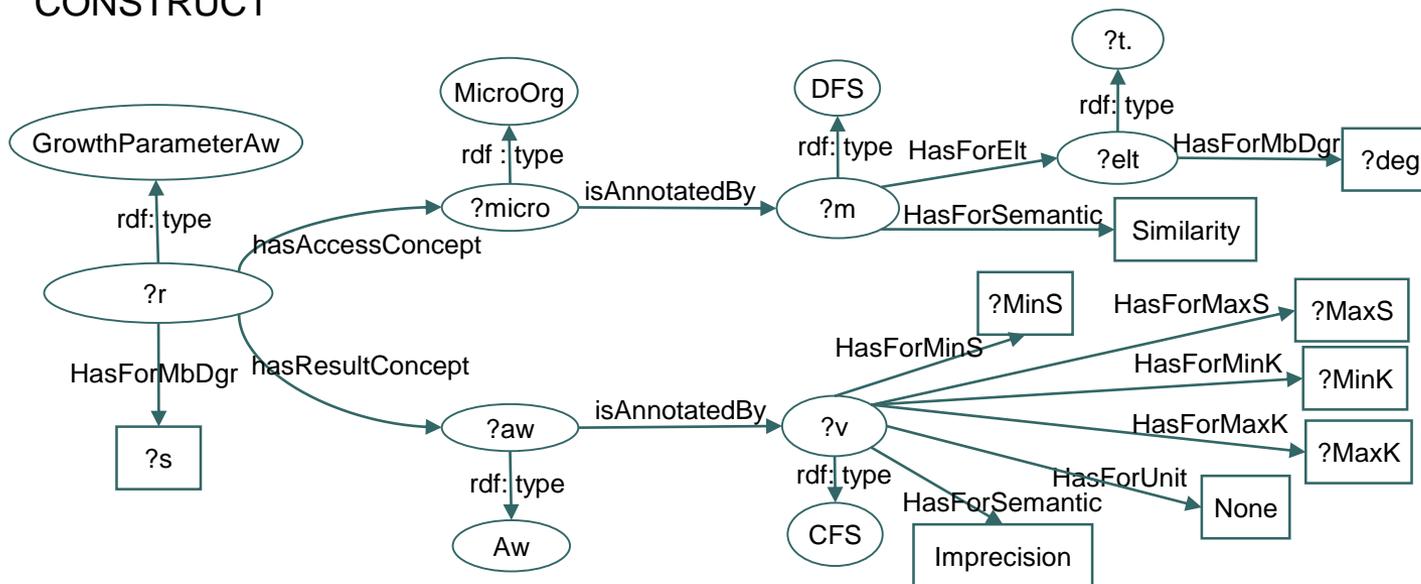
$\{ \text{Microorganism, aw} \mid \text{GrowthParameterAw}(\text{Microorganism, aw})$
 $\wedge (\text{Microorganism} \approx \text{MicroPref}) \wedge (\text{aw} \approx \text{awPref}) \}$

Attributs de
sélection

*MicroPref**AWPref*

Construction du pattern de la requête

CONSTRUCT



Transformation de la requête floue

- Les prédicats flous sont transformés en prédicat booléen (α -cut avec $\alpha=0$)

awPreferences = {0.9, 0.94, 0.97, 0.99}

0-cut

awInterval = [0.9, 0.99]

Transformation de la requête floue

- Les prédicats flous sont transformés en prédicat booléen(α -cut avec $\alpha=0$)

microorganismPreferences = {1.0/Gram+ + 0.5/Gram-}

fermeture

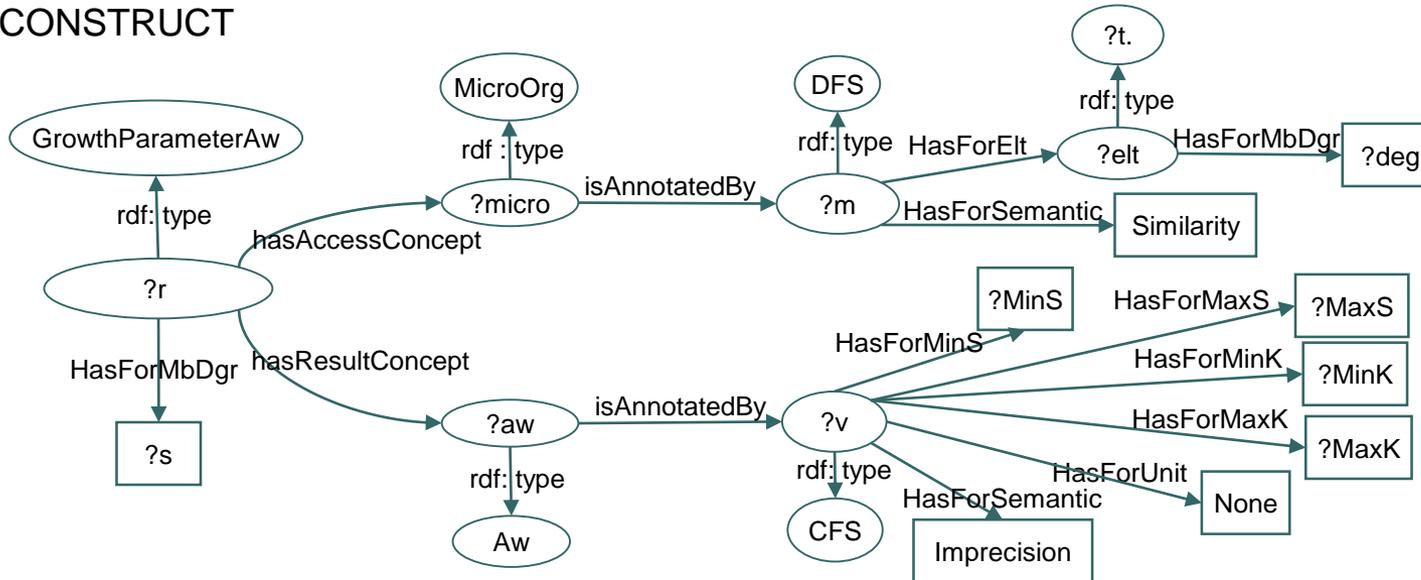
microorganismPreferences = {1.0/Gram+ + 0.5/Gram- + 1.0/
Clostridium botulinum + 1.0/ Staphylococcus Spp. + 0.5/ Salmonella}

0-cut

microorganismsList = {Gram+, Gram- , Clostridium botulinum,
Staphylococcus Spp., Salmonella}

Génération de la requête Sparql

CONSTRUCT



WHERE <idem CONSTRUCT>

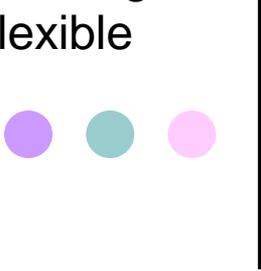
FILTER (xsd:float (?s) >= 0.5)

FILTER (?t = 'Gram+' || ?t = 'Gram-' || ?t = 'Clostridium_botulinum' || ?t = 'Staphylococcus_Spp.' ||
?t = 'Salmonella')

