Distributed artificial intelligence and knowledge management:

ontologies and multi-agent systems for a corporate semantic web

Fabien GANDON - ACACIA Team INRIA

University of Nice - Sophia Antipolis
Ph.D. defence, Thursday the 7th of November 2002

Fabien.Gandon@sophia.inria.fr

Introduction and plan of the defence

Introduction

- **Domain** of application:
  - Organisations **need** to adapt to an ever changing world
  - **Nervous system**: capture and diffuse knowledge
  - **Persistent memory**: store and/or index knowledge
  - Problematics of **corporate memory**

You are here

Plan

- **Positioning**:
  - Concerned **domains of research**
  - Application context: the **CoMMA** IST Project
- Thesis **contributions**:
  - **Corporate semantic web** & **ontology engineering**
  - **Multi-agent** architecture & **distribution of annotations**
- Evaluation, discussion, conclusions, perspectives

Multidisciplinary facets of my work

Corporate / organisational **memory** in this work:

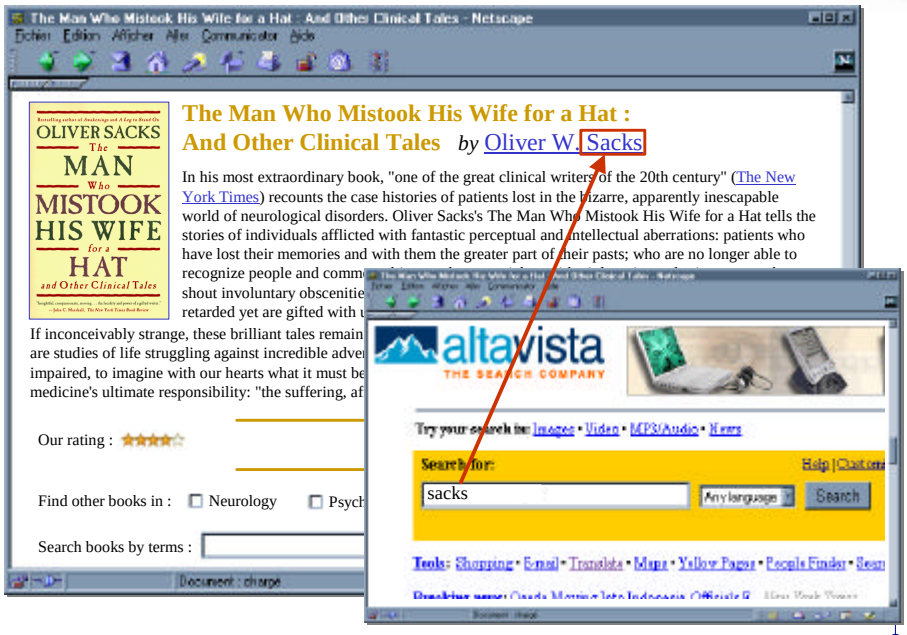
definition

explicit, disembodied, persistent representation and indexing of knowledge and information or their sources in an organisation, in order to facilitate their access, share and reuse by members of the organisation, for their individual and collective tasks.

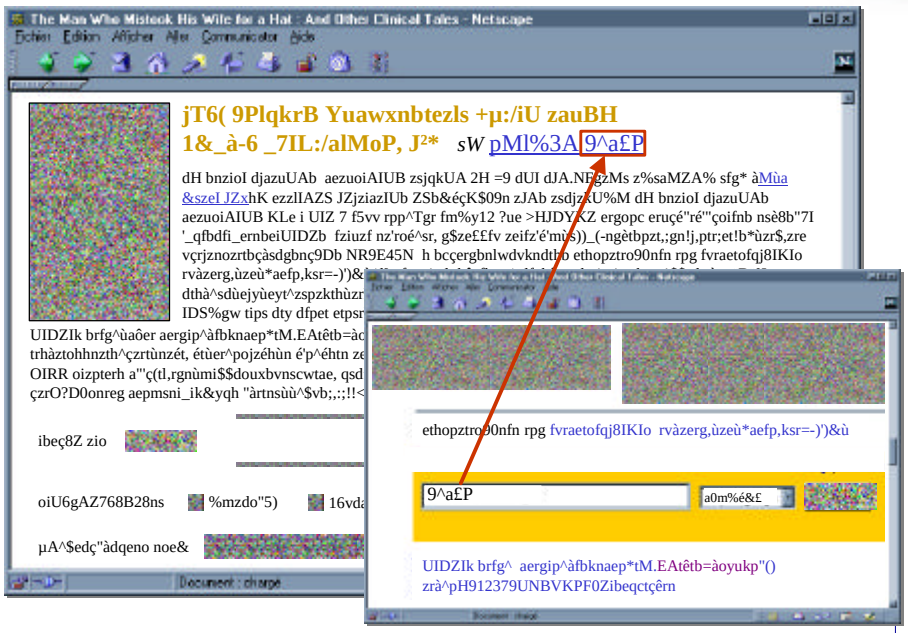
Facets of **my research**:

- Knowledge **management** (capitalise, share, create)
- Information **retrieval** (access right information at right time)
- Knowledge **engineering** (knowledge acquisition & representation)
- **Distributed** system (assist local and global tasks)

The Web to humans



The Web to computers...



Positioning (semantic Web & ontology)

- Internet & **Web** (shared infosphere *for humans*)
- Intranets & **intrawebs**
  - Current trend: **reuse** internet and **web** technologies
  - Same advantages and same **drawbacks**
- **Semantic Web** (shared infosphere *for machines*)
  - **XML**: W3C standard for **structuring** data and documents
  - **RDF(S)**: W3C standard for metadata / **semantic annotations**

Hyp: corporate memories as **semantic intrawebs**

- Knowledge engineering & **ontology** (modularity, reuse)
  - **Assertional** knowledge e.g., "*Hugo* wrote *Notre-Dame de Paris*"
  - **Ontological** knowledge e.g., "*Authors* write *books*"

definition

an **explicit**, **partial account** of the **semantic structure** encoding the rules that constrain our representation of **reality**

Hyp: **ontology** to support semantic intraweb

Positioning (DAI) and context (CoMMA)

Distributed artificial intelligence

- Multi-agent systems: study design of **artificial societies** of intelligent agents

definition

clearly **identifiable artificial entity**; **situated** in an environment which it **senses**, **reacts** to and **acts** upon; **social abilities** to interact with other agents / humans; **self-control** of its behaviour

- Multi-agent information systems: **situated** in & distributed over an **information network**

Hyp: **multi-agent system** to manage semantic intraweb

CoMMA **Co**rporate **M**emory **M**anagement through **A**gents

Two application and trial **scenarios**:

- Assist **new employee** integration
- Support **technology monitoring** activities

Positioning and pointers

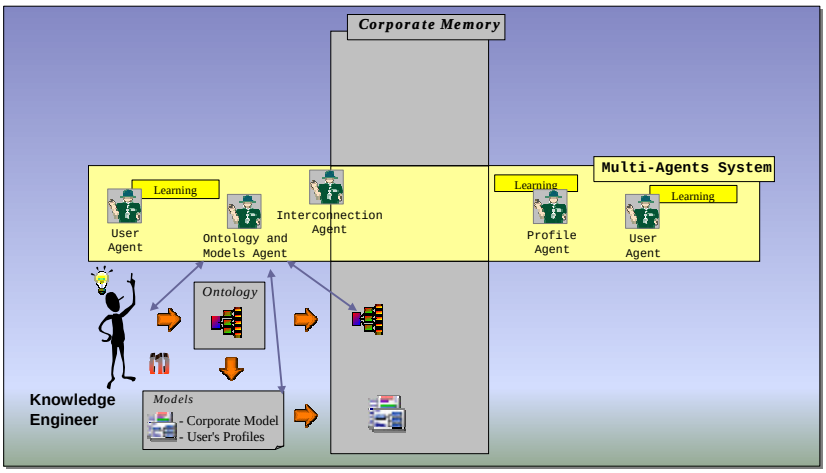
Dynamically **integrating heterogeneous sources** of information OBSERVER [Mena *et al.*, 1996] InfoSleuth [Nodine *et al.*, 1999] Carnot [Collet *et al.*, 1991] InfoMaster [Genesereth *et al.*, 1997] SIMS [Arens *et al.*, 1996] RETSINA [Decker & Sycara, 1997] Manifold [Kirk *et al.*,1995]

