



Annotation sémantique de documents

Application au e-recrutement

Présenté par : Leila Yahiaoui

Dirigé par : Pr. Zizette Boufaïda

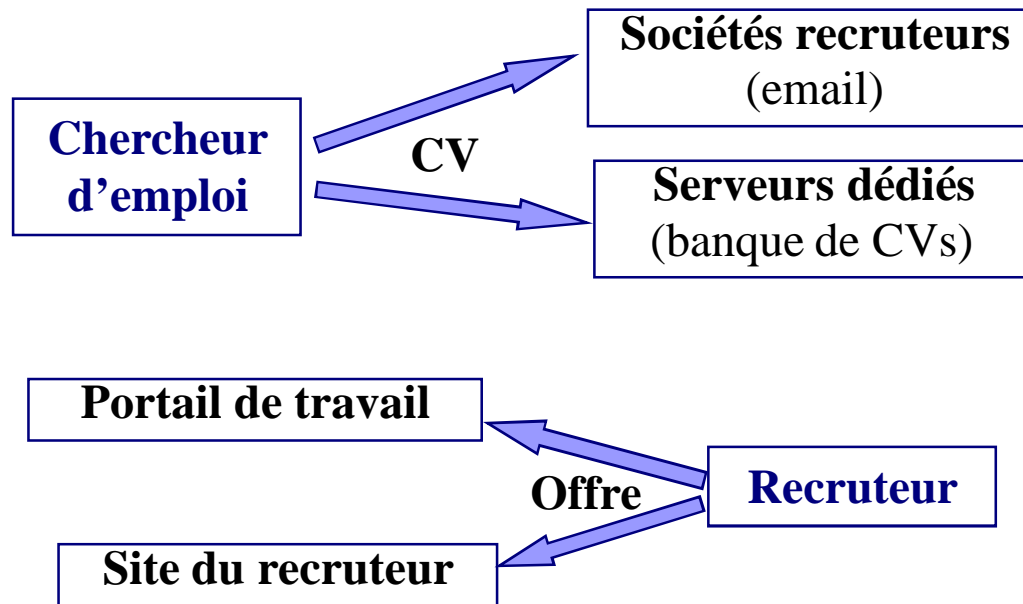


Le plan

- Introduction
- Le e-recrutement sur le Web
- Généralités sur le Web sémantique
- Gestion des CVs /offres d'emploi dans le cadre du WS
- Objectif du travail
- Architecture du système proposé
- Génération des documents à annoter
- Modélisation du contenu sémantique à base d'ontologie
- Appariement sémantique entre CVs et offres d'emploi
- Etude de cas
- Conclusion et perspectives

Introduction

- ✓ Méthodes de recrutement classiques obsolètes ⇒ **Internet** : une nouvelle gestion (documents électroniques)

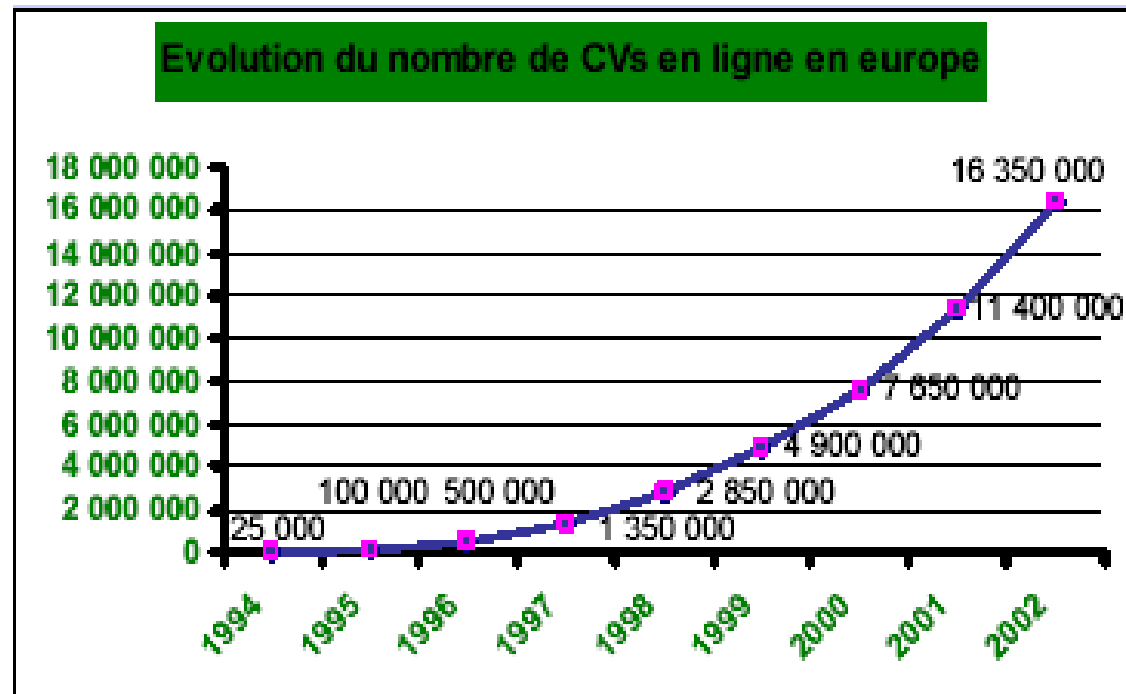


- ✓ **Le Web** : un vecteur de diffusion d'informations important

Le e-recrutement sur le Web

- ✓ **Le e-recrutement** : application typique d'une approche de gestion de connaissances à travers le Web.

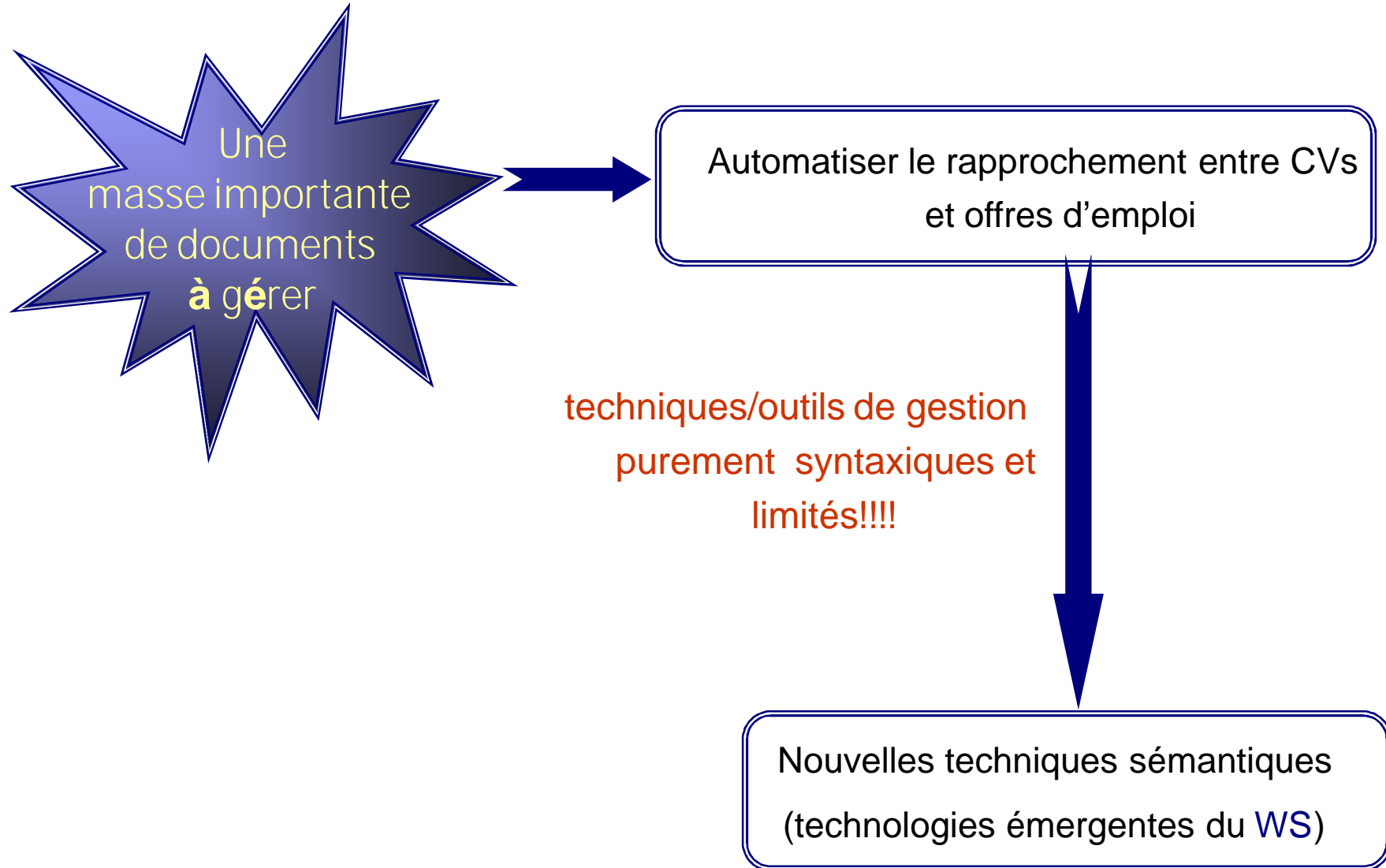
(Computer Economic –2001)



E 13.4 billion\$ en 2005 avec un taux de croissance de 53%.

E Grands sites de recrutement : plus de 50,000 CVs dans leurs BDDs.

Le e-recrutement sur le Web



Généralités sur le Web sémantique



✓ **Le Web sémantique** : Tim Berners Lee (W3C)

Utilisation de **connaissances formalisées** en plus du contenu informel du Web actuel
⇒ **contenu interprétable** par **machine**.

Ontologie = élément crucial du WS

- ∅ **Définition** : spécification **formelle** est **explicite** d'une **conceptualisation** [Gruber,95].
- ∅ **Composants** :
 - Les concepts (terme, notion, extension)
 - Les propriétés (valuées)
 - La relation de subsomption « Is-a » (généralisation/spécialisation)
 - Les autres relations (terme, signature)
- ∅ **Intérêts** : communication , interopérabilité, spécification des systèmes, indexation et recherche de documents . . .



Gestion des CVs/offres dans le cadre du WS

✓ **Approches de modélisation du contenu sémantique :**

∅ L'indexation sémantique de documents :

- Construire un index d'un site dont la structure correspond à celle d'une ontologie
- Dédiée à l'extraction de documents concernant une recherche spécifique

Inconvénients :

- E** Les documents sont juste associés aux concepts des ontologies
- E** Techniques de TALN difficiles à maîtriser.

∅ L'annotation sémantique de documents :

- Enrichir explicitement les documents, à base d'ontologies, avec des méta-données
- Dédiée aux systèmes «request/answer ».

Gestion des CVs/offres dans le cadre du WS

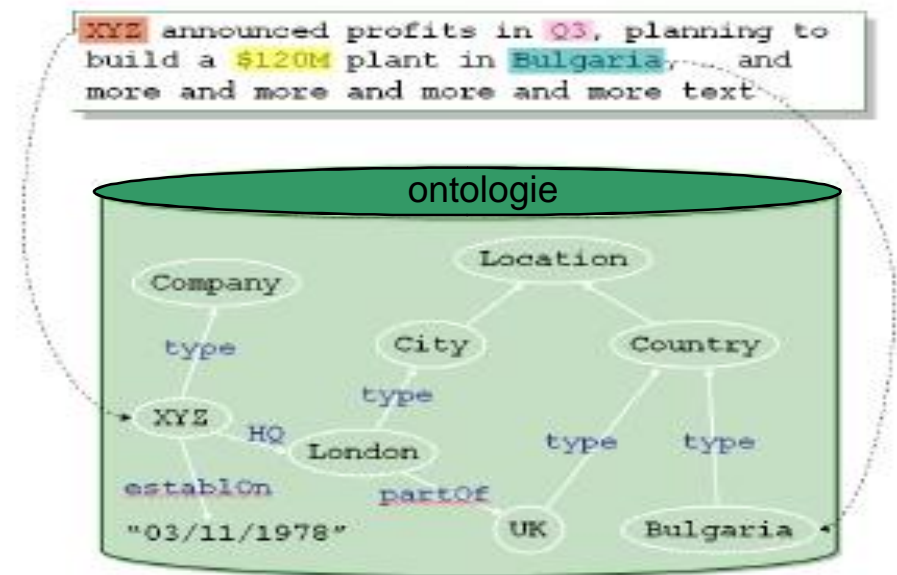
✓ Annotation classique :

Annotation = information graphique/textuelle attachée à un document [Desmontils02]

✓ Annotation sémantique : assigner aux entités d'un document des liens à leurs descriptions sémantiques [Kiryakov, 05].

∅ Processus :

- **Repérer** : placer les références aux concepts dans le document
- **Instancier** : valuer les attributs des concepts
- **Enrichir** : ajouter des informations (attributs non valués)





Gestion des CVs/offres dans le cadre du WS

✓ Travaux existants basés annotation sémantique:

§ *Le projet COMMONCV* intitulé « recherche&emploi »

§ *Un processus de recrutement basé ontologie* [Bizer05]

§ *Os-Skill ...*

∅ *Avantage* : approche facile et naturelle + disponibilité des standards/outils du WS.

∅ *Manques dans ces travaux :*

E Incomplétude ou niveau de détail donné insuffisant

E Concentration sur les compétences relatives à l'expérience professionnelle

E Appariement sémantique partielle (basé compétence seulement)

E Simplicité relative

Objectif du travail

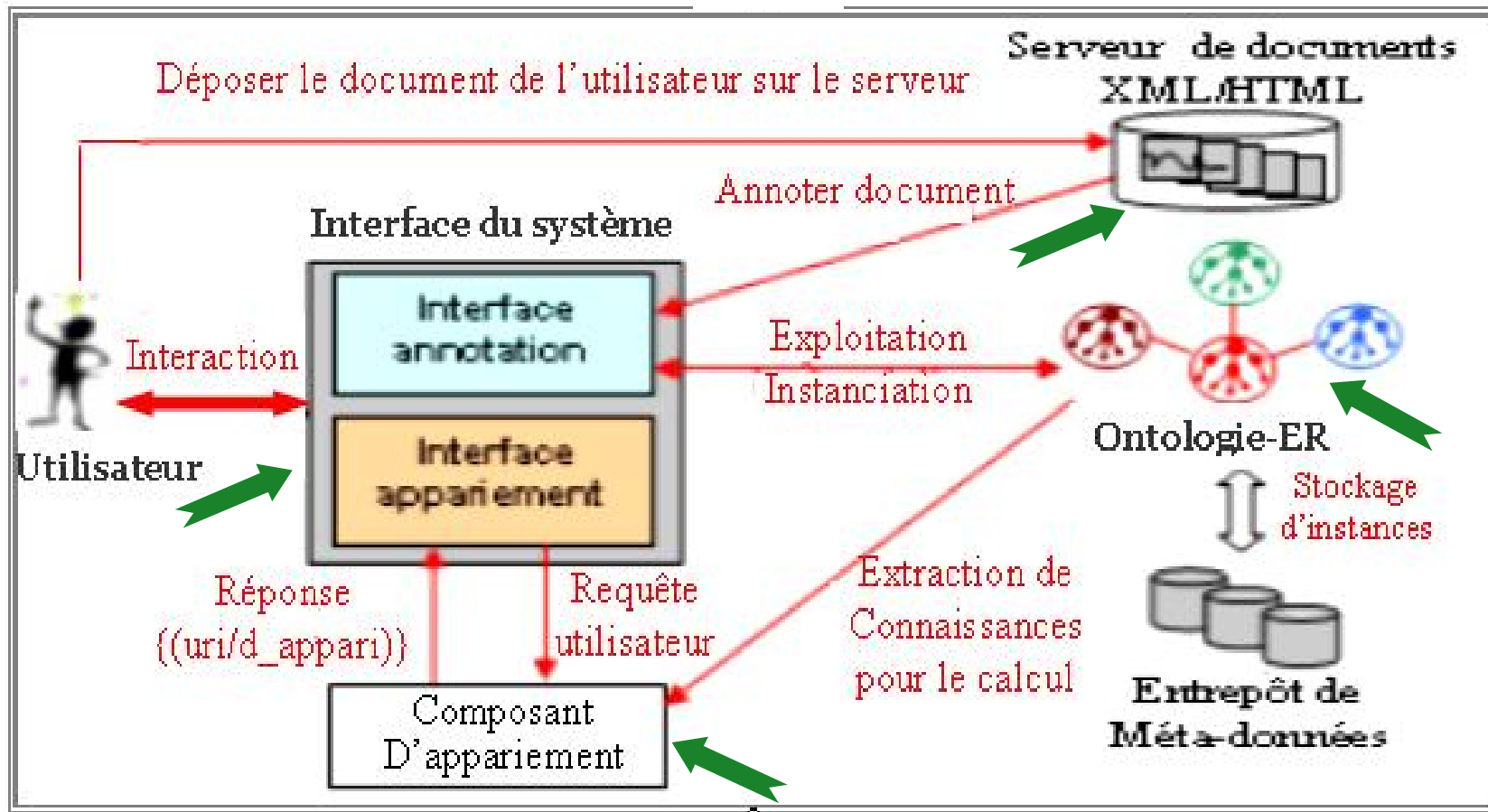


une solution simple, détaillée
et suffisamment complète,
basée annotation sémantique,
pour automatiser le
rapprochement entre les
CVs/ offres d'emploi

- ∅ Génération des documents à annoter (XML)
- ∅ Construction d'une ontologie destinée à la modélisation du contenu sémantique des documents (acquis/requis)
- ∅ Proposition d'un processus d'annotation sémantique de documents.
- ∅ Description des services d'appariement sémantique à appliquer entre les CVs et les offres d'emploi

Architecture du système proposé

Le système d'annotation et d'appariement sémantiques





Génération des documents à annoter

√ Etapes :

E Proposition d'une structure pour chacun des types de documents
(CV/offre d'emploi) → Monster, JobPilot et CVGEN + des particularités