FILTER ((xsd:float (?minS) >= 0.9 && xsd:float (?minS) <= 0.99) || (xsd:float (?maxS) >= 0.9 &&
xsd:float (?maxS) <= 0.99 || (xsd:float (?maxS) <= 0.99 && xsd:float (?minS) >= 0.9))

Conclusion sur l'interrogation flexible avec SPARQL

- Utilisation des standards du Web sémantique (SPARQL) pour interroger les tableaux (en XML) et leurs annotations (en RDF)
- Implémentation : le système @Web



Logiciel @Web

Query Summary

Query scope

Ontology

IC2ACV

Topics

"Bioref-PM", "Bioref-PM-PC-UFM-PS", "Bioref-PM-UFM"

Relation

Composition en glucose d'une biomasse

Preferences

Mandatory

(1) Biomasse : [Herbes et plantes énergétiques : 1]

Optional

(1) Taux de Glucose : [50 ; 50 ; 100 ; 100] - unit : Pour cent

Parameters

Run query



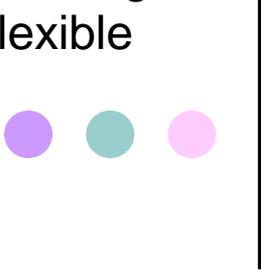
Logiciel @Web



ontology : IC2ACV - relation: Composition en glucose d'une biomasse

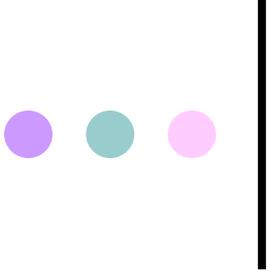
constraints wishes

rank	reliability score		Biomasse	Taux de Glucose	Etat de la biomasse	Numéro de l'expérience
row 0_2534	1		Paille de riz	[5.333e+01;5.600e+01],Pourcentage	Biomasse non traitée	[1.000e+00],1
row 0_2546	2		Paille de riz	[4.522e+01],Pourcentage	Biomasse non traitée	[1.000e+00],1
row 0_2557	3		Paille de blé	[1.180e+01],Pourcentage	Biomasse non traitée	[1.000e+00],1



Références bibliographiques

- Patrice Buche, Juliette Dibie-Barthelemy, Liliana Ibanescu and Lydie Soler (2013). Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource. TKDE - IEEE Transactions on Knowledge and Data Engineering, vol. 25(4), IEEE Computer Society, pp. 805-819
- Patrice Buche, Olivier Couvert, Juliette Dibie-Barthélemy, Gaëlle Hignette, Eric Mettler, Lydie Soler (2011). Flexible querying of Web data to simulate bacterial growth in food. International Journal of Food Microbiology, vol. 28(4), Elsevier, pp. 685-693
- G. Hignette, P. Buche, O. Couvert, J. Dibie-Barthélemy, D. Doussot, O. Haemmerlé, E. Mettler, L. Soler (2008). Semantic annotation of Web data applied to risk in food. International Journal of Food Microbiology, vol. 128(1), Elsevier, pp. 174-180.
- P. Buche, J. Dibie-Barthélemy, O. Haemmerlé and G. Hignette (2006). Fuzzy semantic tagging and flexible querying of XML documents extracted from the Web. Journal of Intelligent Information Systems 26(1), 2006, Kluwer, pp 25-40
- P. Buche, J. Dibie-Barthélemy, O. Haemmerlé and R. Thomopoulos (2006). Fuzzy concepts applied to the design of a database in predictive microbiology. Fuzzy Sets and Systems 157(9), 2006, Elsevier, pp 1188-1200

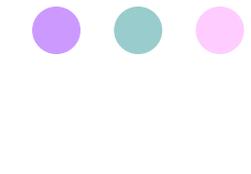


Sommaire

- Ontologie: construction et modélisation
- Annotation sémantique
- Interrogation flexible
- Evolution d'ontologie
- **Alignements d'ontologies et LOD**
- Conclusion et perspectives

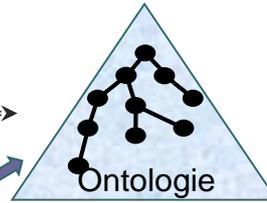
Le système ONDINE

(Ontology based Data INtEgration)



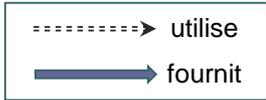
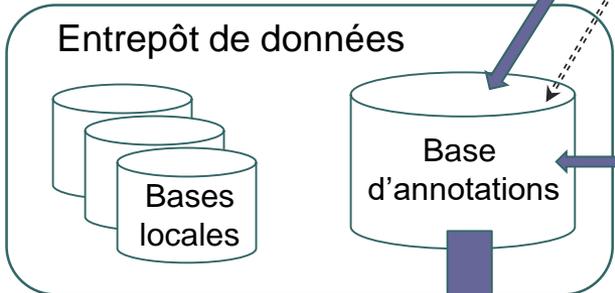
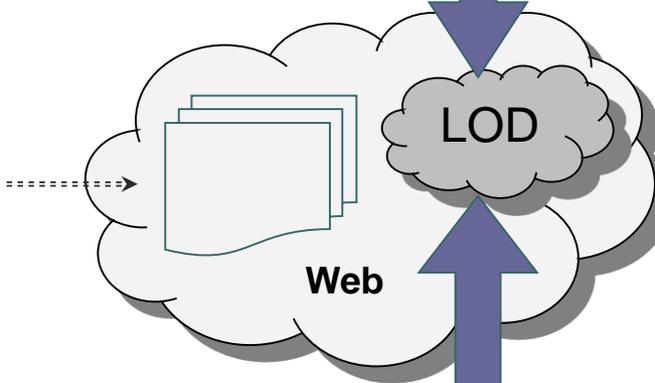
requête

réponse



Publication

Evolution



Publication



La RTO naRyQ sur les emballages

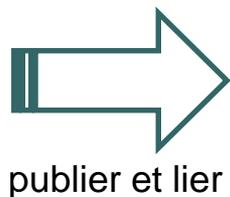
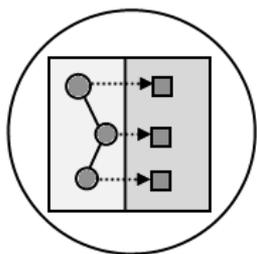
- Projet ANR MAP'OPT
- Représente « la composition et la dynamique des gaz ainsi que l'optimisation de la protection des denrées dans les emballages sous atmosphère modifiée »
 - les produits alimentaires (460 concepts),
 - les microorganismes (180 concepts)
 - les emballages (150 concepts)



