



CORESE

moteur de recherche
sémantique pour RDF

Olivier.Corby@sophia.inria.fr

<http://www.inria.fr/acacia/corese>



2

L'équipe de recherche ACACIA

- Modèles, méthodes et outils pour la gestion des connaissances.
- Ingénierie des connaissances et Web sémantique pour construire et exploiter une mémoire d'entreprise
- Approche Web Sémantique d'entreprise
- Scénarios :
 - Mémoire de projet (Renault, CSTB)
 - Veille technologique (CSTB)
 - Gestion des compétences (Telecom Valley)

Les problèmes de recherche

- Multi expertise, points de vue multiples
- Mémoire distribuée (projet européen CoMMA)
- Acquisition des connaissances à partir de textes
- Recherche d'information guidée par des modèles de connaissance
- Conception et évaluation suivant la méthode des scénarios

Web Sémantique & CORESE

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

– Tim Berners-Lee, James Hendler, Ora Lassila,
The Semantic Web, Scientific American, May 2001

- Web Sémantique d'entreprise
- CORESE: moteur de recherche sémantique
 - Recherche information & Knowledge Management
 - Standards W3C (langages RDF, RDFS, OWL)



5

Caractéristiques de CORESE

- Développement:
 - Depuis 1999
 - ~ 10 années/homme de R&D
 - Une Opération Développement Logiciel INRIA

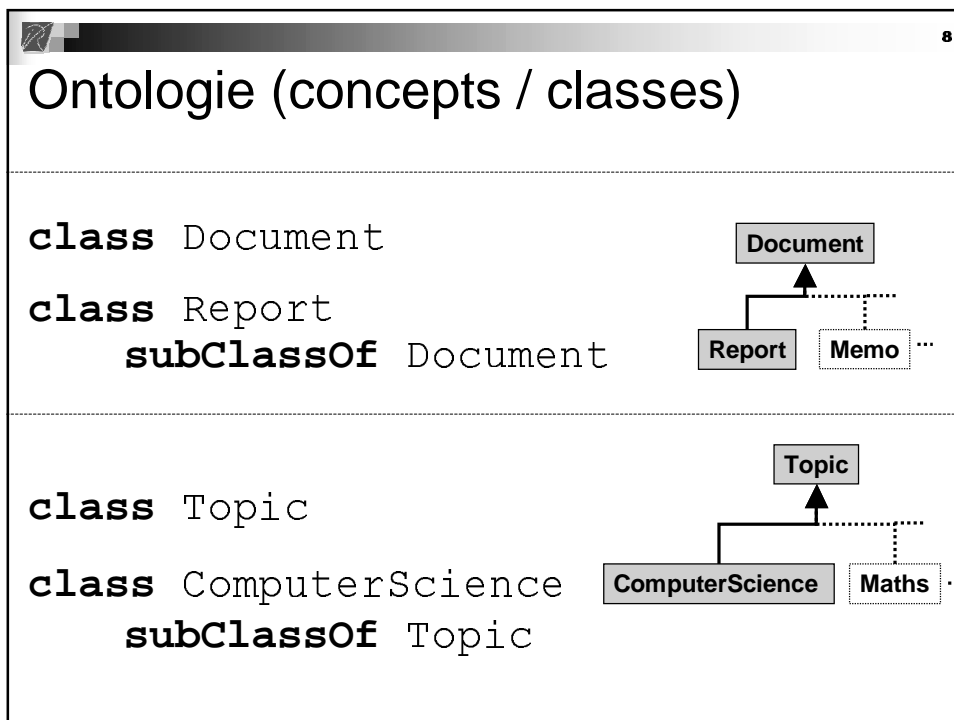
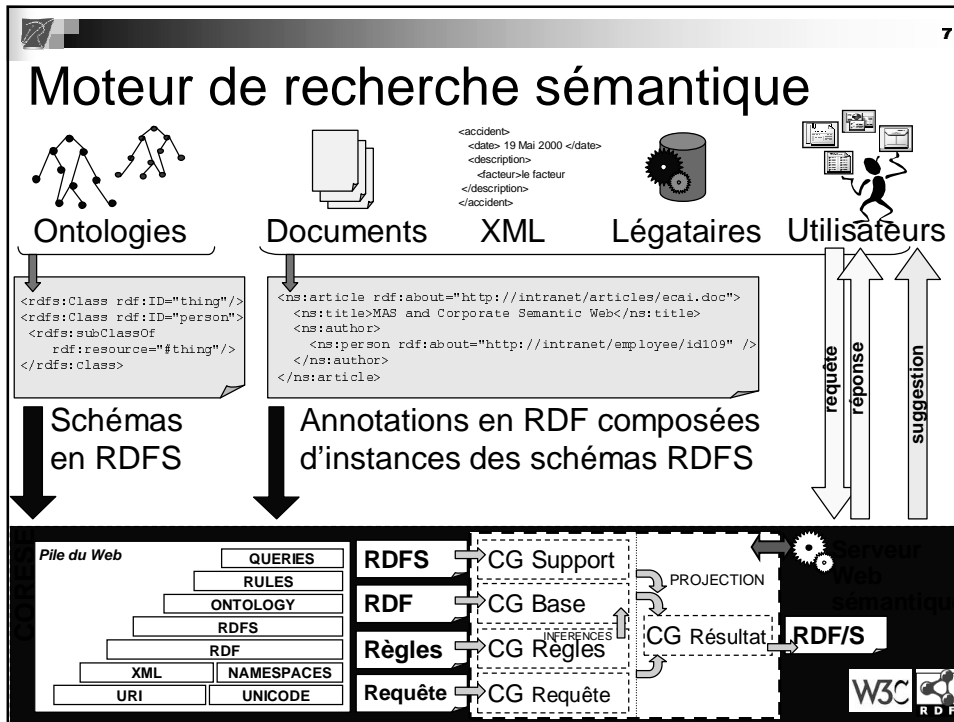
- Deux facettes au package CORESE:
 - Modules pour un Serveur Web Sémantique
 - Programmation avec API, intégrée et intégrable



