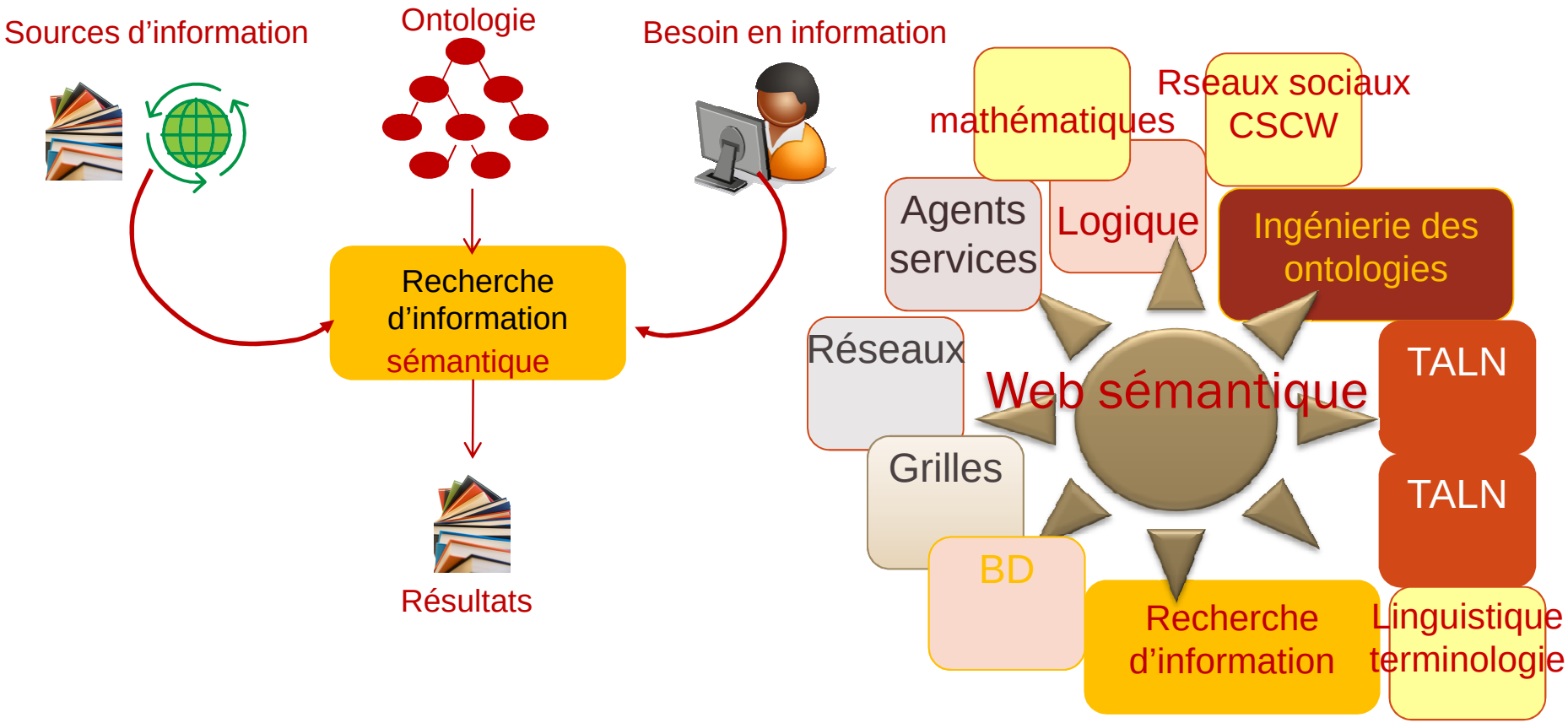
Ontologies et annotation sémantique, enjeux linguistiques et terminologiques

Nathalie Aussenac-Gilles
Équipe IC3
IRIT – Toulouse (F)
aussenac@irit.fr

Recherche d'information sémantique



Analyse et représentation des contenus
Dimension sémantique

Institut de Recherche en Informatique de Toulouse

Plan

- Recherche d'information et sémantique
  - Enjeux liés à l'articulation entre représentations et langue, place du TAL
- Annotation sémantique de documents
  - Repérer des éléments d'ontologie dans des textes : enjeux linguistiques et TAL
- Formuler des requêtes
  - Guider la formulation
- Rapprocher requête et documents
  - distances sémantiques et alignement

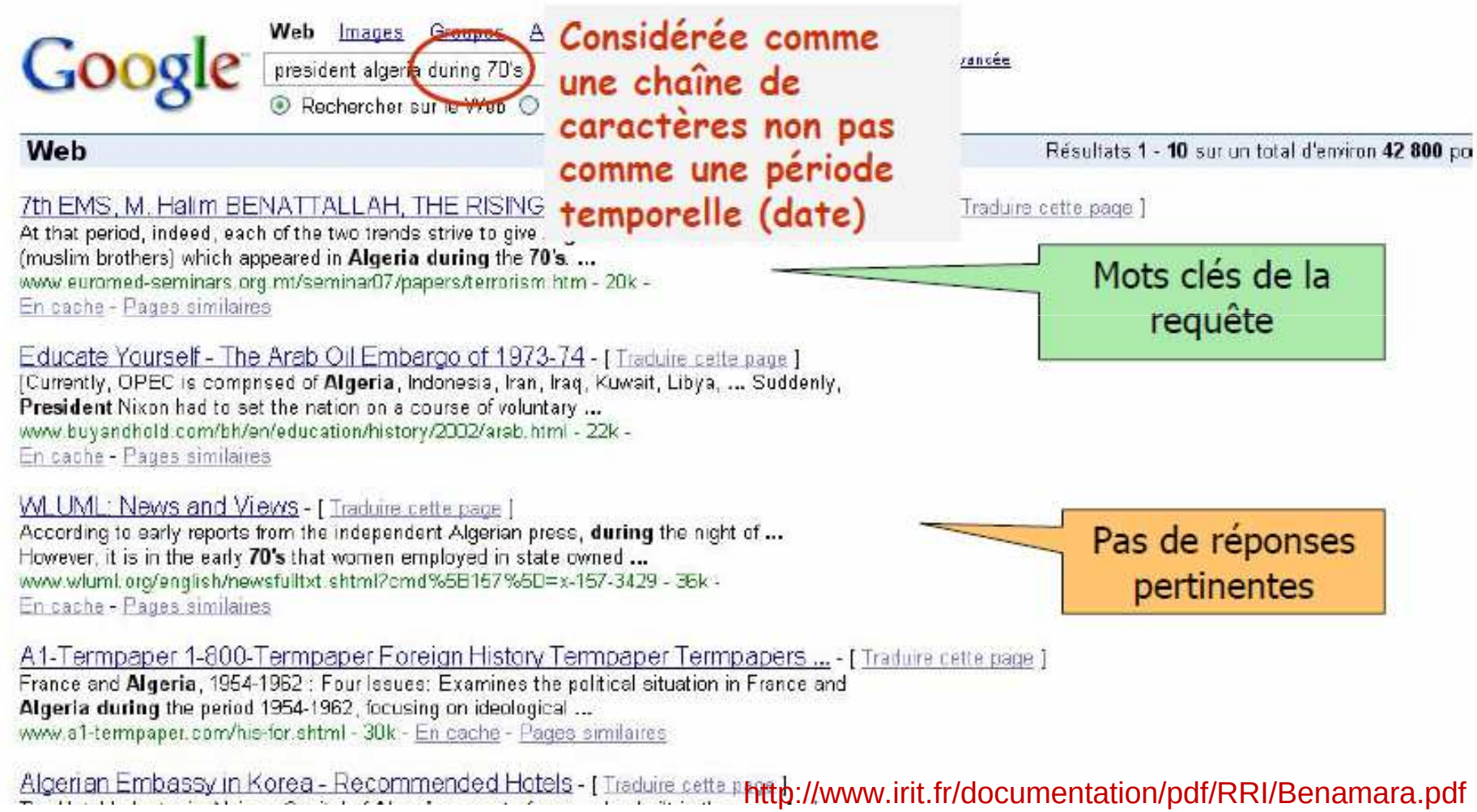
Institut de Recherche en Informatique de Toulouse

Différents types de recherche d'information

- Moteurs de recherche généraux
- Recherche d'informations précises et spécialisées
- Organisation des collections, navigation au sein de collections
- Extraction d'information
- Questions-réponses
- Analyse d'opinion

Institut de Recherche en Informatique de Toulouse

Recherche par mots clés



Considérée comme une chaîne de caractères non pas comme une période temporelle (date)

Mots clés de la requête

Pas de réponses pertinentes

http://www.irit.fr/documentation/pdf/RRI/Benamara.pdf

Institut de Recherche en Informatique de Toulouse

Questions-réponses

http://www.irit.fr/documentation/pdf/RRI/Benamara.pdf

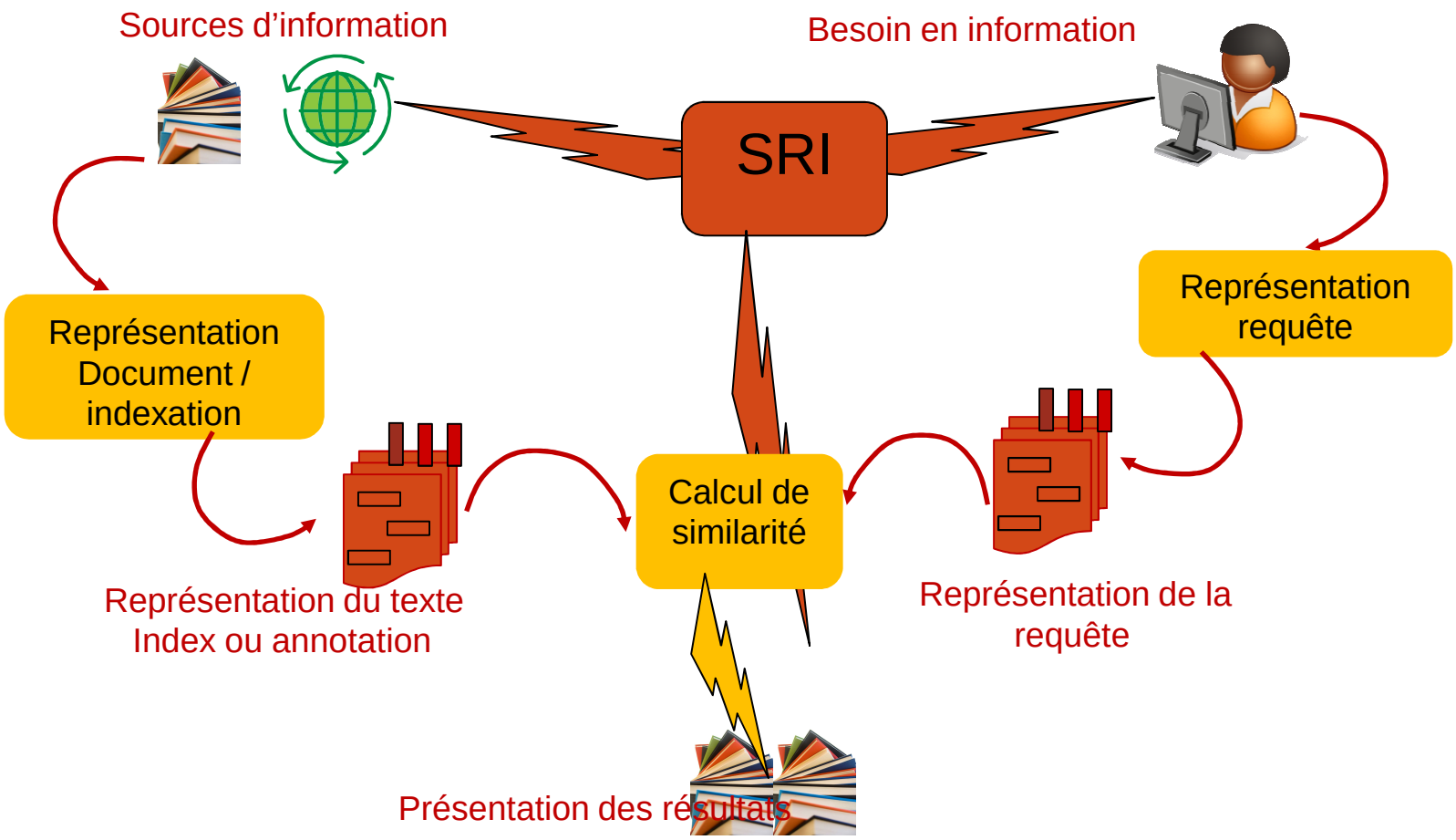
Institut de Recherche en Informatique de Toulouse