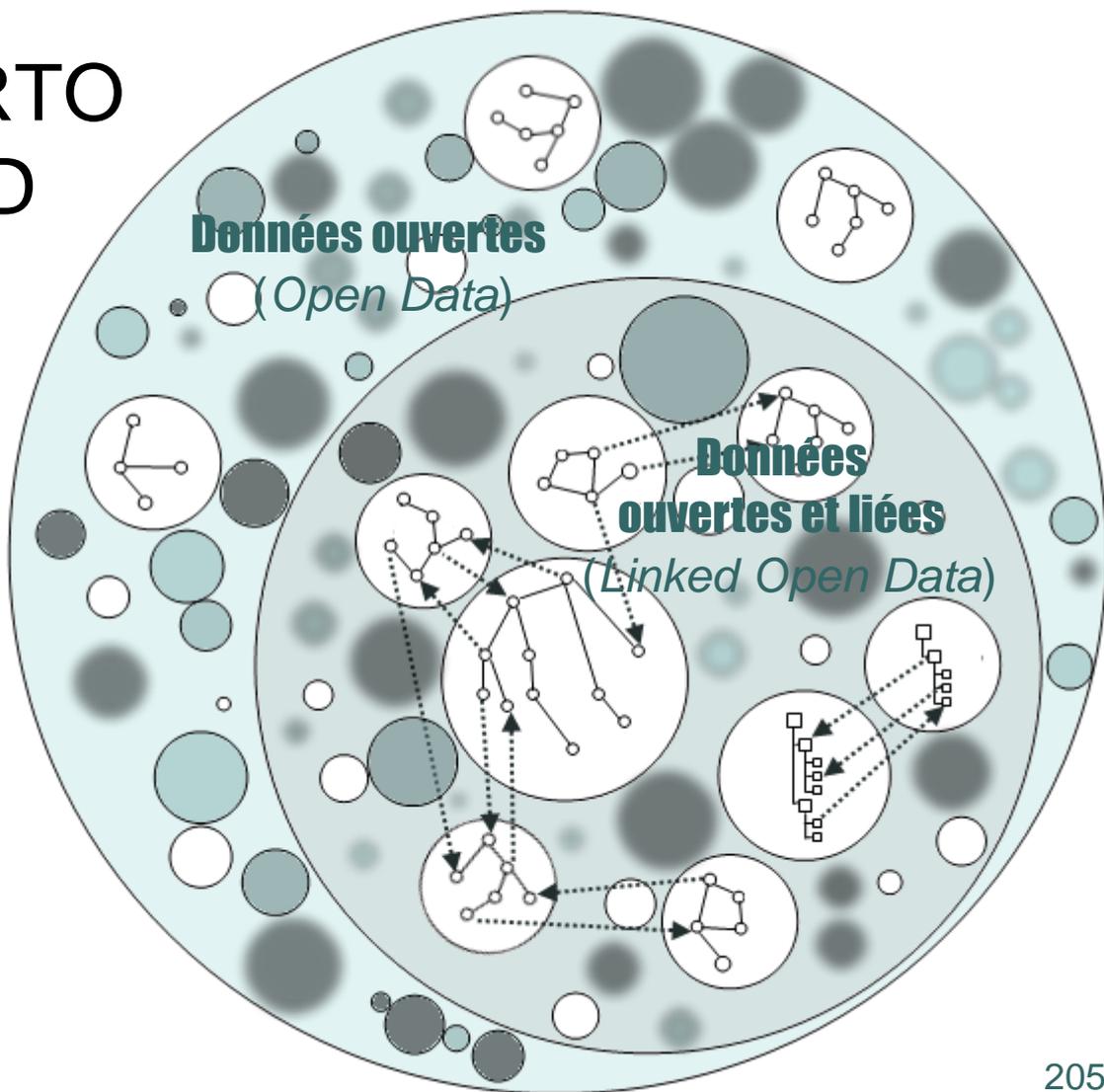
Objectif

- Publier et lier la RTO naRyQ sur le LOD

RTO naRyQ



publier et lier

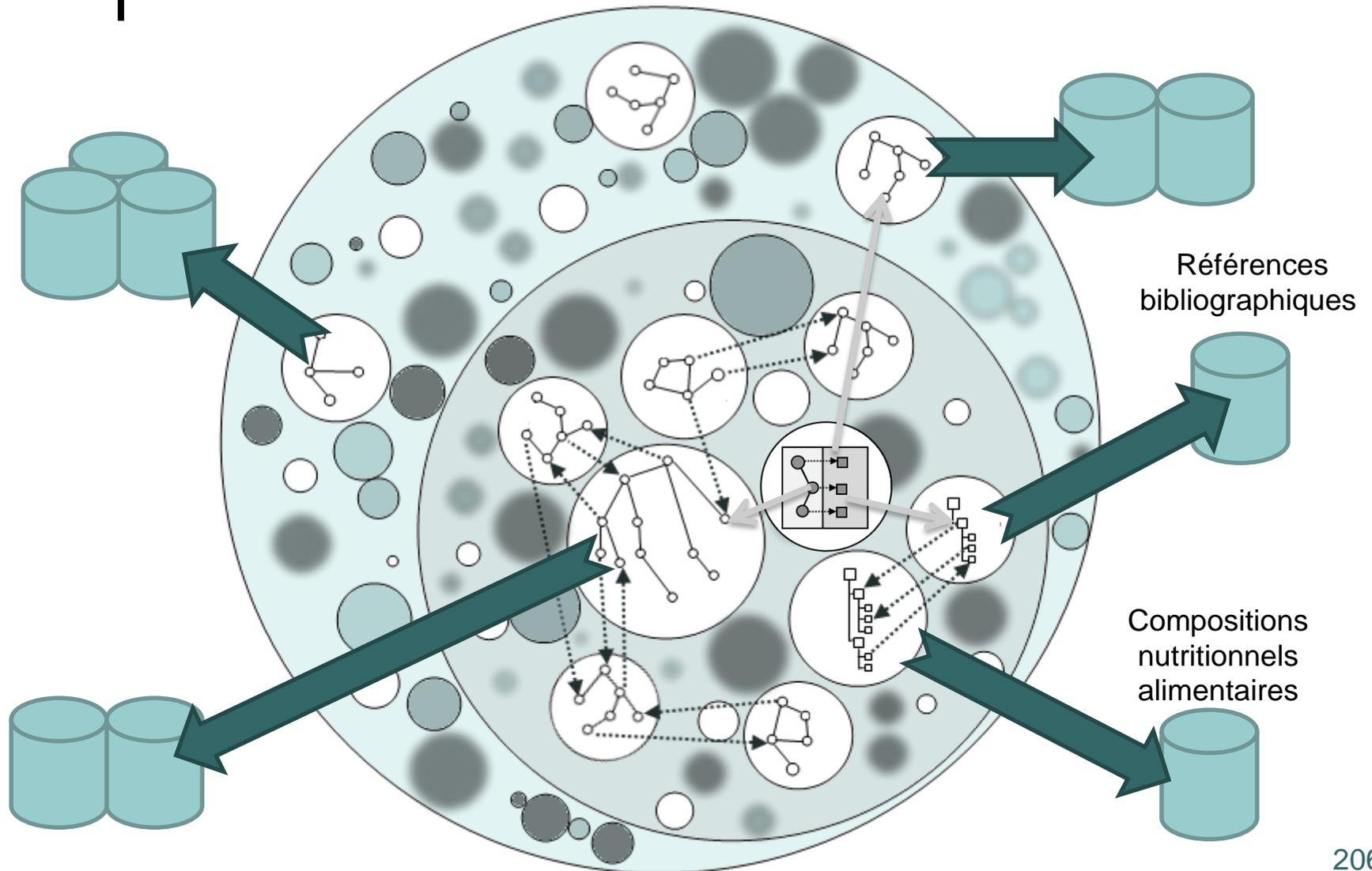


Alignement
d'ontologie