E Modélisation de ces structures par des diagrammes de classes **UML**

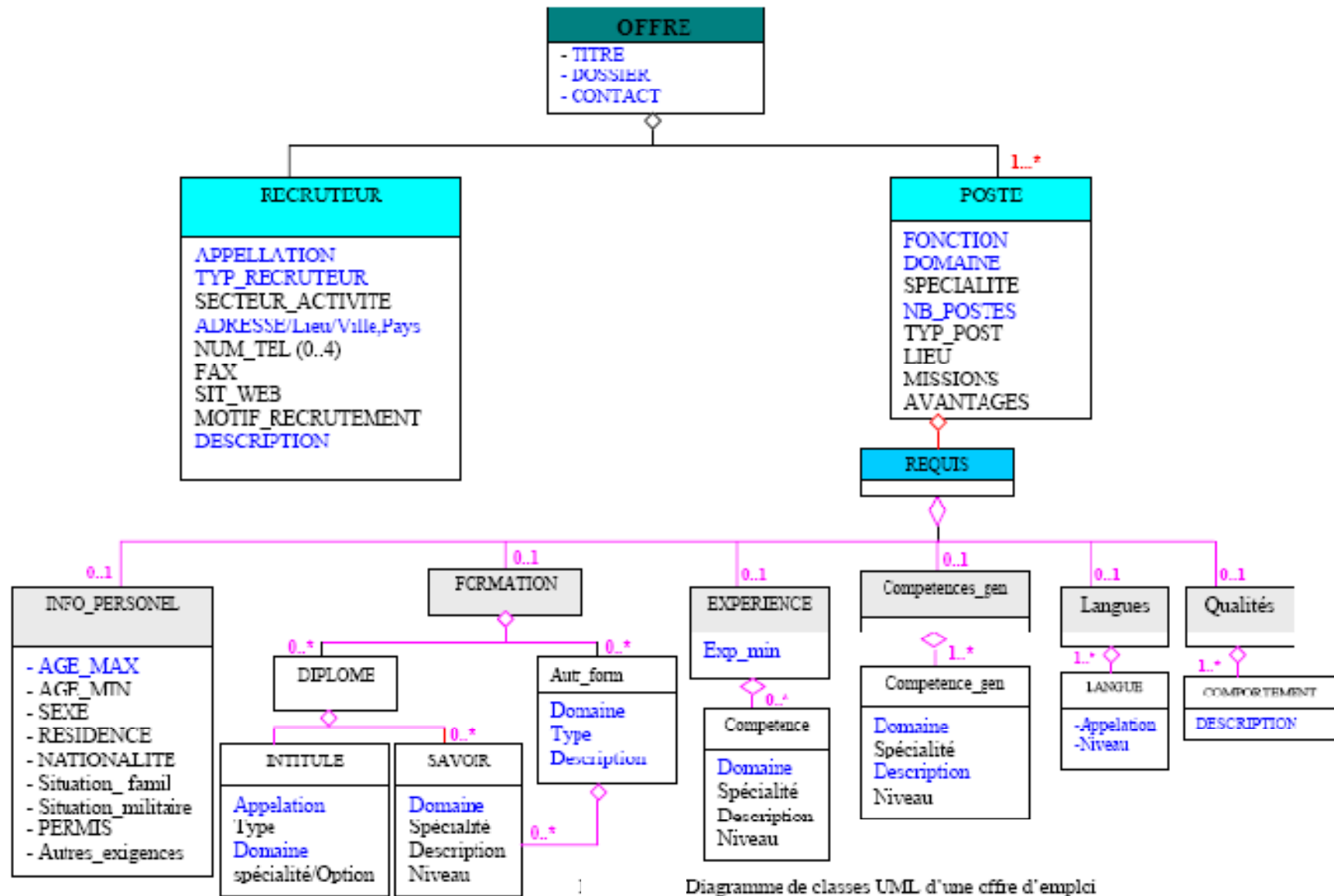
E Codification des diagrammes UML en **Schéma XML**

E Mise en page des documents XML générés (feuilles de styles **XSL**)



Documents stockés dans des bases de données XML

Génération des documents à annoter



Génération des documents à annoter

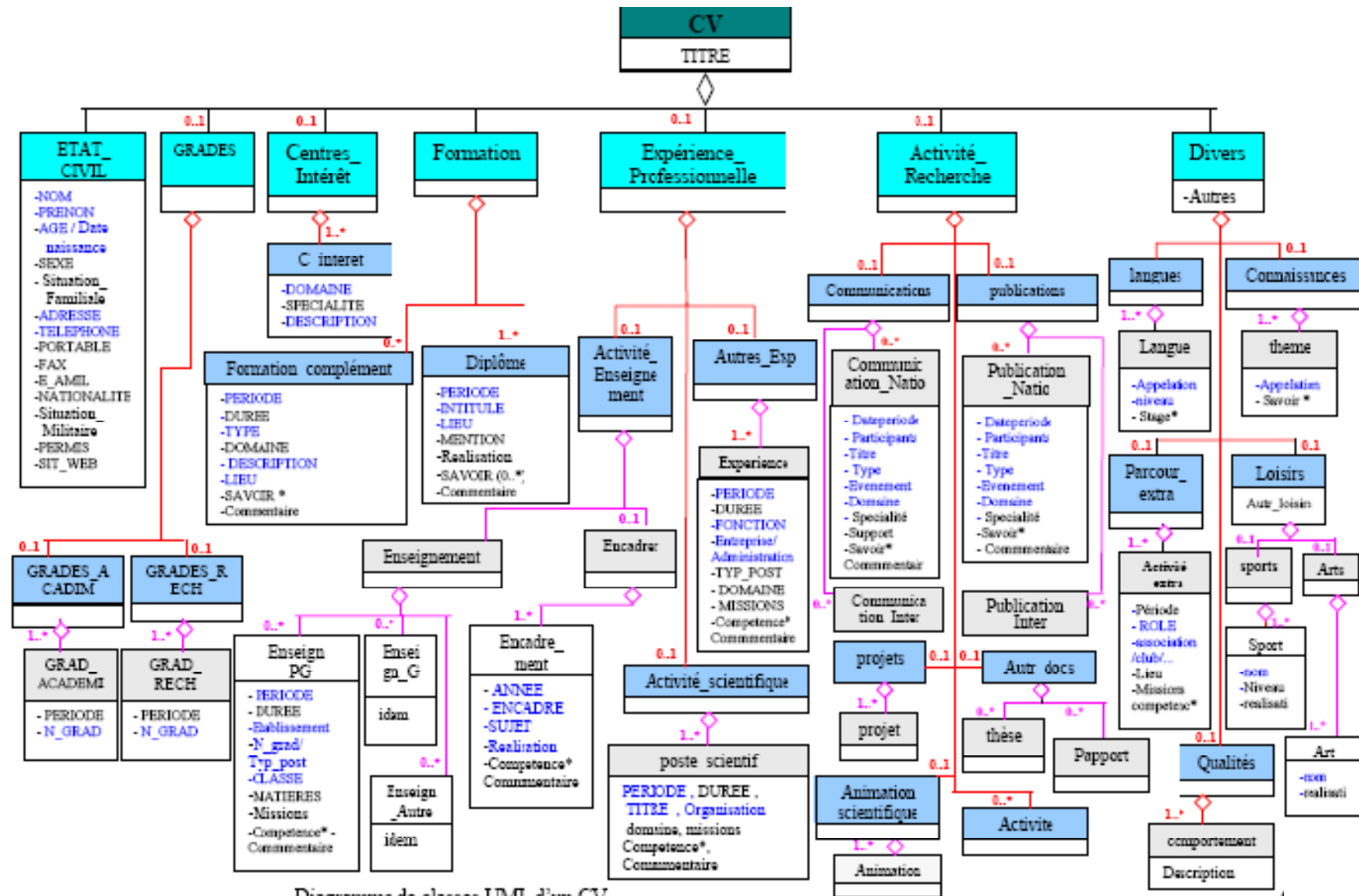
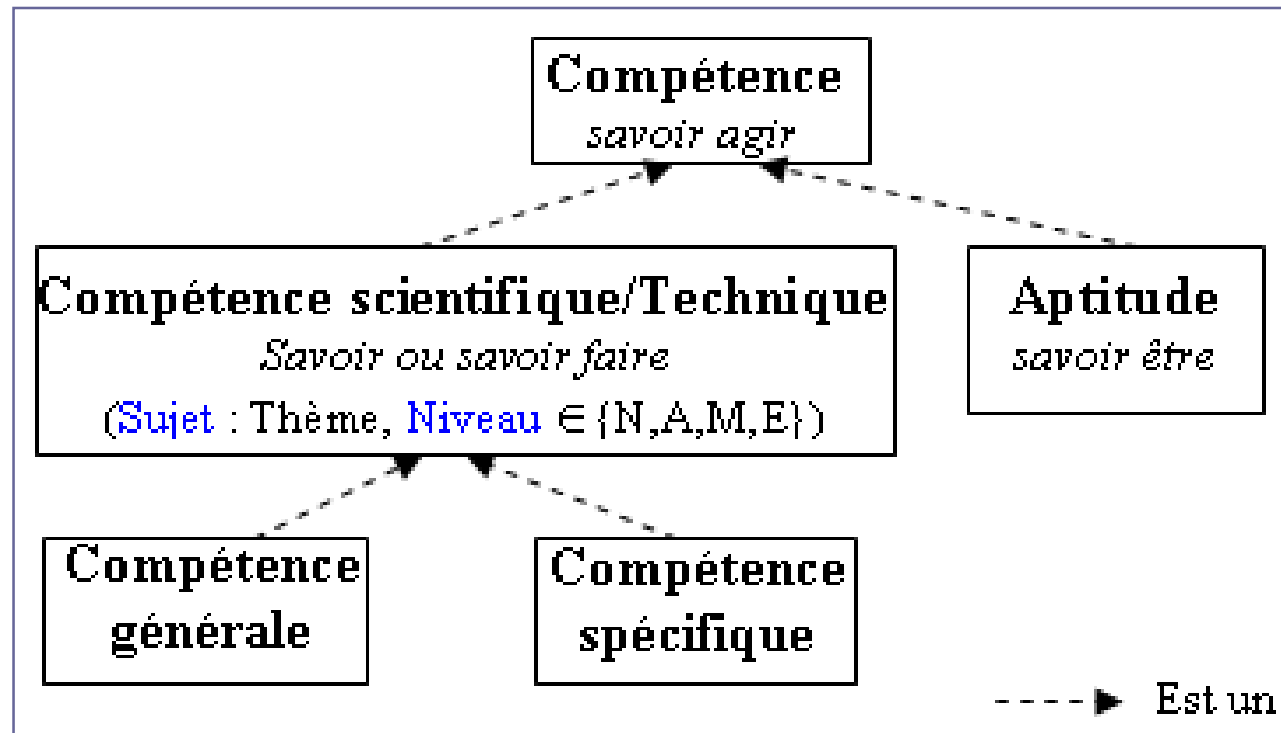


Diagramme de classes UML d'un CV

Modélisation du contenu sémantique

✓ Le modèle de la compétence :



∅ Domaine : Informatique et télécommunications

E Le sujet : « Thème » ou « Aspect logiciel ».

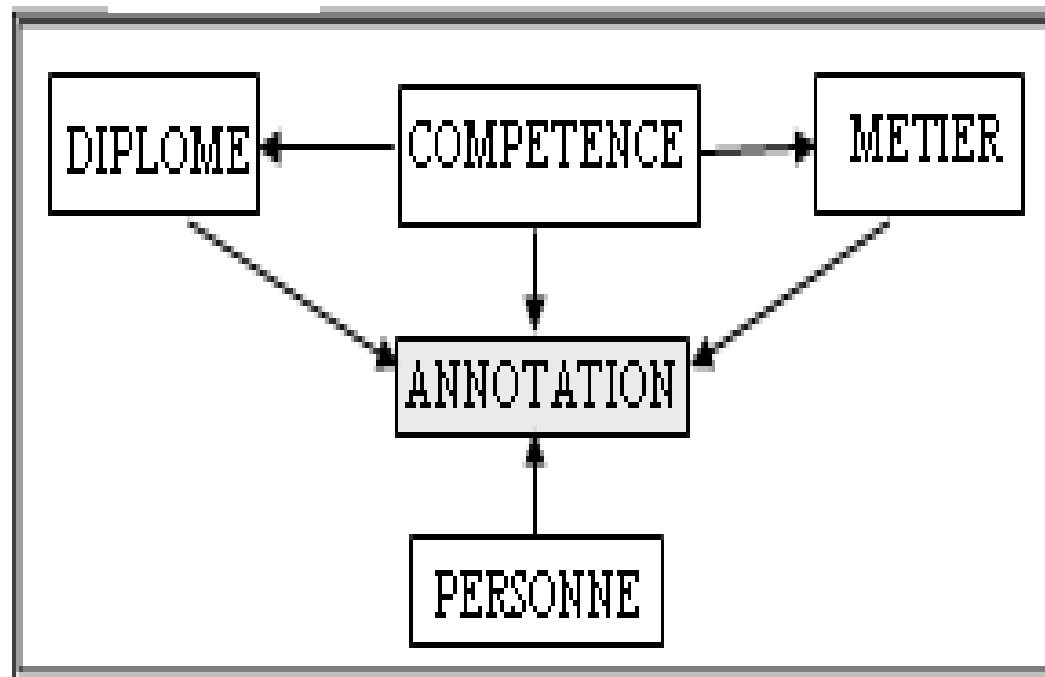
E Le niveau: Notion (20%), Application (50%), Maîtrise (70%) ou Expert (90%).

Modélisation du contenu sémantique

✓ Architecture globale de l'ontologie-ER :

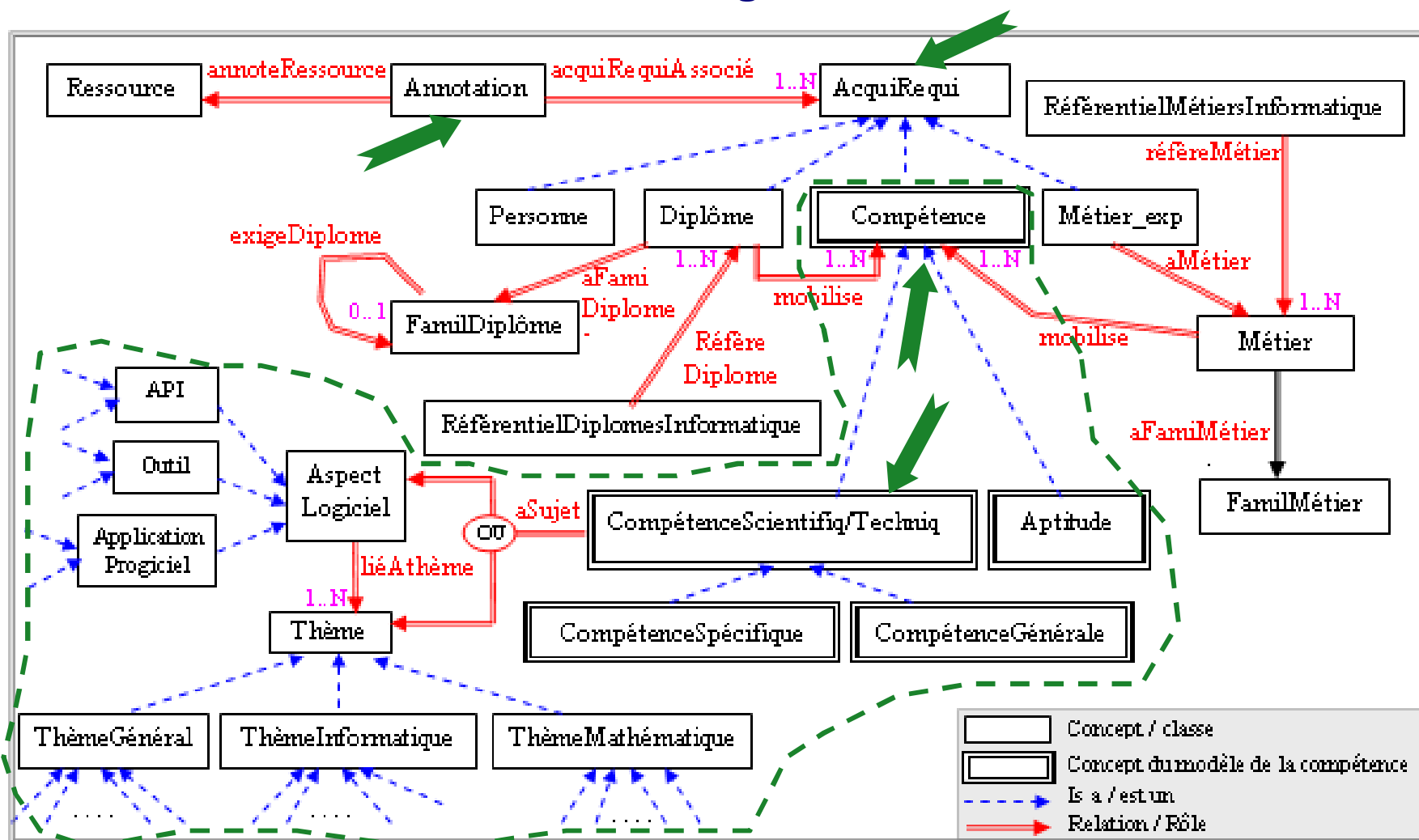
E Domaine : «Informatique et Télécommunications»

E Constituée de 5 sous-ontologies



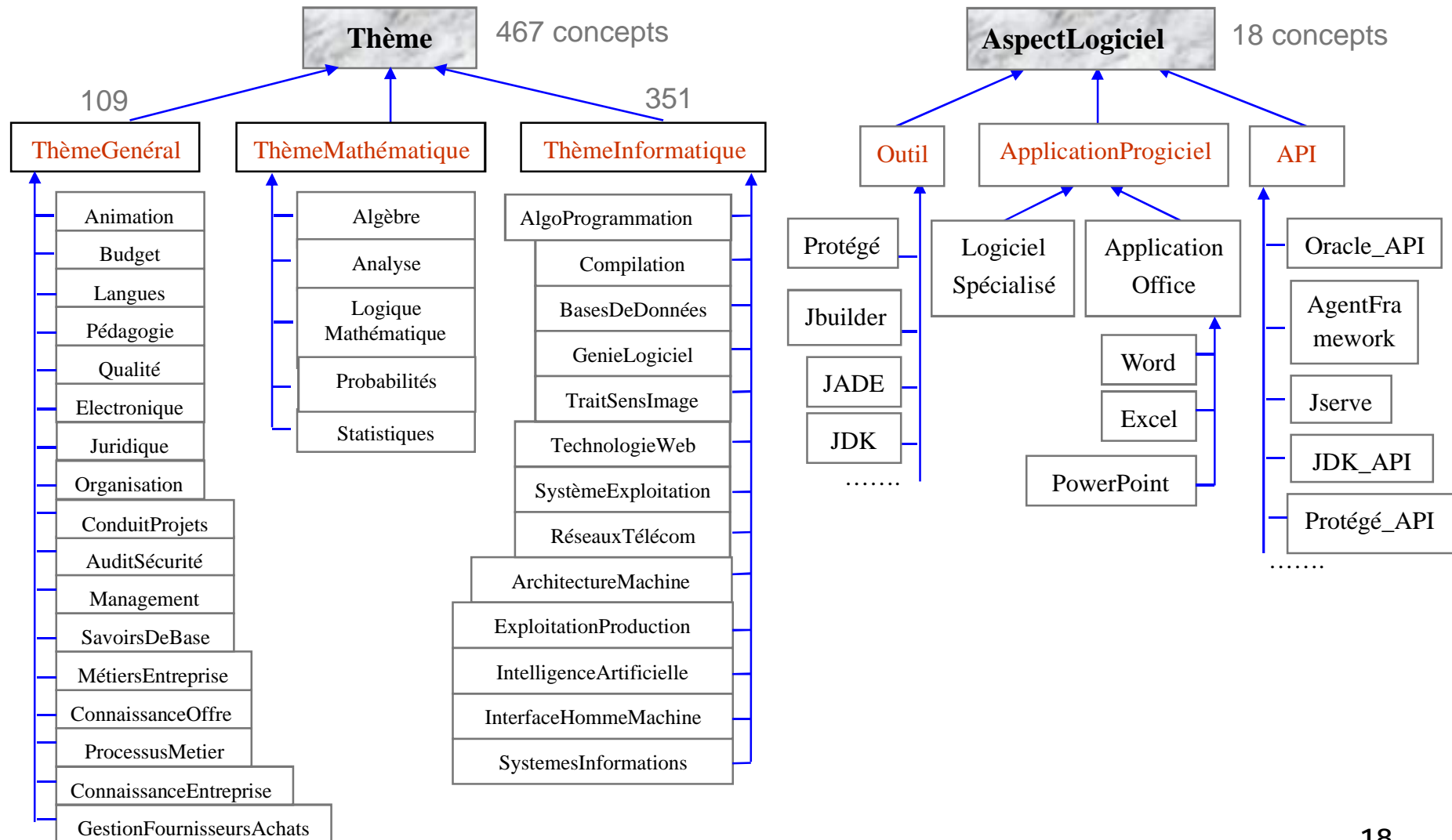
Modélisation du contenu sémantique

Architecture détaillée de l'ontologie-ER : (510 concepts/17 relations/20 attributs)