Assist the management of digital **libraries**
SAIRE [Odubiyi *et al.*, 1997] UMDL [Weinstein *et al.*, 1999]

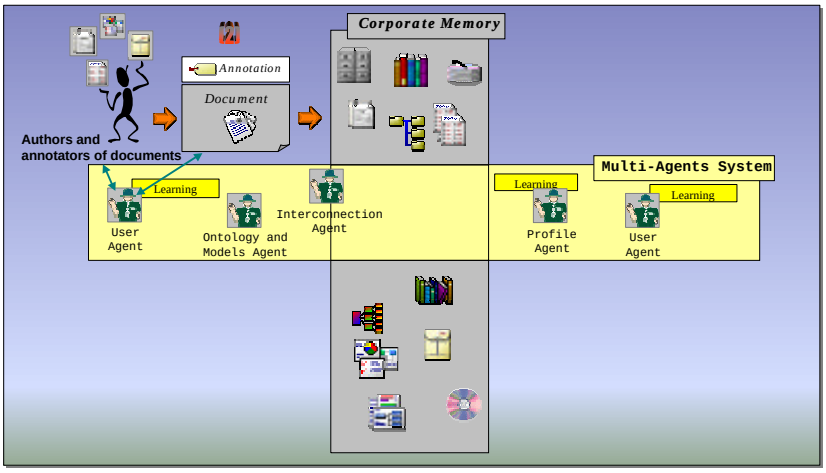
Organisational knowledge management:

- **Collaborative** gathering, filtering and profiling
  CASMIR [Berney & Ferneley, 1999] Ricochet [Bothorel & Thomas, 1999]
- **Mobile** access & domain model for **document classification**
  KnowWeb [Dzbor *et al.*, 2000]
- **Taxonomy** of topics, profiling and **push** RICA [Aguirre *et al.*, 2000]
- **Ontology** and **corporate memory**:
  multiple ontologies FRODO [Van Elst & Abecker, 2001]
  semantic intraweb, ontology, user profiling **CoMMA**

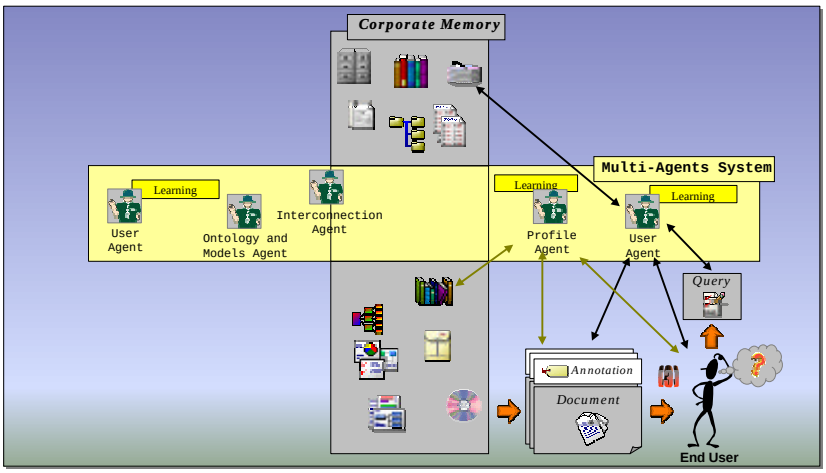
Overall schema: knowledge modelling



Overall schema: populating the memory

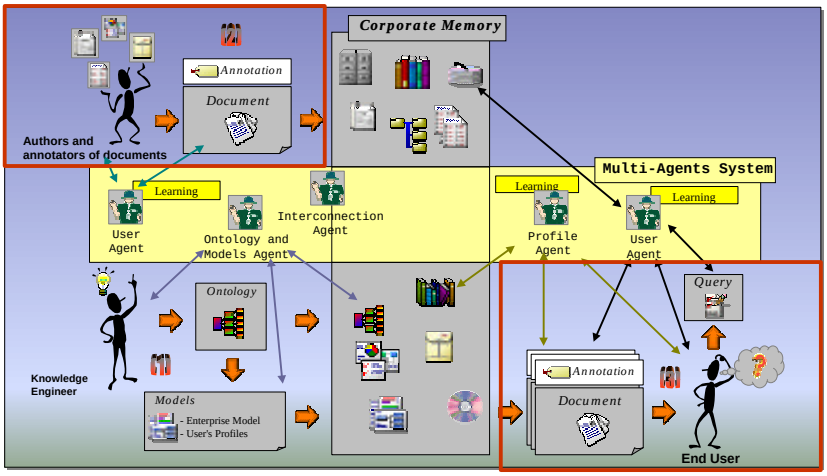


Overall schema: querying the memory



Functionalities of CoMMA: annotate, pull and push

annotation



pull and push

Positioning and focusing on my contribution

Technical choices:

- **Materialisation** of memory
  RDF(S) and its XML syntax (manipulated with **CORESE**)
- **Exploitation** of memory
  Multi-agent system and machine learning techniques
  (implemented with) (implemented with **WEKA**)

Plan

Positioning: ✓

- Concerned **domains of research** ✓
- Application context: the **CoMMA** IST Project ✓

Thesis **contributions**:

- **Corporate semantic web** & **ontology engineering** ←
- **Multi-agent** architecture & **distribution of annotations**

Evaluation, discussion, conclusions, perspectives

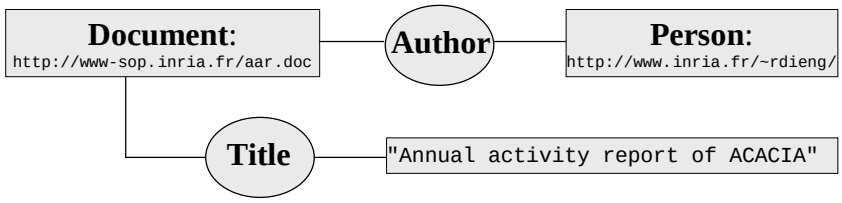
Annotate documents (content awareness)

RDF(S): annotated world for software to make inferences & help users to exploit corporate memory

- RDF = Resource Description Framework (annotation model)
  - annotation:

```
<INRIA:Document rdf:about="http://www-sop.inria.fr/aar.doc">
 <INRIA:Title>Annual activity report of ACACIA</INRIA:Title>
 <INRIA:Author>
  <INRIA:Person rdf:about="http://www.inria.fr/~rdieng/" />
 </INRIA:Author>
</INRIA:Document>
```

  - meaning:



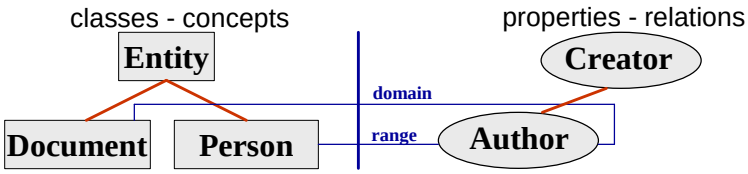
Annotation schema (semantic awareness)

RDF(S): annotated world for software to make inferences & help users to exploit corporate memory

- RDF = Resource Description Framework (annotation model)
- RDFS = RDF Schema (annotation vocabulary / ontology)
  - schema: (e.g. `http://www.inria.fr/schema#`)

```
<rdfs:Class rdf:ID='Entity'/>
<rdfs:Class rdf:ID='Document'>
  <rdfs:subClassOf rdf:resource='#Entity' />
</rdfs:Class>    ...
<rdf:Property rdf:ID='Author'>
 <rdfs:subPropertyOf rdf:resource='#Creator' />
  <rdfs:domain rdf:resource='#Document' />
  <rdfs:range rdf:resource='#Person' />
</rdf:Property>  ...
```

  - meaning:



Annotate persons (context awareness)