Objectif

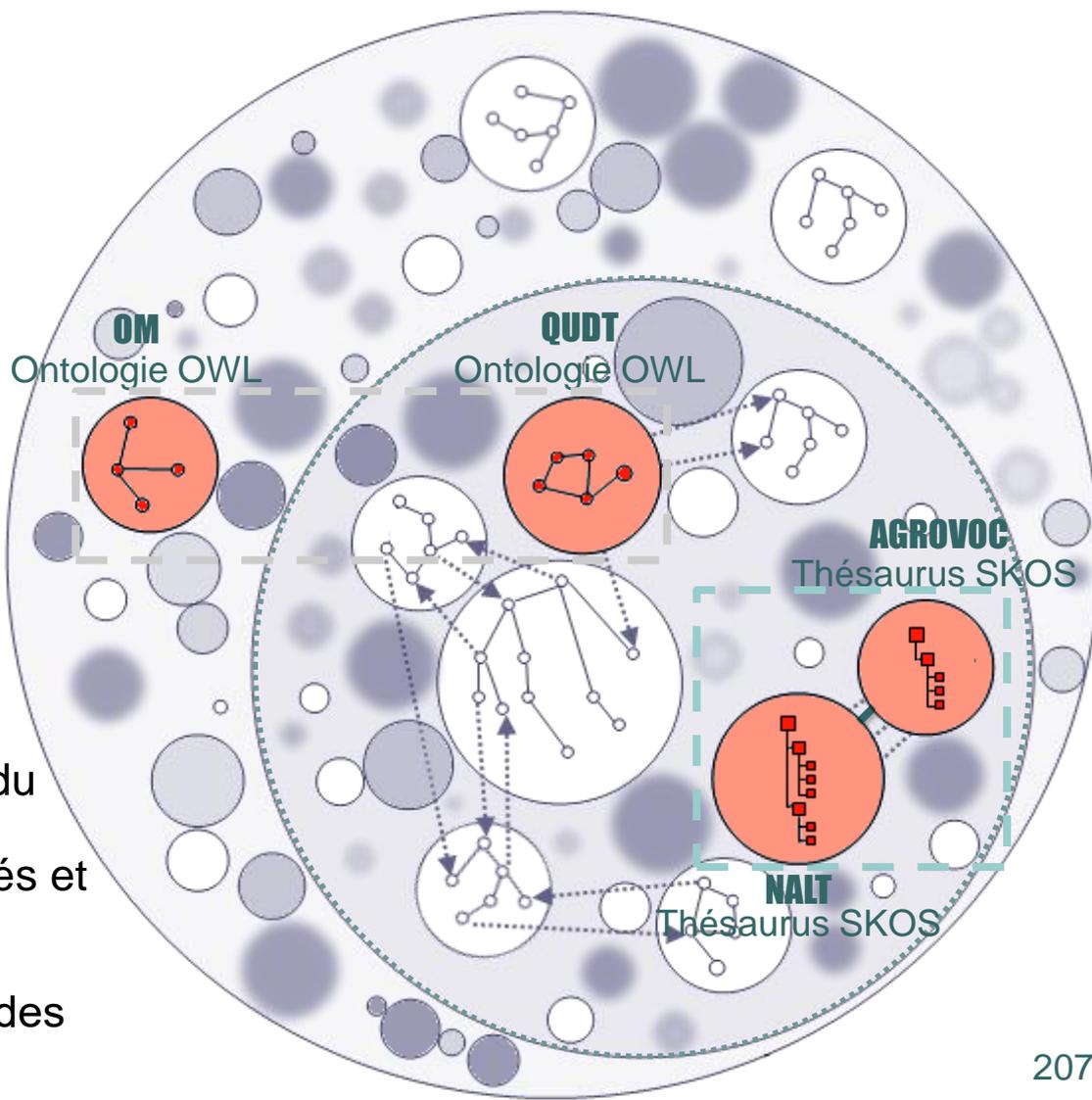
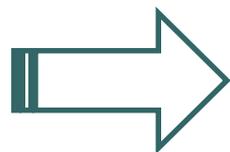
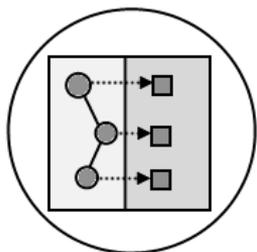
- Rendre visible et enrichir la RTO naRyQ sur le LOD