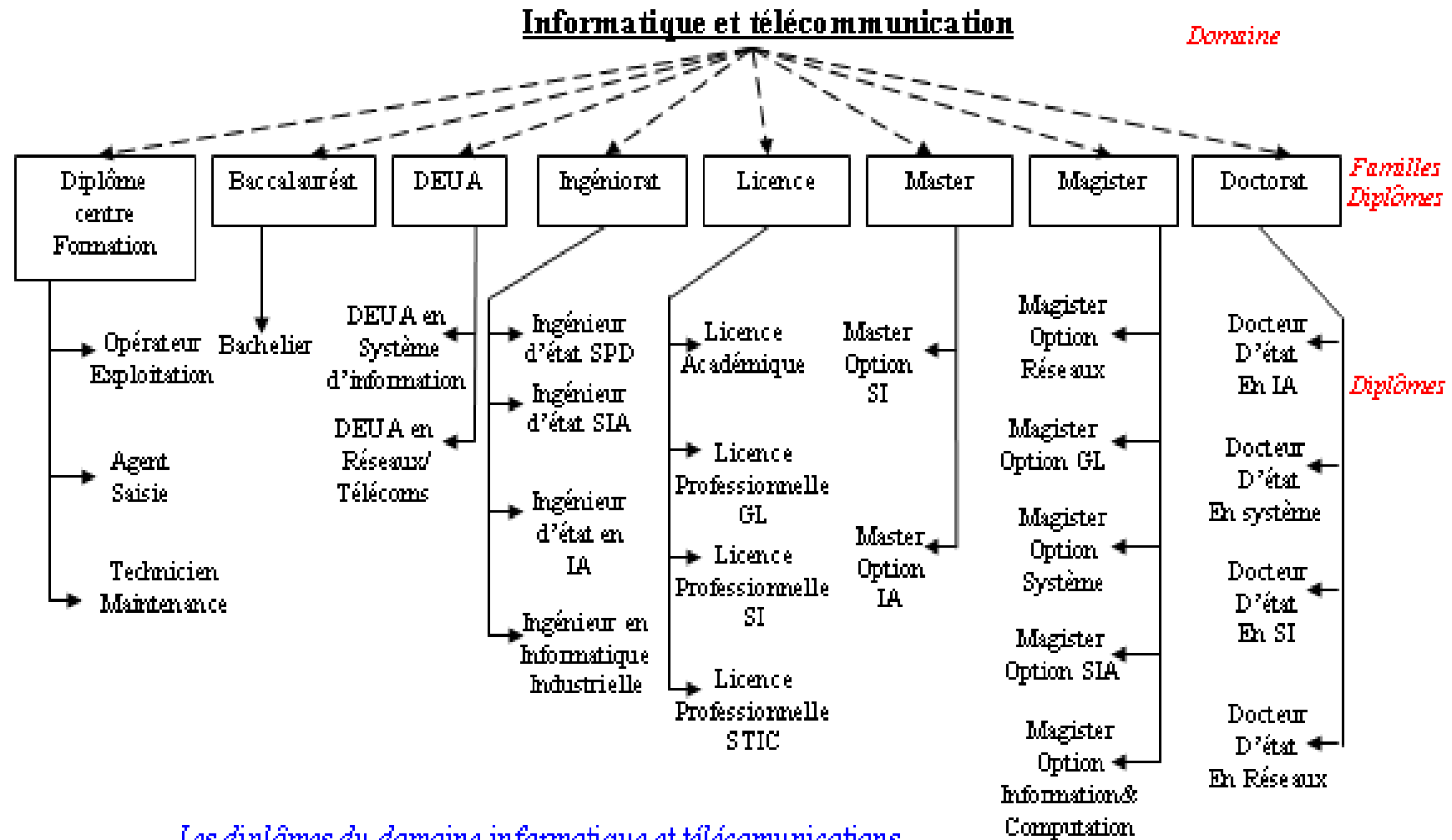
Modélisation du contenu sémantique

Architecture détaillée de l'ontologie-GRH:



Modélisation du contenu sémantique

Architecture détaillée de l'ontologie-ER

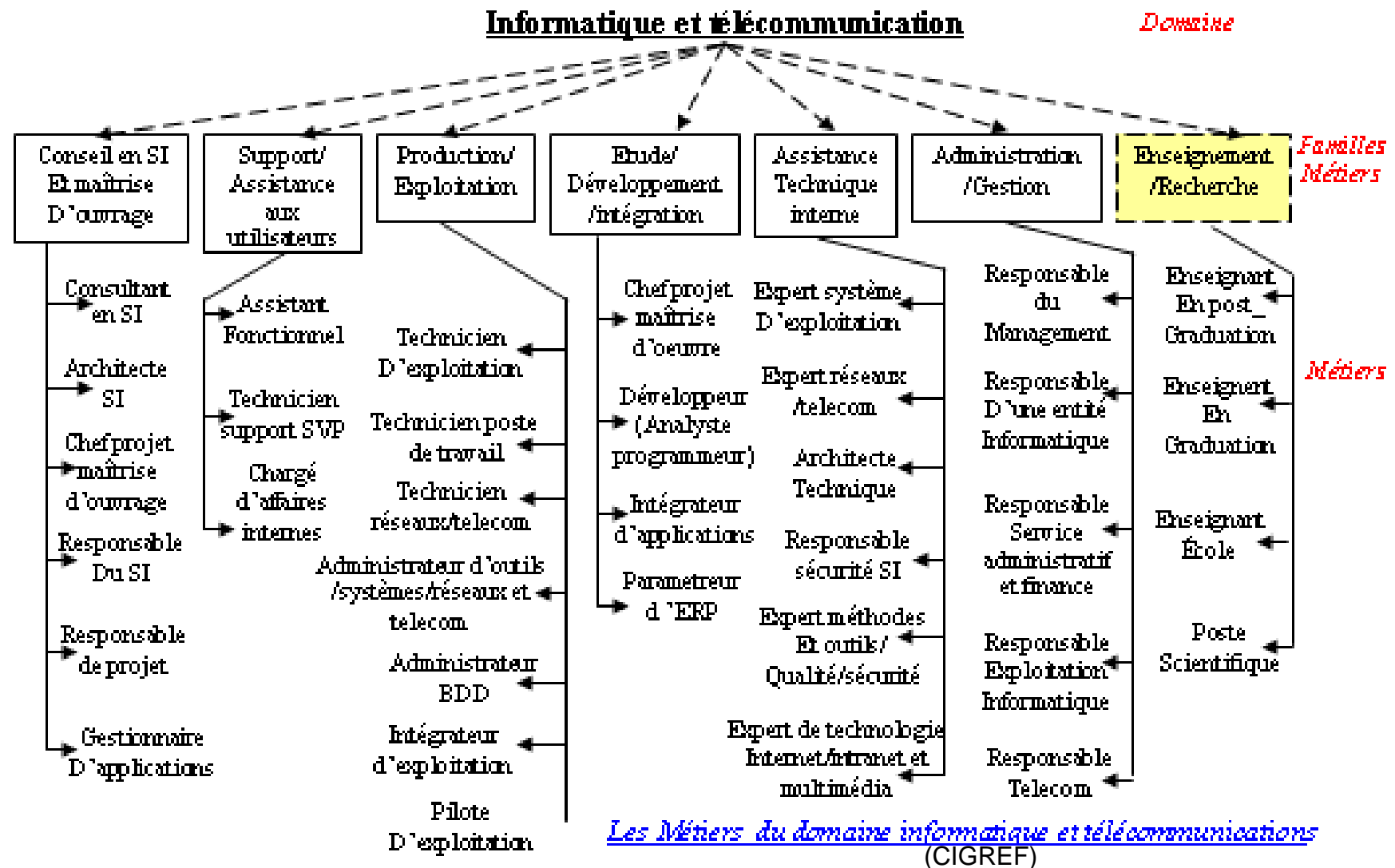


Les diplômes du domaine informatique et télécommunications

(Université de Constantine et autres - Algérie)

Modélisation du contenu sémantique

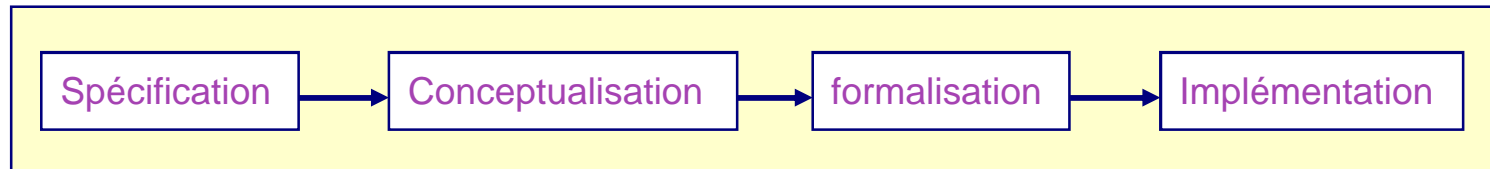
Architecture détaillée de l'ontologie-ER



Modélisation du contenu sémantique

✓ Construction de l'ontologie-ER

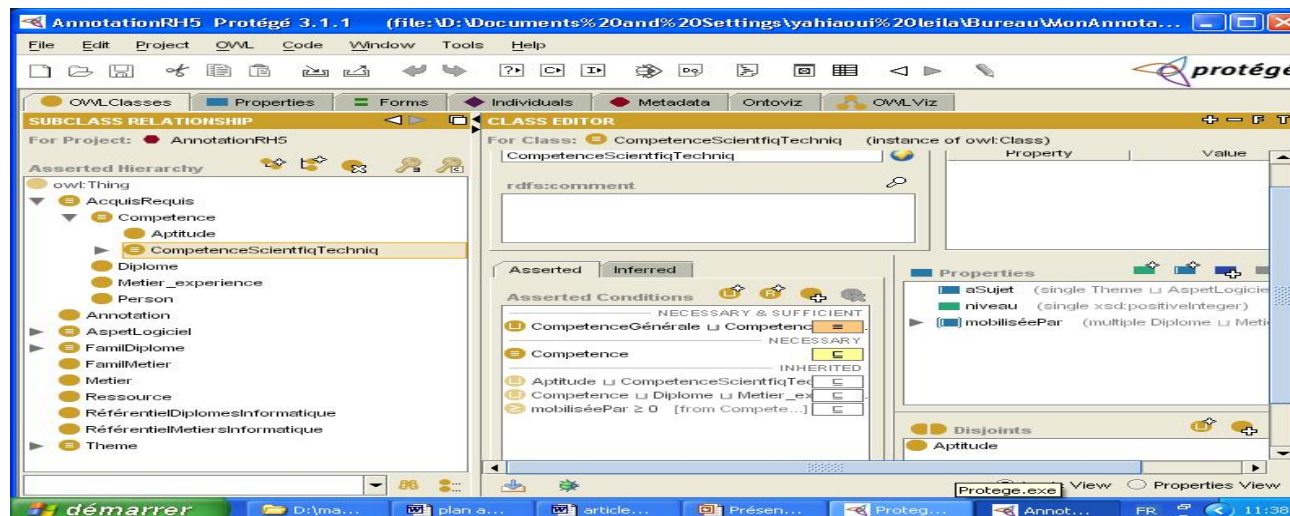
∅ Méthodologie de construction : inspirée de «METHONTOLOGY»



∅ Formalisme de représentation : la logique de description (shoin-D)

∅ Langage de codification : OWL-DL

∅ Outil : Protégé-OWL (version 3.11)





Processus d'annotation sémantique

- ✓ **Rôle de l'expert** : Instancier à l'avance les sous-ontologies «**METIER**», «**DIPLOME**» et «**COMPETENCE**»
 - E** Une instance de «**Thème**»/«**AspectLogiciel**» \Rightarrow 4 instances de «**Compétence**»
 - E** Lier les instances de «**Métier**» et «**Diplôme**» aux instances de «**Compétence**» qu'ils mobilisent

- ✓ **Rôle de l'utilisateur** :
 - ∅ Instancier la classe «**Ressource**» \Rightarrow décrire le document à annoter (URI et type) .
 - ∅ Instancier la classe «**Person**» (qualifications personnelles)
 - ∅ Instancier la classe «**Métier_exp**» (expériences professionnelles)
 - ∅ Instancier la classe «**AcquiRequi**» avec les acquis du candidat (requis de l'offre)
 - ∅ Instancier la classe «**Annotation**» \Rightarrow lier les acquis/requis à la ressource annotée



Appariement sémantique entre CVs et Offres

✓ **Principe d'appariement :**

E Calcul d'un degré de rapprochement (pourcentage)

E Deux techniques différentes mais complémentaires

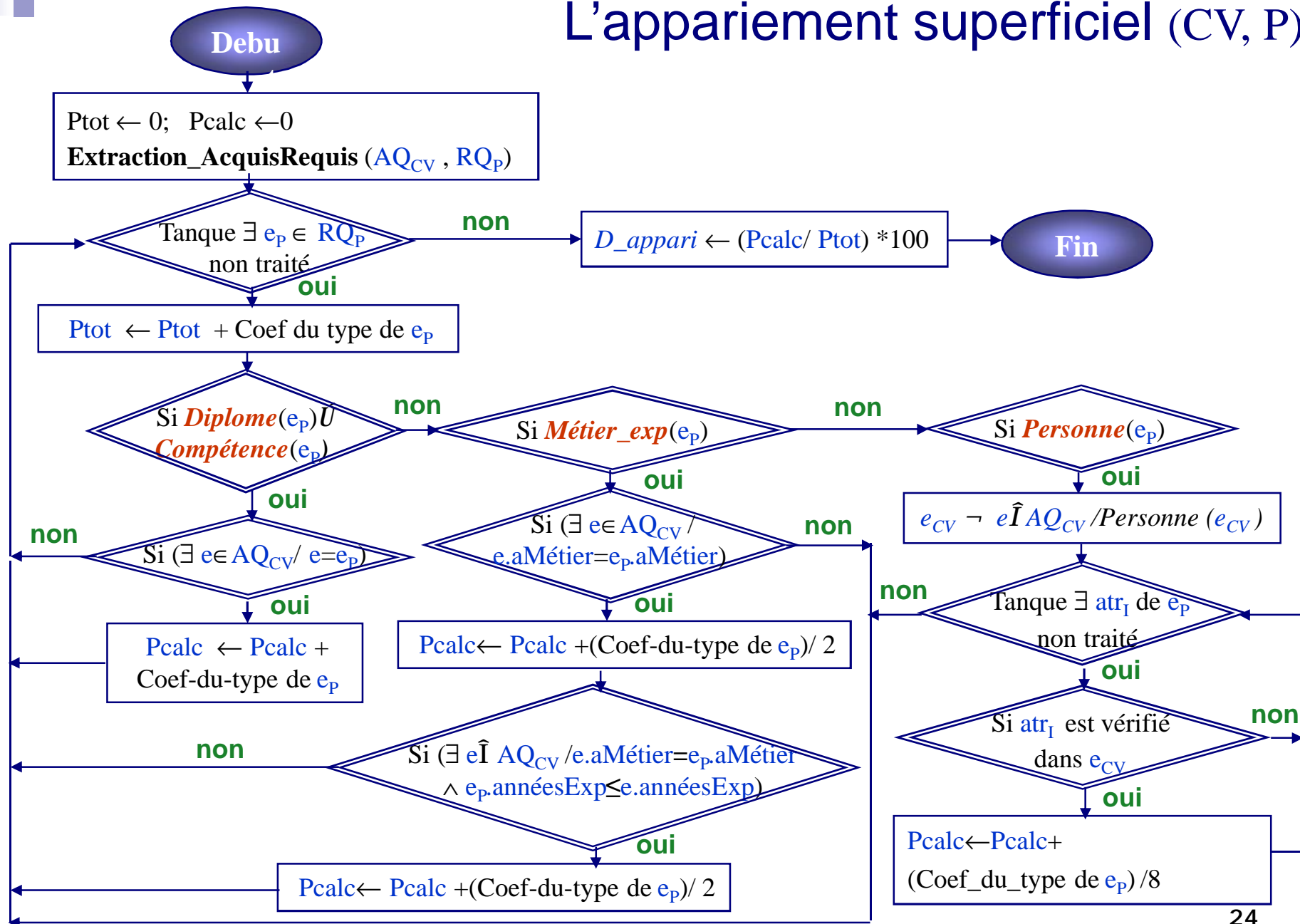
∅ *Appariement sémantique superficiel* : prend en considération les requis/acquis, avec lesquels les documents ont été annotés, à un niveau superficiel

∅ *Appariement sémantique basé compétence* : s'intéresse aux compétences sous-jacentes aux documents annotés.