User modelling - "annotating persons"

extract of my user profile:

```
<CoMMA:Engineer rdf:about="http://www-sop.inria.fr/acacia/personnel/ Fabien.Gandon/
 <CoMMA:FamilyName>GANDON</CoMMA:FamilyName>
 <CoMMA:FirstName>Fabien</CoMMA:FirstName>
 <CoMMA:BirthDate>31-07-1975</CoMMA:BirthDate>
  (...)
 <CoMMA:HasForWorkInterest>
  <CoMMA:MultiAgentSystemTopic rdf:about="http://www.inria.fr/acacia/comma#...
 </CoMMA: HasForWorkInterest >
  (...)
 <CoMMA:HasForPersonalInterest>
  <CoMMA:HumanScienceTopic rdf:about="http://www.inria.fr/acacia/comma#Human...
 </CoMMA:HasForPersonalInterest>
  (...)
</CoMMA:Engineer>

<CoMMA:Employee rdf:about="http://www-sop.inria.fr/acacia/personnel/ Fabien.Gandon/
 <CoMMA:HireDate>1999-11-02</CoMMA:HireDate>
 <CoMMA:EmployedBy>
  <CoMMA:LocalOrganizationGroup rdf:about="http://www.ac-nice.fr/" />
 </CoMMA:EmployedBy>
 <CoMMA:EmploymentContract> <CoMMA:Temporary/> </CoMMA:EmploymentContract>
 <CoMMA:HasForActivity>
    <CoMMA:Research rdf:about="http://www.inria.fr/acacia/comma#Research"/>
 </CoMMA:HasForActivity>
 (...)
</CoMMA:Employee>
```

Annotate organisation (context awareness)

Corporate modelling - "annotating organisations"

extract of INRIA model:

```
<CoMMA:ProjectGroup rdf:about="http://www-sop.inria.fr/acacia/">
 <CoMMA:Designation>Projet Acquisition des Connaissances pour ...
 <CoMMA:HasForActivity> <CoMMA:Research rdf:about="http://www...
 (...)
 <CoMMA:IsInterestedBy>
  <CoMMA:KnowledgeEngineeringTopic rdf:about="http://www.inria.fr/acacia/com...
 </CoMMA:IsInterestedBy>
 (...)
 <CoMMA:Situated>
  <CoMMA:Location>
   <CoMMA:PhoneNumber>(33) (0)4 92 38 77 80</CoMMA:PhoneNumber>
   <CoMMA:FaxNumber>(33) (0)4 92 38 77 83</CoMMA:FaxNumber>
  </CoMMA:Location>
 </CoMMA:Situated>
 <CoMMA:Include>
  <CoMMA:PhDStudent rdf:about=" http://www-sop.inria.fr/acacia/personnel/ Fab...
 </CoMMA:Include>
 (...)
</CoMMA:ProjectGroup>

<CoMMA:LocalOrganizationGroup rdf:about="http://www-sop.inria.fr/">
 <CoMMA:Designation>UR Sophia Antipolis de l'INRIA: Institut National de ...
 <CoMMA:Include><CoMMA:ProjectGroup rdf:about="http://www-sop.inria.fr/acacia...
 (...)
```


Methodological steps in ontology engineering (1 & 2)

Ontology building in five steps

Step 1 - **Data collection** and **analysis**

- Scenario-driven analysis: scenario **reports** from users & grid
- Motivate data collection **internal** and **external** to organisation
- Capture aspects of **conceptualisation** to assist scenarios

extracts

"... wonder if there are **technical reports** about **UMTS**, then I..."

"... what this **manager** or one of his **colleagues** wrote for..."

Step 2 - Build a **lexicon**

- Capture **terms** and their **definitions**
- First **intermediary representation** of the ontology
- Constraint: **one** and only one occurrence of a definition
- **Disambiguate** terms, e.g.:

definition

COLLEAGUE n. (lat. *collega*) someone who shares the same profession || one of a group of people who work together.

Step 3 - Enriching lexicon structure

- **Split** concepts, properties and attributes into different tables
- Augment with **relevant semantic** aspects (e.g. subsumption)
- Enrich, augment, refine, ... for both **humans** and **machines**
- **Taxonomic** skeleton: top-down / bottom-up / middle-out

extracts

Class	View	Super class	Other Terms	Natural Language Definition	Pr
Manager	Organization; Person;	Professional;	director;	**Professional** whose primary job is to manage other people, directing their work activity. A Manager tells his or her subordinate workers what to do.	Cy
...	...	...	...	...	...
UMTS	Domain;	MobilePhone Protocols;	U.M.T.S.; universal mobile telecommunications system	**Mobile phone protocol** of the 3G technology that delivers broadband information at speeds up to 2Mbit s/sec.	Us
...	...	...	...	...	...
Technical Report	Document;	Report;	;	**Report** presenting technical details on a specific topic.	Us
...	...	...	...	...	...

Relation	Domain	Range	View	Super Relation	Other Terms	Natural Language Definition	Sy	Tr	Re	Pr
Colleague	Person	Person	Organisation;	Acquaintance;	co-worker;	Acquaintance between two persons who work together	X			Us
...	...	...	...	...	...	...	...	...	...	...

Step 4 - Script translating tables into RDFS

Example on a **primitive concept**:
e.g. *UMTS* sub-topic of *Mobile phone protocols*

code

```
UMTSTopic(x)⊃ MobilePhoneProtocolsTopic(x)
```

```
<rdfs:Class rdf:ID="UMTSTopic">
<rdfs:subClassOf rdf:resource="#MobilePhoneProtocolsTopic"/>
<rdfs:comment xml:lang="en"/>
<rdfs:comment xml:lang="fr"/>
<rdfs:label xml:lang="en">UMTS</rdfs:label>
<rdfs:label xml:lang="en">U.M.T.S.</rdfs:label>
<rdfs:label xml:lang="en">universal mobile
 telecommunications system</rdfs:label>
<rdfs:label xml:lang="fr">UMTS</rdfs:label>
<rdfs:label xml:lang="fr">U.M.T.S.</rdfs:label>
<rdfs:label xml:lang="fr">systeme universel de
 telecommunications mobiles</rdfs:label>
</rdfs:Class>
```

Browsing the ontology through stylesheets

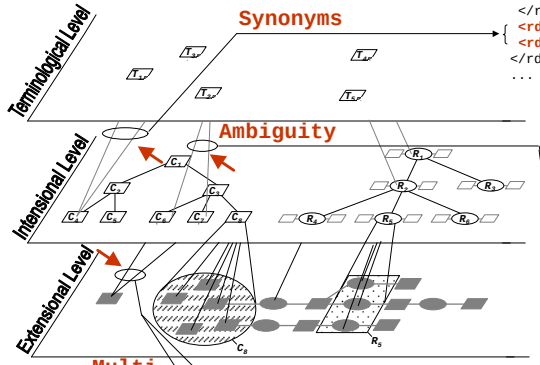
XSLT:

- Visualise during **design** and with **end-users**
- An advantage of using **XML**
- **Reuse**: independent of ontology

snapshot

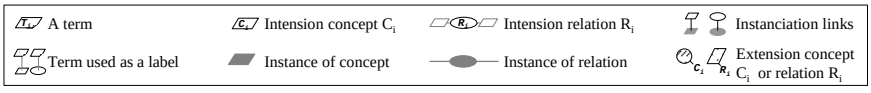
Analysis of the three levels present in RDF(S)

Up to that step RDF(S) has the required expressiveness




```
...
<rdfs:Class rdf:ID="Department">
 <rdfs:subClassOf
rdf:resource="#OrganizationPart"/>
 <rdfs:comment xml:lang="en">
 Organization part which is a sub division of a
 Research Direction, corresponding to sub interest
 field e.g mobile transmission
 </rdfs:comment>
 <rdfs:label xml:lang="en">department</rdfs:label>
 <rdfs:label xml:lang="en">division</rdfs:label>
</rdfs:Class>
...
<rdfs:Class rdf:ID="E-Mail">
 <rdfs:subClassOf rdf:resource="#Mail"/>
 <rdfs:comment xml:lang="en">
   Mail sent in electronic format over a
   computerized world-wide communication
system
 </rdfs:comment>
 <rdfs:label xml:lang="en">e-
mail</rdfs:label>
 <rdfs:label xml:lang="en">electronic mail
                          </rdfs:label>
 <rdfs:label xml:lang="en">mail</rdfs:label>
</rdfs:Class>
...
<rdfs:Class rdf:ID="PostMail">
 <rdfs:subClassOf rdf:resource="#Mail"/>
 <rdfs:comment xml:lang="en">
   Mail transmitted via the post office
 </rdfs:comment>
 <rdfs:label xml:lang="en">mail</rdfs:label>
 <rdfs:label xml:lang="en">post
mail</rdfs:label>
</rdfs:Class>
...
```

```
<CoMMA:Memo rdf:about="http://www.inria.fr/comma.rtf"/>
<CoMMA:E-Mail rdf:about="http://www.inria.fr/comma.rtf"/>
```



Step 5 - Factorising knowledge (when needed)