Extraction d'information

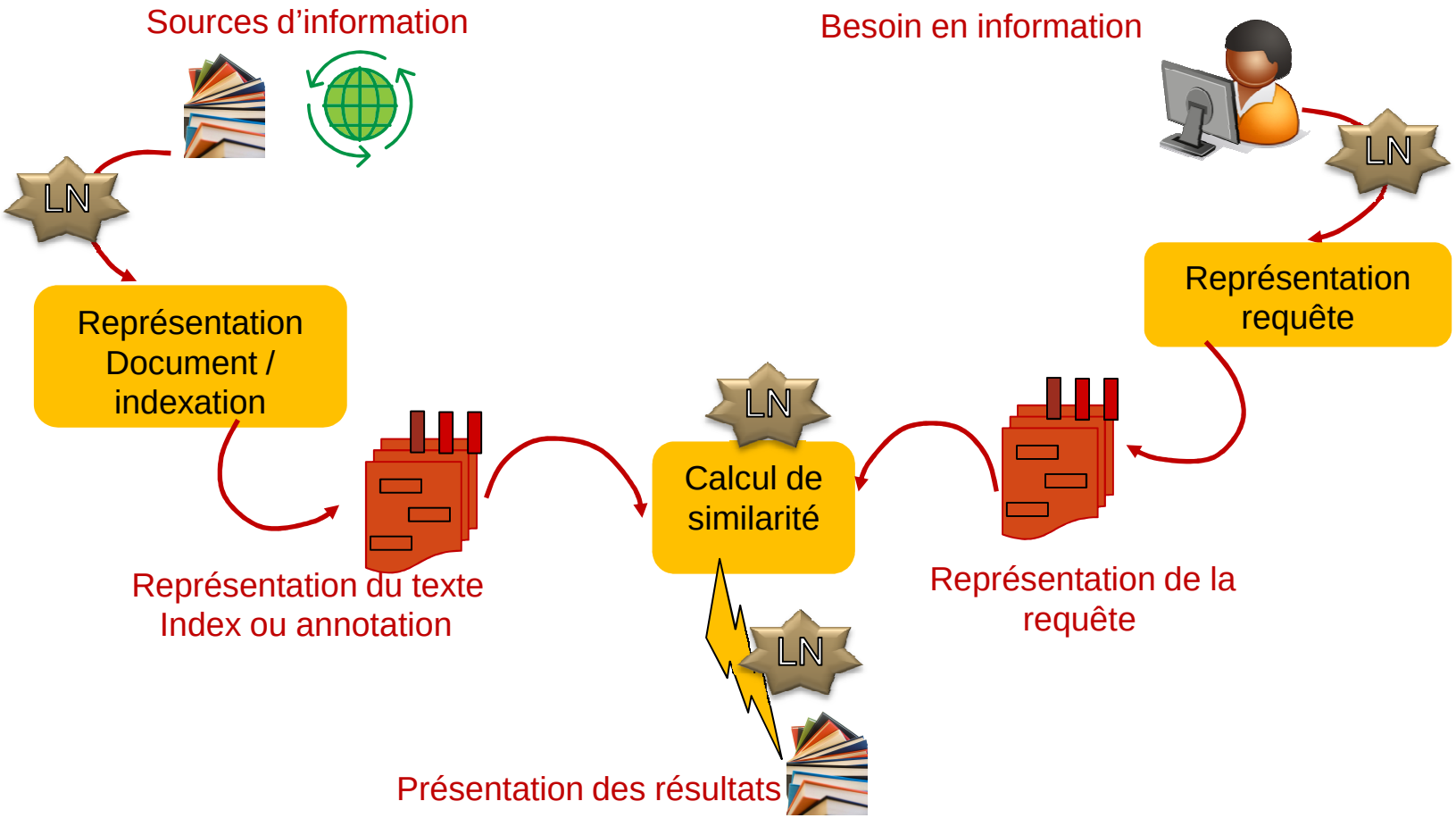
- Exemple : lister tous les records sportifs par discipline sous la forme Nom/discipline
  - Bidule a battu le record du 100 m.
  - Bidule a explosé l'ancien record de Machin.
  - Le 100 mètre a un nouveau recordman : Bidule.
  - Bidule a presque battu le record du monde détenu par Machin.
  - Bidule a fait forte impression au 100m en battant le record du monde.
  - La compagnie Truc produit plus de 100 mètres de couturières par jour, un record.

http://www.irit.fr/documentation/pdf/RRI/Benamara.pdf

Schéma général de recherche d'information



Enjeux liés au langage naturel



Institut de Recherche en Informatique de Toulouse

Apports du traitement automatique des langues à la RI

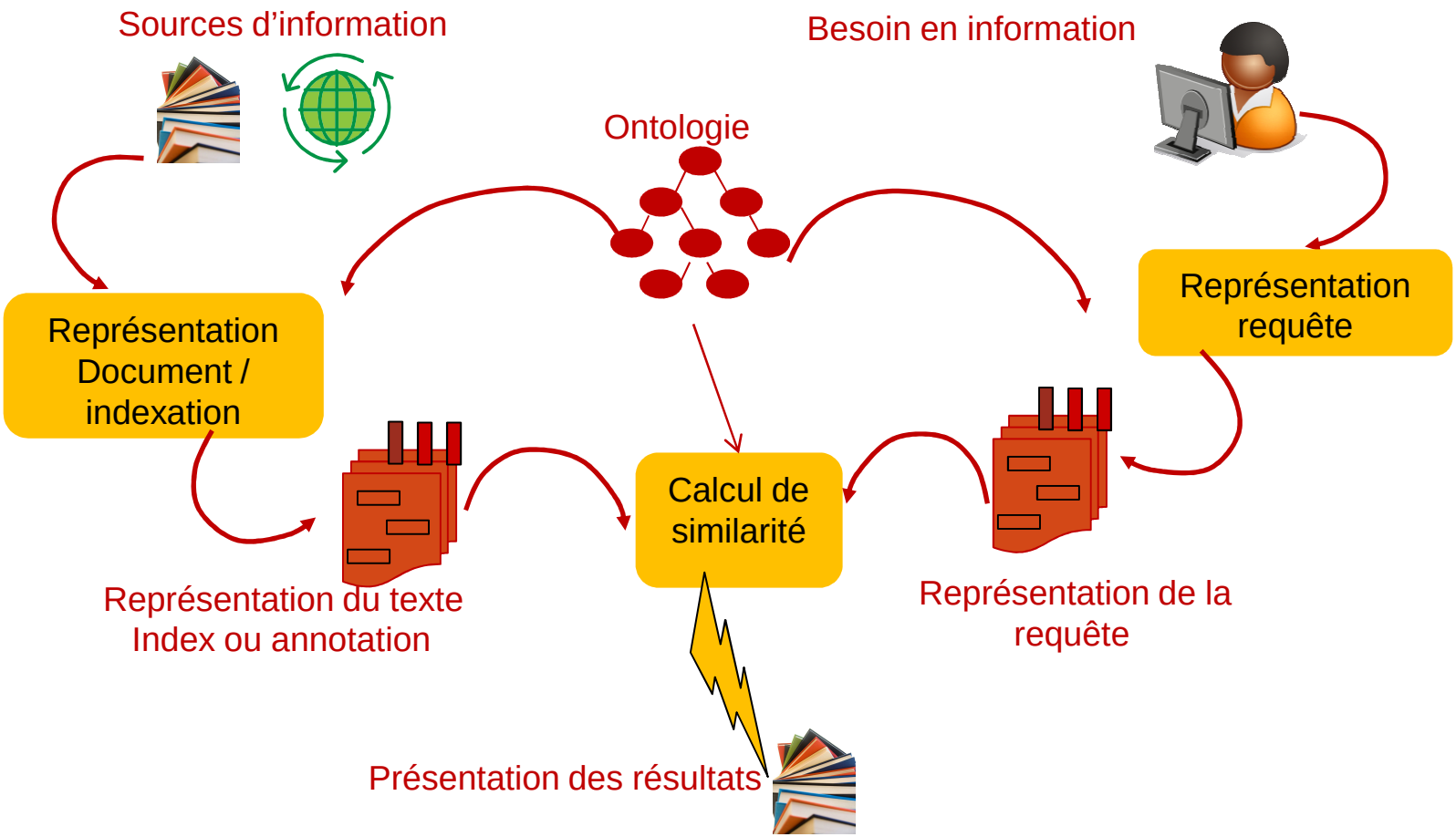
- Approches statistiques pour la fouille de textes
  - Texte = sac de mots
  - Exploiter les voisinages des mots :
    - colocations -> identifier des classes de mots - > concepts
    - segments répétés -> repérage de syntagmes
- Traitements linguistiques élémentaires (de surface)
  - Élimination des mots vides
  - Repérage des noms propres -> gazeteer
  - Stemming (supporting, supported -> support)
  - Tokenization, découpage en phrases
  - Identification de catégories grammaticales / lemmes (POS)

Institut de Recherche en Informatique de Toulouse

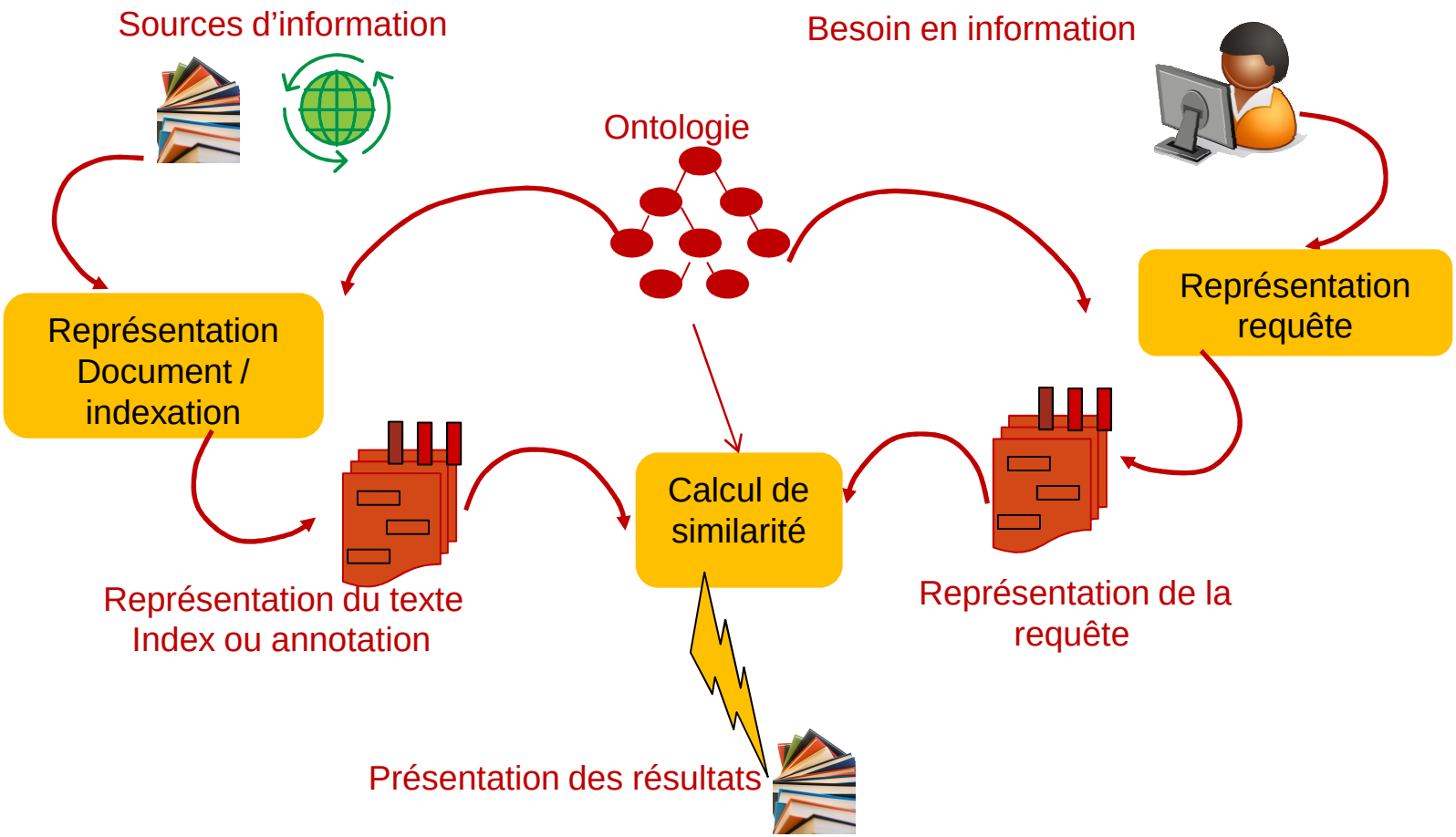
Apport du traitement automatique des langues à la RI

- Traitements linguistiques élaborés
  - Repérer des éléments sémantiques -> exploiter des ontologies / thésaurus
  - Exploiter des éléments sémantiques -> reformuler des requêtes
  - Enchaîner des traitements

Recherche d'information et sémantique



Traitement Automatique des langues, RI et technologies sémantiques



Institut de Recherche en Informatique de Toulouse

Annotation sémantique

- Annotation
  - Information associée à une page existante
  - Information manipulable par un système informatique
  - Information qui peut être séparée des données annotées

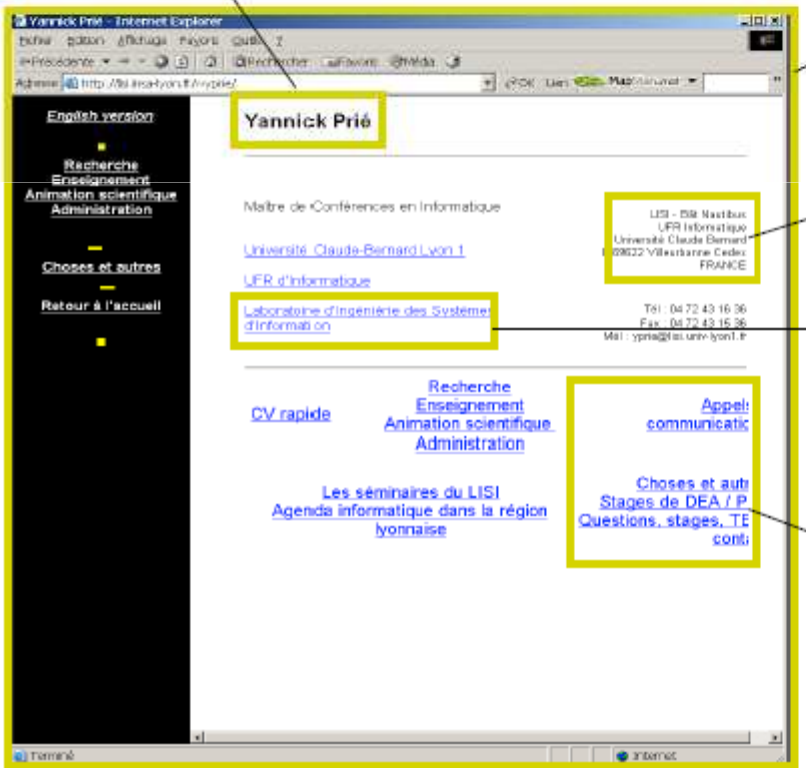
- Différents types d'annotations
  - Les annotations produites par le TAL
  - Annoter pour l'extraction d'information
  - Annoter pour la RI sémantique

Institut de Recherche en Informatique de Toulouse

Exemples d'annotations

Prénom : Yannick
Nom : Prié

Auteur : Y. Prié
Date de création :
Mots-clé : page perso, informatique, annotation