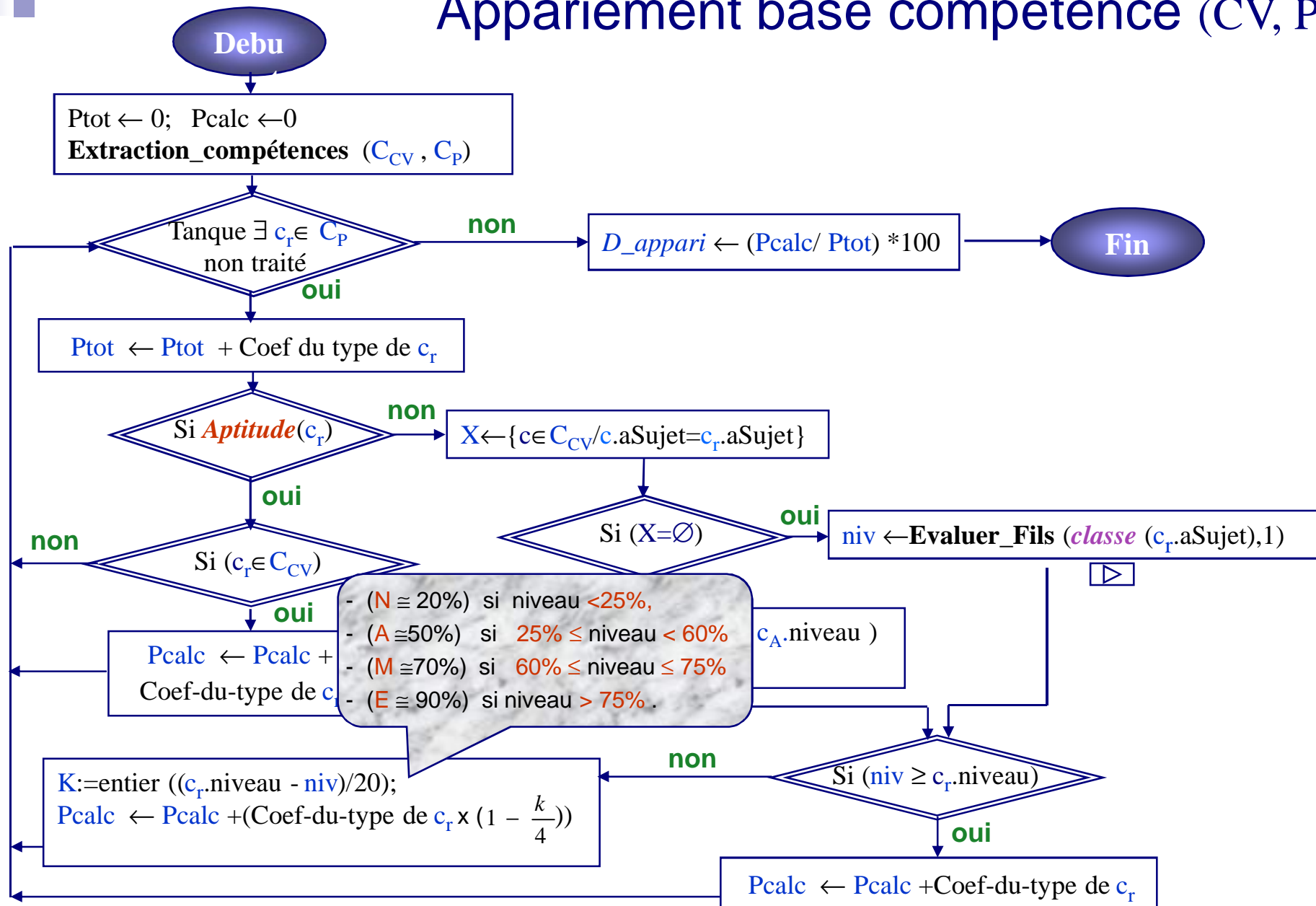
✓ **conventions** : algorithmes entre un CV (r_{CV}) et un poste P (r_P)

- **C(i)** : i est un individu/instance du concept/classe C \Rightarrow classe (i) =C.
- **i.nomAtr** : valeur de l'attribut «nomAtr» de l'individu i / ensemble des individus avec lesquels i est relié par le rôle «nomAtr».
- **A[®] C** : le concept C subsume le concept A.

L'appariement superficiel (CV, P)



Appariement basé compétence (CV, P)



Appariement sémantique entre CVs et Offres

✓ Aspects implémentatoires : *RACER*

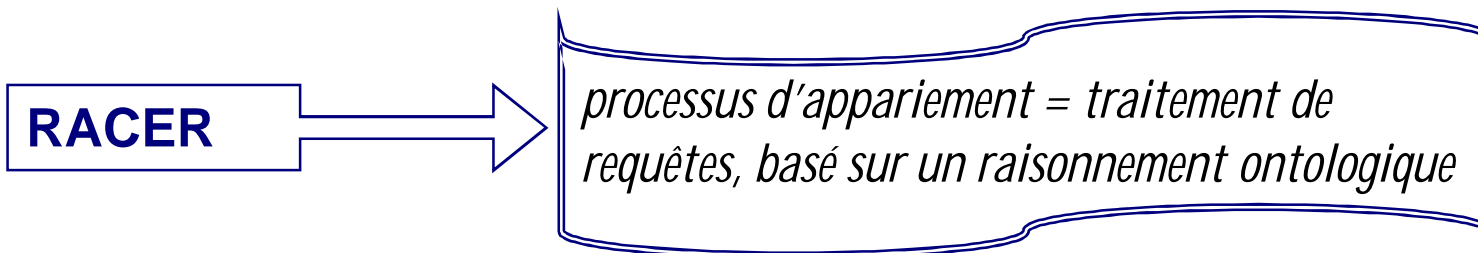
Un système de représentation de connaissances, qui implémente un calcul de tableaux hautement optimisé pour une LD très expressive.

E *Au niveau terminologique* : différentes possibilités

- Vérifier la consistance d'un concept \Rightarrow validation de l'ontologie
- Contrôler les relations de subsomption des concepts \Rightarrow validation d'ontologie
- Déterminer les descendants/parents d'un concept \Rightarrow parcourir hiérarchie «Thème»

E *Au niveau assertional* : autres requêtes

- Vérifier la consistance et tester les instances \Rightarrow validation ontologie
- Calculer le type direct d'un individu/instance \Rightarrow implémenter les tests de type
- Extraire les instances d'une classe particulière, selon des critères (RQL) \Rightarrow implémentation des fonctions d'extraction



Etude de cas

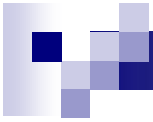
✓ Le rapprochement entre les diplômes et les métiers : (basé compétence)

	Technicien R/T	Gestionnaire applications	Développeur	Administrateur BDDs	Expert SE
Licence académique	60,50%	61,36%	59,32%	54,38%	42,60%
Licence GL	55,63%	67,74%	80,79%	55,21%	37,00%
Licence SI	57,60%	90,66%	72,60%	73,18%	38,80%
Licence STIC	76,71%	66,41%	70,48%	65,33%	48,40%



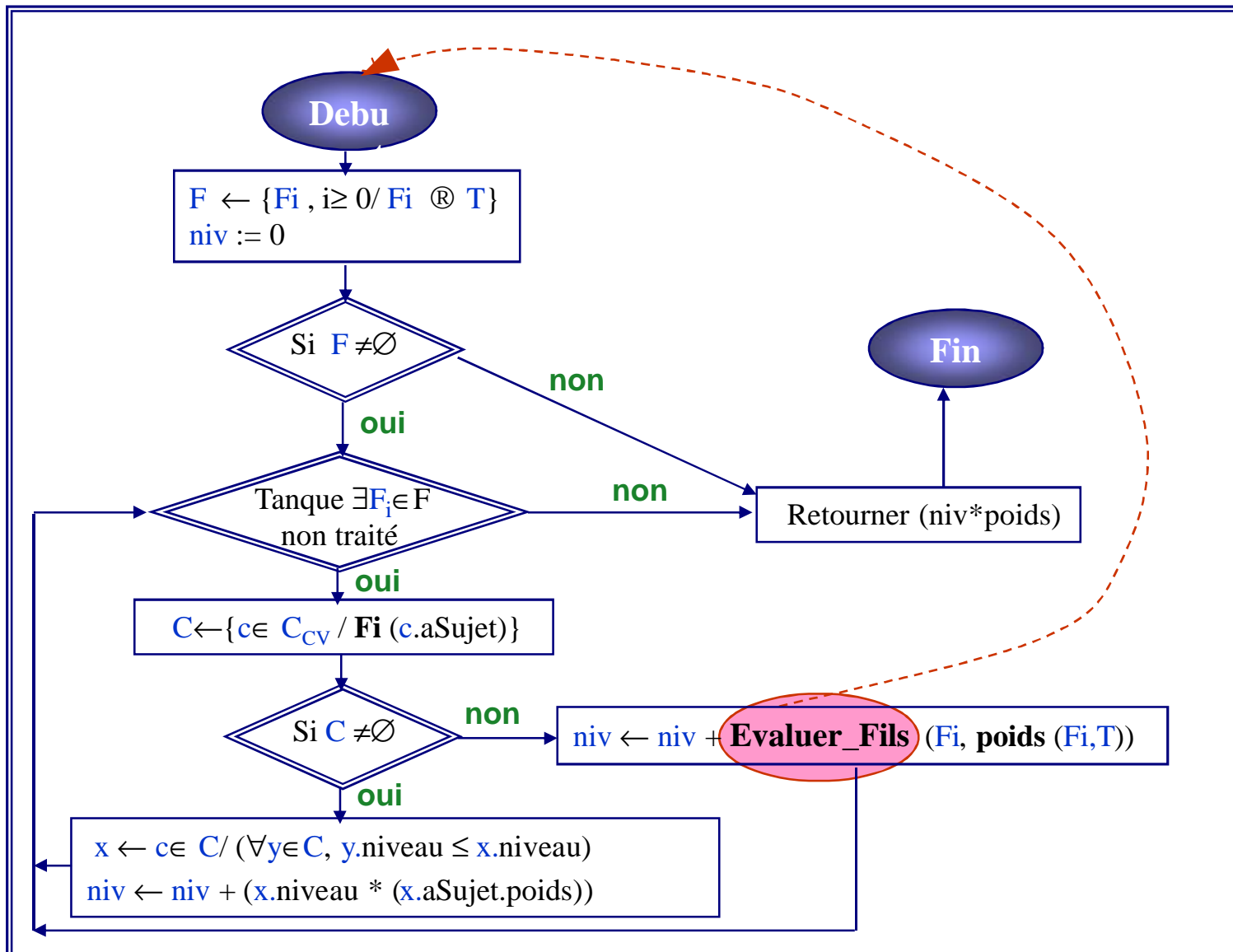
Conclusion et perspectives

- Proposition d'une approche simple et suffisamment complète, basée annotation sémantique, pour automatiser le rapprochement entre CVs et offres d'emploi.
- Modélisation du contenu sémantique à base d'une ontologie partagée.
- Une gestion basée compétence, assurée par une modélisation formelle de cette notion.
- Deux techniques de rapprochement (originaux) différentes mais complémentaires
- La possibilité de rapprocher les compétences des nouveaux diplômés, à celles exigées par les métiers du domaine.
- Une Implémentation assurée de l'approche (standards et technologies du WS)
- Le travail futur vise à
 - ∅ Implémenter les interfaces utilisateur du système et les algorithmes d'appariements
 - ∅ Valider cette approche sur des données réelles
 - ∅ Généraliser l'ontologie construite à d'autres domaines.



Merçi

Evaluer_Fils (T,coef) ◀





Les technologies du Web sémantique

Au service du e-recrutement

Leila Yahiaoui, Zizette Boufaïda et Yannick prié

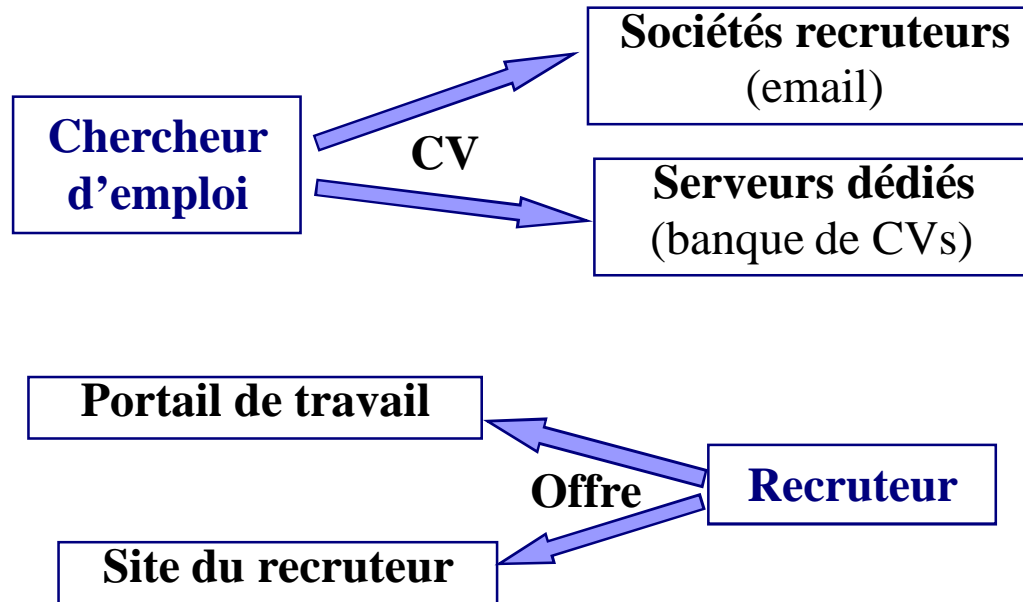


Le plan

- **Introduction**
- **Le e-recrutement sur le Web**
- **Généralités sur le Web sémantique**
- **Gestion des CVs /offres d'emploi dans le cadre du WS**
- **Objectif du travail**
- **Architecture du système proposé**
- **La génération des documents à annoter**
- **La modélisation du contenu sémantique à base d'ontologie**
 - ∅ Construction et architecture de l'ontologie-ER
 - ∅ Processus d'annotation sémantique des documents
- **Appariement sémantique entre CVs et offres d'emploi**
 - ∅ Appariement sémantique superficiel
 - ∅ Appariement sémantique basé compétence
- **Etude de cas**
- **Conclusion et perspectives**

Introduction

- ✓ Méthodes de recrutement classiques obsolètes ⇒ **Internet** : une nouvelle gestion (documents électroniques)

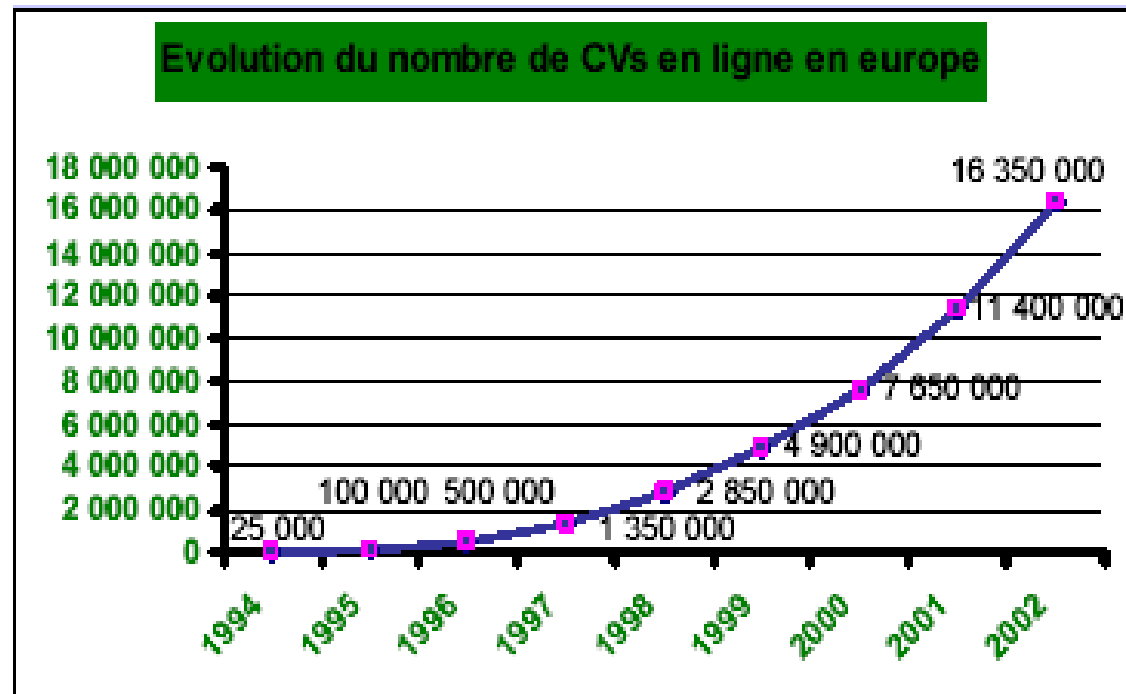


- ✓ **Le Web** : un vecteur de diffusion d'informations important

Le e-recrutement sur le Web

- ✓ **Le e-recrutement** : application typique d'une approche de gestion de connaissances à travers le Web.

(Computer Economic –2001)



E 13.4 billion\$ en 2005 avec un taux de croissance de 53%.

E grands sites de recrutement : plus de 50,000 CVs dans leurs BDDs.



Le e-recrutement sur le Web

✓ Avantages :

- E** Génération automatique des documents
- E** Réactivité pour la MAJ et meilleure gestion (stockage et recherche)
- E** Réduction importante en coût/délais
- E Recruteur** : masse de diffusion importante + nombre important de candidats avec une facilité/rapidité pour les contacter (email)
- E Chercheur d'emploi** : Facilité pour postuler

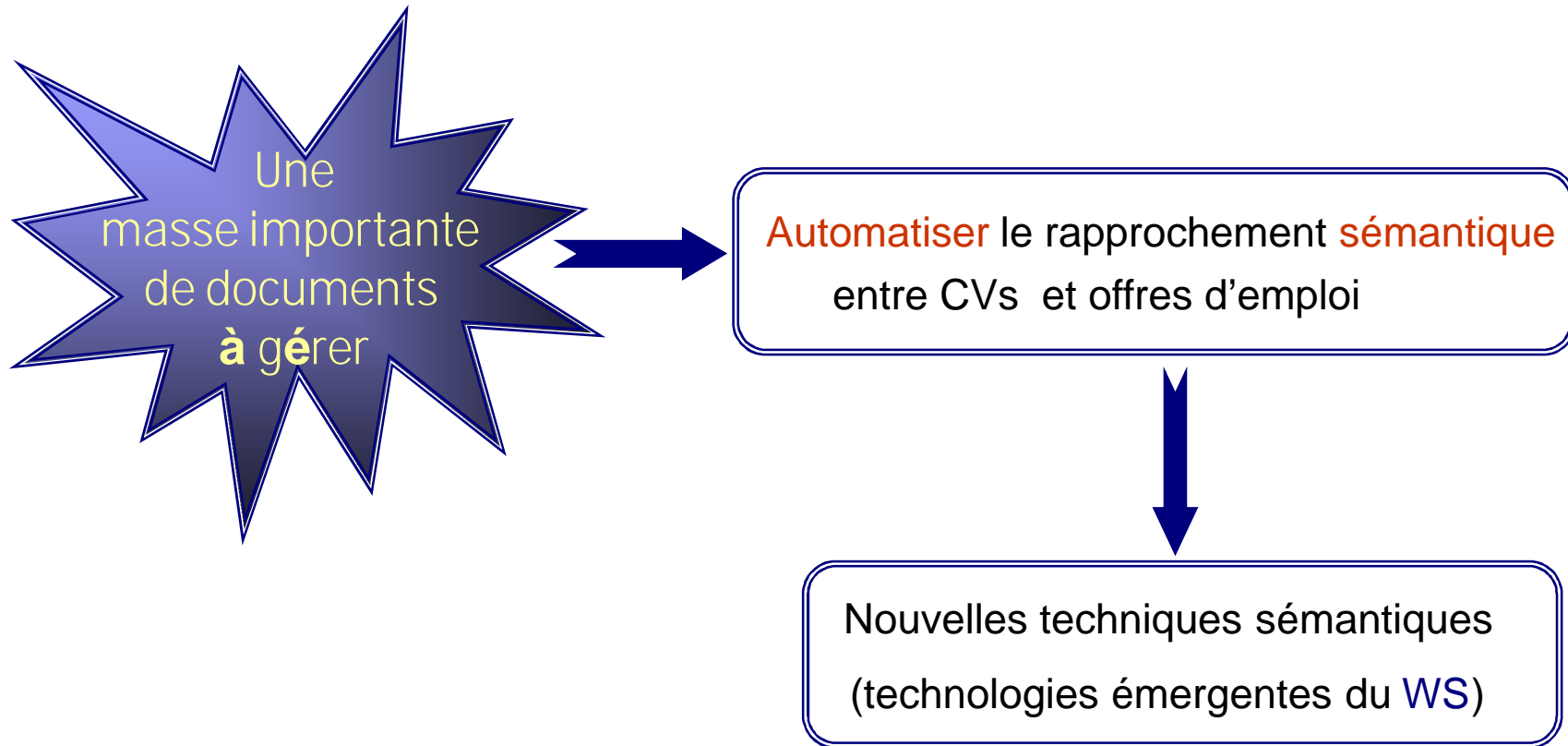
✓ Problèmes affrontés :

- E** Coût de publication élevé et problèmes des portails de travail
- E** Nombre énorme de CVs à traiter par les recruteurs (échec des formulaires prédéfinis : 26% vont jusqu'au bout (HRM News 2001))
- E** Inefficacité des moteurs de recherche actuels (chercheur d'emploi)
- E** Difficulté de formaliser les acquis/requis (CV/offre)
- E** Hétérogénéité et distribution des sources de données

Le e-recrutement sur le Web

- ✓ **Synthèse** : techniques/outils de gestion de documents purement syntaxiques et limités (gestion des compétences/connaissances).