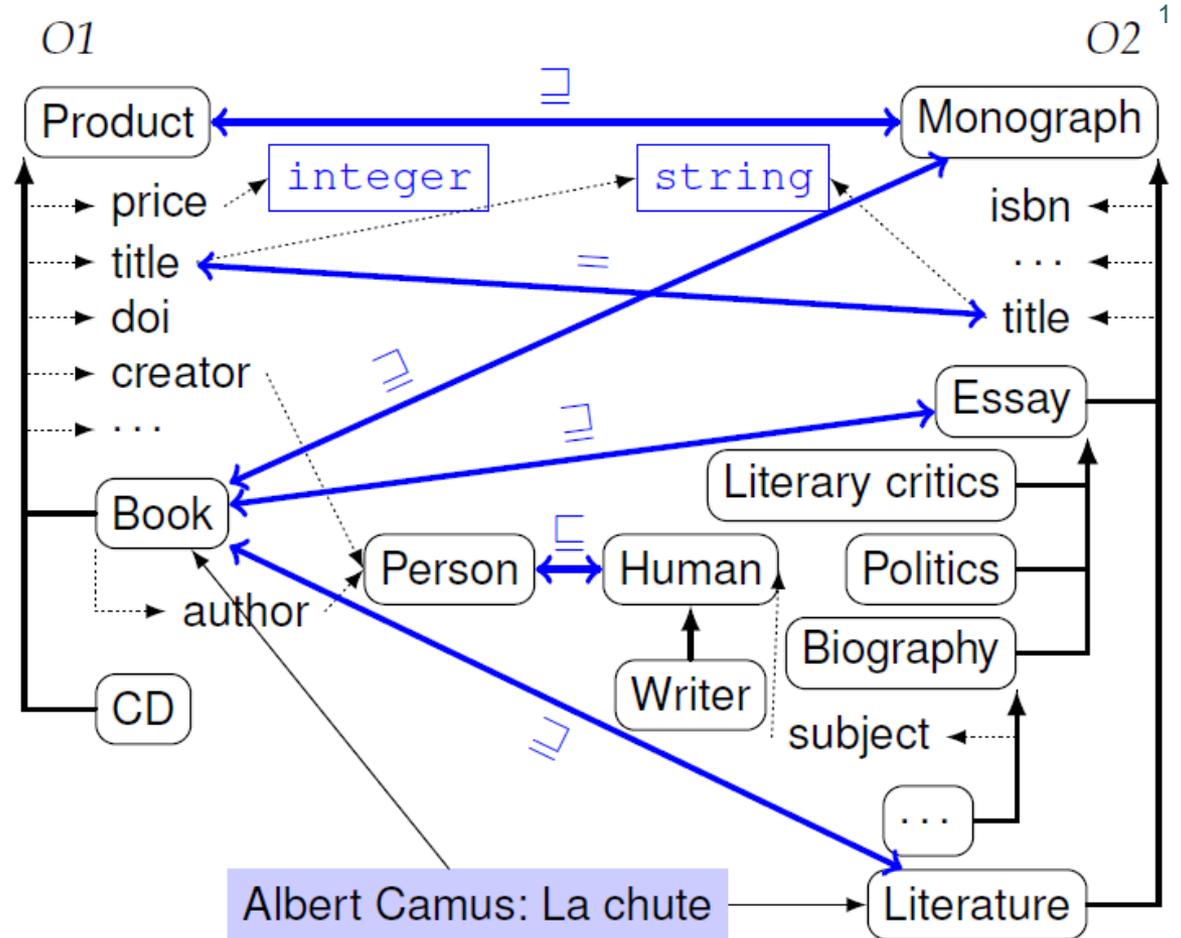
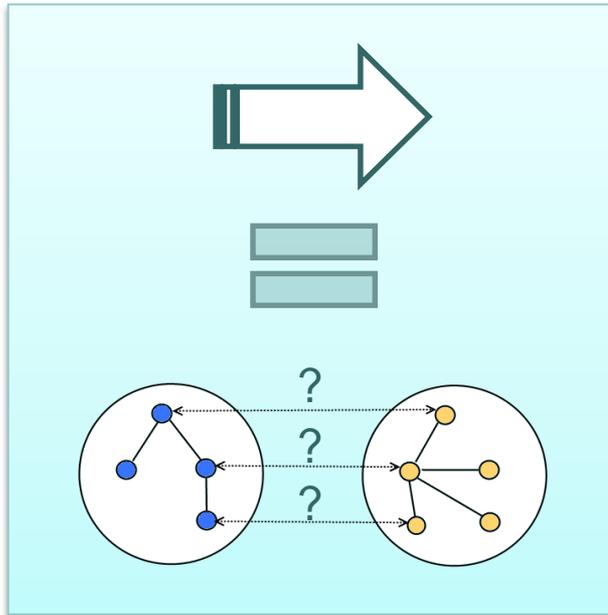
Les cibles

RTO
OWL2/SKOS



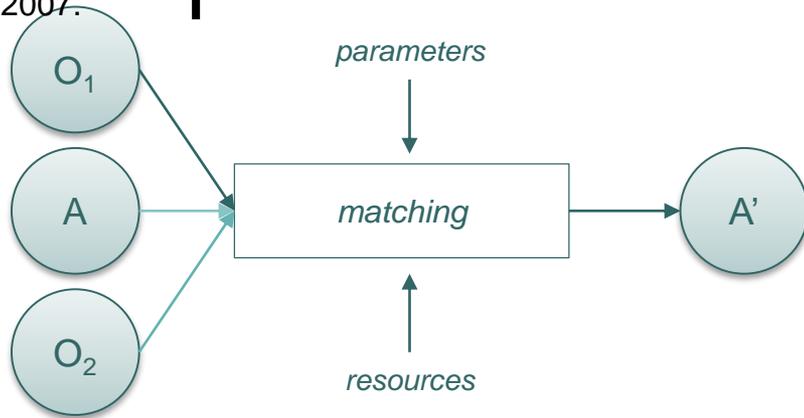
- Quatre ressources-cibles issues du Web de données
 - deux ontologies des quantités et unités de mesure
 - deux thésaurus liés aux domaines de l'agriculture et des sciences du vivant

L'alignement d'ontologies

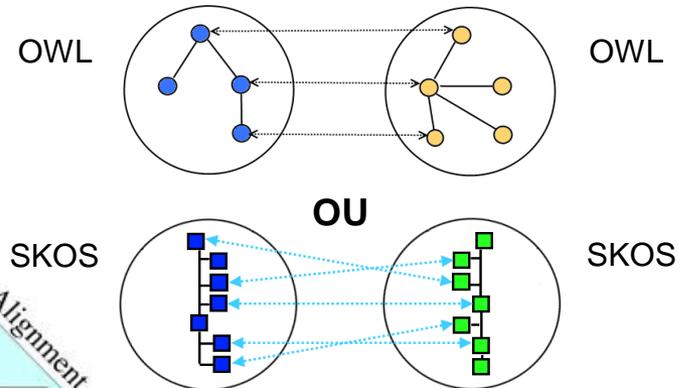


Réutilisation

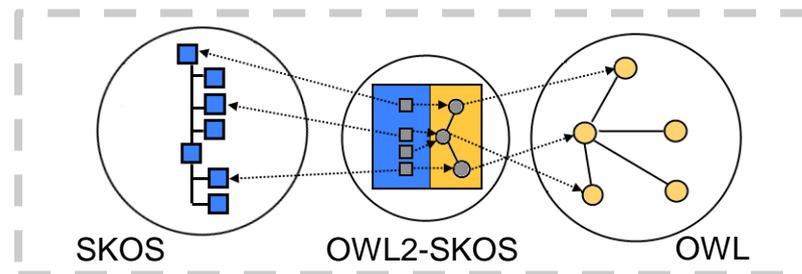
Euzenat, J and Shvaiko, P. *Ontology matching*. Springer, 2007.