Type : adresse

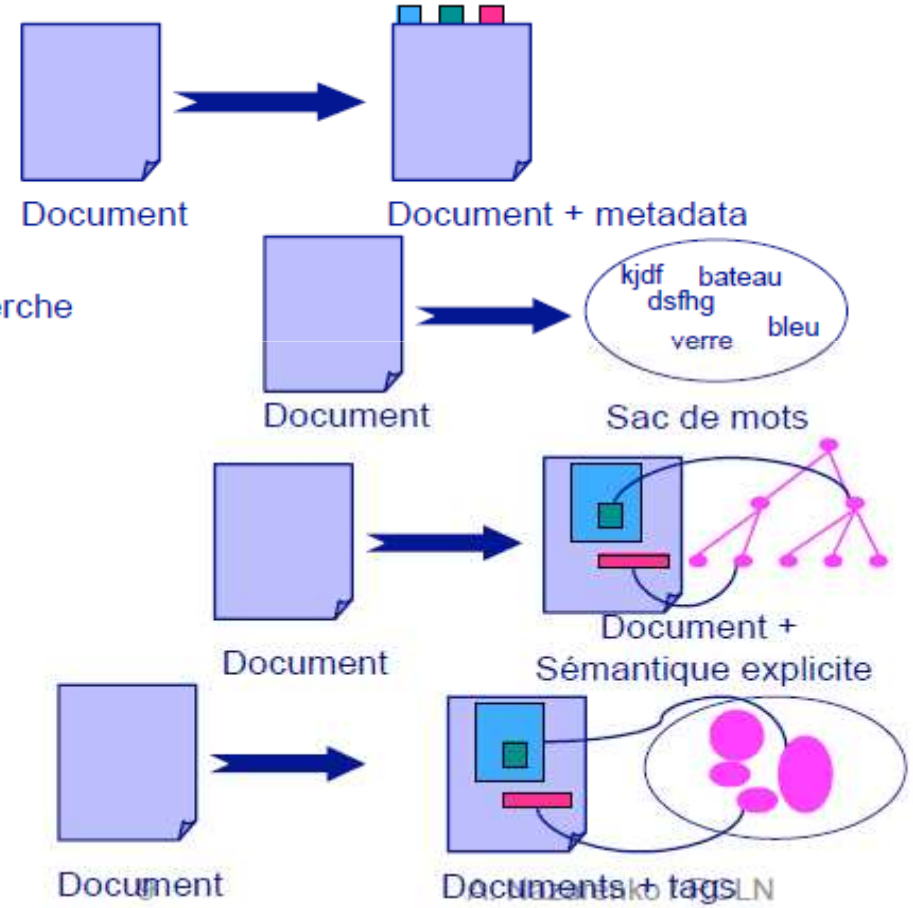
« A actualiser »

SG-03/01/03
« J'aime pas cette partie. »

Emprunté à Y. Prié, cours M2R 2005-2006

Différents types d'annotations

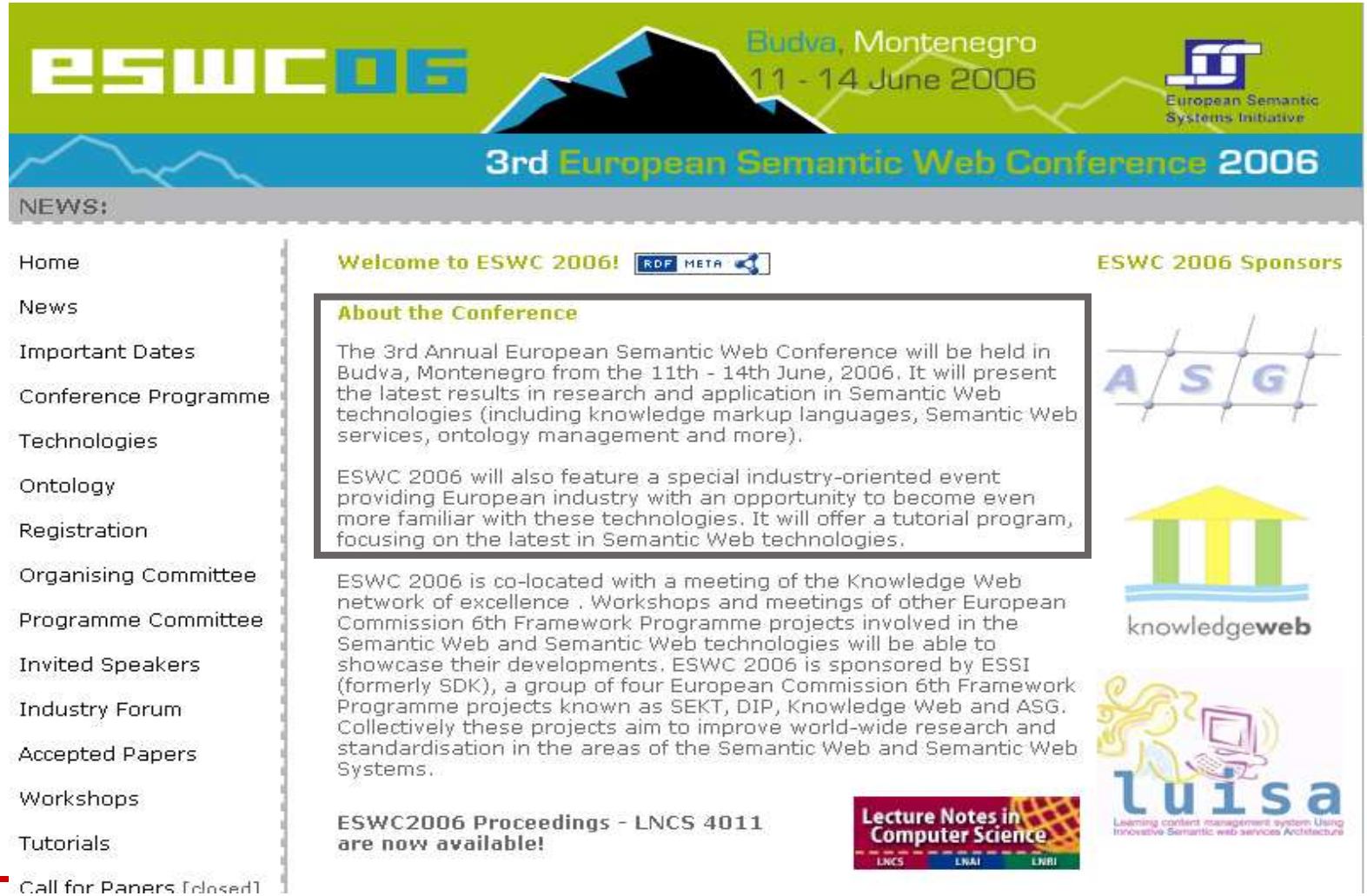
- Indexation traditionnelle
- Indexation des moteurs de recherche
- Approche "web sémantique"
- Indexation Web 2.0



Séminaire de l'IRIT (6 nov. 2008)

A. Nazarenko

Approche sémantique en RI



Institut de Recherche en Informatique de Toulouse

Approche sémantique en RI

Indexation classique

3rd European Semantic Web Conference (ESWC2006)

The 3rd Annual European Semantic Web Conference (ESWC2006) will be held in Budva, Montenegro from the 11th - 14th June, 2006.

It will present the latest results in research and application in Semantic Web technologies (including knowledge markup languages, Semantic Web services, ontology management and more).

ESWC 2006 will also feature a special industry-oriented event providing European industry with an opportunity to become even more familiar with these technologies. It will offer a tutorial program, focusing on the latest in Semantic Web technologies.

Approche sémantique en RI

Indexation classique -> indexation sémantique

Europe
Semanti
Web
Confere
Annual
Europe
Semanti
Web
Confere
Budva
Montene
present
results
Researc
Applica
Semanti
Web

Technol
Includi
Knowled
Markup
Languag
Semanti
Web
Service
Ontolog
managem
Feature
Special
Industr
Oriente
event

…

European [geographic area:Europe]
Semantic
Web [semantic Web]
Conference [conference]
Annual
European [geographic area:Europe]
Semantic
Web
Conference [conference]
Budva [City:Budva]
Montenegro [geographic area:Montenegro]
present [to present]
results [result]
Research [research]
Application [application]

Semantic
Web [semantic Web]
Technology [technology]
Including [to include]
Knowledge [knowledge]
Markup
Language [Markup Language]
Semantic
Web [semantic Web]
Service
Ontology [Ontology]
management [management]
Feature [to feature]
Special
Industrial [industrial]
Oriented
Event [event]

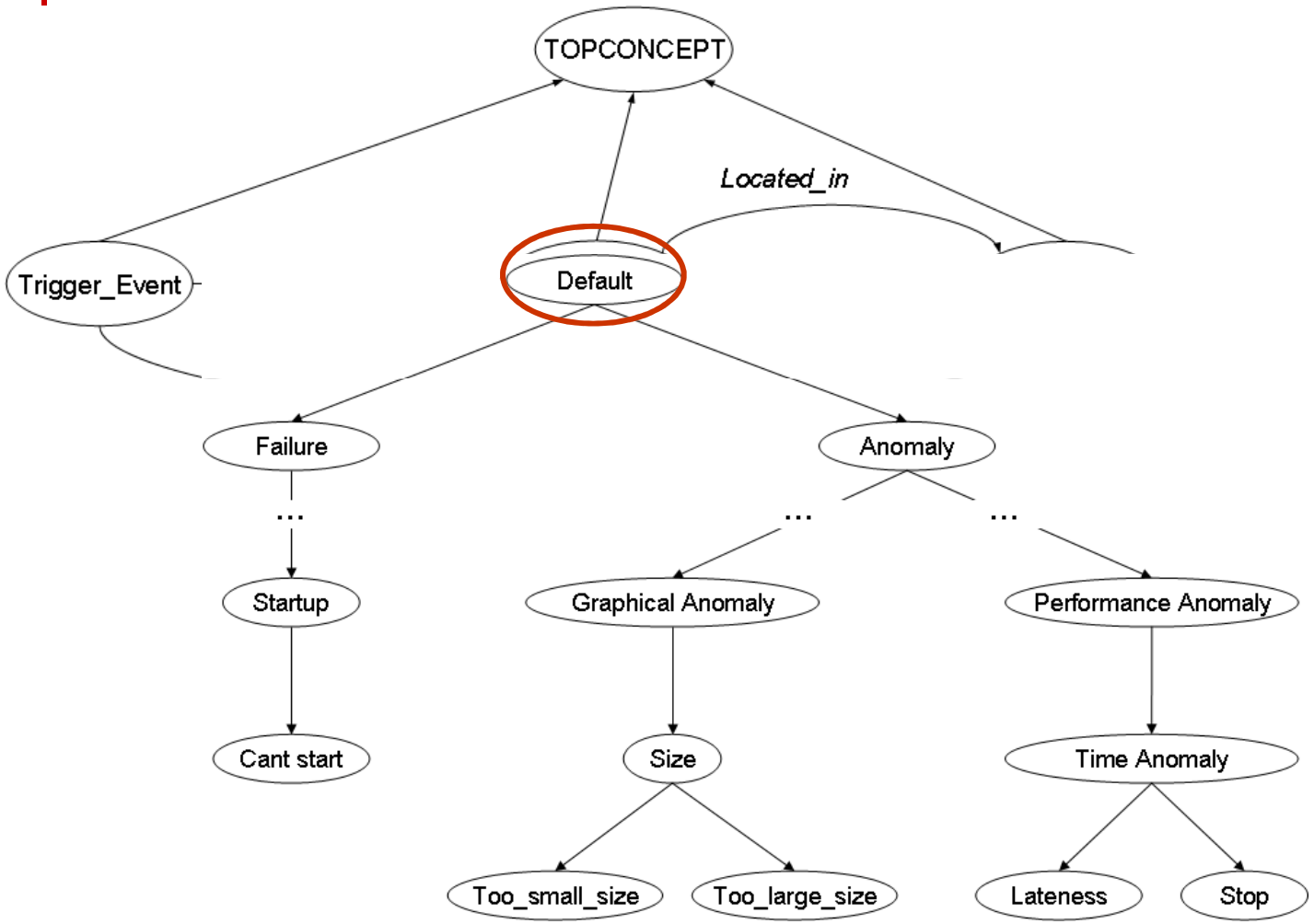
…

Institut de Recherche en Informatique de Toulouse

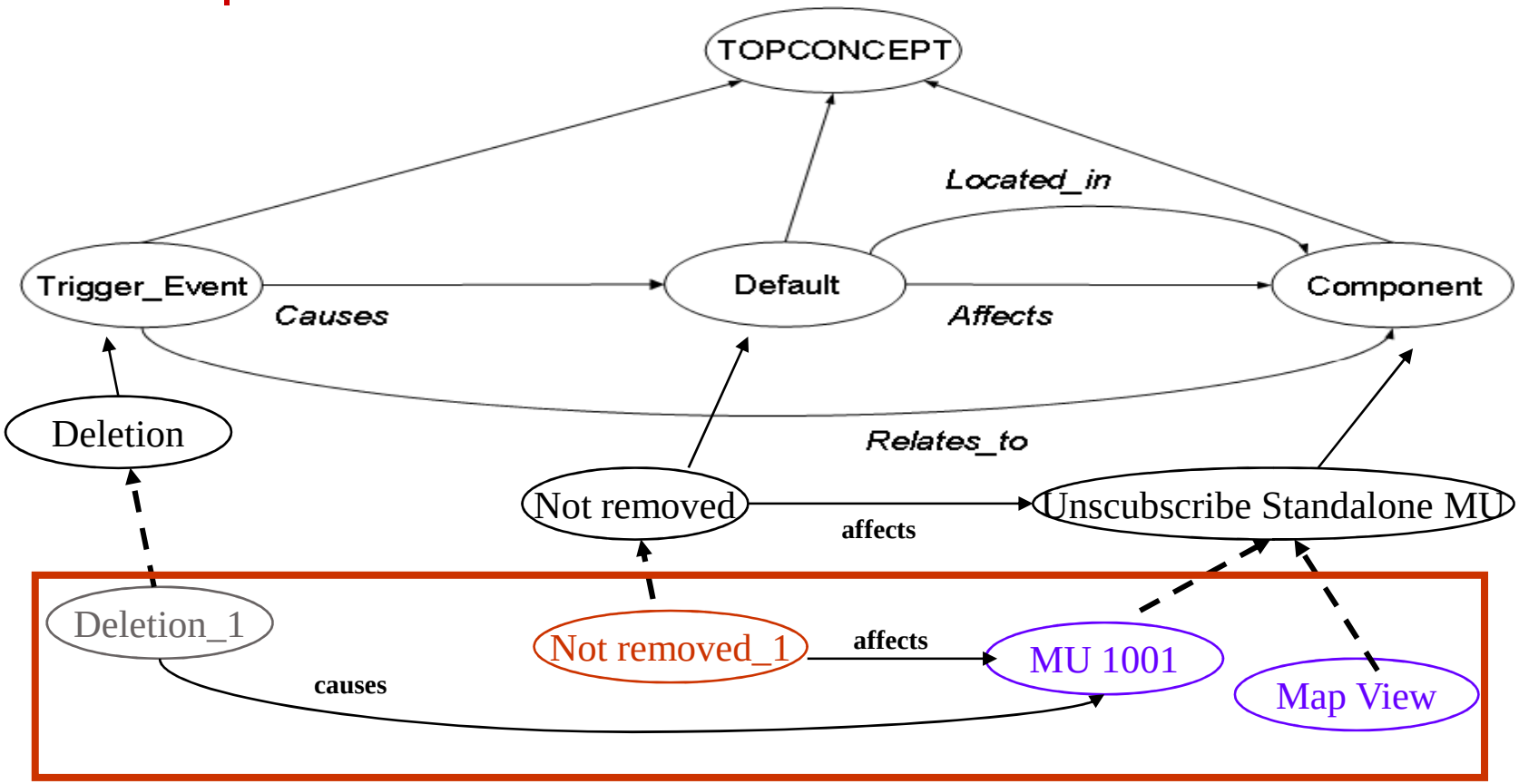
Approche sémantique en RI

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<rdf:RDF
<eswc:ConferenceEvent rdf:about="http://www.eswc2006.org/#">
  <dc:title>3rd European Semantic Web Conference (ESWC2006)</dc:title>
  <dc:description>The 3rd Annual European Semantic Web Conference (ESWC2006) will be held in Budva,