⇒ Ressources mal exploitées



Généralités sur le Web sémantique



✓ Le Web sémantique : Tim Berners Lee (W3C)

- ∅ Utilisation de **connaissances formalisées** en plus du contenu informel du Web actuel
⇒ **contenu interprétable** par **machine**.

Ontologie = élément crucial du WS

- ∅ **Définition** : spécification **formelle** est **explicite** d'une **conceptualisation** [Gruber,95].
- ∅ **Intérêts** :
 - Communication entre êtres humains ou entre humains et machines,
 - Spécification de systèmes,
 - Interopérabilité,
 - Indexation et la recherche de documents . . .



Généralités sur le Web sémantique

√ Les ontologies :

∅ Composants :

- Les concepts (terme, notion, extension)
- Les propriétés (valuées)
- La relation de subsomption « Is-a » (généralisation/spécialisation)
- Les autres relations (terme, signature)

∅ Formalismes de représentation des connaissances :

- Réseaux sémantiques
- Langages de frames
- Graphes conceptuels
- Logiques de description (logique des prédicats + langages de frames + RS)
 - 2 niveaux : T_box et A_box
 - types LDs : FL, SHOIN, SHIQ ...
 - raisonneurs : Classic, Loom, Racer..



Généralités sur le Web sémantique

√ Les ontologies :

∅ Langages : RDF, RDFs, DAML+OIL et OWL

- **OWL** : standard recommandé par W3C.

→ Inspiré des LDs

→ grand nombre de constructeurs : OWL LITE, OWL DL et OWL FULL

∅ Types : de représentation, générique, d'une méthode, de domaine ou d'application.

∅ Outils de construction d'ontologies : Ontoedit, Oiled, Webode, Protégé-OWL ...

∅ Méthodologies de construction : Enterprise, Tove, Methontology ...



Gestion des CVs/offres dans le cadre du WS

✓ **Approches de modélisation du contenu sémantique :**

Ø L'indexation sémantique de documents :

- Construire un index d'un site dont la structure correspond à une ontologie
- Dédiée à l'extraction de documents concernant une recherche spécifique

Inconvénients :

- E** Les documents sont juste associés aux concepts des ontologies
- E** Techniques de TALN difficiles à maîtriser.

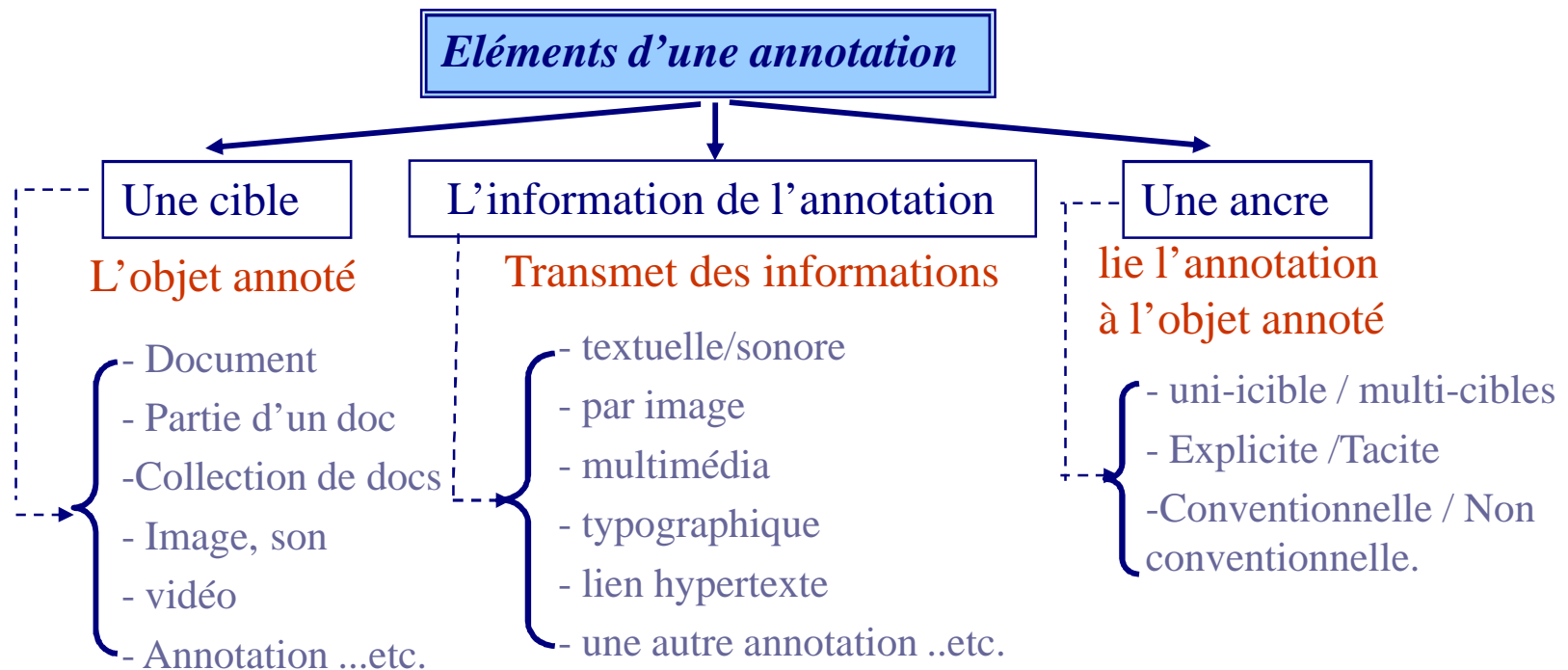
Ø L'annotation sémantique de documents :

- Enrichir explicitement les documents, à base d'ontologies, avec des méta-données
- Dédiée aux systèmes «request/answer ».

Gestion des CVs/offres dans le cadre du WS

✓ Annotation classique :

- ∅ Définition : information graphique/textuelle attachée à un document [Desmontils, 02]



Gestion des CVs/offres dans le cadre du WS

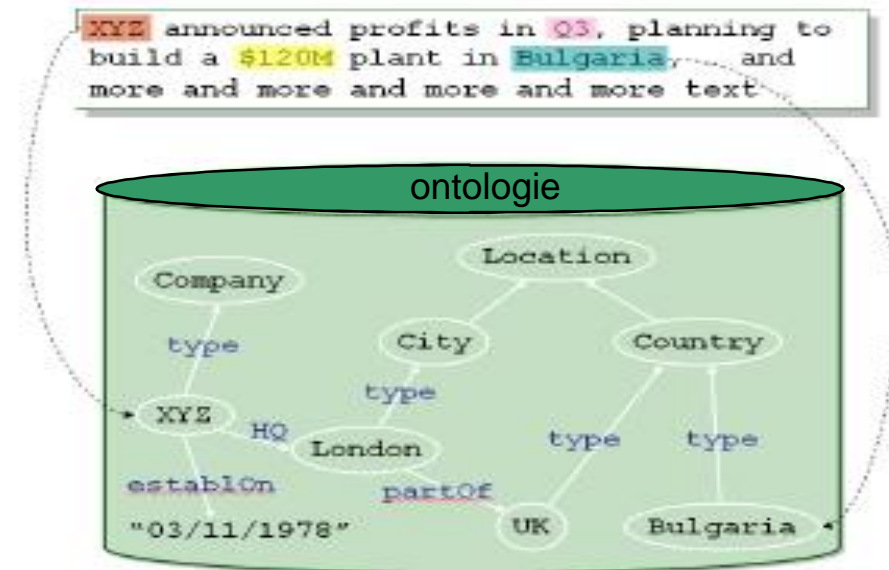
✓ Annotation sémantique :

∅ Définition : assigner aux entités d'un document des liens à leurs descriptions sémantiques [Kiryakov, 05].

→ des informations de classes et d'instances liées à une ontologie

∅ Processus :

- **Repérer** : placer les références aux concepts dans le document
- **Instancier** : valuer les attributs des concepts
- **Enrichir** : ajouter des informations (attributs non valués)



∅ But : ressources Web définies et liées, munis d'un sens explicitement interprétable par machines (recherche de documents, résumé, composition ...)



Gestion des CVs/offres dans le cadre du WS

✓ Travaux existants basés annotation sémantique:

§ *Le projet COMMONCV* intitulé « recherche&emploi »

§ *Un processus de recrutement basé ontologie* [Bizer05]

§ *Os-Skill ...*

∅ *Avantage* : approche facile et naturelle + disponibilité des standards/outils du WS.

∅ *Manques dans ces travaux* :

E L'incomplétude ou le niveau de détail insuffisant donné

E Simplicité relative.

E La concentration sur les compétences relatives à l'expérience professionnelle

E Un appariement sémantique partielle (basé compétence seulement)

Objectif du travail

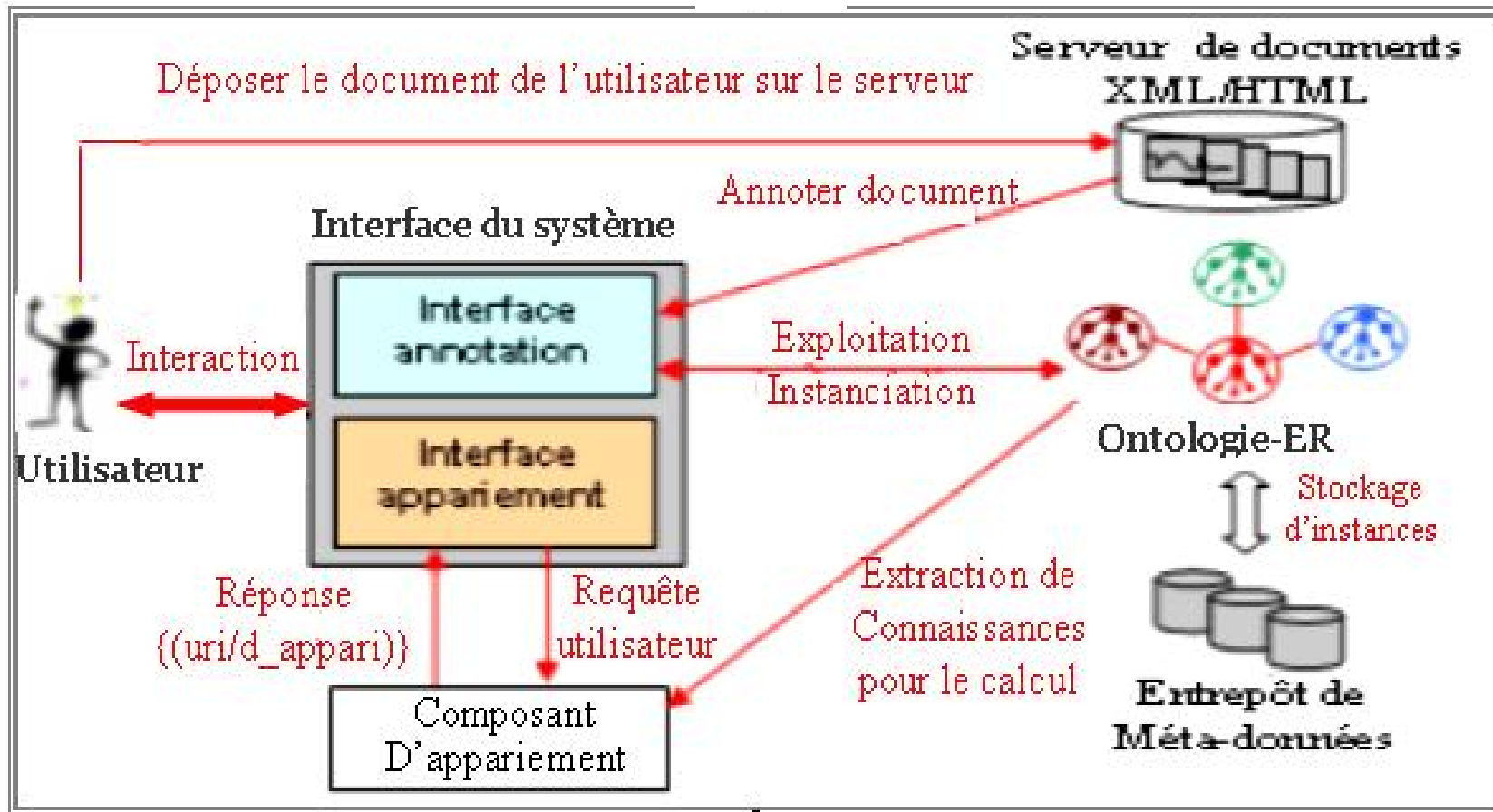


proposer une solution **simple**,
détaillée et suffisamment
complète, basée
annotation sémantique, pour
automatiser l'appariement
sémantique entre les
CVs/ offres d'emploi

- ∅ Génération des documents à annoter (XML)
- ∅ Construction d'une **ontologie** destinée à la modélisation du contenu sémantique des documents (acquis/requis)
- ∅ Proposition d'un processus d'annotation sémantique de documents.
- ∅ Description des services d'appariement sémantique à appliquer entre les CVs et les offres d'emploi

Architecture du système proposé

Le système d'annotation et d'appariement sémantique





Génération des documents à annoter

- E proposition d'une structure** pour chacun des types de documents
(CV/offre d'emploi) → Monster, JobPilot et CVGEN + des particularités
- E modélisation** de ces structures par des diagrammes de classes **UML**
- E Codification** des diagrammes UML en **Schéma XML**
- E Mise en page** des documents XML générés (feuilles de styles **XSL**)



Documents stockés dans des bases de données XML

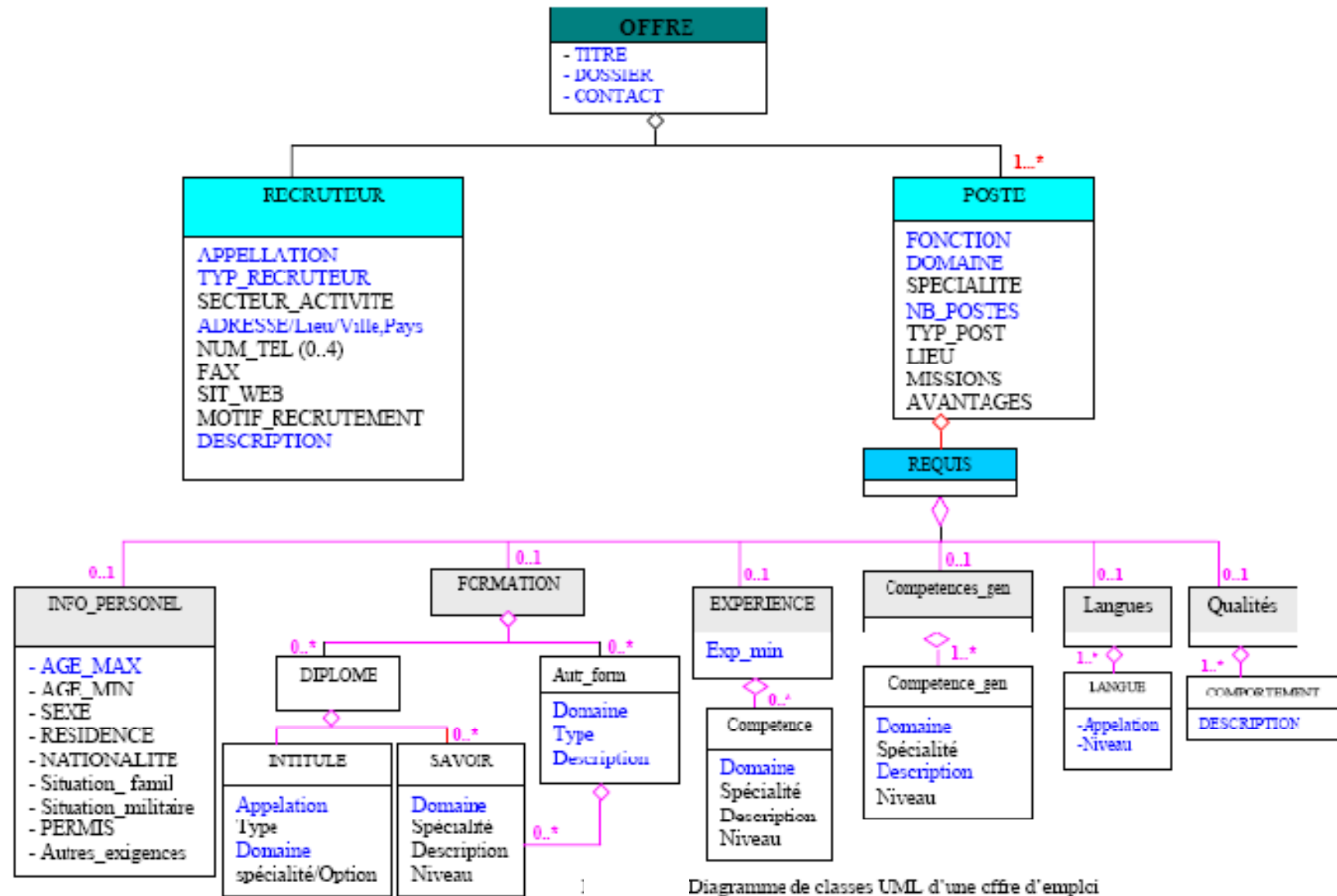


Génération des documents à annoter

- ✓ **Caractéristiques d'un CV** : document court , syntaxiquement pauvre mais bien structuré + caractéristiques lexicales/syntaxiques/typographiques traditionnelles
 - Ø Données signalétiques
 - Ø Formation
 - Ø Expérience professionnelle
 - Ø Activités extra-professionnelles (facultative)
 - Ø Divers : langues, centres d'intérêts, connaissances en nouvelles technologies ...