- Declare **algebraic properties** of relations (symmetric / transitive / reflexive relations)

code

```
colleague(x,y)Þ  acquaintance(x,y) ∧ colleague(y,x)
```

```
<rdf:Property rdf:ID="Colleague">
 <rdfs:subPropertyOf rdf:resource="#Acquaintance"/>
 <rdfs:range rdf:resource="#Person"/>
 <rdfs:domain rdf:resource="#Person"/>
 <cos:symmetric>true</cos:symmetric>
 <rdfs:comment xml:lang="en">acquaintance  between two
persons who work together</rdfs:comment>
 <rdfs:comment xml:lang="fr">accointance entre deux personnes
travaillant ensemble.</rdfs:comment>
 <rdfs:label xml:lang="en">colleague</rdfs:label>
 <rdfs:label xml:lang="en">co-worker</rdfs:label>
 <rdfs:label xml:lang="fr">collegue</rdfs:label>
 <rdfs:label xml:lang="fr">collegue de travail</rdfs:label>
</rdf:Property>
```


Step 5 - Factorising knowledge (when needed)

- No one generates all the instances of *colleague* **by hand**
- "I am a colleague of X because I work in the same group as X" (inference)
- Encode axiomatic knowledge in **rules** and **definitions**

code

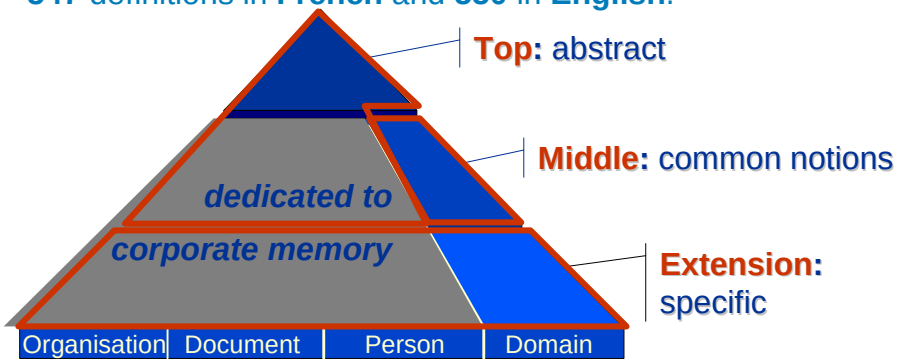
```
colleague(x,y)Û  person(x) ∧ person(y) ∧
                 ($z group(z) Ù include(z,x) Ù include(z,y))
```

```
IF                                (rule for sufficient condition)
     Group
      Include
        Person ?x
      Include
        Person ?y
THEN
      Person ?x
        Colleague
          Person ?y
```

- **Coverage** of ontology: **exhaustivity**, **specificity**, and **granularity** against usage scenarios

Resulting ontology: **O'CoMMA**

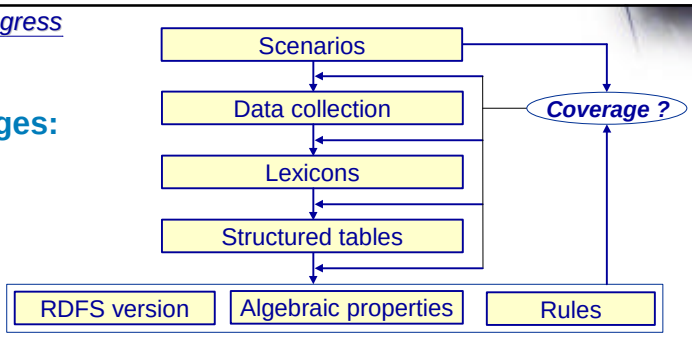
- **470 concepts** (taxonomy depth = 13 subsumption links).
- **79 relations** (taxonomy depth = 2 subsumption links).
- **715** terms in **English** and **699** in **French**.
- **547** definitions in **French** and **550** in **English**.



- Abstract top & middle layer for corporate memory: reusable
- Middle layer for domain: reusable in same domain
- Extension layer: usable but not reusable
- Reuse e.g., CSELT→CSTB, APROBATIOM, KMP

Summary

- Design **stages:**



- In CoMMA this method provided **O'CoMMA**

Plan

Positioning: ✓

- Concerned **domains of research** ✓
- Application context: the **CoMMA** IST Project ✓

Thesis **contributions**:

- **Corporate semantic web** & **ontology engineering** ✓
- **Multi-agent** architecture & **distribution of annotations** ←

Evaluation, discussion, conclusions, perspectives

A multi-agent architecture for CoMMA

Problem = handle information **distribution**

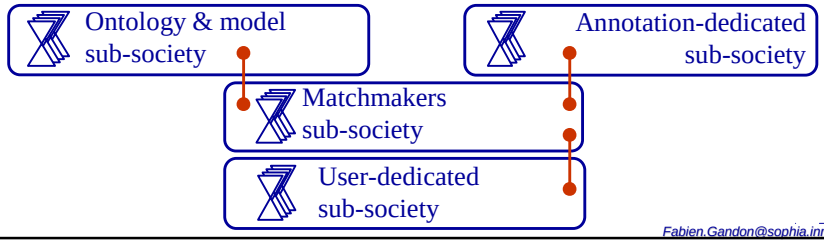
- Handle fact information & knowledge are **naturally scattered**
- Assist **diffusion** of captured information and knowledge

Follow **multi-agent** paradigm:

- **Artificial societies** collaborating for global capitalisation
- **Artificial individual intelligence**, able to locally adapt

Step 1 - **Sub-societies** identification

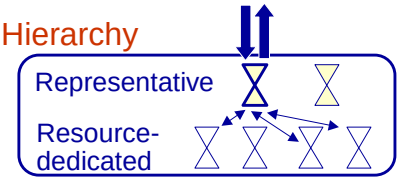
- Started from the **tasks to be performed** (provide ontology, manage annotations, manage users and matchmaking)
- Thus four **sub-societies** to handle these four tasks:



Organising resource-dedicated sub-societies

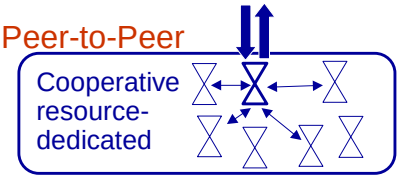
Step 2 - Analyse possible organisation **structures**:

Hierarchy



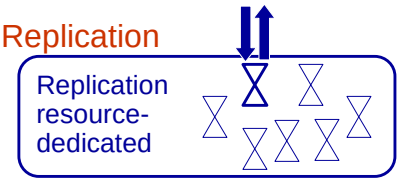
Distribute roles & resource +
distribute workload +
increase network load –

Peer-to-Peer



Distribute resource only
workload & network load +/-
redundancy +

Replication



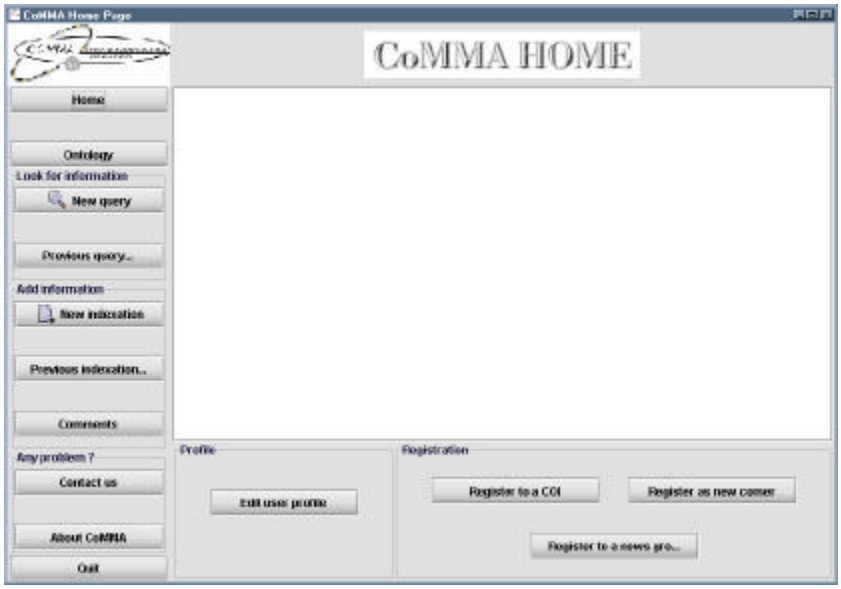
Complete replication
redundant & less network load +
content replicated –