6

Fonctionnalités de CORESE

- Manipuler, valider & résoudre requêtes basées sur des ontologies
- Améliorer la recherche d'information avec sémantique & raisonnement
- Portail & Serveur sémantique
- Interfaces dynamiques et formulaires, mise en page et présentation
- Intégration XHTML + XML + XSLT + RDF



9

Ontologie (relations / propriétés)

property author

domain Document

range Person

```

graph LR
    Document[Document] -.-> author((author))
    author --> Person[Person]
            
```

property concern

domain Document

range Topic

```

graph LR
    Document[Document] -.-> concern((concern))
    concern --> Topic[Topic]
            
```

10

Ontologie RDFS / XML

```

<rdfs:Class rdf:ID='Document' />

<rdfs:class rdf:ID='Report' >
  <rdfs:subClassOf rdf:resource='#Document' />
</rdfs:Class>

<rdf:Property rdf:ID='author'>
  <rdfs:domain rdf:resource='#Document' />
  <rdfs:range rdf:resource='#Person' />
</rdf:Property>
            
```



Annotation

Le rapport RR-1834 a été écrit par le chercheur Olivier Corby et porte sur le langage de programmation JAVA

Report <http://www.inria.fr/RR-1834.html>

author <http://www.inria.fr/o.corby>

concern <http://www.inria.fr/acacia#Java>

Researcher <http://www.inria.fr/o.corby>

name "Olivier Corby"



Exemple de Requête

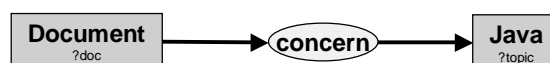
Exprimée dans le vocabulaire de l'ontologie, exemple :

Trouver des documents sur Java

```
?doc rdf:type c:Document
```

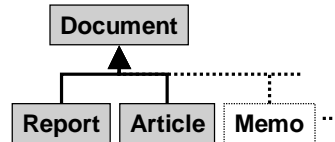
```
?doc c:concern ?topic
```

```
?topic rdf:type c:Java
```



Requêtes & exploitation ontologie

- Les rapports et les articles sont des documents, ...