- ✓ **Caractéristiques d'une offre d'emploi** : document court, syntaxiquement pauvre, structurée, monoposte/ multipostes, formalisé en texte.
 - ⇒ description du recruteur+ description de chacun des postes de travail offerts
 - Ø description générale du poste
 - Ø requis personnels du poste
 - Ø formation exigée
 - Ø expérience professionnelle exigée (profession avec nombres d'années d'expertise)
 - Ø autres requis

Génération des documents à annoter



Génération des documents à annoter

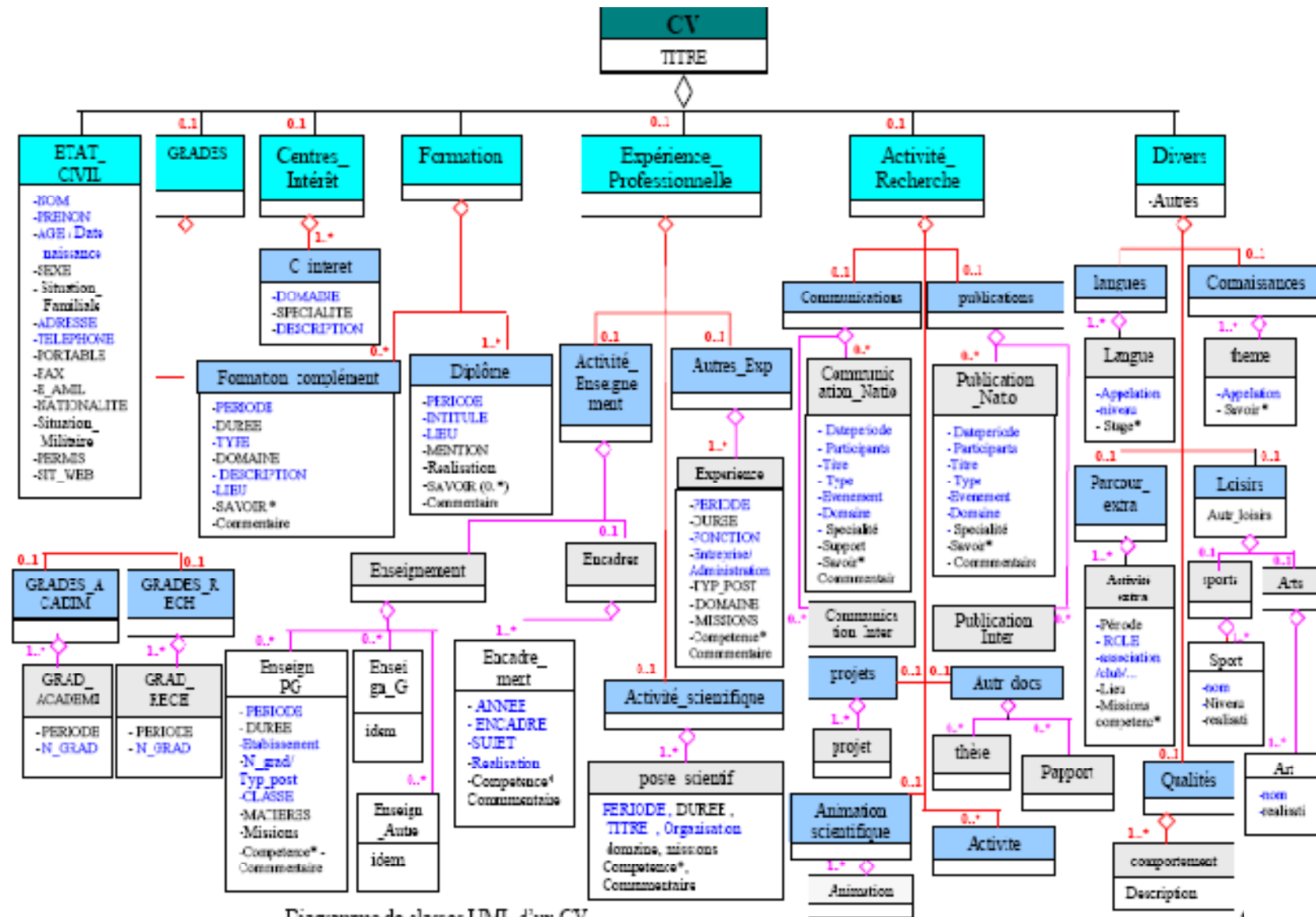


Diagramme de classes UML d'un CV

Génération des documents à annoter

The screenshot displays the Oxygen XML Editor interface. The main window shows a diagram of an XSD schema for an 'OFFRE' (job offer). The diagram illustrates a sequence of elements: 'OFFRE' (root), followed by a choice between a document icon and a box icon. The box icon leads to a sequence of 'TITRE', 'recruteur', and 'POSTE' (with a '1..oo' cardinality). The 'Composants Schema' (Schema Components) panel on the right lists simple types: 'moncp', 'numéroDeTéléphone', 'typeAdresseElectronique', 'typePermis', 'typeSitWeb', 'typeSituFAMIL', and 'typeSituMilitaire'. Below the diagram, the XML code is visible, showing an annotation and the definition of the 'OFFRE' complex type with its sequence of elements.

```
304 <xsd:documentation>l'offre inclue des infor sur le recruteur,et sur les postes ainsi que des requis dans les candidats</xsd:
305 </xsd:annotation>
306 <xsd:complexType>
307 <xsd:sequence>
308 <xsd:element name="TITRE" type="xsd:string" minOccurs="1" />
309 <xsd:element ref="recruteur" />
310 <xsd:element ref="POSTE" minOccurs="1" maxOccurs="unbounded"/>
311 <xsd:element name="DOSSIER" type="xsd:string" minOccurs="1" />
312 <xsd:element name="CONTACT" type="xsd:string" minOccurs="1" />
```

Le schéma XML de l' offre d'emploi (éditeur Oxygen)

Génération des documents à annoter

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSLTransform" version="1.0">
3   <xsl:output method="html"/>
4   <xsl:template match="/">
5     <html>
6       <body>
7         <xsl:for-each select="CV">
8           <h2> <center> <xsl:value-of select="TITRE"/> </center> </h2>
9           <xsl:apply-templates select="ETAT_CIVIL"/>
10          <xsl:apply-templates select="GRADES"/>
11          <xsl:apply-templates select="CENTRES_INTERET"/>
12          <xsl:apply-templates select="FORMATION"/>
13          <xsl:apply-templates select="EXPERIENCE_PROFESSIONNELLE"/>
14          <xsl:apply-templates select="ACTIVITE_RECHERCHE"/>
15          <xsl:apply-templates select="DIVERS"/>
16        </xsl:for-each>
17      </body>
18    </html>
19  </xsl:template>
20  <xsl:template match="ETAT_CIVIL">
21    <h4 style="color:red"> ETAT CIVIL </h4>
22    <xsl:text > Nom et Prénom : </xsl:text>
23    <span style="font-style:italic; padding-right:3pt" > [2 lines]
24    <span style="font-style:italic; padding-right:3pt" > [2 lines]

```

Feuille
de style