$$c = \langle id, e_s, e_c, r, n \rangle$$



Théorie, littérature, formalismes, outils d'alignement et API existants

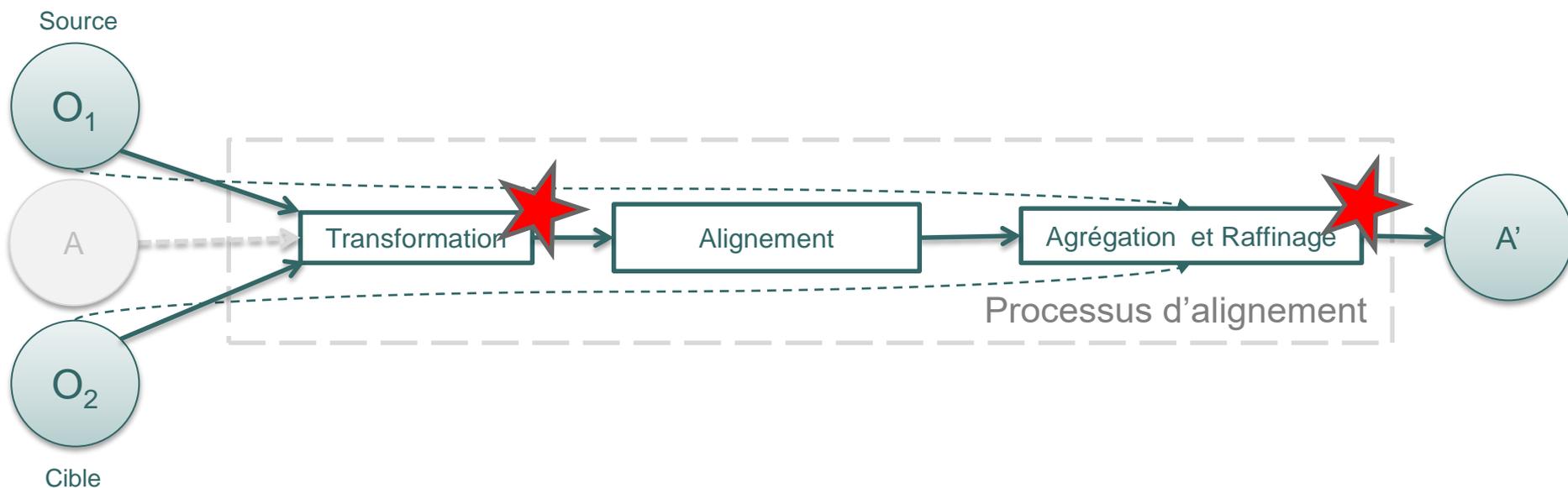


Méthode d'alignement

Adapter les
ressources aux
outils d'alignement
existants

Combiner les
résultats puis les
raffiner

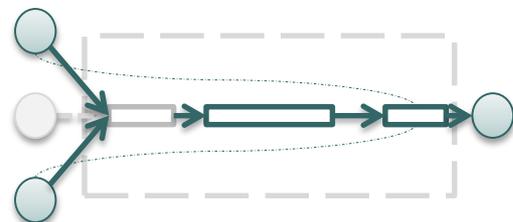
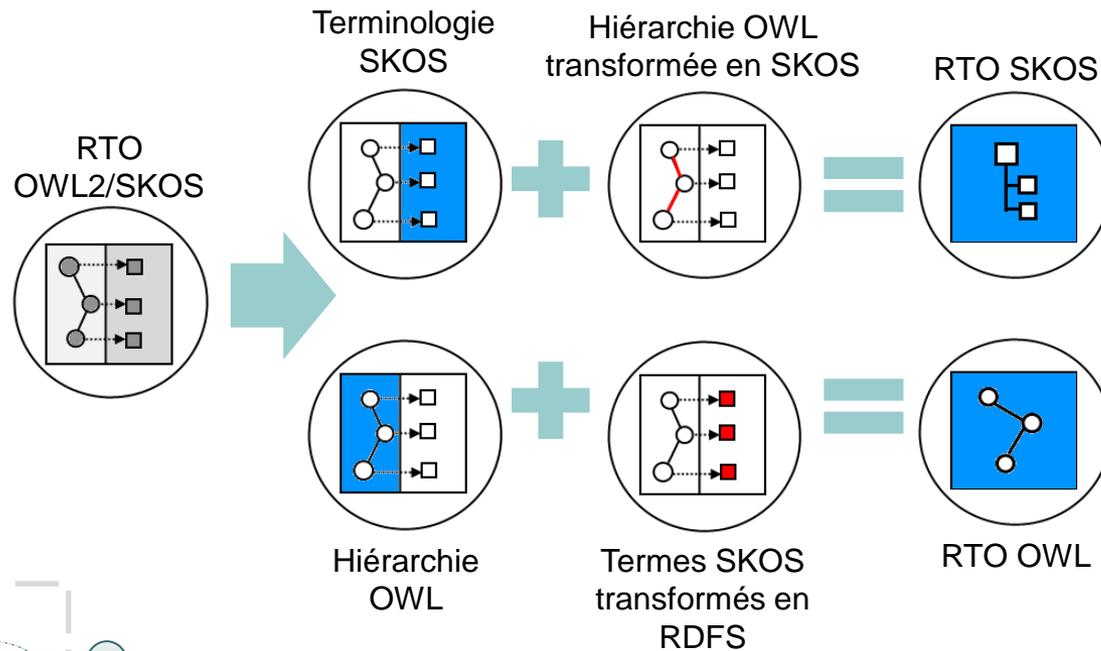
Evaluer les
alignements
raffinés





Transformation

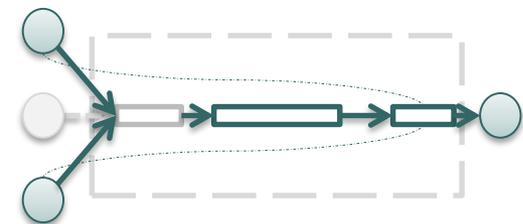
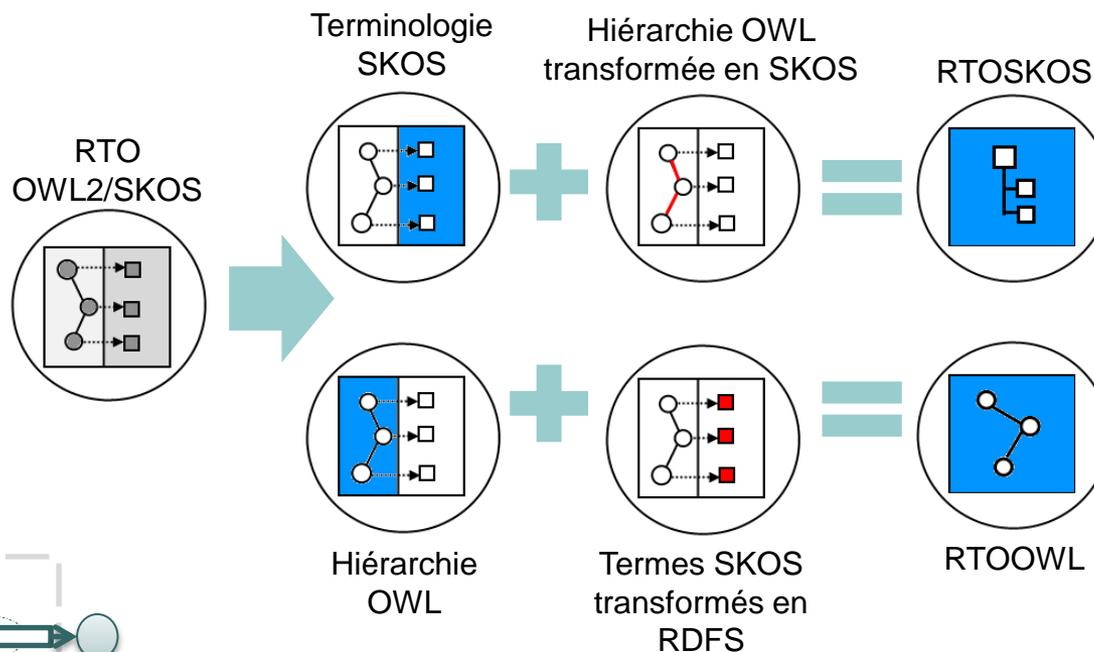
- Variantes des ontologies
 - En OWL
 - En SKOS