  <eswc:hasLocation rdf:resource="http://www.eswc2006.org/places/#Maestral" />
  <eswc:hasStartDateTime>2006-06-11T09:00:00+02:00</eswc:hasStartDateTime>
  <eswc:hasEndDateTime>2006-06-14T17:45:00+02:00</eswc:hasEndDateTime>
  <eswc:hasProgramme rdf:resource="http://www.eswc2006.org/conference_programme.html" />

  <eswc:hasProceedings rdf:resource="http://www.springer.com/uk/home/computer/database?SGWID=3-4010

  <eswc:hasProgramme rdf:resource="http://www.eswc2006.org/conference_programme.html" />

  <eswc:hasProceedings rdf:resource="http://www.springer.com/uk/home/computer/database?SGWID=3-40109-22-173660166-0" /

  <!--
  <eswc:hasCall rdf:resource="" />
  preferably add roles here
  <eswc:hasRole rdf:resource="" />
  -->
```

Annotation par un graphe d'instances

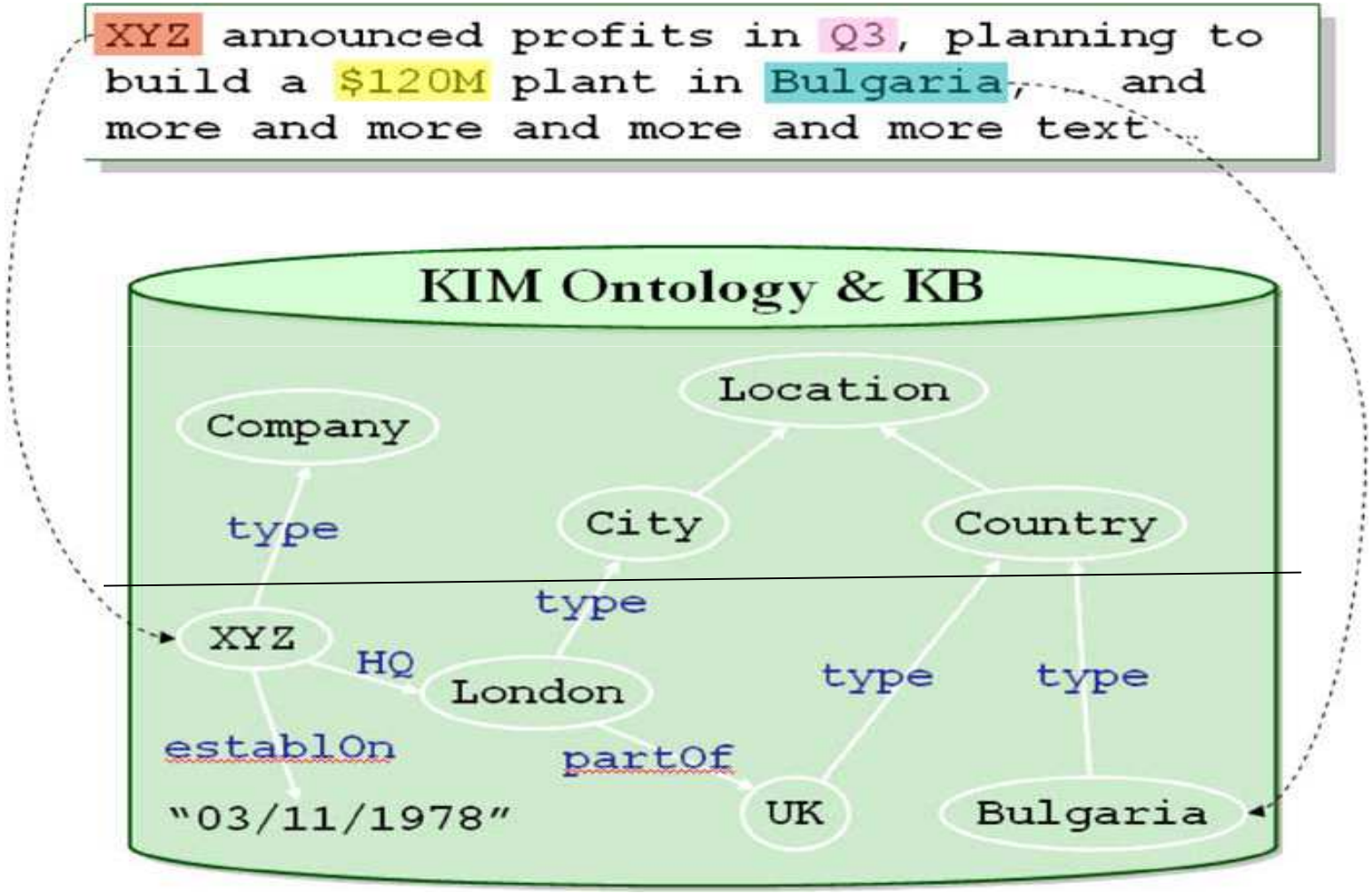


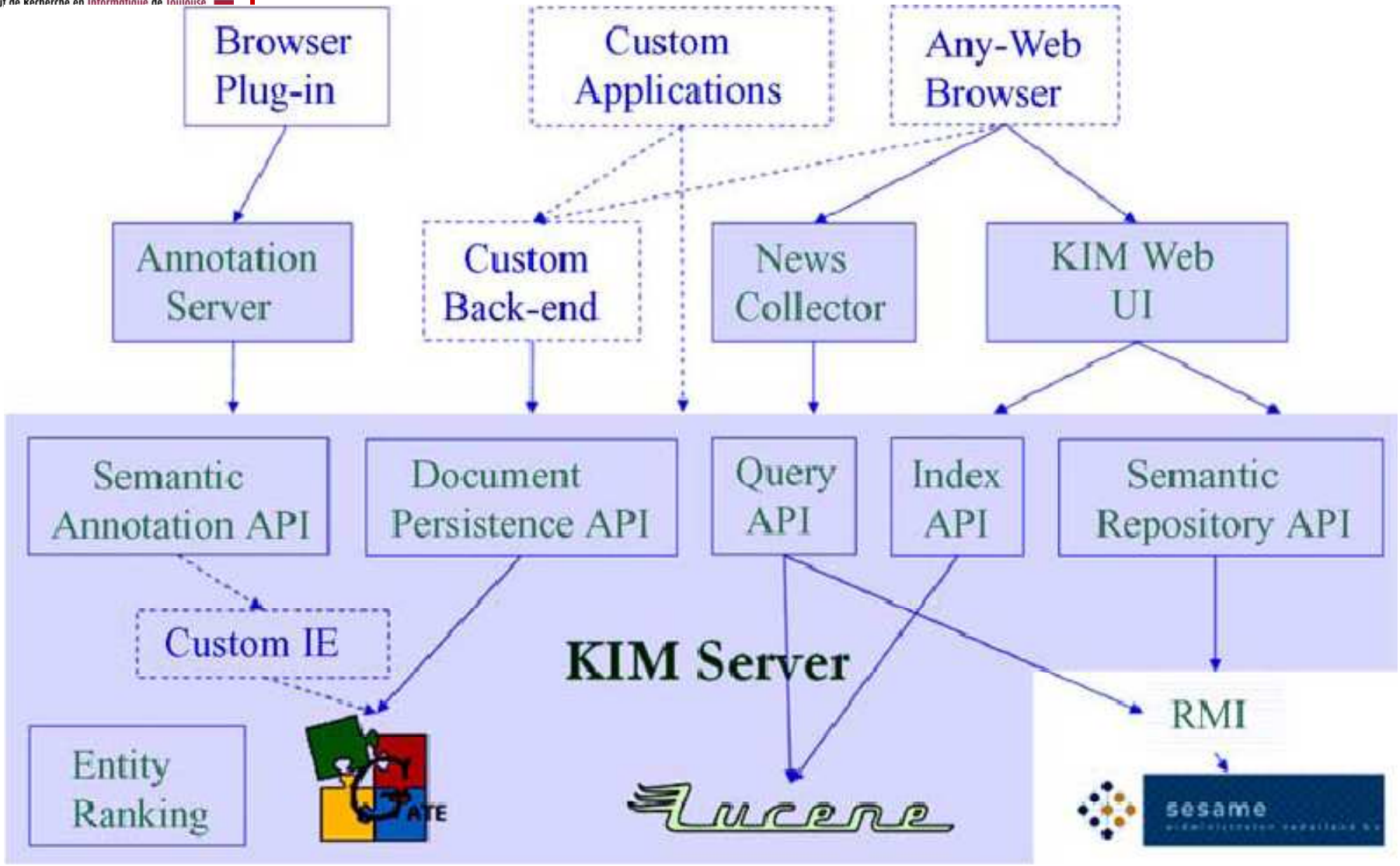
Annotation par un graphe d'instances

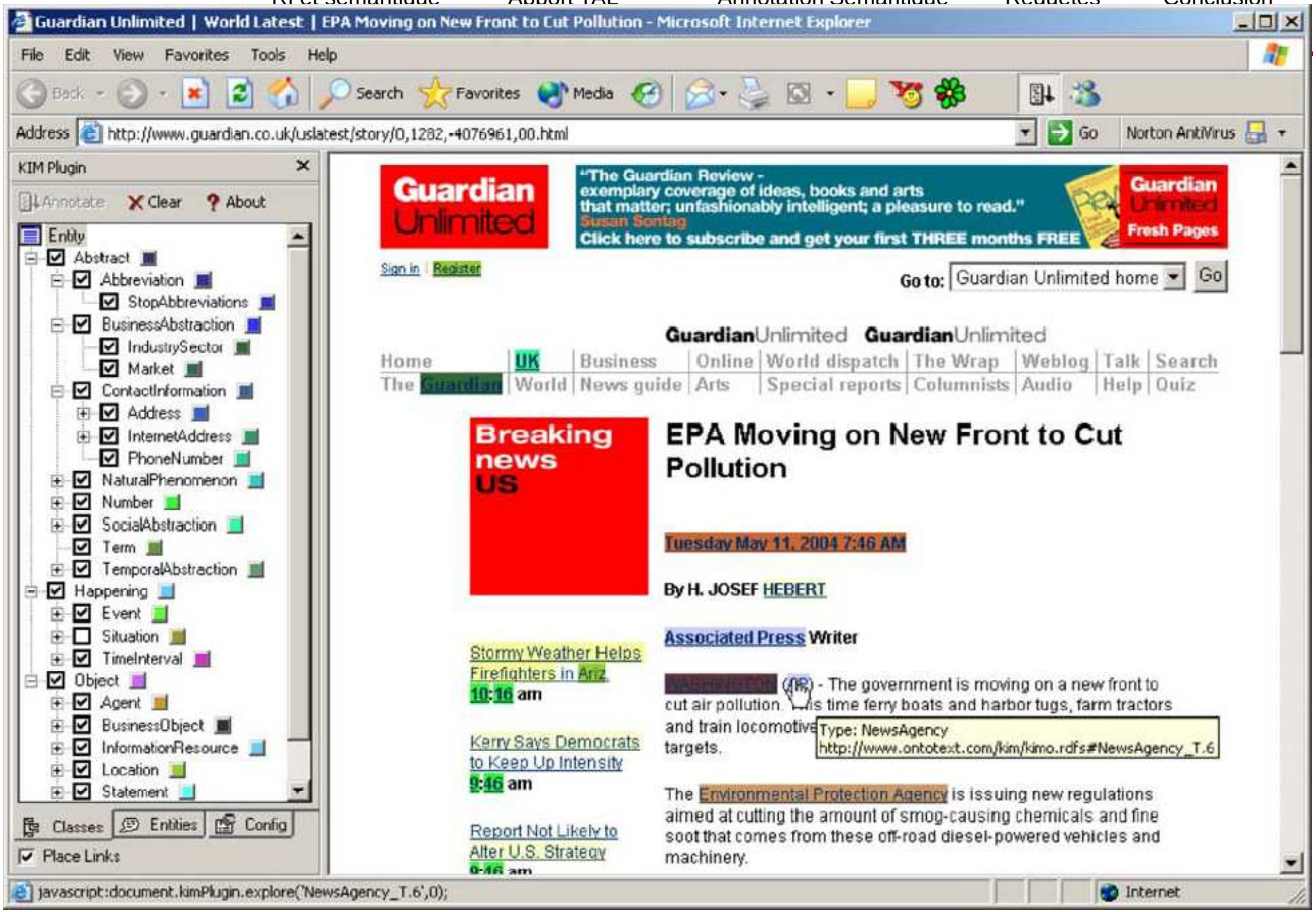


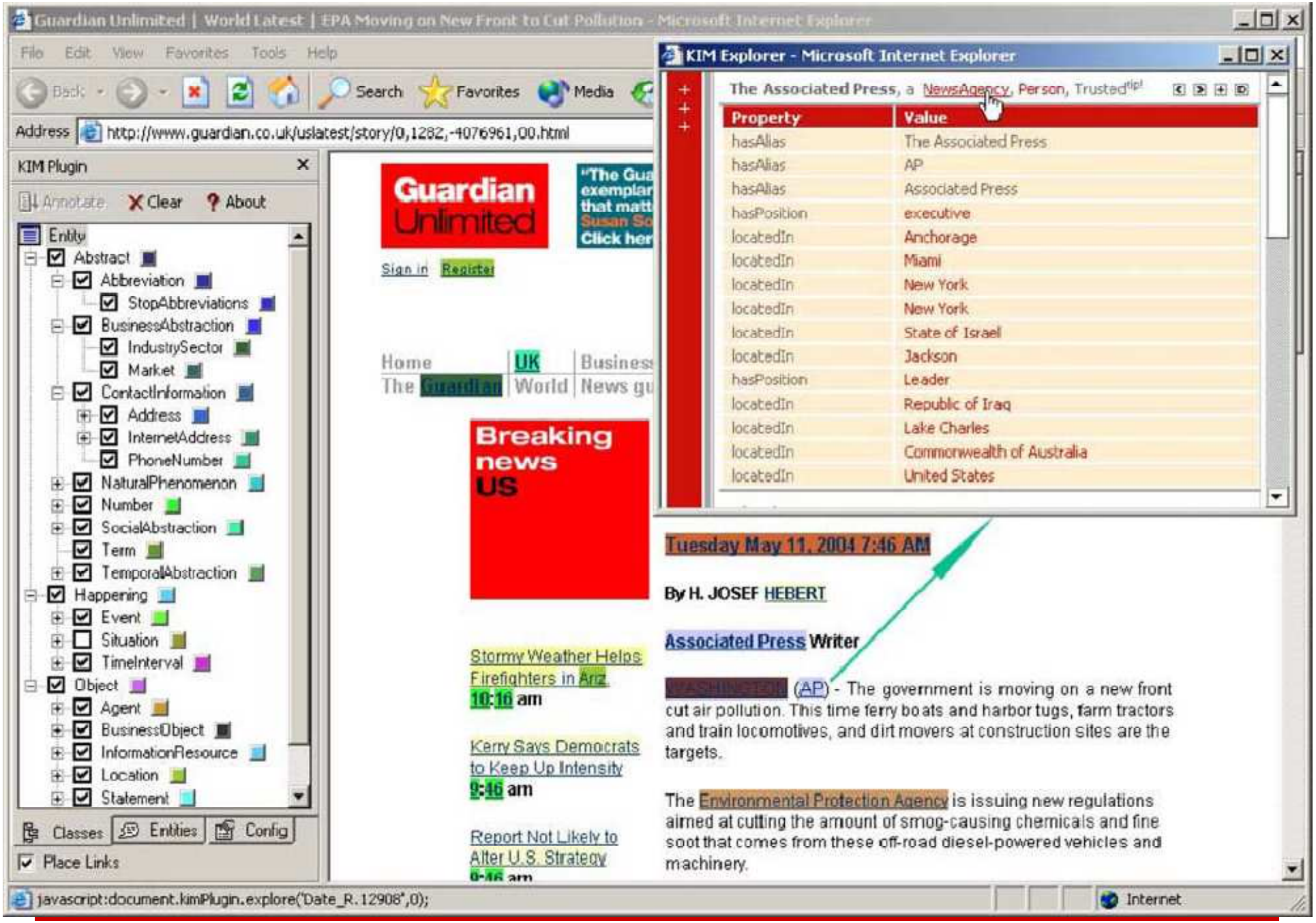
MU101 still visualized in map view after having been deleted

KIM : plate-forme d'extraction d'information




Browser Plug-in
Custom Applications
Any-Web Browser
Annotation Server
Custom Back-end
News Collector
KIM Web UI
Semantic Annotation API
Document Persistence API
Query API
Index API
Semantic Repository API
Custom IE
KIM Server
Entity Ranking
RMI
sesame


Guardian Unlimited | World Latest | EPA Moving on New Front to Cut Pollution - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.guardian.co.uk/uslatest/story/0,1282,-4076961,00.html
KIM Plugin
Annotate Clear About
Entity
Abstract
Abbreviation
StopAbbreviations
BusinessAbstraction
IndustrySector
Market
ContactInformation
Address
InternetAddress
PhoneNumber
NaturalPhenomenon
Number
SocialAbstraction
Term
TemporalAbstraction
Happening
Event
Situation
TimeInterval
Object
Agent
BusinessObject
InformationResource
Location
Statement
Classes Entities Config
Place Links
Guardian Unlimited
"The Guardian Review - exemplary coverage of ideas, books and arts that matter; unfashionably intelligent; a pleasure to read."
Susan Sontag
Click here to subscribe and get your first THREE months FREE
Fresh Pages
Sign in Register
Go to: Guardian Unlimited home Go
Home UK Business Online World dispatch The Wrap Weblog Talk Search
The Guardian World News guide Arts Special reports Columnists Audio Help Quiz
Breaking news US
EPA Moving on New Front to Cut Pollution
Tuesday May 11, 2004 7:46 AM
By H. JOSEF HEBERT
Associated Press Writer
Stormy Weather Helps Firefighters in Ariz. 10:16 am
Kerry Says Democrats to Keep Up Intensity 9:46 am
Report Not Likely to Alter U.S. Strategy
- The government is moving on a new front to cut air pollution. This time ferry boats and harbor tugs, farm tractors and train locomotives
Type: NewsAgency
http://www.ontotext.com/kim/kimo.rdfs#NewsAgency_T.6
The Environmental Protection Agency is issuing new regulations aimed at cutting the amount of smog-causing chemicals and fine soot that comes from these off-road diesel-powered vehicles and machinery.
javascript:document.kimPlugin.explore('NewsAgency_T.6',0);
Internet


Guardian Unlimited | World Latest | EPA Moving on New Front to Cut Pollution - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.guardian.co.uk/uslatest/story/0,1282,-4076961,00.html
KIM Plugin
Annotate Clear About
Entity
Abstract
Abbreviation
StopAbbreviations
BusinessAbstraction
IndustrySector
Market
ContactInformation
Address
InternetAddress
PhoneNumber
NaturalPhenomenon
Number
SocialAbstraction
Term
TemporalAbstraction
Happening
Event
Situation
TimeInterval
Object
Agent
BusinessObject
InformationResource
Location
Statement
Classes Entities Config
Place Links
Guardian Unlimited
Sign in Register
Home UK Business
The Guardian World News gu
Breaking news US
Stormy Weather Helps Firefighters in Ariz. 10:16 am
Kerry Says Democrats to Keep Up Intensity 9:46 am
Report Not Likely to Alter U.S. Strategy 9:46 am
KIM Explorer - Microsoft Internet Explorer
The Associated Press, a NewsAgency, Person, Trusted
Property Value
hasAlias The Associated Press
hasAlias AP
hasAlias Associated Press
hasPosition executive
locatedIn Anchorage
locatedIn Miami
locatedIn New York
locatedIn New York
locatedIn State of Israel
locatedIn Jackson