```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="cv_stylesheets.xml"?>
3 <CV xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4   xsi:noNamespaceSchemaLocation="file:/D:/Documents%20and%20Set
5   <TITRE>Demande d'emploi</TITRE>
6   <ETAT_CIVIL> [17 lines]
24  <GRADES> [9 lines]
34  <CENTRES_INTERET> [11 lines]
46  <FORMATION>
47    <DIPLOME>
48      <PERIODE>
49        <DATPERIODE>1994-06</DATPERIODE>
50      </PERIODE>
51      <INTITULE>
52        <APPELATION>baccalauréat</APPELATION>
53        <TYPE>bac</TYPE>
54        <DOMAINE>science </DOMAINE>
55        <SPECIALITE>science de la nature et de la vie</SPECIALITE>
56      </INTITULE>
57      <LIEU>Constantine - Algérie</LIEU>
58      <MENTION>bien</MENTION>
59    </SAVOIR>
60    <DOMAINE>science</DOMAINE>
61    <SPECIALITE>physique</SPECIALITE>
62    <DESCRIPTION>apprendre bien</DESCRIPTION>

```

Document
XML

ETAT CIVIL.

Nom et Prénom : *yahiaoui leila*
 Date de naissance/ Age : 30
 Sexe : *feminin*
 État civil : *célibataire*
 Adresse : *cité des frères abbès rue "s" n°= 7 Constantine 25000 Algérie*
 Tél / portable : *031 66 25 25 /*
 Fax :
 e_mail : *leila@yahoo.fr*
 Nationalité : *algérienne*
 Situation militaire :
 Permis de conduire : *A*
 Sit Web :

GRADES

Grades Pédagogiques :
 2005 *maître assistant stagiaire*

CENTRES D'INTÉRÊT

web sémantique / construction et évolution d'ontologies (création d'ontologies de domaines dans le cadre de la gestion des ressources humaines)
bases de données / oracle (développement de logiciels sous oracle Si et l'extension vers le Web)

DIPLÔMES ET FORMATION

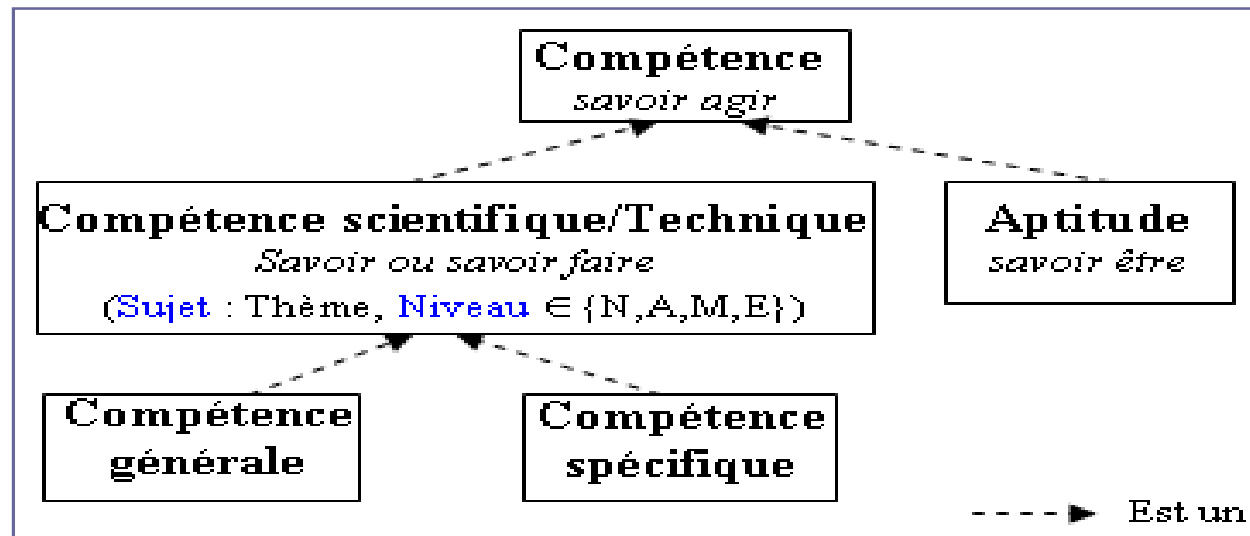
DIPLÔMES:

Période	Intitule	Lieu	Mention	Réalisations	Acquis
1994-06/	baccalauréat , type bac , spécialité:science /science de la nature et de la vie	Constantine - Algérie	bien		-- Domaine: science /physique (apprendre bien),niveau : bon -- Domaine: chimie /organique (tout ou rien),niveau : excellent (la vie est très difficile et elle est très collante de la science à l'organisation)
					-- Domaine: http://www.formation.formation.com

Document HTML (CV)

Modélisation du contenu sémantique

✓ Le modèle de la compétence :



∅ **Domaine** : Informatique et télécommunications

E Le sujet : un « Thème » ou un « Aspect logiciel ».

E Le niveau:

- Notion : N ou 20%
- Application : A ou 50%
- Maîtrise : M ou 70%
- Expert : E ou 90%

Modélisation du contenu sémantique

✓ Construction de l'ontologie-ER :

∅ Importance :

- E** Référentiel commun entre recruteurs et chercheurs d'emploi
- E** Description du contenu sémantique (non ambiguë et précise).
- E** Acquisition de nouvelles connaissances (de l'ontologie)
- E** La formalisation \Rightarrow un raisonnement automatique.

∅ **Eléments** : inspirés des parties communes, les plus significatives des documents

- E** certaines informations personnelles (attestées/exigées)
- E** les métiers (acquis/requis)
- E** des diplômes (acquis/requis)
- E** compétences explicites et implicites (acquises/requises)



Modélisation du contenu sémantique

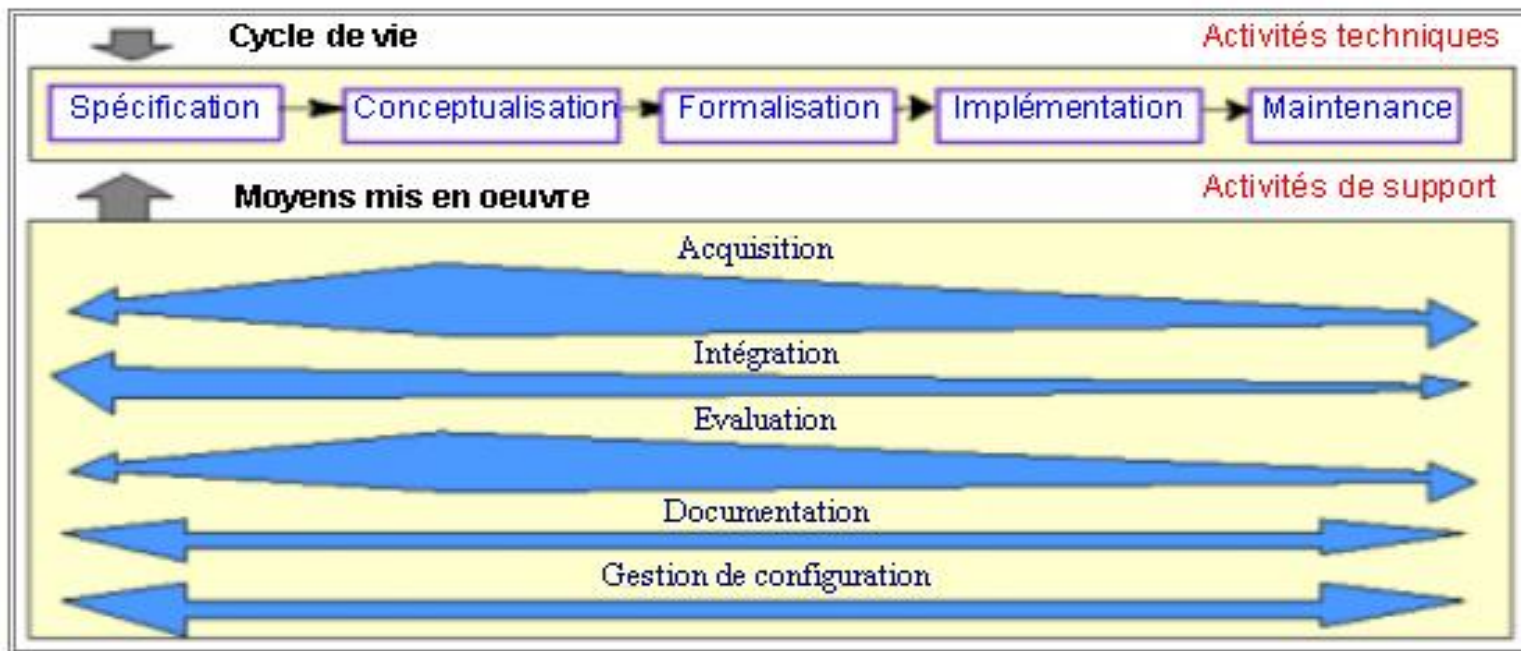
✓ Construction de l'ontologie-ER : Méthodologie

Inspirée de «METHONTOLOGY» (Fernandez et Gomez)

E Cadrage : cerner le domaine de l'ontologie

E Conceptualisation : Identification et structuration des connaissances (tables/graphes)

E Implémentation : formalisation (LDs) + codification (OWL..).



Modélisation du contenu sémantique

✓ Construction de l'ontologie-ER : formalisme de représentation

Les logiques de description (SHOIN D)

- E** Pouvoir d'expression + mécanismes d'inférences
- E** Faire apparaître les notions : d'**identité** (individus), de **classes** (concepts), de **sous-typage** (subsomption) ...
- E** Recommandation du W3C

concepts	
Universel/absurde	\top / \perp
Atomique	A, B
Non	$\neg C$
Et	$C \sqcap D$
Ou	$C \sqcup D$
Existe	$\exists R.C$
Pour tous	$\forall R.C$
Au moins	$\geq n R.C$ ($\geq n R$)
Au plus	$\leq n R.C$ ($\leq n R$)
énumération	$\{i_1, \dots, i_n\}$
Rôles	
Atomique	R
Inverse	R^{-}

Axiomes de concept (T_Box)	
Sous-classe	$C \sqsubseteq D$
Equivalence	$C \equiv D$

Axiomes de Rôle (R_Box)	
Sous rôle	$R \sqsubseteq S$
transitivité	Trans (S)

Axiomes assertionels (A_Box)	
Instance	$C(a)$
Rôle	$R(a,b)$
même	$a = b$
différents	$a \neq b$

Modélisation du contenu sémantique

✓ Construction de l'ontologie-ER : langage d'implémentation

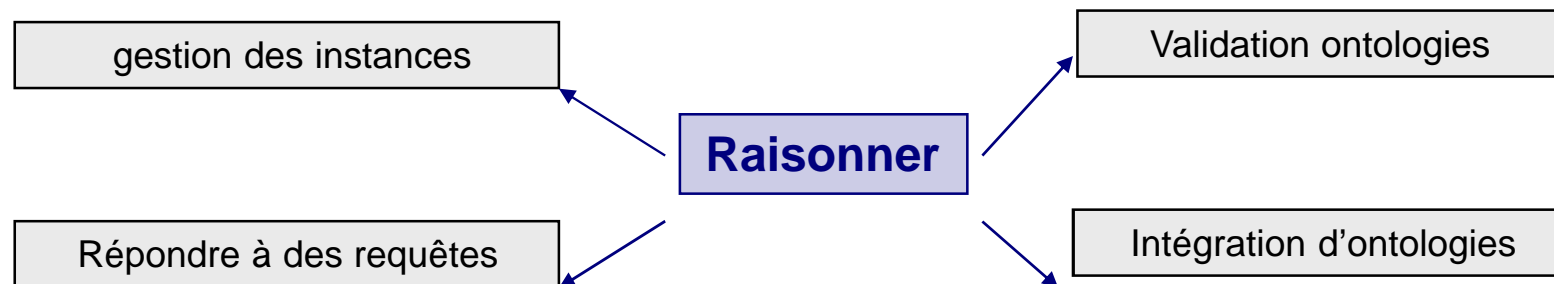
OWL-DL (Ontology Web Language)

E Pouvoir d'expression

E Puissance des services d'inférences (raisonneur RACER)

E Recommandation du W3C

E Correspondance directe avec SHOIN D

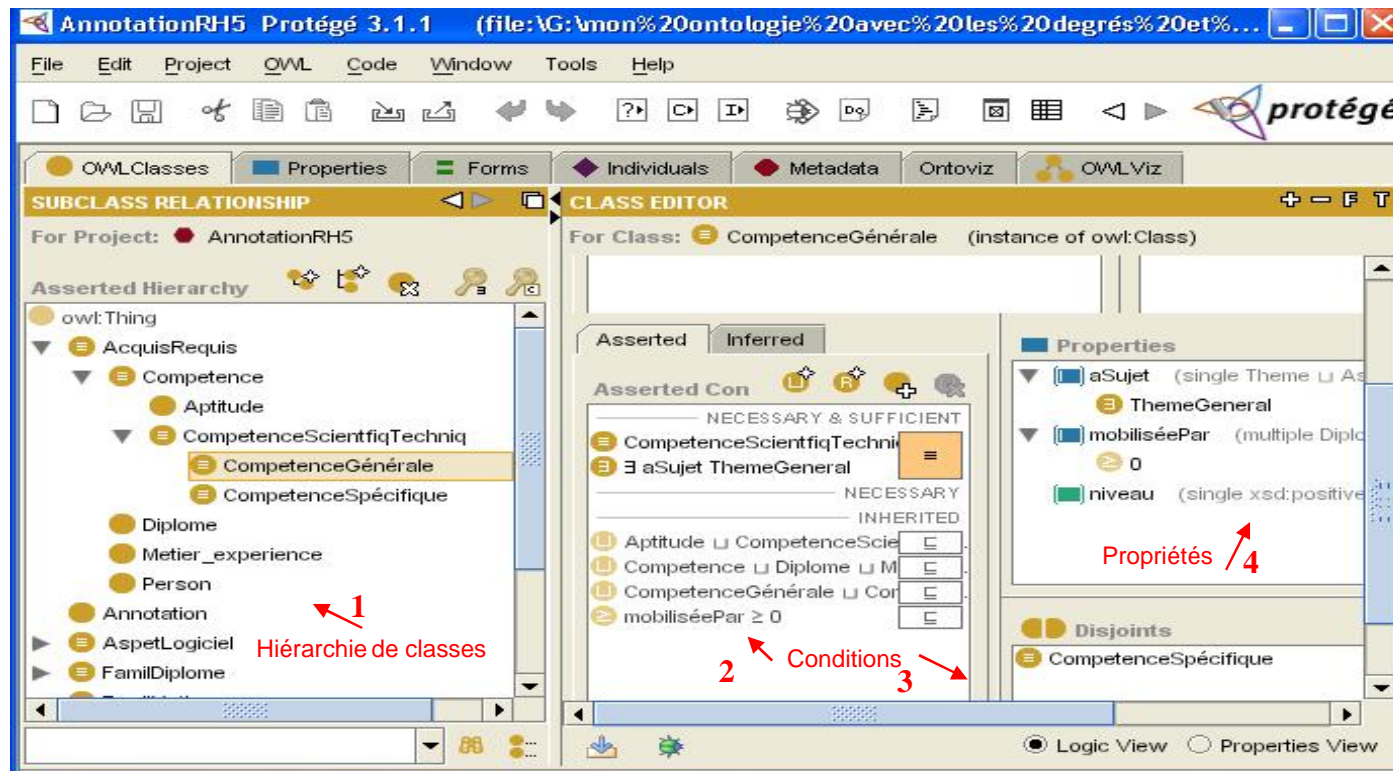


Modélisation du contenu sémantique

✓ Construction de l'ontologie-ER : Outil de construction

Protégé-OWL (version 3.1.1)

- E** Charger et sauvegarder des ontologies (OWL)
- E** Editer/visualiser les classes/propriétés
- E** Exprimer des contraintes OWL, éditer les individus...
- E** Exécuter des raisonneurs

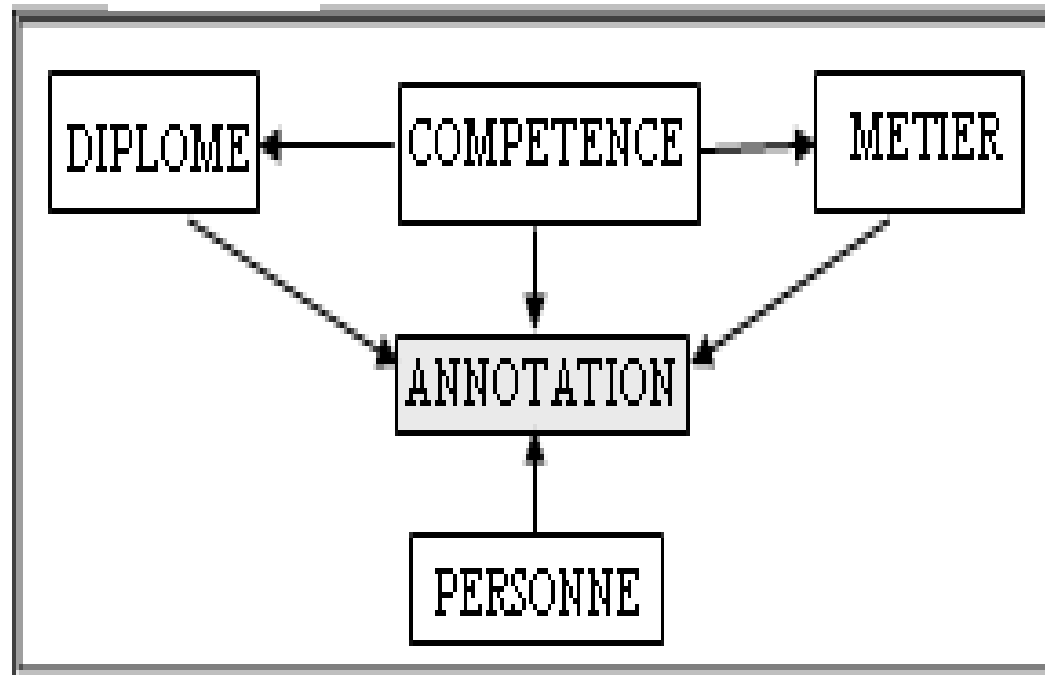


Modélisation du contenu sémantique

✓ Architecture globale de l'ontologie-ER :

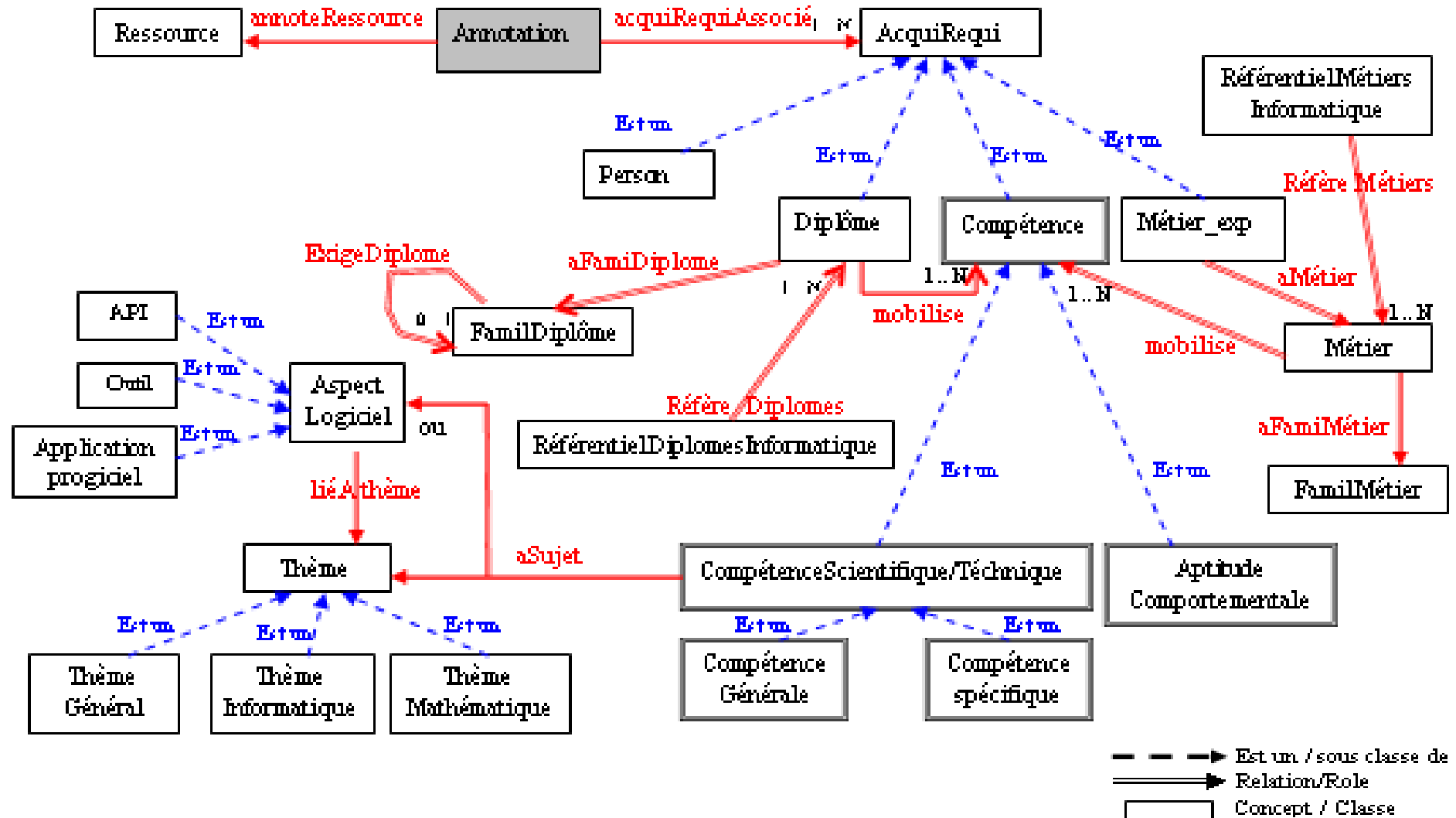
E Domaine : «Informatique et Télécommunications»

E Constituée de 5 sous-ontologies



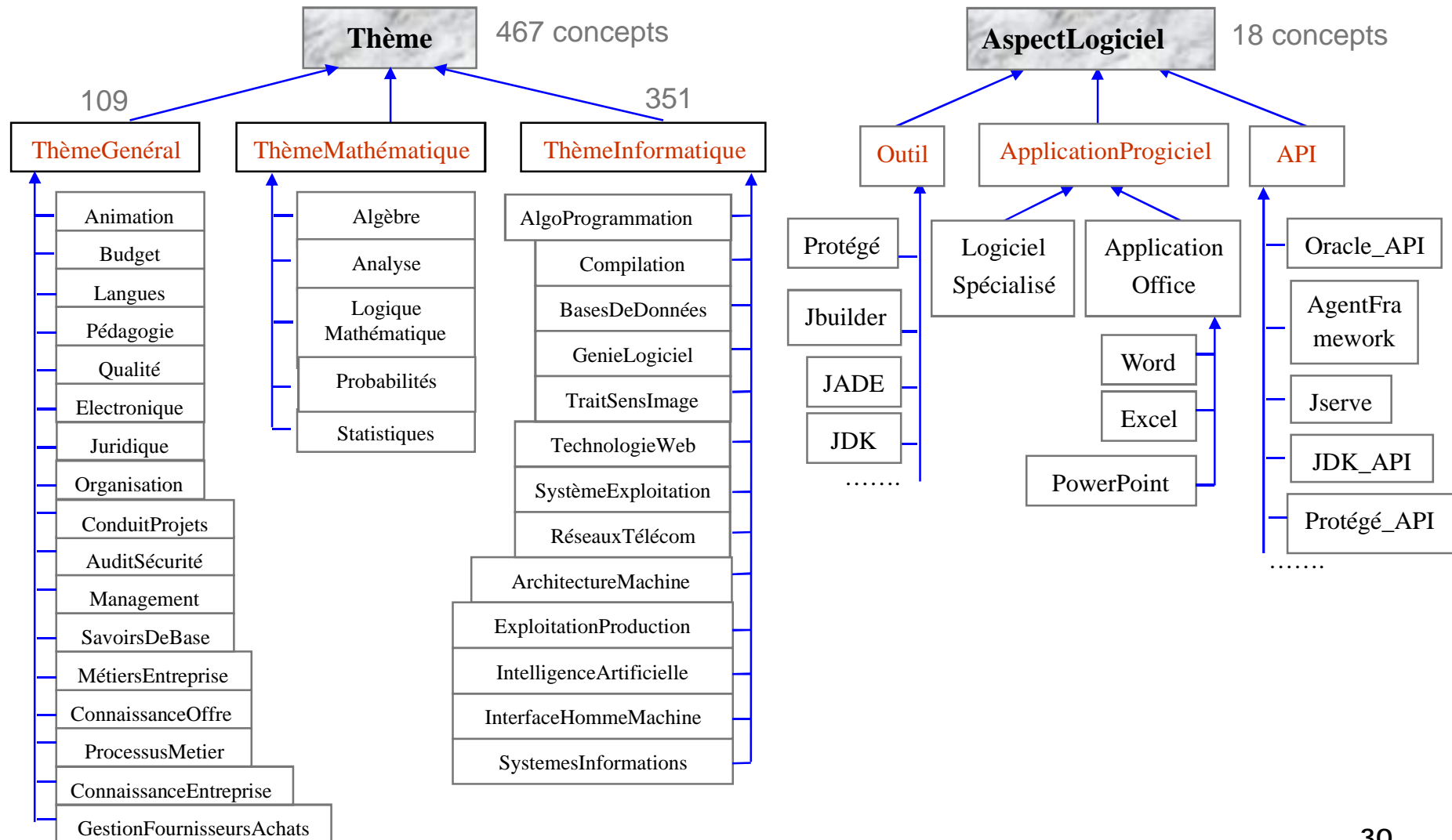
Modélisation du contenu sémantique

✓ Architecture détaillée de l'ontologie-ER : (510 concepts/17 relations/20 attributs)



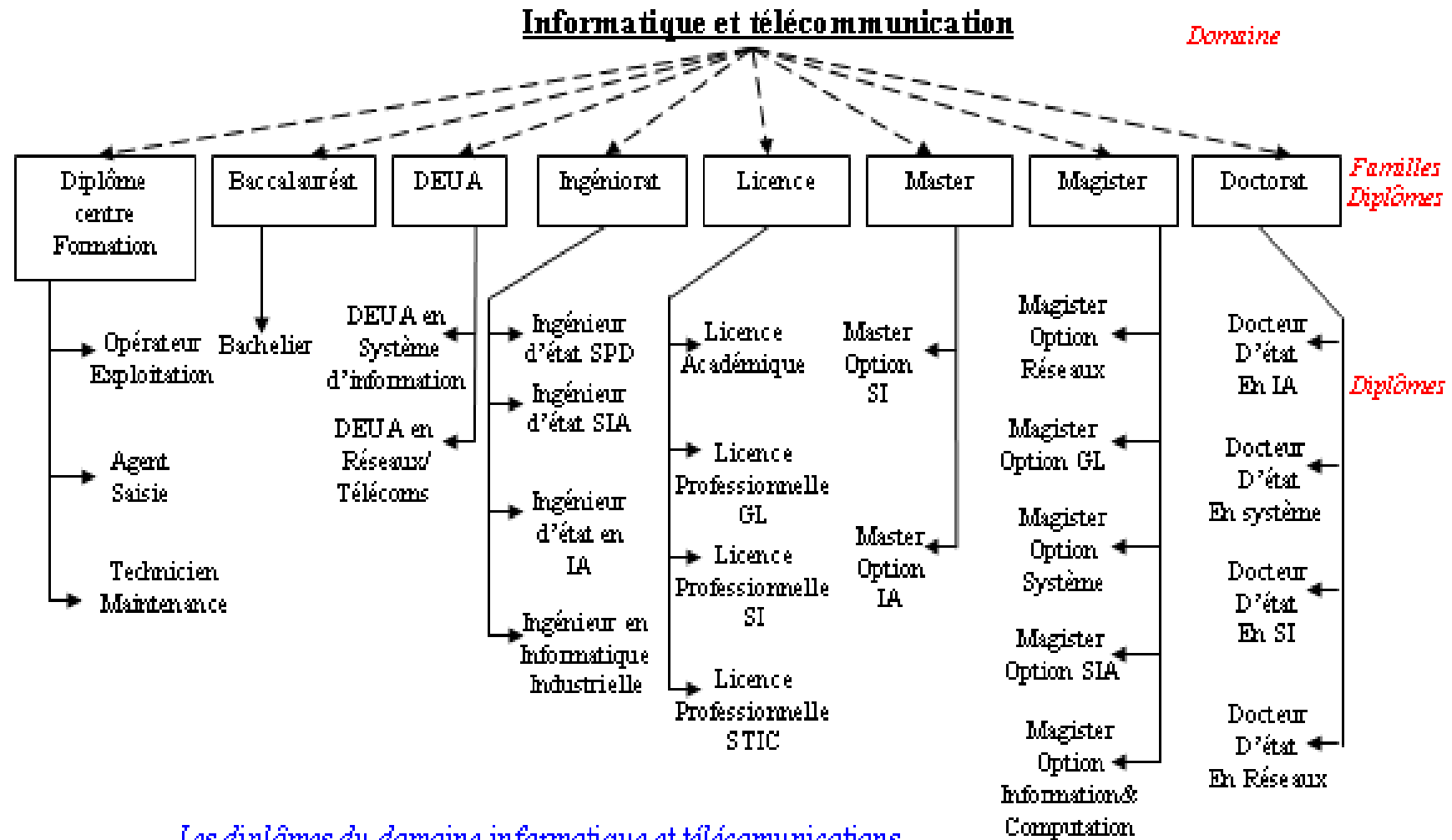
Modélisation du contenu sémantique

Architecture détaillée de l'ontologie-GRH:



Modélisation du contenu sémantique

Architecture détaillée de l'ontologie-ER

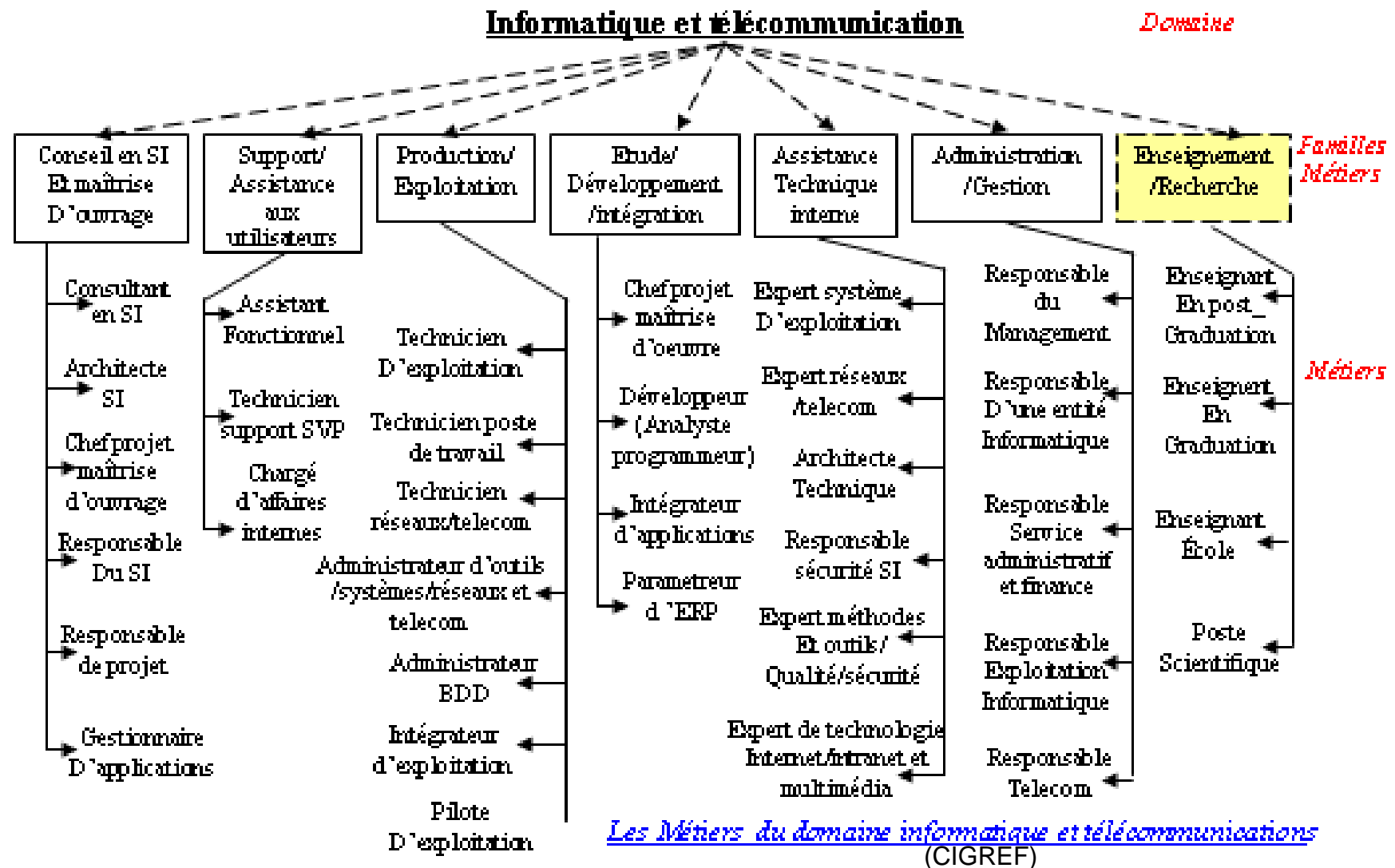


Les diplômes du domaine informatique et télécommunications

(Université de Constantine et autres - Algérie)

Modélisation du contenu sémantique

Architecture détaillée de l'ontologie-ER





Processus d'annotation sémantique

- ✓ **Rôle de l'expert** : Instancier à l'avance les sous-ontologies «**METIER**», «**DIPLOME**» et «**COMPETENCE**»
 - E** Une instance de «**Thème**»/«**AspectLogiciel**» \Rightarrow 4 instances de «**Compétence**»
 - E** Lier les instances de «**Métier**» et «**Diplôme**» aux instances de «**Compétence**» qu'ils mobilisent

- ✓ **Rôle de l'utilisateur** :
 - ∅ Instancier la classe «**Ressource**» \Rightarrow décrire le document à annoter .
 - ∅ Instancier la classe «**Person**» (qualifications personnelles)
 - ∅ Instancier la classe «**Métier_exp**» (expériences professionnelles)
 - ∅ Instancier la classe «**AcquiRequi**» avec les acquis du candidat (requis de l'offre)
 - ∅ Instancier la classe «**Annotation**» \Rightarrow lier les acquis/requis à la ressource annotée



Appariement sémantique entre CVs et Offres

✓ **Principe d'appariement :**

E Calcul d'un degré de rapprochement (pourcentage)

E Deux techniques différentes mais complémentaires

∅ **Appariement sémantique superficiel** : prend en considération les requis/acquis, avec lesquels les documents ont été annotés, à un niveau superficiel

∅ **Appariement sémantique basé compétence** : s'intéresse aux compétences sous-jacentes aux documents annotés.

✓ **conventions** : algorithmes entre un CV (r_{CV}) et un poste P (r_P)

- **C(i)** : i est un individu/instance du concept/classe C \Rightarrow classe (i) =C.
- **i.nomAtr** : la valeur de l'attribut «nomAtr» de l'individu i ou l'ensemble des individus avec lesquels i est relié par le rôle «nomAtr».
- **A[®] C** : le concept C subsume le concept A.

Appariement Superficiel (r_p, r_{cv})

Poids_tot \leftarrow 0 /*cumul coefficients des requis*/

Poids_calc \leftarrow 0 /*cumul coefficients des acquis*/

Extraction_AcquisRequis (AQ_{cv}, RQ_p)

Pour chaque $e_p \in RQ_p$ Répéter /*chaque requis*/

Si **Person** (e_p) alors /* e_p est de type « person »*/

 Poids_tot \leftarrow Poids_tot + Coef_du_type (**Person**)

$e_{cv} \leftarrow e \in AQ_{cv}$ tel que **Person**(e_{cv})

Pour chaque attribut de e_p vérifié dans e_{cv} répéter

 Poids_calc \leftarrow Poids_calc + Coef_du_type (**Person**) / 3

Fin répéter

Si **Diplôme** (e_p) alors

 Poids_tot \leftarrow Poids_tot + Coef_du_type (**Diplôme**)

Si ($\exists e \in AQ_{cv}$ tel que $e = e_p$) alors

 Poids_calc \leftarrow Poids_calc + Coef_du_type (**Diplôme**)

Si (**Métier_exp** (e_p) $\wedge e_p$.annéesExp \geq 0) alors

 Poids_tot \leftarrow Poids_tot + Coef_du_type (**Métier_exp**)

Si ($\exists e \in AQ_{cv}, e = e_p \vee (e.aMétier = e_p.aMétier \wedge e_p.annéesExp \leq e.annéesExp)$) alors

 Poids_calc \leftarrow Poids_calc + Coef_du_type (**Métier_exp**)

Simon Si ($\exists e \in AQ_{cv}$ tel que $e.aMétier = e_p.aMétier$) alors

 Poids_calc \leftarrow Poids_calc + Coef_du_type (**Métier_exp**) / 2.

Si **Compétence** (e_p) alors

 Poids_tot \leftarrow Poids_tot + Coef_du_type (**Compétence**)

Si ($\exists e \in AQ_{cv}$ tel que $e = e_p$) alors

 Poids_calc \leftarrow Poids_calc + Coef_du_type (**Compétence**)

Fin Répéter