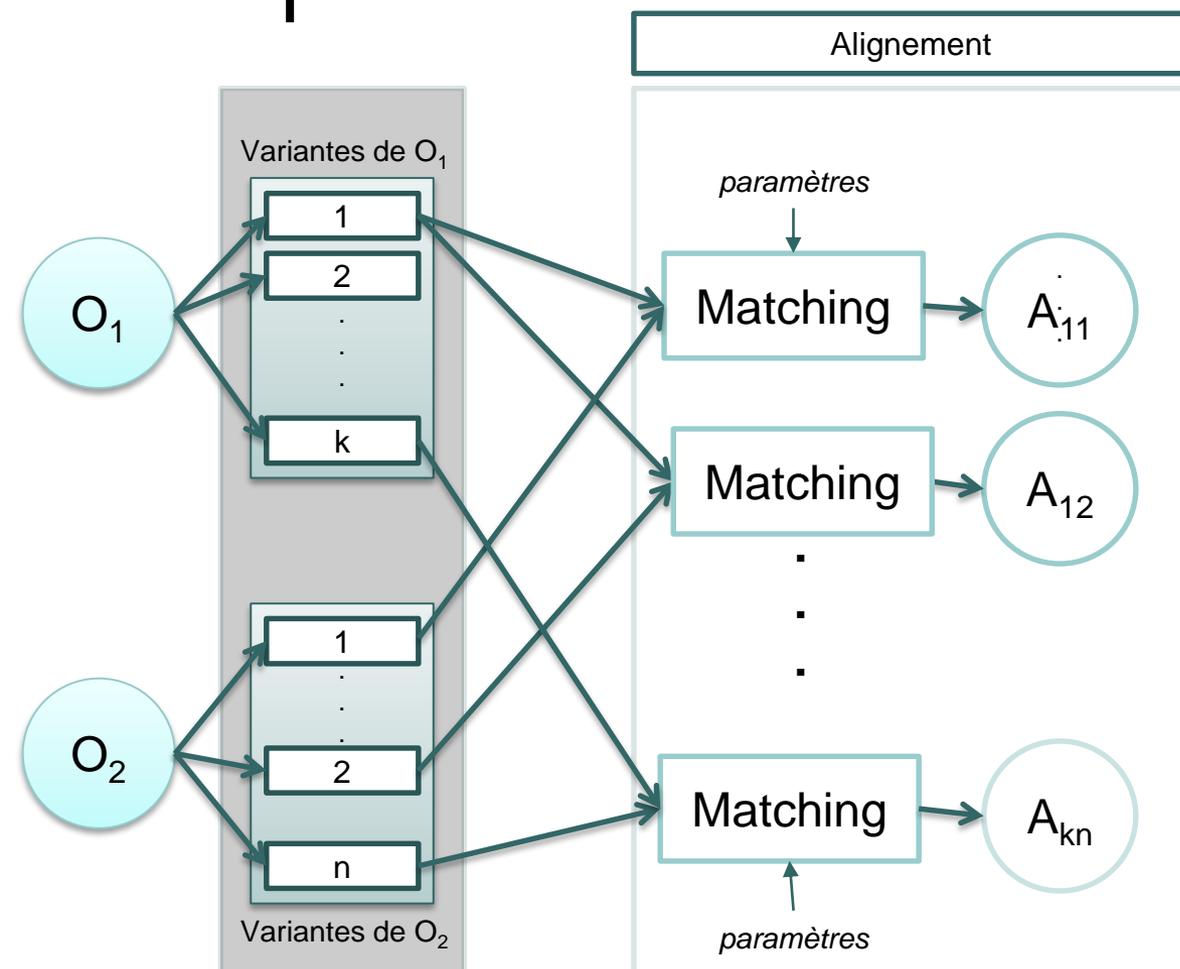


Transformation

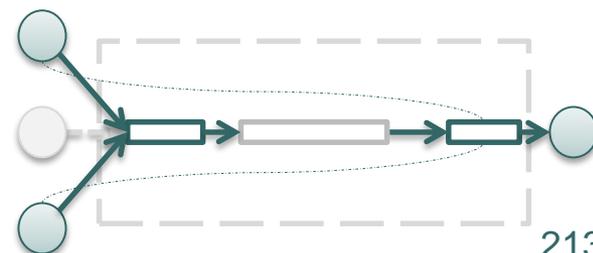
- Expérimentations :
 - 4 variantes AgroVoc
 - 2 variantes NALT
 - 3 variantes pour naRyQ



Alignement



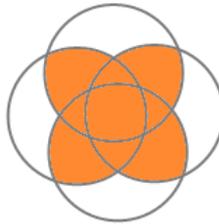
- Expérimentations
 - 4 variantes AgroVoc
 - 2 variantes NALT
 - 3 variantes pour naRyQ
- Expérimentations
 - LogMap
<http://www.cs.ox.ac.uk/isg/projects/LogMap/>
 - Aroma
<http://aroma.gforge.inria.fr/>



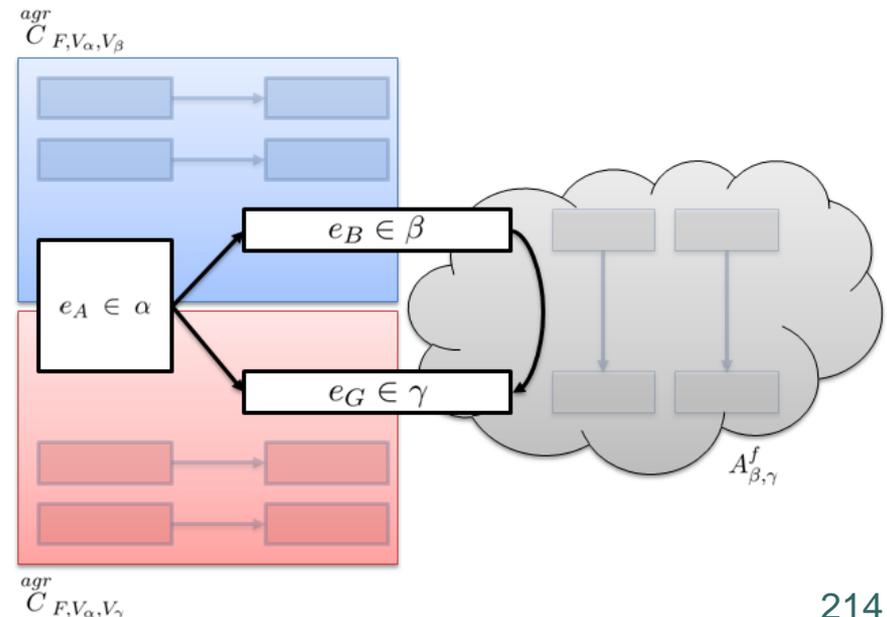
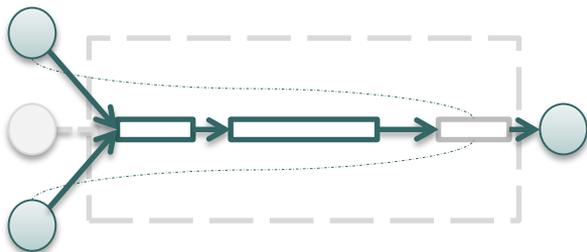
Raffinage

- Supprimer des correspondances ambiguës
- Identification des correspondances potentiellement correctes

● Recouvrement



● Recouvrement du LOD





Expérimentation

- Outils: LogMap et Aroma
- Aligner naRyQ à AgroVoc
- Aligner naRyQ à Nalt
- Evaluation
 - Alignement de référence
 - Résultats

Alignement de référence

○ Production:

- 1 453 correspondances
- Validées par 2 experts, 4h
- Outil de visualisation
- RAgrovoc, RNalt

○ Amélioration

- R+Agrovoc, R+Nalt

○ Alignements de référence partiels(!)