hasPosition Leader
locatedIn Republic of Iraq
locatedIn Lake Charles
locatedIn Commonwealth of Australia
locatedIn United States
Tuesday May 11, 2004 7:46 AM
By H. JOSEF HEBERT
Associated Press Writer
(AP) - The government is moving on a new front cut air pollution. This time ferry boats and harbor tugs, farm tractors and train locomotives, and dirt movers at construction sites are the targets.
The Environmental Protection Agency is issuing new regulations aimed at cutting the amount of smog-causing chemicals and fine soot that comes from these off-road diesel-powered vehicles and machinery.
javascript:document.kimPlugin.explore('Date_R.12908',0);
Internet

Institut de Recherche en Informatique de Toulouse

Annotation sémantique : défis

Limites de la collocation

Iran **invaded** Iraq **on September 22, 1980**. **After eight long years** of fighting, …
UN resolution 598ended the war

How long *did Iraq* ***fight with*** *Iran?*
??? **September 22, 1980 ?**??

Besoins

- Ressource lexicale : ***fight with*** = *invaded*
- Calculer les rattachements : **on** introduit **September 22, 1980, after** répond à **how long** **et** introduit **eight long years**

Institut de Recherche en Informatique de Toulouse

Annotation sémantique : défis

Résolution des anaphores, relations exprimées sur plusieurs phrases

Andorre est un petit pays enclavé entre le France et l'Espagne dans le sud-Ouest de l'Europe,
…
Le tourisme est le secteur principal de ce petit pays avec une économie saine, qui compte pour environ 80% de son PNB
…

Quel est le secteur économique majeur de l'Andorre ?

Identifier des éléments d'ontologie à partir de textes : types de connaissances

TOUT x, Pays(x) -> <il existe y, capitale(y) et Etre-Capitale(y,x)

Disjoint (rivière, montagne)

ETRE-CAPITALE < SITUED-DANS

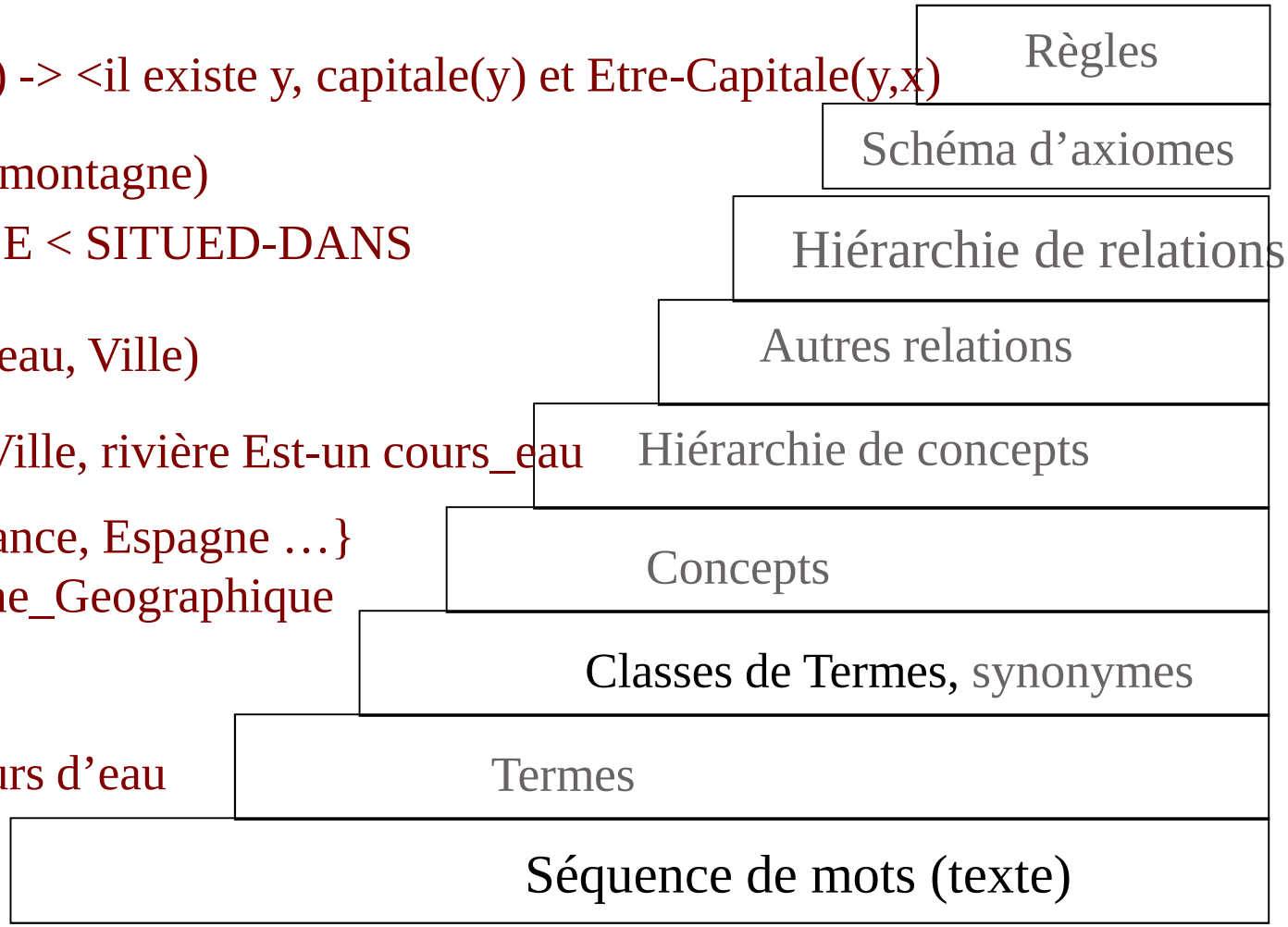
Traverse (cours_eau, Ville)

Capitale Est-un Ville, rivière Est-un cours_eau

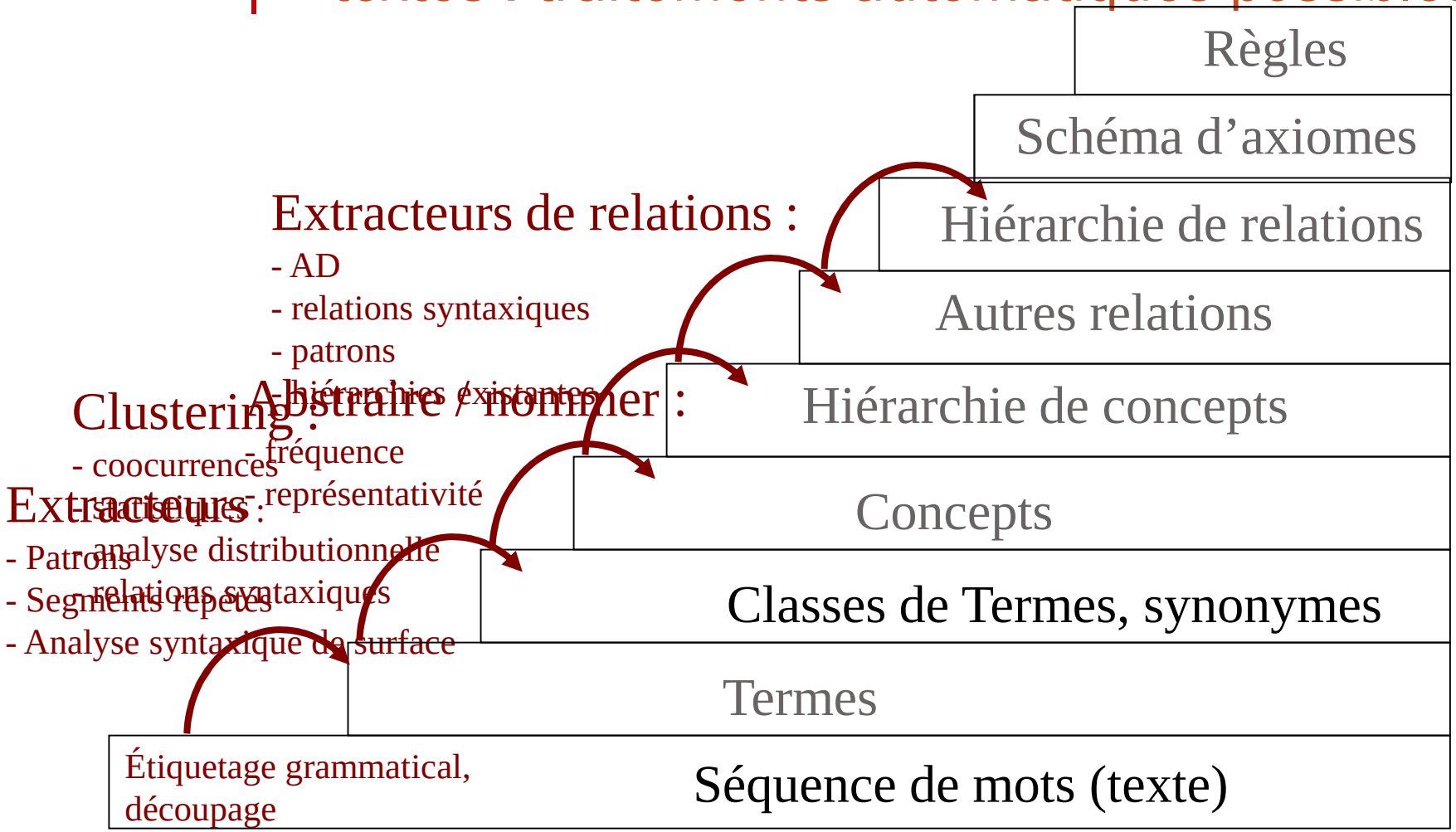
C := pays := {France, Espagne …}
:= type-de Zone_Geographique

{nation, pays}

Ville, rivière, cours d'eau



Identifier des éléments d'ontologie à partir de textes : traitements automatiques possibles



Institut de Recherche en Informatique de Toulouse

Repérage de concepts

- Extracteurs de termes
  - TerminoWeb, TermoStat (univ. Montreal), TermExtractor (Univ. La Sapienza), Nomino, Ana
  - Yatea (LIPN)
  - Calcul de variantes : FASTER
  - Regroupement en classes : Asium,
- À partir des termes extraits (Noms ou SN)
  - Fréquence, Productivité
  - Synonymes de termes très fréquents
  - Terme apparaissant dans les titres des chapitres du livre
- A partir de la prise en compte des besoins
  - Connaissance superficielle du corpus et du domaine
  - Avis des experts

Institut de Recherche en Informatique de Toulouse

Syntex (Bourigault, 2000)

- Extracteur de termes et analyseur syntaxique
- A partir d'un corpus étiqueté, produit :
  - un corpus analysé syntaxiquement
  - un réseau de syntagmes
    - syntagmes verbaux, nominaux, adjectivaux
    - réseau structuré par les relations Tête et Expansion
- Interface de consultation OntoTerm
- Principes de base
  - Analyse syntaxique et analyse distributionnelle

Institut de Recherche en Informatique de Toulouse

Etude des relations à partir d'un concept

- Des relations lexicales aux relations sémantiques
  - Déterminer les concepts reliés et le type des relations
  - Les représenter à l'aide de rôles ou de relations EST-UN
  - Décider des propriétés des relations : héritage, …
- Des relations à une hiérarchie de concepts
  - Différencier des concepts selon des critères homogènes : corps solide/corps liquide, différents processus de fabrication …
  - Relations révèlent des critères de différentiation
  - Ajout de concepts non terminologiques

Institut de Recherche en Informatique de Toulouse

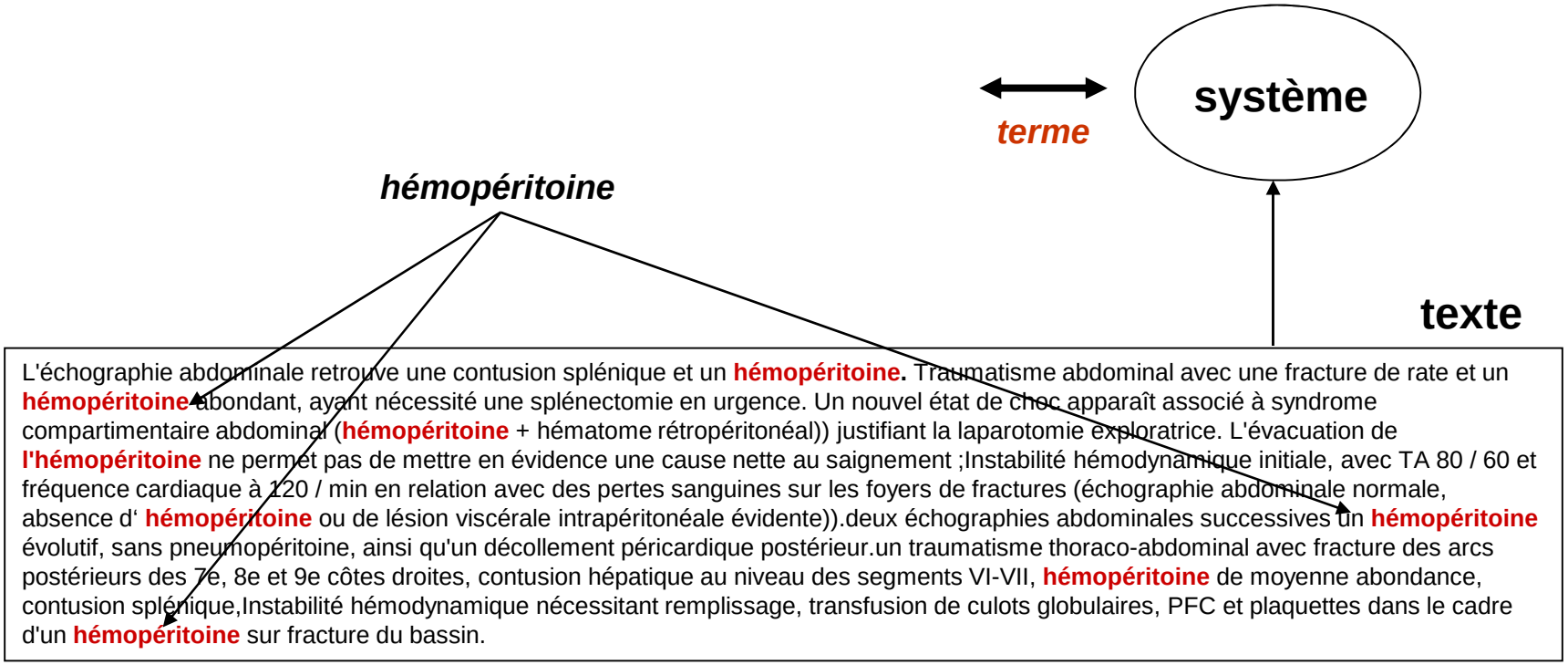
Recherche de relations à l'aide de patrons lexico-sémantiques

- Adaptation de marqueurs génériques
  - ex: SN EST art indef SN …
  - peu précis en français en général, pertinent en anglais sur un corpus didactique
- Identification de nouveaux marqueurs
  - Régularités syntaxiques dans les corpus :
    - Dans le corpus « brevets », un marqueur pour repérer un nom de process ***NOM_Process according to the invention***