D_appari \leftarrow (Poids_calc / Poids_tot) * 100 /*degré appari*/

Fin

Appariement basé compétence (C_P, C_{CV})

Poids_tot \leftarrow 0 /*cumul coefficients des requis*/

Poids_calc \leftarrow 0 /*cumul coefficients des acquis*/

Extraction_compétences (C_{CV}, C_P)

Pour chaque $C_T \in C_P$ répéter

Si **Aptitude** (C_T) **alors** /*compétence comportementale*/

Poids_tot \leftarrow Poids_tot + Coef_du_type (**Aptitude**)

Si $C_T \in C_{CV}$ **alors** /*la compétence est acquise*/

Poids_calc \leftarrow Poids_calc + Coef_du_type (**Aptitude**)

Simon /*compétence scientifique&technique*/

Si **CompétenceGénérale** (C_T) **alors**

$C_1 \leftarrow$ Coef-du-type (**CompétenceGénérale**)

Simon $C_1 \leftarrow$ Coef-du-type (**CompétenceSpécifique**)

Poids_tot \leftarrow Poids_tot + C_1

$X \leftarrow$ { $c \in C_{CV} / c.aSujet = C_T.aSujet$ } /* chercher le même sujet*/

Si $X = \emptyset$ **alors**

niv \leftarrow 0 ; **Evaluer_Fils** (classe($C_T.aSujet$), niv)

Simon $C_A \leftarrow c \in X / (\forall y \in X, y.niveau \leq C_A.niveau)$

niv \leftarrow $C_A.niveau$ /*choix de la meilleure compétence*/

Si (niv $\geq C_T.niveau$) **alors** /*niveau compétence acquise supérieur*/

Poids_calc \leftarrow Poids_calc + C_1

Simon $K :=$ entier (($C_T.niveau - niv$) / 20) /*arrondir le niveau*/

Poids_calc \leftarrow Poids_calc + ($C_1 * (1 - K / 4)$)

Fin Répéter

$D_{appari} \leftarrow$ (Poids_calc / Poids_tot) * 100 /*degré appari*/

Fin

Evaluer_Fils (T, niv)

$F \leftarrow$ { $F_i, i \geq 0 / F_i \rightarrow T$ } /* subsumés du thème T*/

Si $F \neq \emptyset$ **alors** /* Pas de sous_thème de T */

Pour chaque $F_i \in F$ répéter /* pour chaque thème fils*/

$C \leftarrow$ { $c \in C_{CV} / F_i(c.aSujet)$ } /*ensemble des compétences acquises dont le sujet est instance de F_i */

Si $C \neq \emptyset$ **alors** /*choisir la compétence de meilleur niveau*/

$x \leftarrow c \in C / (\forall y \in C, y.niveau \leq x.niveau)$

niv \leftarrow niv + ($x.niveau * (x.aSujet.poids)$)

Simon /*descendre dans la hiérarchie*/

Evaluer_Fils (F_i, niv) /*récursivité*/

Fin Répéter

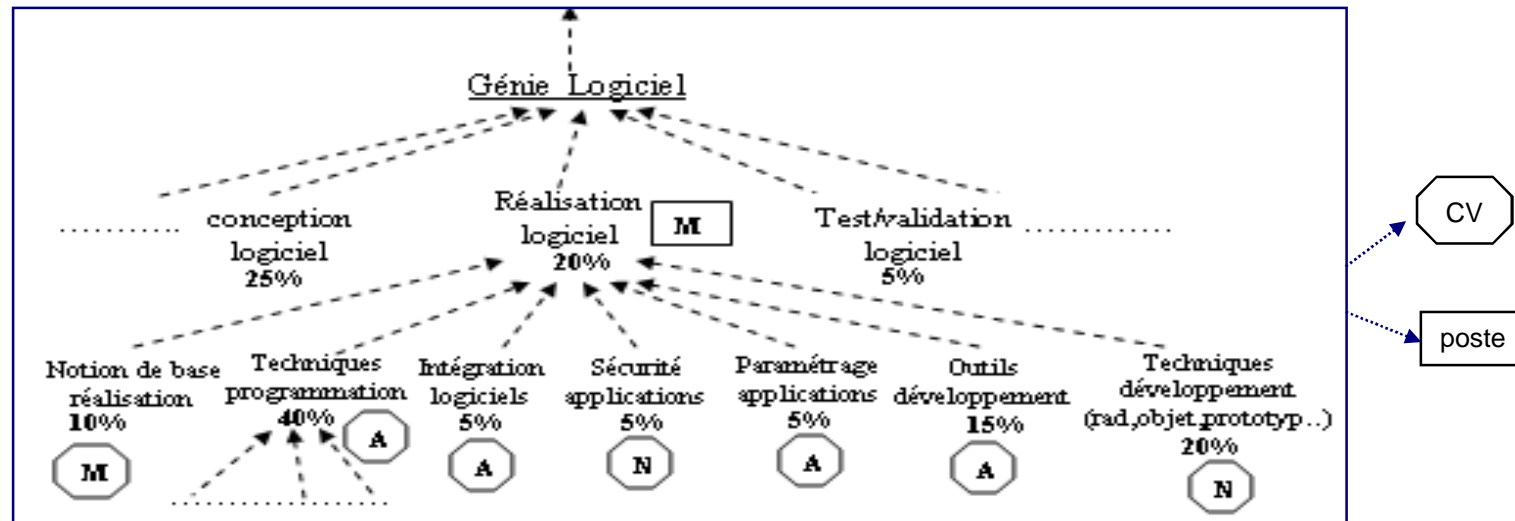
Fin

Le niveau s'évalue à

- (N \cong 20%) si niveau < 25%,
- (A \cong 50%) si 25% \leq niveau < 60%
- (M \cong 70%) si 60% \leq niveau \leq 75%
- (E \cong 90%) si niveau > 75% .

Appariement sémantique entre CVs et Offres

✓ Exemple d'application : appariement basé compétence



E Calcul du niveau du candidat en « réalisation logiciel » :

$$\text{Niv} = (70 * 0.1) + (50 * 0.4) + (50 * 0.05) + (20 * 0.05) + (50 * 0.05) + (50 * 0.15) + (20 * 0.2) = 44.5\% \cong A \text{ (Application)}$$

E Supposons que ce sont les seules compétences :

- Poids_tot = 6 (cas d'une compétence spécifique)
- K = entier ((70 - 50) / 20) = 1
- Poids_calc = 6 * (1 - 1/4) = 4.5
- **D_appari** = (4.5 / 6) * 100 = 75% (degré d'appariement)

Appariement sémantique entre CVs et Offres

✓ Aspects implémentatoires : *RACER*

Un système de représentation de connaissances, qui implémente un calcul de tableaux hautement optimisé pour une LD très expressive.

E *Au niveau terminologique* : différentes possibilités

- Vérifier la consistance d'un concept ⇒ validation de l'ontologie
- Contrôler les relations de subsomption des concepts ⇒ validation d'ontologie
- Déterminer les descendants/parents d'un concept ⇒ parcourir hiérarchie «Thème»

E *Au niveau assertional* : autres requêtes

- Vérifier la consistance et tester les instances ⇒ validation ontologie
- Calculer le type direct d'un individu/instance ⇒ implémenter les tests de type
- Extraire les instances d'une classe particulière, selon des critères (RQL) ⇒ implémentation des fonctions d'extraction

RACER








processus d'appariement = traitement de requêtes, basé sur un raisonnement ontologique





Etude de cas

✓ Le rapprochement entre les diplômes et les métiers :

E 5 postes de travail (différents métiers) :

-  Expert système d'exploitation
-  Technicien réseaux/télécoms
-  Développeur (analyste programmeur)
-  Gestionnaire d'applications
-  Administrateur de BDDs

E 4 candidats (différents diplômes) : 4 CVs

-  Licence académique
-  Licence professionnelle GL
-  Licence professionnelle SI
-  Licence professionnelle STIC

Etude de cas (Poids des sous-thèmes)

Algorithmique & programmation 10	Abstraction de données 10	
	Programmation 50	Techniques de programmation 90
		Programmation événementielle 5
		Programmation Fonctionnelle 5
		Principes prog fonctionnelle 50
		Langage Fonctionnelle 50
		Lisp 40
		ML 40
		Autre LF 20
		Programmation Logique 10
Principes prog logique 50		
Langage logique 50		
Prolog 40		
Autre LL 20		
Programmation OO 50		
Principes prog OO 50		
Langage de classes 40		
C++ 60		
JAVA 60		
Autre LO 30		
Langage de frames 5		
Langage d'acteurs 5		
Programmation procédurale 30		
Principes prog procédurale 50		
Langage procédural 50		
C 50		
Pascal 50		
Autre LP 30		
Sémantiques des langages de Programmation 10		
Sémantique axiomatique 30		
Sémantique dénotationnelle 40		
Sémantique opérationnelle 30		
Thème de recherche en Algo/programmation 10		
Architecture machine 10	Analyse performances machine 5	
	composants matériels 20	Connaissance composante matérielle 40
		Connectique 30
		Entretien installations techniques 20
		Systèmes de câblage 10
	Architecture fondamentale de l'ordinateur 25	
	Interfaçage matériel/logiciel 20	
	La microprogrammation 5	
	Types architecture machine 10	
	Intégration de matériel 5	

Etude de cas (compétences sous-jacents aux métiers)

Algorithmique & programmation	Chapitre et sous-chapitre		
	Abstraction de données		
	Programmation	Techniques de programmation	Programmation événementielle Programmation fonctionnelle Programmation logique Programmation OO Programmation procédurale
		Sémantiques des langages de programmation	Sémantique axiomatique
			Sémantique dénotationnelle
			Sémantique opérationnelle
		Thème de recherche en Algo/programmation	
	Architecture machine	Analyse performances machine (A)	
		composants matériels	Connaissance composante matérielle (M)
			Connectique (E)
Entretien installations techniques			
Systèmes de câblage (E)			
Architecture fondamentale de l'ordinateur (M)			
Interfaçage matériel/logiciel (F)			
La microprogrammation			
Types architecture machine (A)			
Intégration de matériel (M)			
Thème de recherche en architecture machine			

Etude de cas (compétences sous-jacents aux diplômes)

Algorithmique & programmation		Abstraction de données	A	A	A	A	
	Programmation	Techniques de programmation	Programmation événementielle	N	N	A	A
			Programmation Fonctionnelle	Principes: prog fonctionnelle A			
				Langage Fonctionnel	Lisp A	ML A	
					Axiom LP		
		Programmation logique	Principes: prog logique A A A A				
			Langage logique	Prolog M N A M			
				Axiom LL			
		Programmation OO	Principes: prog OO A M M M				
			Langage de classes	C++ N A A N			
				JAVA A M A M			
	Axiom LO						
		Langage de frames		A	N		
	Langage d'acteurs		A	N			
Programmation procédurale	Principes: prog procédurale E E M M						
	Langage procédural	C M M A A					
		Pascal M M M M					
Axiom LP A A A A							
Sémantiques des langages program	Sémantique axiomatique		N				
	Sémantique dénotationnelle		N				
	Sémantique opérationnelle		N	N			
Thème de recherche en Algo./programmation							
Architecture machine	Analyse performances machine						N N
	composants matériels	Connaissance composante matérielle		A	A	A	A
		Connectique		N	N	N	A
		Entretien installations techniques					
		Systèmes de câblage					
	Architecture fondamentale de l'ordinateur			A	A	A	A
	Interfaçage matériel/logiciel			M	M	A	A
	La microprogrammation			A	N		
	Types architecture machine			A	A	N	N
	Estimation de matériel						

Etude de cas (résultat du calcul des degrés d'appariement)

	Technicien R/T	Gestionnaire applications	Développeur	Administrateur BDDs	Expert SE
Licence académique	60,50%	61,36%	59,32%	54,38%	42,60%
Licence GL	55,63%	67,74%	80,79%	55,21%	37,00%
Licence SI	57,60%	90,66%	72,60%	73,18%	38,80%
Licence STIC	76,71%	66,41%	70,48%	65,33%	48,40%



Conclusion et perspectives

- Proposition d'une approche simple et suffisamment complète, basée annotation sémantique, pour automatiser le rapprochement entre CVs et offres d'emploi
- Modélisation du contenu sémantique à base d'un référentiel partagé
- Une gestion basée compétence, assurée par une modélisation formelle de cette notion.
- Deux techniques de rapprochement (originaux) différentes mais complémentaires
- La possibilité de rapprocher les compétences des nouveaux diplômés, à celles exigées par les métiers du domaine.
- Une Implémentation assurée de l'approche (standards et technologies du WS)
- Le travail futur vise à
 - ∅ Implémenter les interfaces du systèmes et les algorithmes d'appariements proposés
 - ∅ Valider cette approche sur des données réelles
 - ∅ Généraliser l'ontologie construite à d'autres domaines.



Merçi