Sub-societies of CoMMA

- Ontology & model society: **replication** — ACACIA
  - **Ontologist**: to **store** and **provide** ontology
  - **Corporate model archivist**: to **store** and **provide** structural model of the (human) organisation
- Annotations society: **hierarchy** **(mediator/archivist)** — ACACIA
- Matchmakers society: **peer-to-peer** — University of Parma
  - **Directory facilitator**: yellow pages service
  - Federable co-operating matchmakers
  - **Agent management system**: white pages service
- User-dedicated sub-society: **three main roles**
  - Handle the users' profile:
    - **User profile manager**: **machine learning** techniques — LIRMM
    - **User profile archivist**: **archiving** and **secured login** — ATOS
  - **Interface controller**: graphical user interface... — ATOS

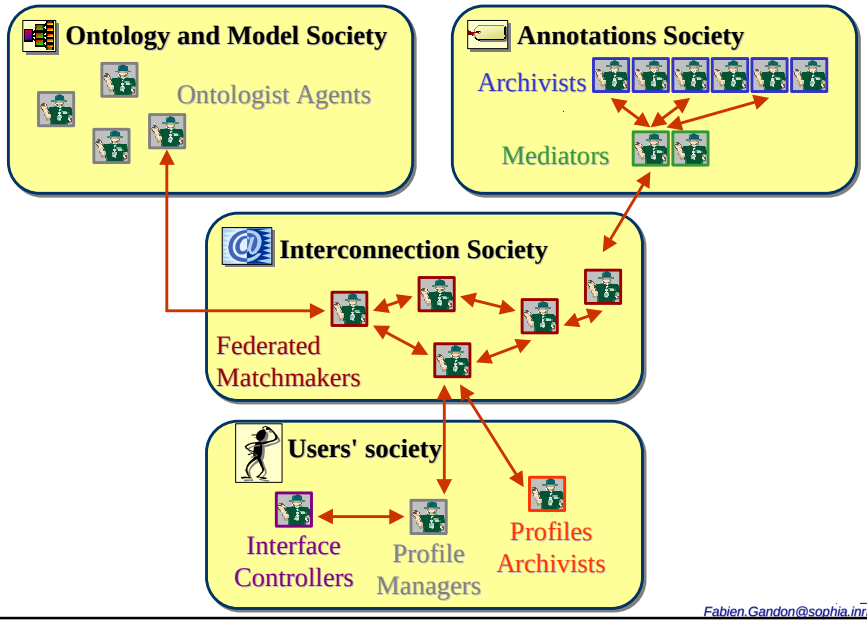
Interface Agent: ontology-guided query on the corporate semantic web

snapshot



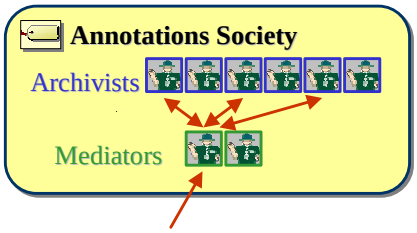
specified by INRIA, implemented by ATOS-Origin

Step 3 - Roles and interactions



The **annotation**-dedicated sub-society

- Mediators & archivists
- Manage local sources of annotations distributed over the organization network



Propose annotation-related services to other agents

- **Archive** new annotations on documents of the memory
- **Search** & retrieve references matching queries
- **Notify** registered agents of the arrival of a new annotation

Present 2 **roles** and then 2 interaction **protocols**

Annotation-dedicated sub-society (mediator)

Mediator role (supervises social interactions)

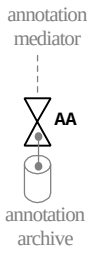
role model	Annotation Mediator role in the Annotation-dedicated society
responsibilities	handle distribution of annotations over the archivists both for new annotation submissions and query solving processes
collaborators	Directory facilitator, User Profile Manager, Ontology Archivist, Annotation Archivist, Corporate Model Archivist
external interfaces	RDF annotation manipulation interface
relationships	-
expertise	query and submission management
interactions	Query-Ref, Contract-Net, Subscribe, Request; FIPA ACL
others	-

- **Contact point** for other sub-societies
- Supervising **distribution** of tasks for query solving
- **Allocating** a new annotation to an archive
- **Notifying** arrival new annotations to trigger push functions

Annotation-dedicated sub-society (archivist)

Archivist role (manage a local archive)

role model	Annotation Archivist role in the Annotation-dedicated society
responsibilities	store and query the annotations of the memory
collaborators	Directory facilitator, Annotation Mediator, Ontology Archivist
external interfaces	RDF annotation manipulation interface
relationships	also part of the roles in Corporate Model Archivist and User Profile Archivist
expertise	annotation archiving an querying
interactions	Query-Ref, Contract-Net; FIPA ACL
others	-



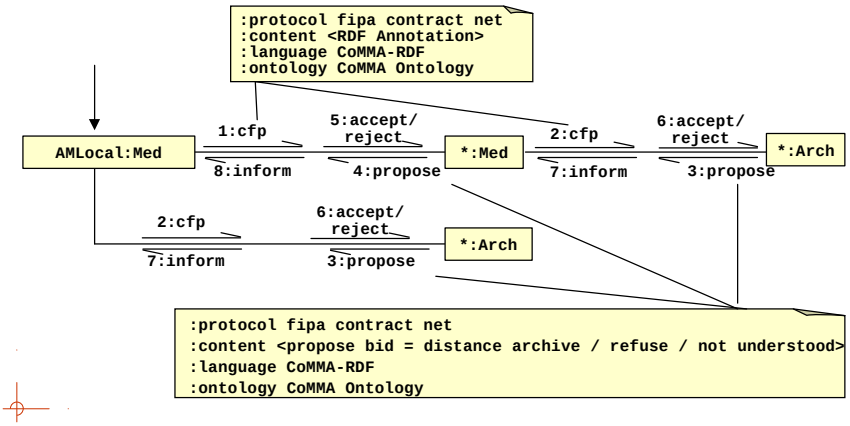
- Attached to & exploits **local base**
- **Answers** to query as much as it can with local knowledge
- Proposes **archiving** services advertising the archive content

Interactions mediator-archivist in allocating an annotation: contract net

Interaction 1 - Annotation allocation

- **Pb:** archives of annotation **distributed** all over organisation
- Mediator & archivists discuss **best archive** for new one
- **Contract-net** (CfP, Proposal, Accept/Reject):

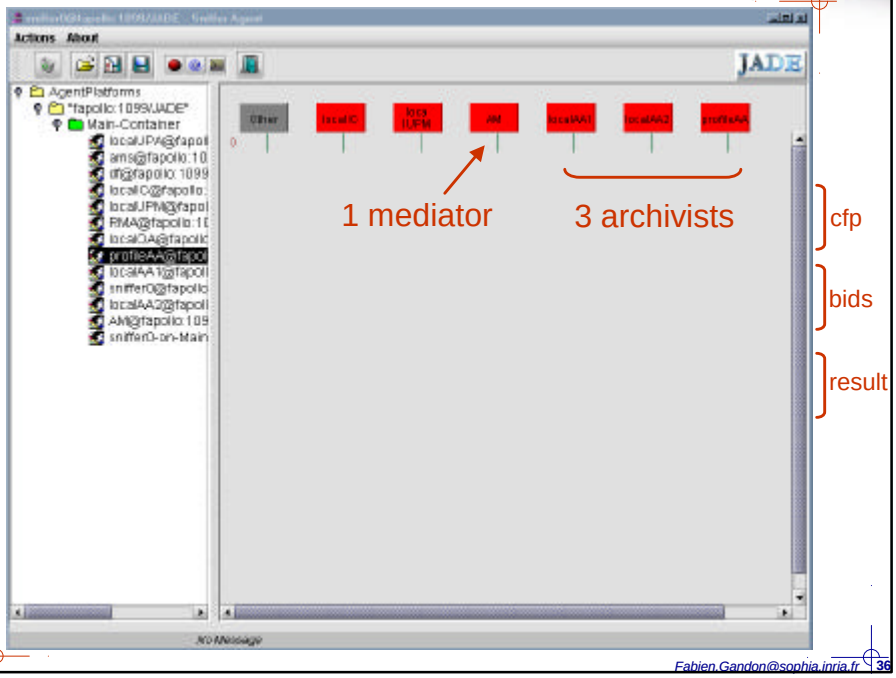
protocol



- Proposals: **semantic distance** new annotation - archive

Message traces for an annotation allocation:

snapshot

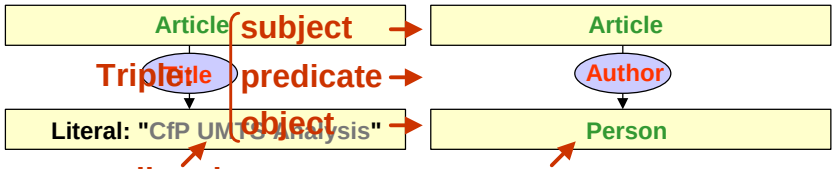


Content of an annotation or an archive

Simple **annotation**

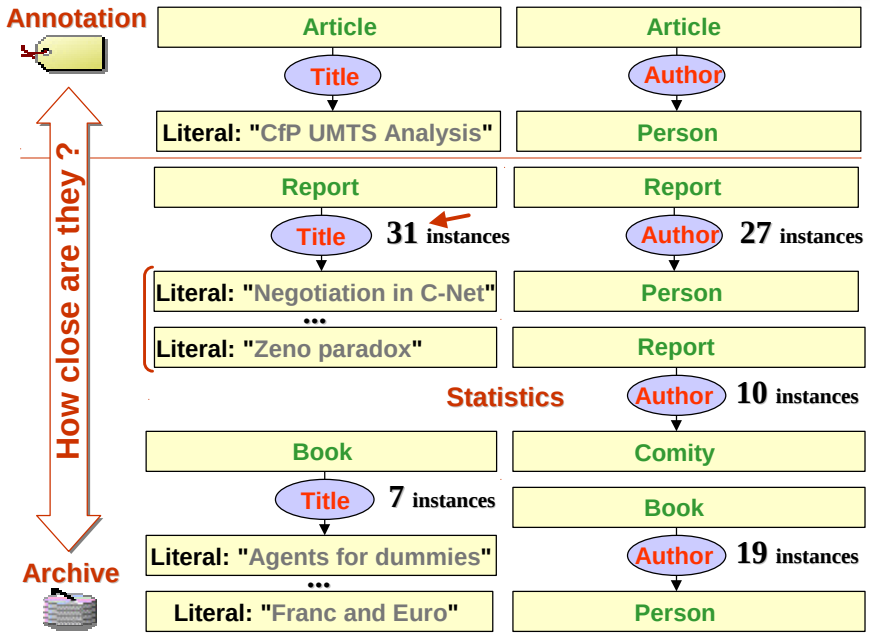
```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:CoMMA="http://www.inria.fr/acacia/comma#">
 <CoMMA:Article rdf:about="http://intranet/reports/R3029">
  <CoMMA:Title>CfP UMTS Analysis</CoMMA:Title>
  <CoMMA:Author>
   <CoMMA:Person rdf:about="http://www.mycorp.com/~fab" />
  </CoMMA:Author>
 </CoMMA:Article>
</rdf:RDF>
```

Corresponding **triples**:



Unstructured set of triple to describe **content** type

Distance between an annotation and an archive



Building the distance

Case 1 - **literal** values

- Lexicographical distance $\mathrm{Dist}_L(\mathrm{Lit}_A, \mathrm{Lit}_B) = \left| \sum_{i=0..s} \frac{C_{B,i} - C_{A,i}}{\mathrm{Max}_L^{\,i}} \right|$
- Distance literal (from annotation) - interval (from archive)

$\mathrm{Dist}_I(\mathrm{Lit}_X, [B_{low}, B_{up}])$

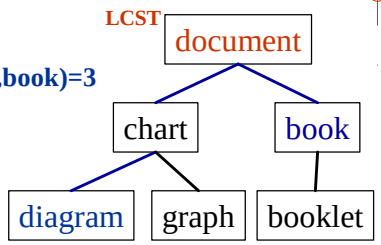
if $\mathrm{Lit}_X \in [B_{low}, B_{up}]$ ***then*** $= 0$

else $= \mathrm{Min}(\mathrm{Dist}_L(\mathrm{Lit}_X, B_{low}), \mathrm{Dist}_L(\mathrm{Lit}_X, B_{up}))$

Case 2 - **concept** types

extract 2 from O'CoMMA

Distance(diagram,book)=3



Semantic distance between types

Distance between **two triples** (conditional weighted sum)

$\mathrm{Dist}_{TFStat}(\mathrm{Triple}_A, \mathrm{Triple}_B) = \mathrm{Dist}_{C1} + \mathrm{Dist}_R + \mathrm{Dist}_{C2}$

$\mathrm{Dist}_{Ci} = W_C * \mathrm{Dist}_H(\mathrm{Type}_1, \mathrm{Type}_2)$ or $W_C * \mathrm{Dist}_H(\mathrm{Type}, \mathrm{Lit})$

or $W_L * N * \mathrm{Dist}_I(\mathrm{Lit}, [B_{low}, B_{up}])$ with $N = \mathrm{Max}_C * 2 / \mathrm{Max}_L$

Distance **annotation « archive**

- triple↔stat: $\mathrm{Dist}_{TStat}(\mathrm{Triple}, \mathrm{Stat}) = \mathrm{Min}_{\mathit{Triple}_i \hat{I}\ \mathit{Stat}}(\mathrm{Dist}_{TFStat}(\mathrm{Triple}, \mathrm{Triple}_i))$
- annotation↔stat: $\mathrm{Dist}_{AStat}(\mathrm{An}_X, \mathrm{Stat}) = \sum_{\mathit{Triple}_j \in \mathit{An}_X} \mathrm{Dist}_{TStat}(\mathrm{Triple}_j, \mathrm{Stat})$
- annotation↔pref: $\mathrm{Dist}_{APref}(\mathrm{An}_X, \mathrm{Pref}) = \sum_{\mathit{Triple}_j \in \mathit{An}_X} \mathrm{Dist}_{TPref}(\mathrm{Triple}_j, \mathrm{Pref})$

  sub-type $\Rightarrow \mathrm{Dist}_H = 0$
- annotation↔archive:

$\mathrm{Dist}(\mathrm{An}_X, \mathrm{Arc}_Y) = \mathrm{Dist}_{AStat}(\mathrm{An}_X, \mathrm{Stat}_Y) + \mathrm{Dist}_{APref}(\mathrm{An}_X, \mathrm{Pref}_Y)$

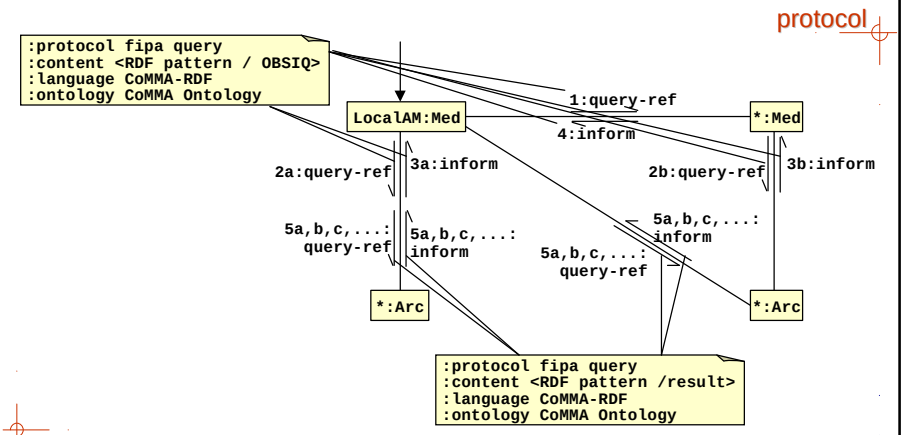
Distance = allocation **criteria** of contract-net

- "and the winner is..." the archivist with the smallest distance
- **Cluster** annotations & specialise archives
- Improve **query solving** & **respect** knowledge distribution

Interactions mediator-archivist in solving a query

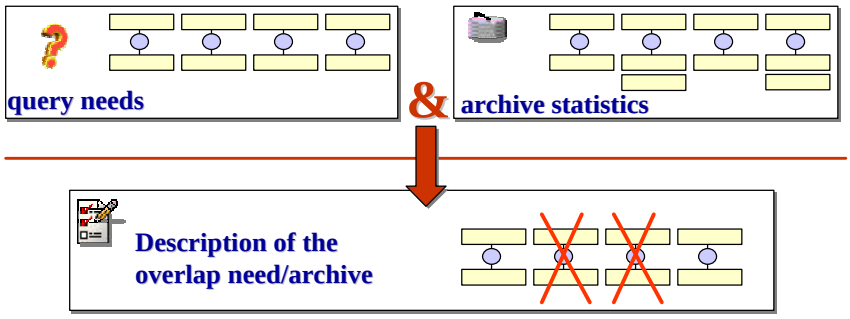
Interaction 2 - **distributed** query **solving**