    - Dans le corpus veille, des propriétés de (chemical) company

    SN chemical company (ex Saint Louis' based chemical company)
  - Résultats de SYNTEX

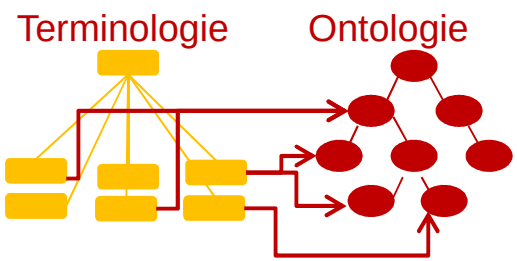
Quelles ontologies pour l'annotation sémantique ?

Défi : retrouver les variantes de termes associées à un concept

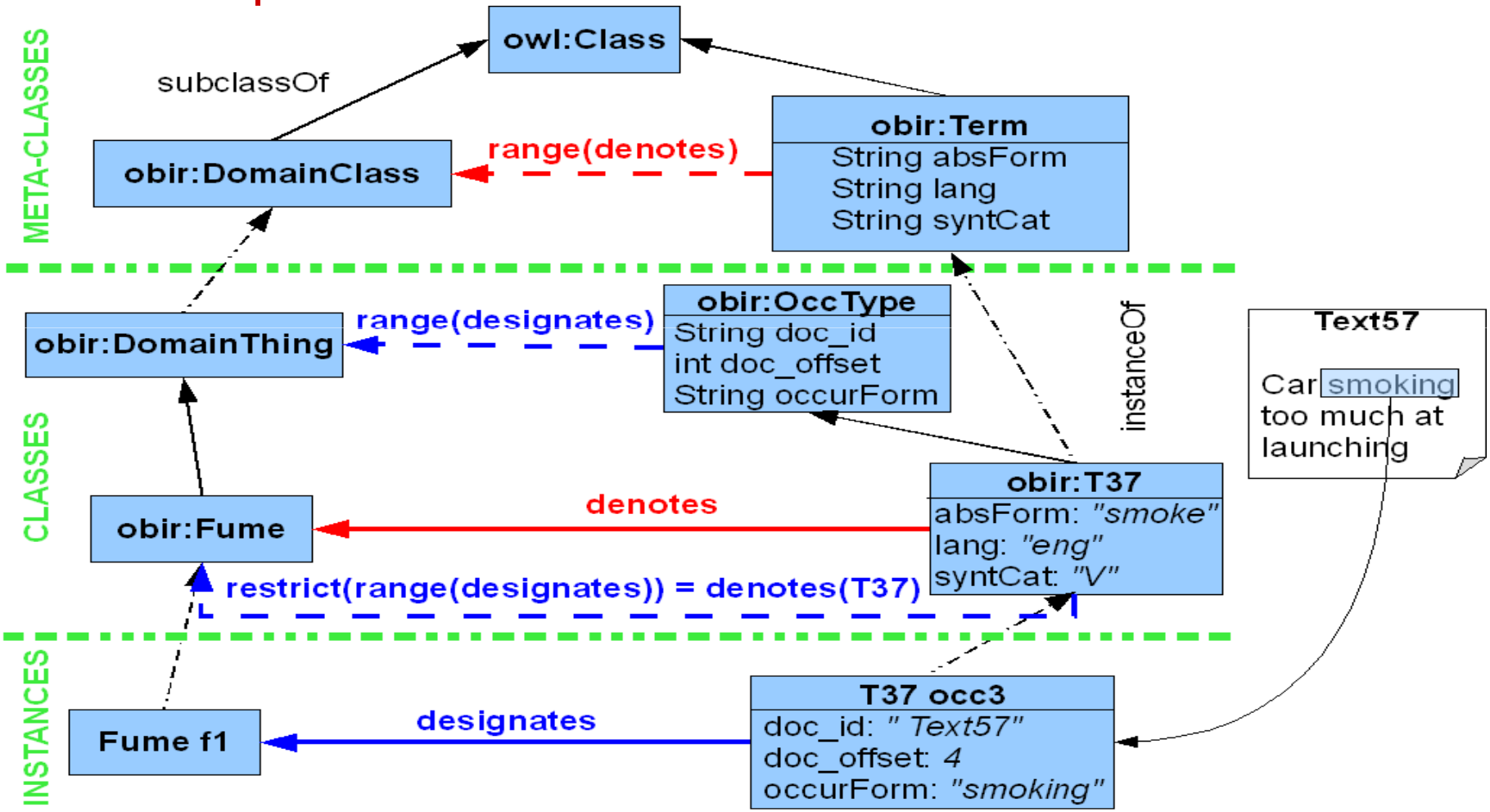


Quelles ontologies pour l'annotation sémantique ?

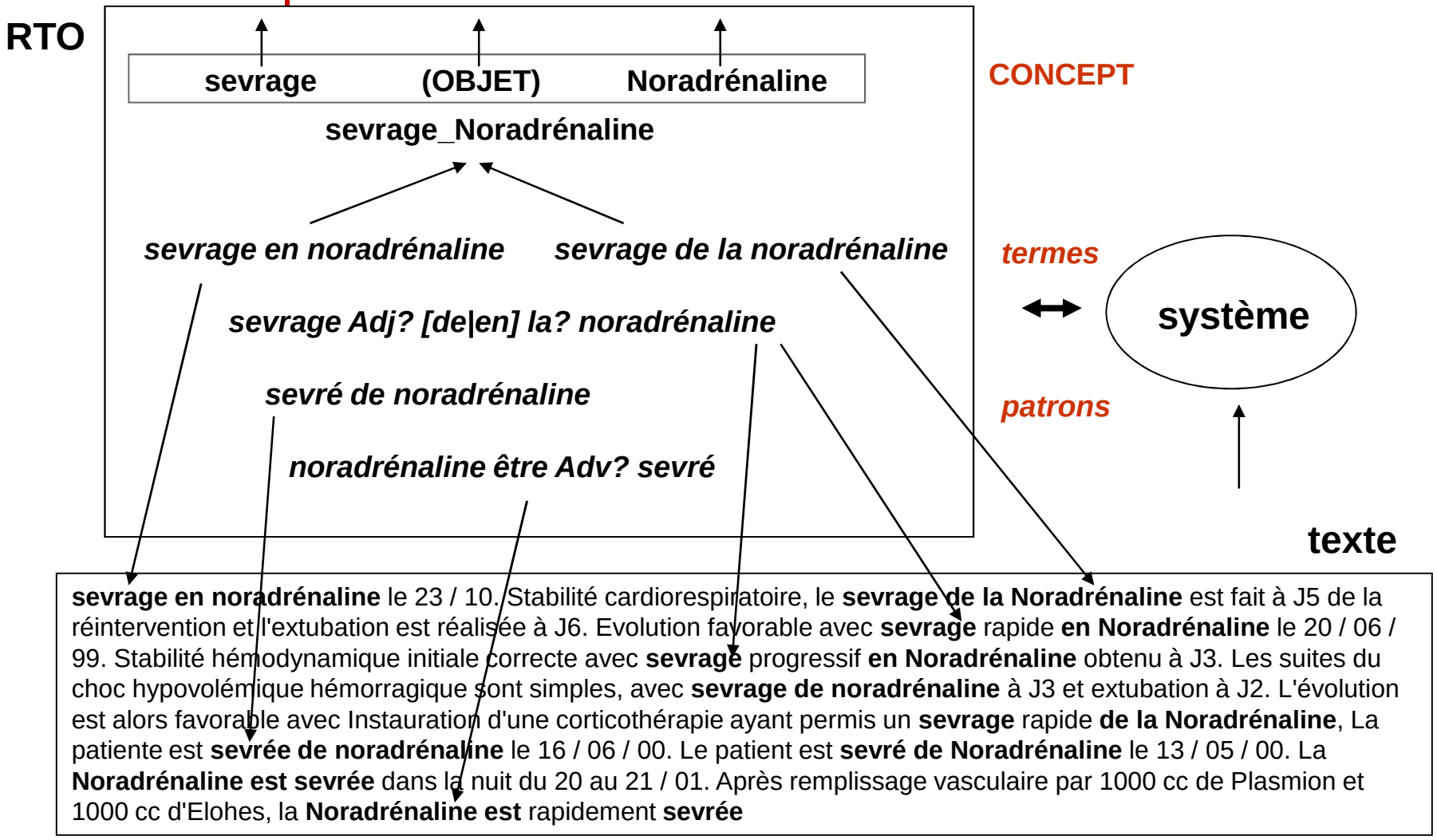
- Intérêt d'une composante lexicale
  - Les termes dénotent les concepts
  - Méta-modèle de RTO *(Reymonet et al, 09)*
  - Modèle d'ontologie de la plate-forme DAFOE *(Charlet et al, 09)*



Méta-modèle de RTO *(Reymonet et al, 09)*



Quelle ontologie pour l'annotation sémantique ?

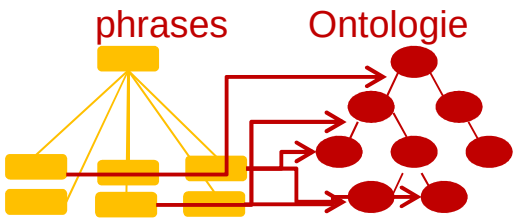


sevrage en noradrénaline le 23 / 10. Stabilité cardiorespiratoire, le **sevrage de la Noradrénaline** est fait à J5 de la réintervention et l'extubation est réalisée à J6. Evolution favorable avec **sevrage** rapide **en Noradrénaline** le 20 / 06 / 99. Stabilité hémodynamique initiale correcte avec **sevrage** progressif **en Noradrénaline** obtenu à J3. Les suites du choc hypovolémique hémorragique sont simples, avec **sevrage de noradrénaline** à J3 et extubation à J2. L'évolution est alors favorable avec Instauration d'une corticothérapie ayant permis un **sevrage** rapide **de la Noradrénaline**, La patiente est **sevrée de noradrénaline** le 16 / 06 / 00. Le patient est **sevré de Noradrénaline** le 13 / 05 / 00. La **Noradrénaline est sevrée** dans la nuit du 20 au 21 / 01. Après remplissage vasculaire par 1000 cc de Plasmion et 1000 cc d'Elohes, la **Noradrénaline est** rapidement **sevrée**

Institut de Recherche en Informatique de Toulouse

Quelles ontologies pour l'annotation sémantique ?

- Le cas des relations
  - Des termes dénotent-ils des relations ?
  - Traces linguistiques des relations
    - Patrons lexico-syntaxiques
    - Indices : verbes (*sert-à, est un*) , préposition (*chez, quand, …)*
    - Phrases exemples
  - Où placer ces informations ?

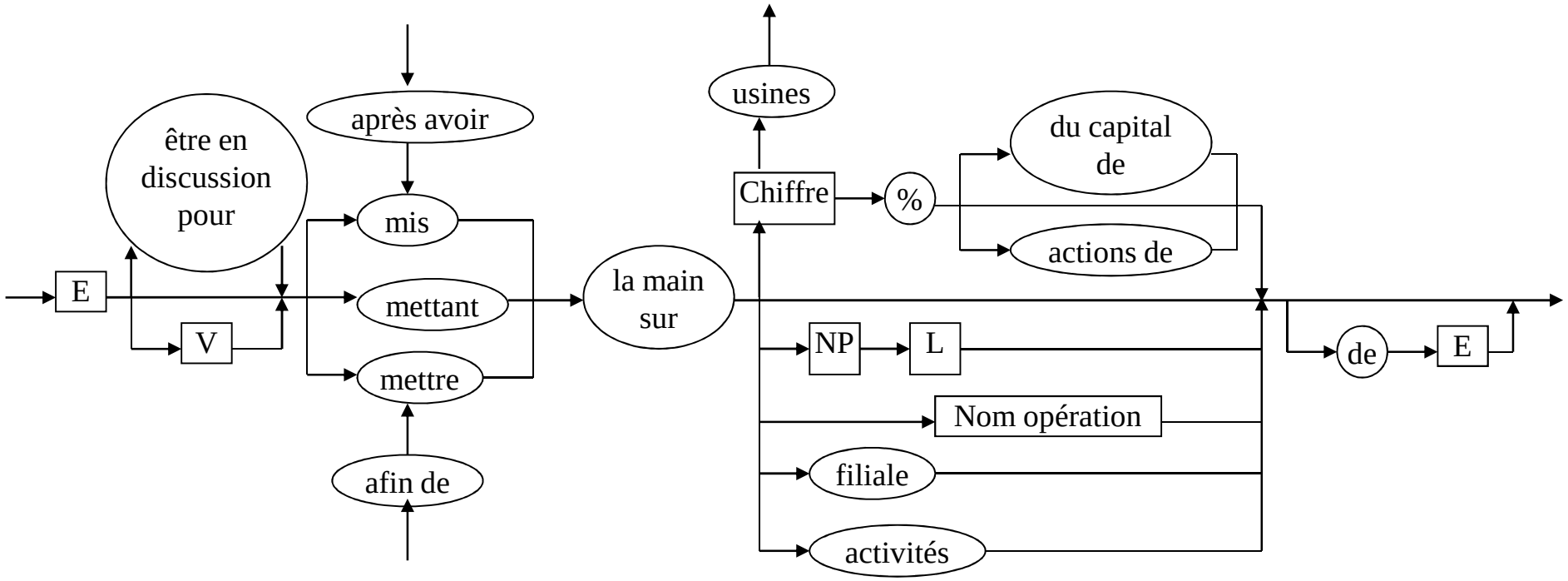


Institut de Recherche en Informatique de Toulouse

Extraction d'information et apprentissage

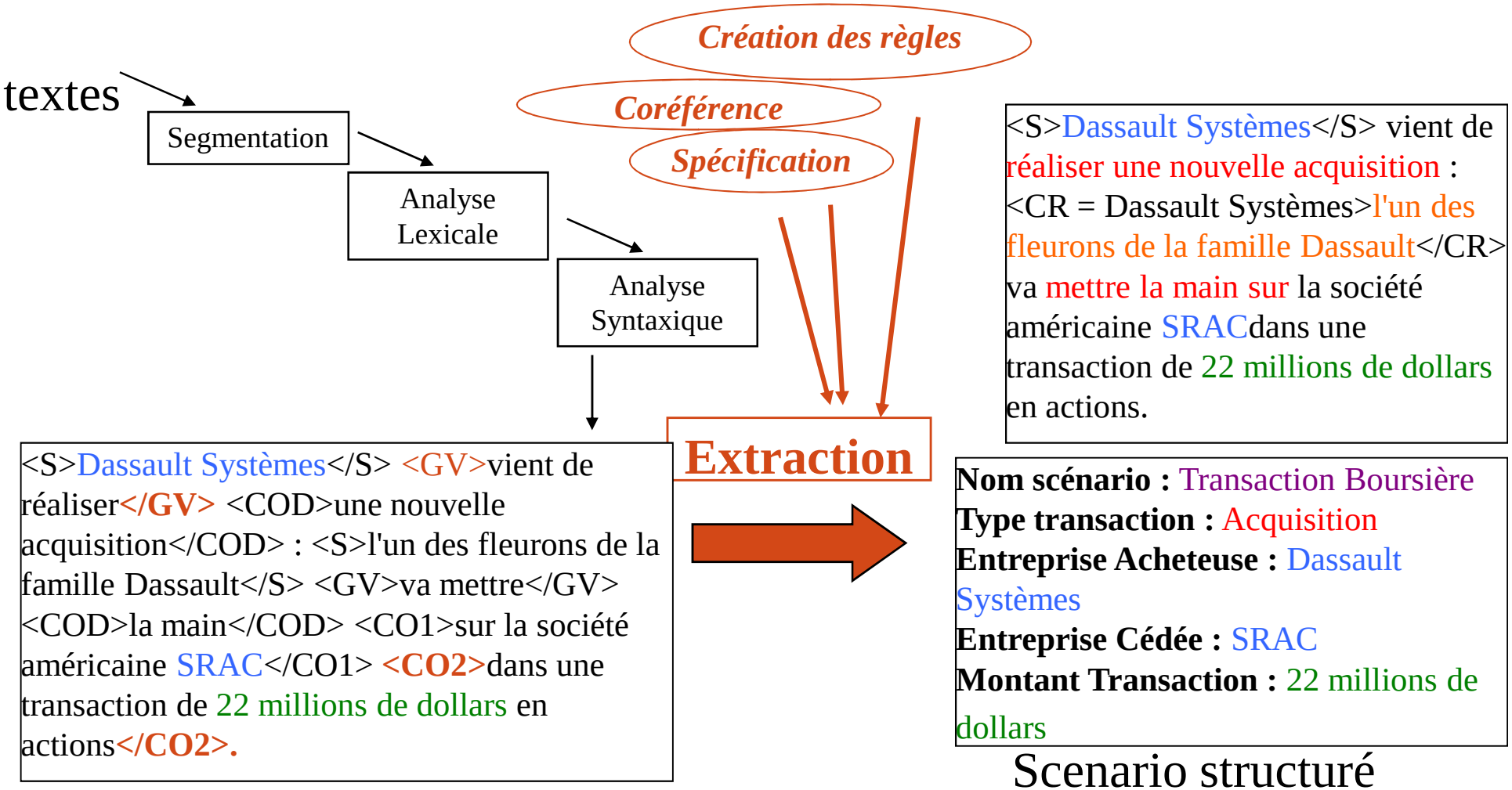
- Extraction d'information : principes
  - Patrons de fouille
  - Caractérisation syntaxique, lexicale et sémantique de la présence de certaines informations
  - Automates appliquant les patrons sur des textes
- Extraction d'information : intérêt
  - Repérage de concepts, de relations entre concepts
  - Repérage d'instances
- Extraction d'information : coût
  - Identification des patrons
  - Mise au point des automates

Extraction d'information : exemple de patron *(Amardeilh, 2002)*



Automate du patron « mettre la main sûr »

Extraction d'information : extraction



Institut de Recherche en Informatique de Toulouse

Intérêt de combiner TAL et apprentissage

- Apprentissage à partir d'exemples
  - Textes étiquetés grammaticalement
  - Puis annotés manuellement (tag) pour
  - Indiquer la présence de concepts et de relations entre concepts

  - Le système « apprend » la caractérisation syntaxique des contextes dans lesquels les concepts sont présents
  - = nouveaux patrons de fouille

- Extraction d'information
  - Utilisation des patrons appris
  - Pour le repérage d'instances de concepts

Institut de Recherche en Informatique de Toulouse

Formuler les requêtes

- Utiliser un langage d'interrogation
  - Ex de SPARQL pour de représentations RDFs
- Interface en LN
- Interface guidée par l'ontologie

Institut de Recherche en Informatique de Toulouse

Formulation en SPARQL

- interroger un ensemble de descriptions RDF à partir d'opérations de mise en correspondance de patterns de graphes
- Pattern de graphe de base **:** un pattern est composé d'une ou plusieurs description(s) RDF dans lesquelles on peut introduire une (ou plusieurs) variable(s) à la place de la ressource/propriété/valeur.

- Exemple : soit la liste de descriptions RDF interrogée

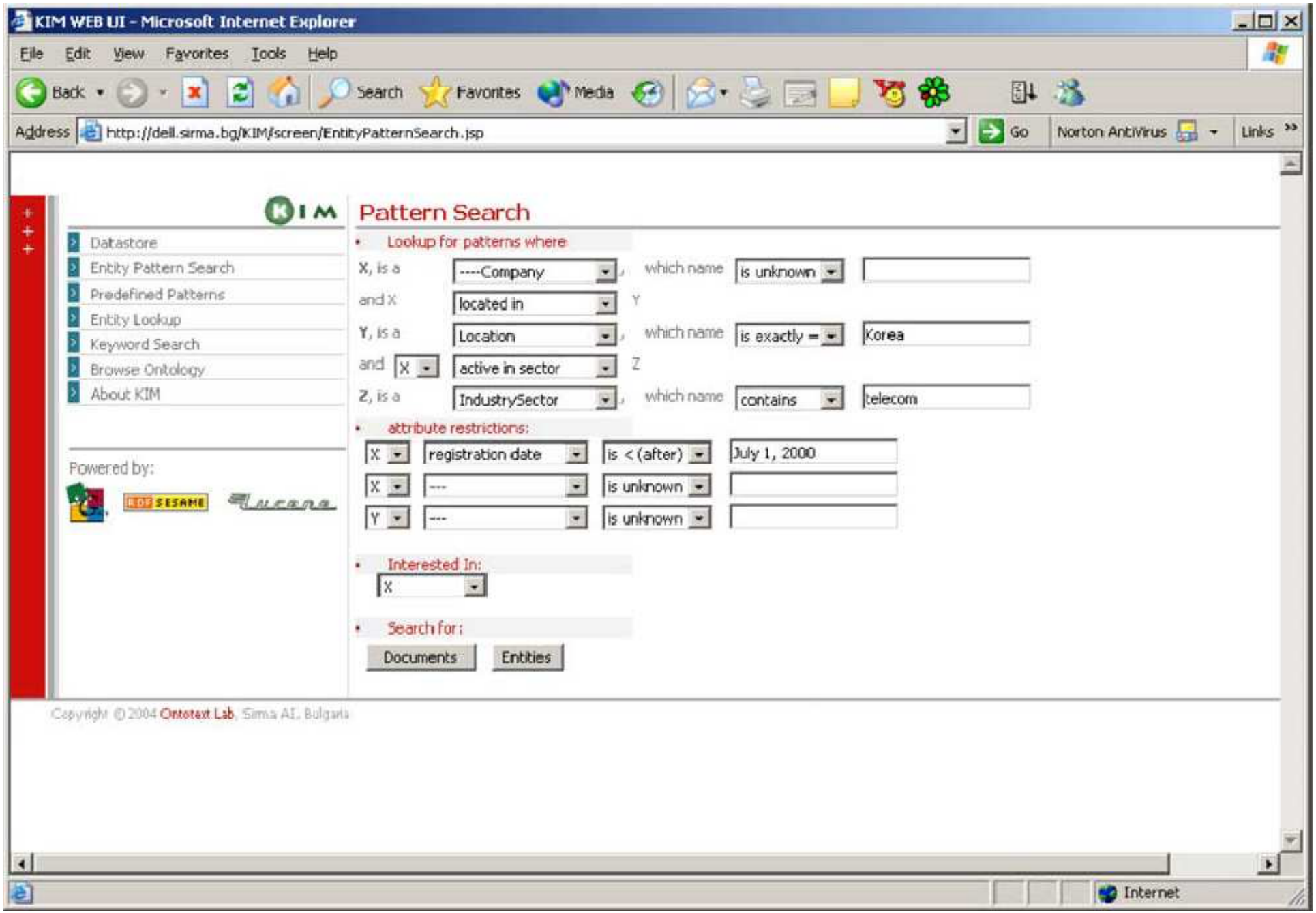
```
@prefix foaf: <http://xmlns.com/foaf/0.1/>
.
_:a foaf:name "Johnny Lee Outlaw" .
_:a foaf:mbox <mailto:outlaw@example.com> .