- Les documents ont des auteurs, qui sont des personnes, ...



- Les personnes ont des centres d'intérêt ...



Langage de requête

select data **where** exp

exp :

<i>resource</i>	<i>property</i>	<i>value</i>
?x	rdf:type	c:Person

<i>resource</i>	<i>operator</i>	<i>value</i>
?name	=	"Olivier"

Exemple de select

Trouver les documents sur Java et retourner le titre et les auteurs :

select ?doc c:title ?person **where**

```
?doc      rdf:type      c:Document
?doc      c:concern     ?topic
?topic    rdf:type      c:Java
?doc      c:title       ?title
?title    ~             "web"
?doc      c:author      ?person
```

Langage de requête

- Opérateurs: = <= ~ != ...
- XML Schema Datatypes : number, boolean, string, date, etc.
- Langage utilisé: `xml:lang="en-us"`
- Combinaison expressions booléennes
and / or
- Non existence d'arc, arc optionnel, chemin
- Interroger RDF Schema

Interroger l'ontologie

Trouver les sous-classes de Personne avec leurs labels / termes en Anglais

```
?x rdfs:subClassOf c:Person
?x rdfs:label ?l@en
?x rdfs:label "homme"@fr
```

Trouver les traductions anglaises des labels de la classe homme

Recherche approchée

- Rechercher la meilleure approximation en fonction de l'ontologie.

- Exemple:

- Requête

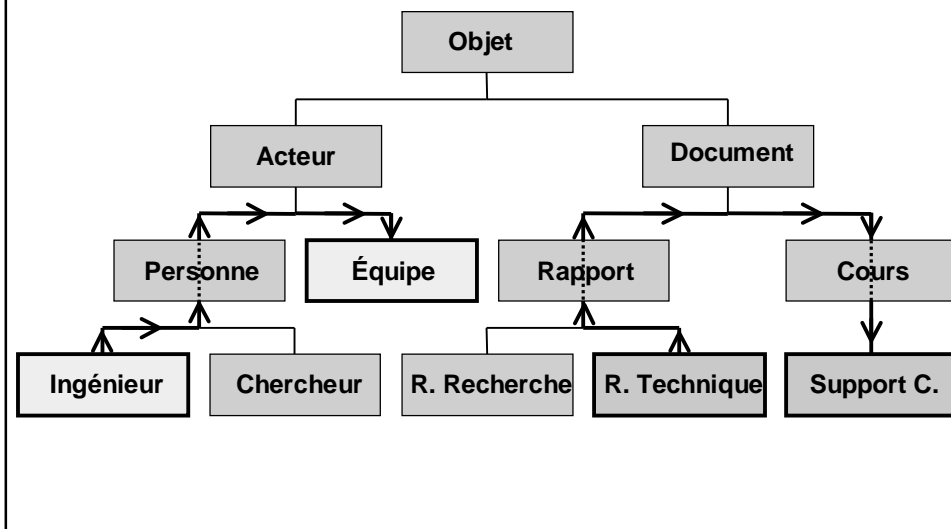
Rapport Technique sur Java écrit par un ingénieur ?