- **Pb:** archives of annotation **distributed** all over organisation
- Mediators & archivists **share knowledge** to solve a query
- **Composition** of Query-Ref protocol for distributed solving



Overlap description

Each **archivist** calculate the **overlap** between query needs and its archive statistics:

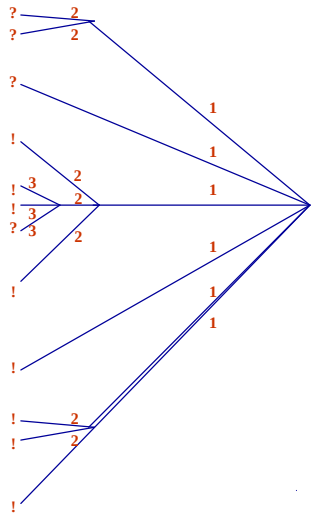


- Archivists refine their **service description** for a query
- Mediators **target** multicast **communications** query-solving
- **Exploit** archive **specialisation** obtained during the allocation of annotations

RDF query tree structure

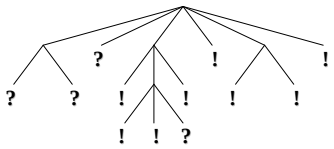
query

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
                 xmlns:CoMMA="http://www.inria.fr/acacia/comma#">
  <CoMMA:Document>
    <CoMMA:CreatedBy>
      <CoMMA:Person>
        <CoMMA:FamilyName>?AuthorName</CoMMA:FamilyName>
        <CoMMA:FirstName>?AuthorFirstName</CoMMA:FirstName>
      </CoMMA:Person>
    </CoMMA:CreatedBy>
    <CoMMA:Title>?DocTitle</CoMMA:Title>
    <CoMMA:Edited>
      <CoMMA:Company>
        <CoMMA:Designation>~smith</CoMMA:Designation>
        <CoMMA:Situated>
          <CoMMA:Location>
            <CoMMA:City>Nice</CoMMA:City>
            <CoMMA:Country>France</CoMMA:Country>
            <CoMMA:Phone>?EditorPhone</CoMMA:Phone>
          </CoMMA:Location>
        </CoMMA:Situated>
        <CoMMA:HasForActivity>
          <CoMMA:Edition />
        </CoMMA:HasForActivity>
      </CoMMA:Company>
    </CoMMA:Edited>
    <CoMMA:CreationDate>2000</CoMMA:CreationDate>
    <CoMMA:ReviewedBy>
      <CoMMA:Person>
        <CoMMA:FamilyName>hofstadter</CoMMA:FamilyName>
        <CoMMA:FirstName>douglas</CoMMA:FirstName>
      </CoMMA:Person>
    </CoMMA:ReviewedBy>
    <CoMMA:Concern>
      <CoMMA:ComputerScienceTopic />
    </CoMMA:Concern>
  </CoMMA:Document>
</rdf:RDF>
```



Decomposition

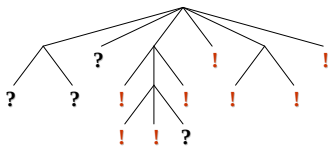
Tree **decomposition** & work **distribution**



- Mediator **prepares** query (compact, cross-references)
- Mediator **decomposes** query
  - **Bottom-up** constraints solving
  - **Top-down** question answering
  - URI as **cut/joint** points
- Mediator sends **sub-queries** to archivists according to overlap

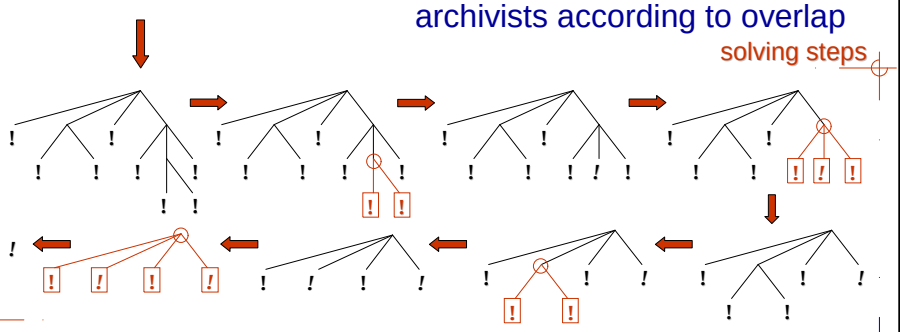
Decomposition (constraints)

Tree **decomposition** & work **distribution**



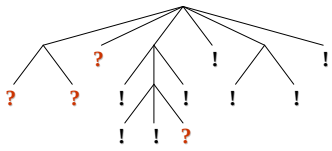
- Mediator **prepares** query
  (compact, cross-references)
- Mediator **decomposes** query
  - **Bottom-up** constraints solving
  - **Top-down** question answering
  - URI as **cut/joint** points
- Mediator sends **sub-queries** to archivists according to overlap

solving steps



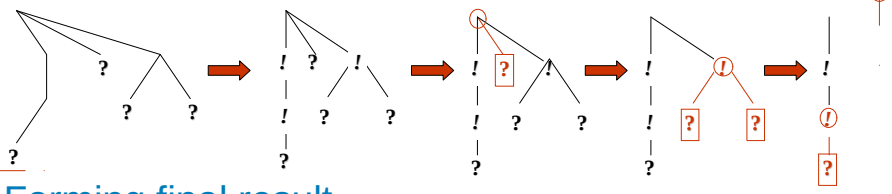
Decomposition (questions) & merging

Tree **decomposition** & work **distribution**



- Mediator **prepares** query
  (compact, cross-references)
- Mediator **decomposes** query
  - **Bottom-up** constraints solving
  - **Top-down** question answering
  - URI as **cut/joint** points
- Mediator sends **sub-queries** to archivists according to overlap

answering steps



Forming final result

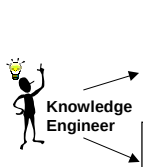
- Mediator merges partial results
- Mediator solves cross-references

Step 4 - Behaviours implementation

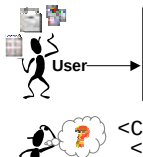
CORESE: COnceptual REsource Search Engine

- RDF(S) for schema, annotations, rules(+) , queries(+)
- Light weight component & API

ACACIA principle



RDF Schema → Support in CGs



RDF Rules

RDF Annotation

example of schema

```
<rdfs:Class rdf:ID='Entity'/>
<rdfs:Class rdf:ID='Document'
  <rdfs:subC                    ity' />
</rdfs:Class
<rdf:Propert
 <rdfs:subPr                    reator' />
  <rdfs:doma                    t'/>
  <rdfs:rang
</rdf:Proper
```

example of rule

```
IF
  Group
   Include
    Person ?x
   Include
    Person ?y
```

example of query annotation

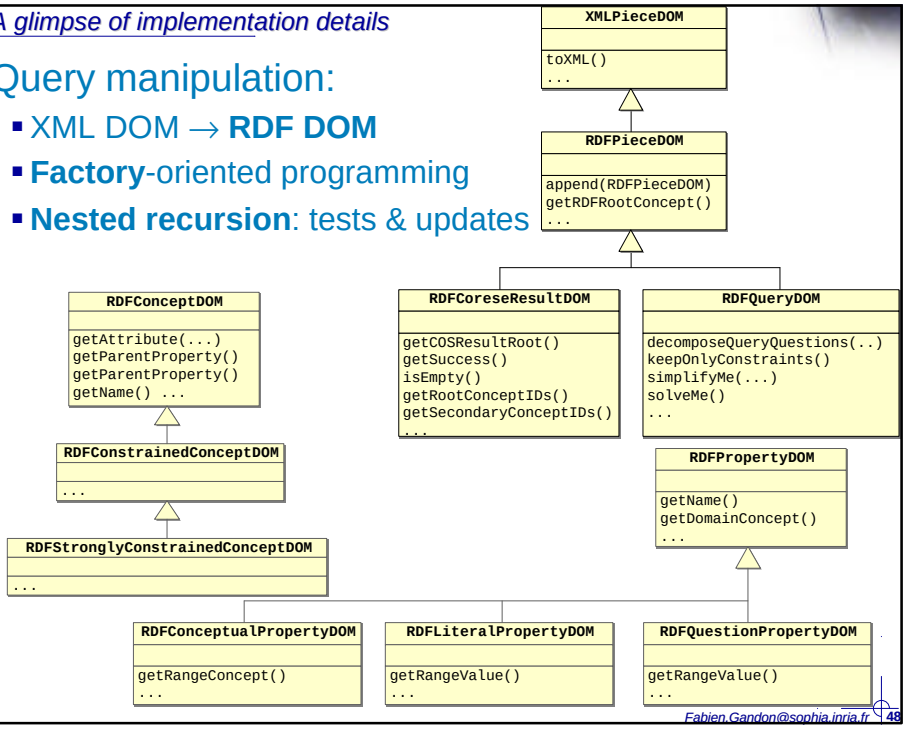
```
<CoMMA:Document rdf:about=
 <CoMMA:Title>Annual activ
 <CoMMA:Author>
  <CoMMA:Person rdf:about=
 </CoMMA:Author>
</CoMMA:Document>
```