_:b foaf:name "A. N. Other" .
_:b foaf:mbox <mailto:other@example.com> .
```

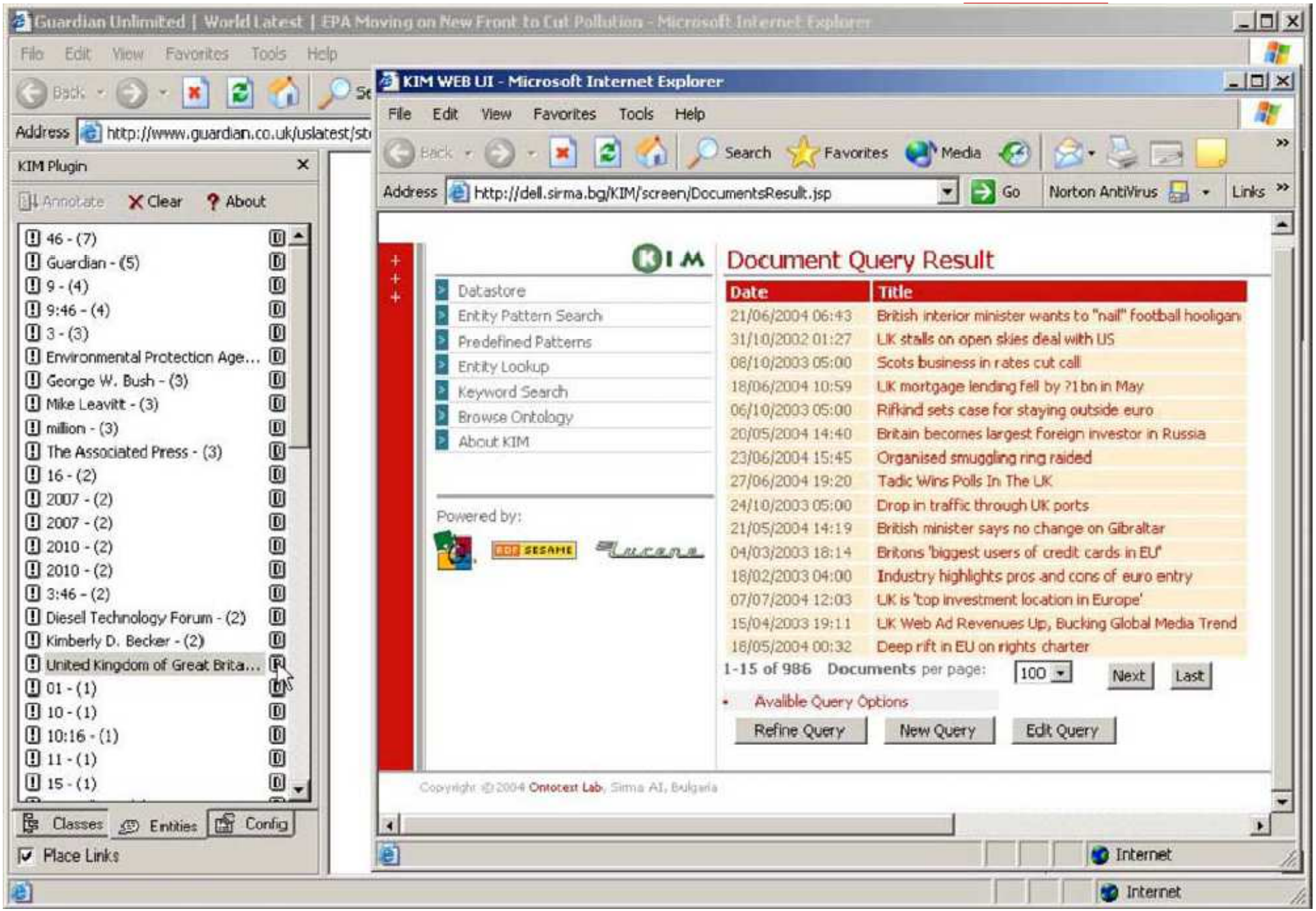
La requête SPARQL suivante :

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?mbox
WHERE
{ ?x foaf:name "Johnny Lee Outlaw" .
?x foaf:mbox ?mbox }
```