- Réponse approchée :

Rapport Technique → Support de cours

Ingénieur → Equipe

Distance dans l'ontologie



Calcul de distances

- Calcul de la distance ontologique
- Distance = longueur des chemins entre concepts approximés
- Minimiser la distance, classer résultats par distance croissante et seuillage

- Syntaxe:

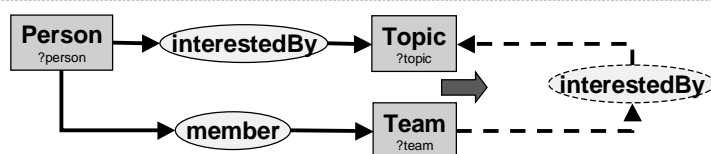
```
select more where exp
```

Inférences & Règles

Exploiter des inférences (règles) pour la recherche d'information

Si un membre d'une équipe a un centre d'intérêt alors l'équipe a aussi ce centre d'intérêt

```
?person interestedBy ?topic
?person member ?team
→
?team interestedBy ?topic
```

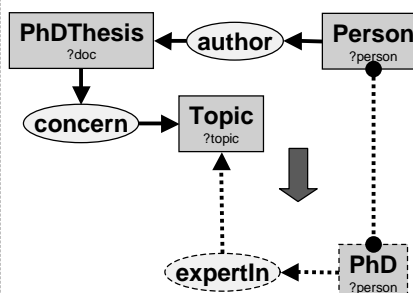


Inférences & Règles (II)

Classer une ressource

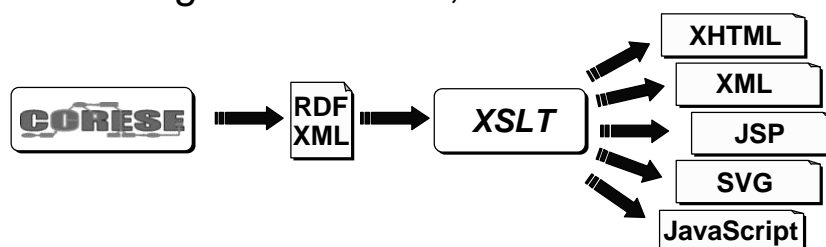
Si une personne a écrit une thèse sur un sujet alors c'est un docteur et un expert du sujet.

```
?person author ?doc
?doc rdf:type PhDThesis
?doc concern ?topic
→
?person expertIn ?topic
?person rdf:type PhD
```



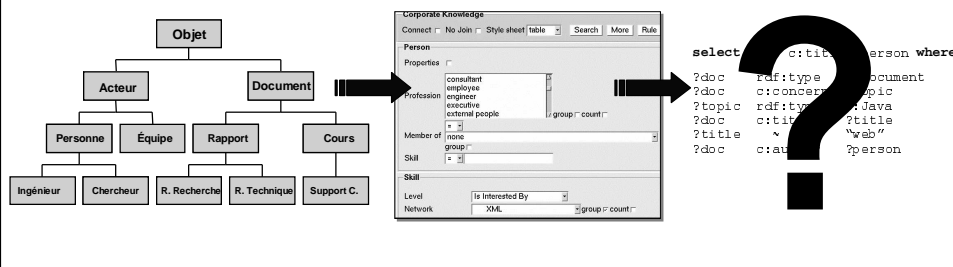
Résultats

- Réponse traduite en RDF/XML
- Traitée par feuille de style XSLT
- Adaptable à l'utilisateur, au contexte
- Peut engendrer HTML, SVG etc.



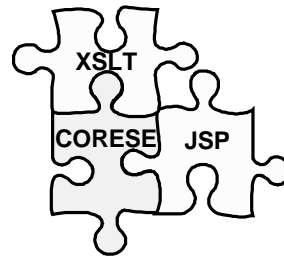
IHM

- Formulaire de requête
- Engendré par connexion avec ontologie et graphe RDF
- Requête prédéfinie éditée par l'utilisateur