					Oui <input type="radio"/> ?
Agro31	rto:cooling	=	95.45	agrovoc:c_36844	<input checked="" type="radio"/> Non <input type="radio"/> Oui <input type="radio"/> ?
Agro32	rto:cooling	=	100	agrovoc:c_1854	<input type="radio"/> Non <input checked="" type="radio"/> Oui <input type="radio"/> ?
Agro33	rto:bacillus_cereus_toxin	=	95.45	agrovoc:c_26384	<input checked="" type="radio"/> Non <input type="radio"/> Oui <input type="radio"/> ?
Agro34	rto:aeration	=	100	agrovoc:c_154	<input checked="" type="radio"/> Non <input type="radio"/> Oui <input type="radio"/> ?
Agro35	rto:enterobacteria	=	96.25	agrovoc:c_2580	<input checked="" type="radio"/> Non <input type="radio"/> ?

rto:cooling	
Propriété	Valeur
skos:prefLabel	Cooling@en
skos:prefLabel	Refroidissement@fr
rdfs:subClassOf	rto:cold
rdf:type	skos:Concept
rdf:type	owl:Class

agrovoc:c_36844	
Propriété	Valeur
skos:prefLabel	Precooling@en
skos:prefLabel	Prérefroidissement@fr
skos:broader	agrovoc:c_1854
rdf:type	skos:Concept



Evaluation naRyQ -> AgroVoc

	# total	# bons	P	R	F-m
<i>BestScore</i>		300	0,90	0,82	0,85
Tous	1 583	366	0,23	0,99	0,37
rec	582	354	0,61	0,96	0,74
Rec LOD	336	254	0,76	0,69	0,72
U	447	344 >	0,77	0,93 >	0,84 =

BestScore: Meilleurs scores d'alignements obtenus avec les deux outils utilisés.

- Rappel = (# correspondances correctes trouvés) / (# correspondances à trouver)
- Précision = (# correspondances correctes trouvés) / (# correspondances trouvés)
- F-mesure = $(2 \times \text{Rappel} \times \text{Précision}) / (\text{Rappel} + \text{Précision})$



Evaluation naRyQ -> Nalt

	# total	# bons	P	R	F-m
<i>BestScore</i>		359	0,80	0,84	0,81
Tous	850	415	0,49	0,97	0,65
rec	480	368	0,77	0,86	0,81
Rec LOD	337	255	0,76	0,59	0,67
U	400	348	0,87 >	0,81	0,84 >

BestScore: Meilleurs scores d'alignements obtenus avec les deux outils utilisés.



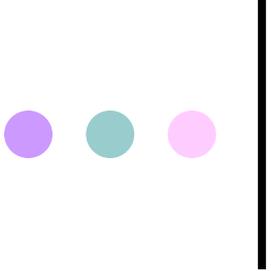
Conclusion sur alignement d'ontologies et LOD

- Une méthode d'alignement d'ontologies
 - Matching with background knowledge [Shvaiko & Euzenat 2013]
 - Combine les méthodes et outils existants
 - Exploite la redondance et les alignements existants
- Expérimentation
 - naRyQ -> AgroVoc
 - naRyQ -> Nalt



Références bibliographiques

- Thomas Hecht, Patrice Buche, Juliette Dibie, Liliana Ibanescu and Cassia Trojahn dos Santos (2017). Ontology Alignment Using Web Linked Ontologies as Background Knowledge. Advances in Knowledge Discovery and Management, vol. 6 of the series Studies in Computational Intelligence pp 207-227, Guillet, Fabrice, Pinaud, Bruno, Venturini, Gilles (Eds.), Springer International Publishing, ISBN 978-3-319-45763-5



Sommaire

- Annotation sémantique
- Interrogation flexible
- Ontologie: modélisation et évolution
- Alignements d'ontologies et LOD
- **Conclusion et perspectives**

Le système ONDINE

(Ontology based Data INtEgration)



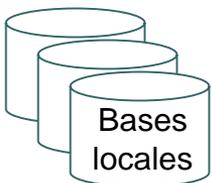
requête

réponse

Interrogation

Interrogation unifié et flexible des différentes bases de l'entrepôt

Entrepôt de données



Bases locales

Base d'annotations

Evolution

Annotation

Annotation sémantique de tableaux et de texte extraits de documents scientifiques du Web pour enrichir des bases locales

Ontologie

Publication

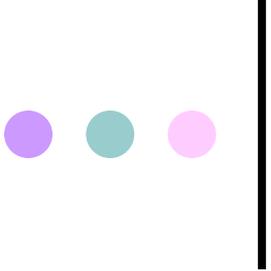
LOD

Web

Publication

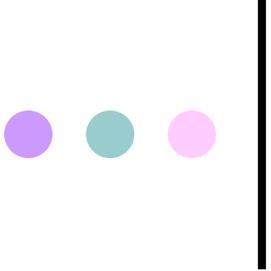
-----> utilise

—————> fournit



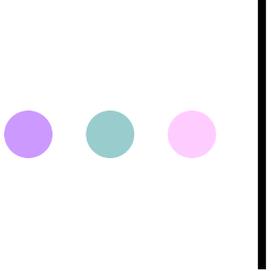
Conclusion

- ONDINE: un système complet d'intégration de données de sources diverses pour enrichir des bases locales
 - Système guidé par une ontologie
 - Intégration de tableaux de données
 - Système allant de l'extraction et l'annotation de tableaux à leur interrogation
 - Prise en compte de données imprécises et de similarités lors de l'annotation et de l'interrogation



Conclusion

- L'ontologie au cœur du système
 - Une ontologie de domaine modélisée pour la **représentation** de relations n-aires entre des données quantitatives expérimentales
 - médiateur entre des connaissances de sources diverses
 - Une ontologie publiable sur le LOD par alignements avec des ontologies existantes
- Un travail de recherche appliqué à des domaines de l'INRA
 - Un logiciel opérationnel : @Web
 - 3 domaines d'études :
 - Emballage
 - Bio raffinerie
 - Blé dur



Travaux en cours et perspectives

- Ontologie
 - Construction de l'ontologie
 - Modélisation d'une ontologie multi-échelles et multi-étapes pour prendre en compte le processus de transformations des produits alimentaires (thèse qui a débuté en oct 2015)
 - Evolution et enrichissement de l'ontologie
 - Alignements d'ontologies et LOD
 - Utilisation de l'ontologie dans le raisonnement...
- Application au domaine des procédés de transformation
 - Production et stabilisation des levures
 - Gels laitiers