a pour résultat:

```
mbox
<mailto:outlaw@example.com>
```


KIM WEB UI - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://dell.sirma.bg/KIM/screen/EntityPatternSearch.jsp
Go Norton AntiVirus Links
KIM
Datastore
Entity Pattern Search
Predefined Patterns
Entity Lookup
Keyword Search
Browse Ontology
About KIM
Powered by:
SESAME
Pattern Search
Lookup for patterns where
X, is a ----Company which name is unknown
and X located in Y
Y, is a Location which name is exactly = Korea
and X active in sector Z
Z, is a IndustrySector which name contains telecom
attribute restrictions:
X registration date is < (after) July 1, 2000
X --- is unknown
Y --- is unknown
Interested In:
X
Search for:
Documents Entities
Copyright © 2004 Ontotext Lab, Sirma AI, Bulgaria
Internet


Guardian Unlimited | World Latest | EPA Moving on New Front to Cut Pollution - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.guardian.co.uk/uslatest/st
KIM Plugin
Annotate
Clear
About
46 - (7)
Guardian - (5)
9 - (4)
9:46 - (4)
3 - (3)
Environmental Protection Age...
George W. Bush - (3)
Mike Leavitt - (3)
million - (3)
The Associated Press - (3)
16 - (2)
2007 - (2)
2007 - (2)
2010 - (2)
2010 - (2)
3:46 - (2)
Diesel Technology Forum - (2)
Kimberly D. Becker - (2)
United Kingdom of Great Brita...
01 - (1)
10 - (1)
10:16 - (1)
11 - (1)
15 - (1)
Classes
Entities
Config
Place Links
KIM WEB UI - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address http://dell.sirma.bg/KIM/screen/DocumentsResult.jsp
Go
Norton AntiVirus
Links
KIM
Datastore
Entity Pattern Search
Predefined Patterns
Entity Lookup
Keyword Search
Browse Ontology
About KIM
Powered by:
SESAME
Document Query Result
Date
Title
21/06/2004 06:43
British interior minister wants to "nail" football hooligan
31/10/2002 01:27
UK stalls on open skies deal with US
08/10/2003 05:00
Scots business in rates cut call
18/06/2004 10:59
UK mortgage lending fell by ?1bn in May
06/10/2003 05:00
Rifkind sets case for staying outside euro
20/05/2004 14:40
Britain becomes largest foreign investor in Russia
23/06/2004 15:45
Organised smuggling ring raided
27/06/2004 19:20
Tadic Wins Polls In The UK
24/10/2003 05:00
Drop in traffic through UK ports
21/05/2004 14:19
British minister says no change on Gibraltar
04/03/2003 18:14
Britons 'biggest users of credit cards in EU'
18/02/2003 04:00
Industry highlights pros and cons of euro entry
07/07/2004 12:03
UK is 'top investment location in Europe'
15/04/2003 19:11
UK Web Ad Revenues Up, Bucking Global Media Trend
18/05/2004 00:32
Deep rift in EU on rights charter
1-15 of 986 Documents per page: 100
Next
Last
Avalible Query Options
Refine Query
New Query
Edit Query
Copyright © 2004 Ontotext Lab, Sirma AI, Bulgaria
Internet
Internet

Institut de Recherche en Informatique de Toulouse

Formuler des requêtes – PICSEL

(Reynaud, Saffa, Roussey et al, 2002)

Spécification des propriétés

Vous pouvez précisez les propriétés pour le concept : sejourSportif

propriété	concept associé	
prixAssocie		Chiffre
dateDebutAssocie		Date:JJMMAA
serviceProduitAssocie	service	Préciser
prestationProduitAssocie	prestationRepas	Préciser
nbreNuitsCombineSejourAssocie		Chiffre
logementProduitAssocie	location	Préciser
sejourTrajetAssocie	trajetTrain	Préciser
activiteAssocie	activiteSportive	Préciser
loisirProduitAssocie	loisirDetente	Préciser
activitePrincipaleAssocie	activiteSportive	Préciser

OK

Go

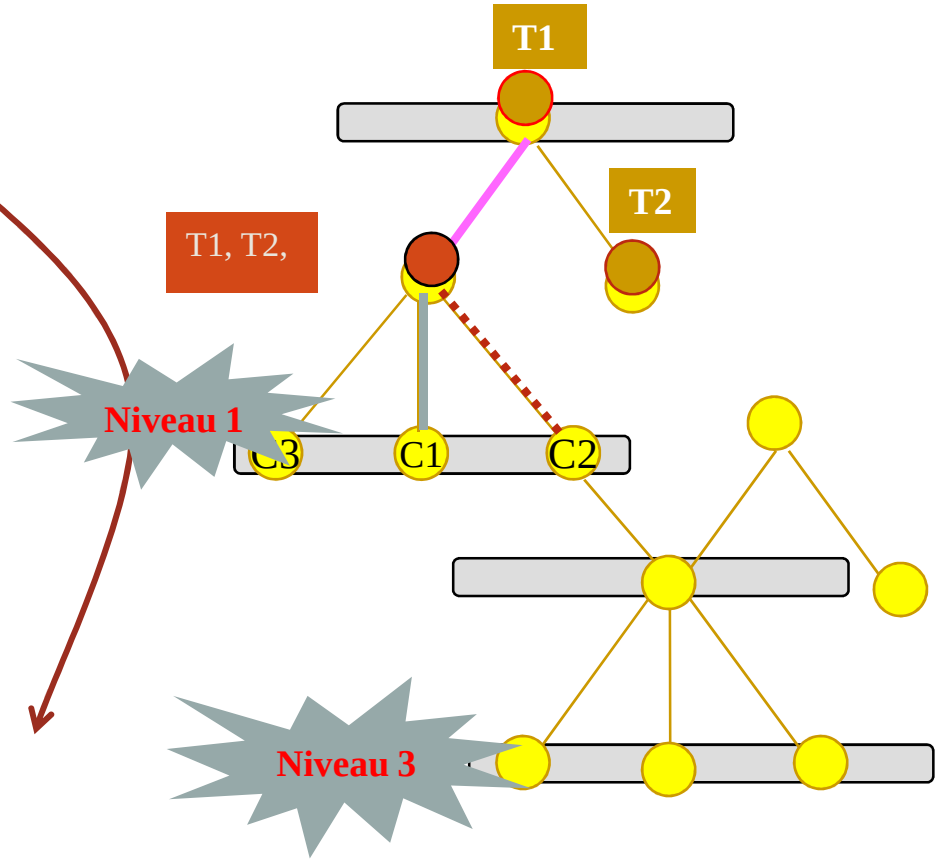
Etendre une requête via une ontologie

Requête initiale :

T1, T2

R1= T1, T2, ..Tn

Ex: hôtel plage à Madeira



Requête étendue :

R2= T1, T2, …Tn,

Ex: hôtel à Madeira

Ou hôtel plage aux Canaries

Liens sémantiques – cas de Wordnet

Généralisation / Spécialisation

Partie-de

Institut de Recherche en Informatique de Toulouse

Entendre via une ontologie

- Désambiguïsation
  - Comment exploiter les éventuels concepts présents dans les requêtes des utilisateurs ?
  - Comment désambiguïser les termes des requêtes ?
- Expansion
  - Quel est l'apport de chaque relation sémantique ?
  - Comment pondérer les termes rajoutés ?
  - À quel point étendre la requête ?

Institut de Recherche en Informatique de Toulouse

Expérimentations réalisées *(Baziz, 2005)*

Recommandations pour le processus d 'expansion :

- Un **concept/relation** pour toute la requête
- Poids des **termes rajoutés** < poids des termes de **la requête initiale**
- **Limiter la taille des concepts** retenus (en nombre de mots)
- Utiliser des termes de base (**lemmes**) améliore la précision.
- Exploiter **plusieurs niveaux** de l'ontologie n'améliore pas la précision.
- La relation ***hyperonymie*** (généralisation) améliore la précision moyenne, alors que la ***synonymie*** améliore la précision pour les premiers documents restitués.

Institut de Recherche en Informatique de Toulouse

Discussion : limites et points critiques

- Limites liées aux ontologies
  - Coût de construction des RTO
  - Qualité, validation des RTO
  - Adéquation modèles/applications/utilisateurs
  - Couverture du contenu des textes
- Limites liées au processus d'annotation sémantique
  - Processus coûteux si supervisé
  - Efficace pour les cas les plus fréquents et simples
  - Annotation exaustive : requiert une analyse TAL très poussée
- Maintenance conjointe des modèles et des ressources
- Passage à l'échelle du Web, pertinence de ressources « générales »
- QUI va construire les ressources ? Annoter les pages Web ?

Institut de Recherche en Informatique de Toulouse

Recherches en cours

- Automatisation :
  - apprentissage d'ontologies (TAL et extraction d'information)
  - réutilisation et adaptation semi-automatique,
  - indexation ou annotation automatisée ...
- Fusion, comparaison, alignement d'ontologies
- Nécessaire inter-disciplinarité
  - Question du sens : linguistique, terminologie, IA, IC, recherche d'information …
  - Question des usages : sciences de l'information, sociologie, ergonomie, IC …

Institut de Recherche en Informatique de Toulouse

Questions d'actualité

- est-ce que cela a du sens de figer des représentations ? sont-elles vraiment des connaissances sur un domaine? ou des ressources utiles à des applications ?
- comment rendre compte de la dynamique du langage, des connaissances, des usages, des corpus de documents dans lesquels on cherche des informations ?
- vers des processus dynamiques de reconstruction régulière des réseaux terminologiques annotant des documents

Institut de Recherche en Informatique de Toulouse

Éléments de bibliographie

- Florence Amardeih
- Reymonet DX 2009
- Cours Yannick Prié
- Charlet J. Laublet P. Reynaud C., *Web sémantique*, rapport final de l'action spécifique 32 du CNRS/STIC. Déc. 2003. http://www.ensib.fr/rtp-doc/
- D. Bourigault, M.-C. L'Homme & C. Jacquemin (eds), *Recent Advances in Computational Terminology*, John Benjamins. 2000.
- Maedche A. , *Ontology learning for the Semantic Web*. Kluwer Academic Publisher. 2002.