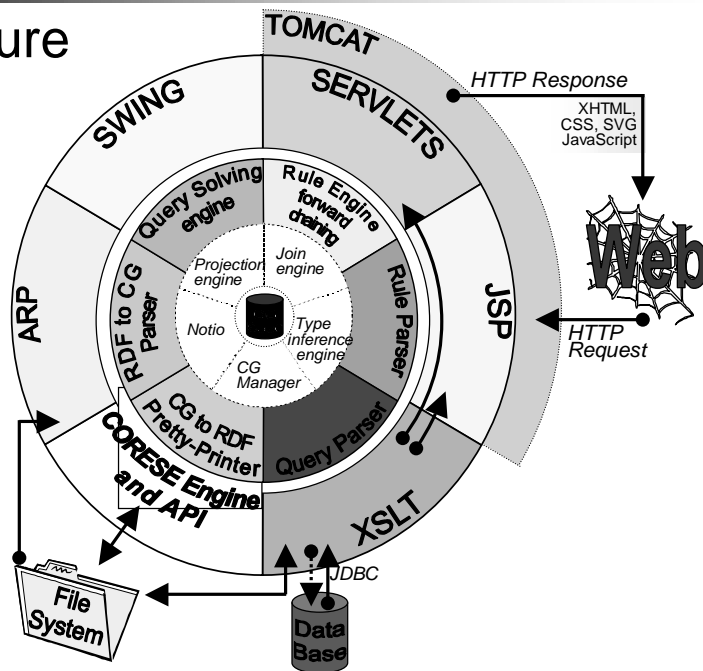


Intégration XHTML+XML+XSLT+RDF

- Dans une feuille de style XSLT :
 - Appel au moteur de recherche,
 - Connexion à une BD : engendrer un schéma RDF ou des annotations
- Intégration du résultat dans le flux de sortie



Architecture résumée



Applications de CORESE

- ESCRIRE : recherche d'information en biologie
- Renault: mémoire projet conception véhicules
- CSTB : mémoire de projet bâtiment, veille
- EADS : mémoire documentaire pour un labo de recherche industrielle
- CoMMA: mémoire distribuée veille et accueil
- MEAT: mémoire d'expériences en Biologie
- KmP: gestion des compétence Telecom Valley
- Ligne de Vie: collaboration dans réseau de soin

Applications CORESE (MEAT)

- Mémoire d'expériences en Biologie: assister les expériences sur les biopuces
- IPMC,
INRIA,
Bayer Crop
Science

The screenshot displays the Gate 2.2 build 1350 interface. The main window shows a text document titled "GATE document_0001E" with the following text:

Keratinocyte and hepatocyte growth factors in the lung roles in lung development, inflammation, and repair. Ware, Lorraine B., and Michael A. Matthay. Keratinocyte and hepatocyte growth factors in the lung: roles in lung development, inflammation, and repair. *Am J Physiol Lung Cell Mol Physiol* 282: R92476949, 2002. doi: 10.1152/ajplung.00459.2001. A growing body of evidence indicates that the epithelial-specific growth factors keratinocyte growth factor (KGF), fibroblast growth factor (FGF-10), and hepatocyte growth factor (HGF) play important roles in lung development, lung inflammation, and repair. The therapeutic potential of these growth factors in lung disease has yet to be fully explored. KGF has been best studied and has impressive protective effects against a wide variety of injurious stimuli when given as a pretreatment in animal models. Whether this protective effect could translate to a beneficial effect in humans with acute lung injury needs to be investigated. FGF-10 and HGF may also have therapeutic potential, but more extensive studies in animal models are needed. Because HGF lacks true animal models, it may have less potential than KGF and