```
<CoMMA:Article rdf:about="?d">
 <CoMMA:Title>?t</CoMMA:Title>
 <CoMMA:Author>?a</CoMMA:Author>
 <CoMMA:Concern>
  <CoMMA:UMTSTopic />
 </CoMMA:Concern>
</CoMMA:Article>
```

Translation Projection Translation

Query manipulation:

- XML DOM → **RDF DOM**
- **Factory**-oriented programming
- **Nested recursion**: tests & updates



Non identified existential quantification

- Treated as **anonymous statements** with existential quantification:

```
<report>
 <title>Effectiveness of Factories in AOP</title>
 <author>
  <person>
   <name>Smith</name>
   <first_name>John</first_name>
  </person>
 </author>
</report>
```

mediator interpretation

```
($ report(x) ...
 ($ person(y)
  author(x,y) ...
```

- Treated as **named statement** on:

```
<report ID="acacia_archivist@1...
  <title>Effectiveness of Factories in AOP</title>
  <author>
    <person ID="acacia_archivist@fapollo:1099/JADE#genID79">
      <name>Smith</name>
      <first_name>John</first_name>
    </person>
  </author>
</report>
```

mediator interpretation

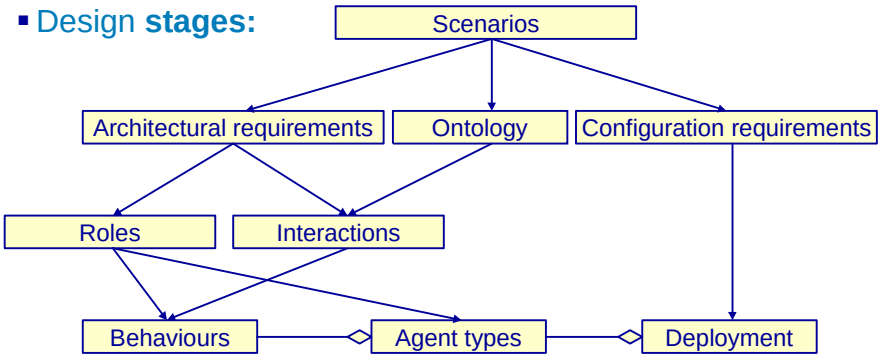
```
acacia_archivist@fapollo:1099/JADE said:
($ report(#genID54) ...
 ($ person(#genID79)
  author(#genID54,#genID79) ...
```

Ordering const

- First algorithm: f
- Second: **most constrained** node a disjunctive **lists of URI**

Summary

- Design **stages:**



- Result:
  - Design and implementation of **society** providing **ontology** and **corporate model**
  - Design and implementation of **annotation-dedicated society**

Progress in the plan and feedback from trial

CoMMA prototype

- Two **trials** mid-project and project end in development spiral
- One public **open-day** demo: industrial & EU commission
- **Usability** and **usefulness** recognised in user/public

Plan

Positioning: ✓

- Concerned **domains of research** ✓
- Application context: the **CoMMA** IST Project ✓

Thesis **contribution**:

- **Corporate semantic web** & **ontology engineering** ✓
- **Multi-agent** architecture & **distribution of annotations** ✓

Evaluation, discussion, conclusions, perspectives

Lessons learnt (ontology)

Ontology as a **cornerstone**:

- Provides **modelling** and representation **consensus**
  - Communication between humans (ontology is a component of the memory, allows annotations of information resources) +
  - Communication between agents (c.f. message) +
- Ontological consensus to **build consensus above**
  - Computational / inferential consensus, e.g.: distance, projection +
  - Co-operation consensus, e.g.: query and allocation protocol +
- **Acquiring** knowledge and **engineering** ontology
  - Complex **design rationale** & many **open questions** in practice (...)
  - Need for **integrated platforms** with all design tools –
- An ontology is a **living** object
  - Support **life cycle** and the maintenance of **consensus** –
  - Support maintenance of what was built **above** ontology –

Lessons learnt (semantic web and agents)

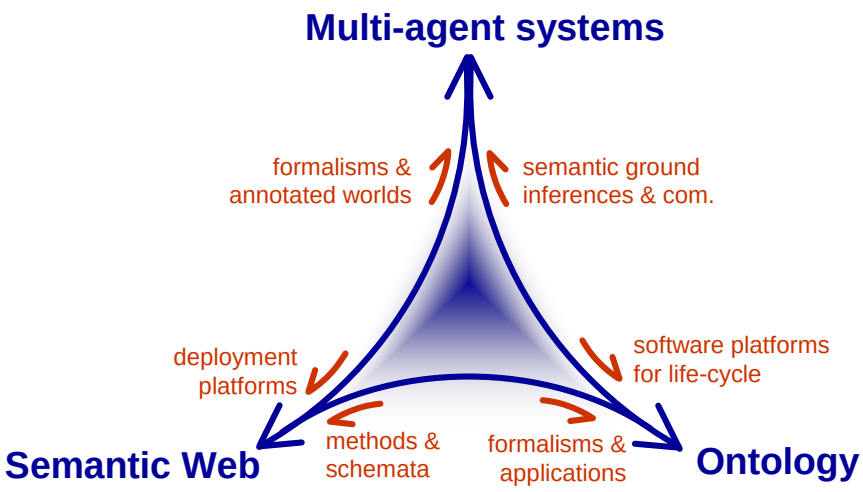
Semantic web framework:

- XML, RDF(S) proved **appropriate** in a 2-year project +
- **Tools** (parsers, API) **available** and **updated** +
- XML format, RDF model, platform-independent standards +
- Real **standard** & adequacy to knowledge representation +
- Layered **extension** needed for expressiveness (OWL) (...)
- Few **edition** and **visualisation** tools –

Multi-agent paradigm:

- **Developers** appreciation of **agent-ontology** approach + loosely-coupled software (specify, design, integrate, deploy)
- Integration of results from **different research domains** +
- **Knowledge-level** message passing, natural in KM +
- **Paradigm** is still under debate (...)
- **Multiplication** of theoretical methods –
- Lack of design and documentation **tools** –

Complementary research domains



Current extensions and perspectives

- This **Ph.D.** provided:
  - **Methodological** steps and **tools** in building ontology, semantic intraweb and associated multi-agent architecture
  - A running prototype as a **proof of concept** on these points
  - Both are **reusable** contributions
- Current **extensions**:
  - **Wrapper society**: semi-automatic annotation of web sources
  - **Extranet gateways**: mapping between semantic intrawebs
- Some **perspectives**
  - Semantic **distance** (multiple criteria, literal distance, length subs.)
  - **Distributed solving** (multi-instanciation, constraint sorting, URI)
  - **Ontology** services (querying, propagate update)
  - **Annotation** services (query registration, edition, update)
  - Ontology-based **interfaces** (filtering, customising, surrogates)
  - Collaborative filtering, security, distributed rules, *etc.*

Acknowledgements

- To **Rose Dieng-Kuntz**, **Olivier Corby**, **Alain Giboin**, and the **ACACIA** team for advice and discussions
- To the assistants **Hortense Hammel** and **Sophie Honnorat** for their help in the day-to-day work
- To all the **partners** of the **CoMMA** IST project for very fruitful discussions
- To the members of the **jury** and the **reporters** for the review work
- To the **SEMIR** for high quality working environment
- To the **librarian department,** for providing me with needed information resources
- To **INRIA** and **University of Nice-Sophia Antipolis**
- To the **European Commission**

Dedicated to my family.