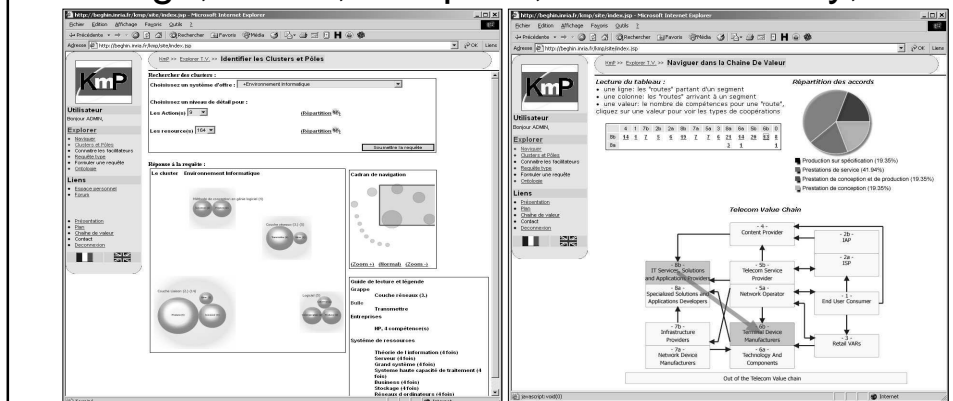
Below the text, a "Features" list is visible:

- eratinocytes, cui=C0022567, term=Keratinocyte)
- Peptide, or Protein, cn=Hepatocyte Growth Factor, cui=C006
- Organ, or Organ Component, cn=Lung, cui=C0024109, term=
- avior, cn=Role, cui=C0035820, term=role)
- Issue Function, cn=lung development, cui=C1160389, term=

The interface also includes a sidebar with "Gate" menu items (Applications, Language Resources, GATE document_0001E, Processing Resources, Data stores) and a "Default annotations" panel on the right with checkboxes for Concept, Playrole, Sentence, SpaceToken, Split, Token, and Paragraph.

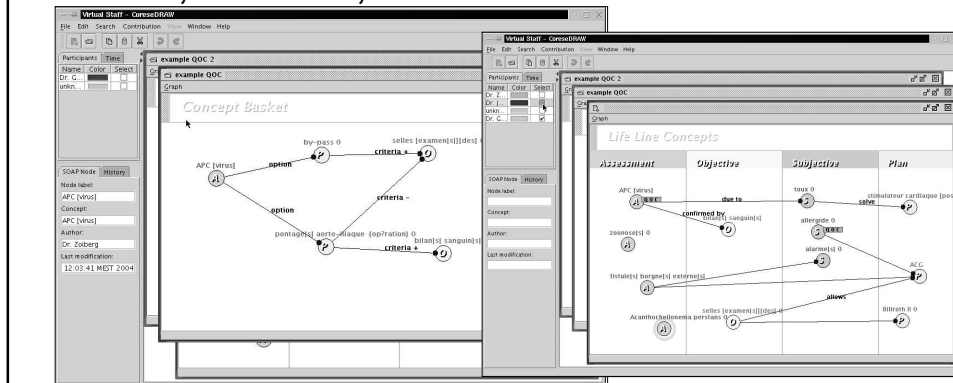
Applications CORESE (KmP)

- Knowledge Management Platform: Serveur Web Sémantique comme portail de gestion des compétences pour partenariats sur Sophia
- Rodige, INRIA, Latapses, Telecom Valley, GET



Applications CORESE (Ligne de Vie)

- Assister collaboration médicale dans un réseau de soin, pour la prévention, le diagnostic, la thérapie, le suivi des patients et l'éducation.
- INRIA, Nautilus, SPIM



Étapes méthodologiques

- Ingrédients: CORESE, intranet, RDF/S, XML, utilisateurs
- Recette méthodologique
 - Analyse par scénarios
 - Réutiliser ou concevoir les ontologies
 - Annoter ressources & intégration sys. légataires
 - Concevoir IHM & feuilles de styles dédiées
 - Bien mélanger dans CORESE
 - Laisser inférer & évaluer
- Servir ... sur le web

Conclusion

- Site de Corese :
<http://www.inria.fr/acacia/corese>
- eCore : Projet de startup autour de la gestion des compétences avec Corese
- ACACIA ouvert aux coopérations ☺