Transitioning Applications to Ontologies

The VigiTerms Use Case

Florence Amardeilh
Mondeca

Talk Outline

Introduction
- VigiTerms Project

Ontology Learrning
- Need for a Pharmacovigilance ontology
- Transitioning Thesaurimed with RDBToOnto

Case report documentation
- VigiTerms Architecture
- CA Manager specialised workflow

Conclusion & Future work

The Pharmacovigilance Issue

⇨ *Drug'sAdverse Effect profile: Right dose of the right drug to the right patient*

◈ Pharmacovigilant Partner:
  - ◈ CRPV de Paris-HEGP. 4 arrondissements de Paris (1, 14, 15, 16) + département 92.
  - ◈ 3,5 doctors and pharmacologists
  - ◈ >600 dossiers de PV/2007, mainly spontaneous notification
  - ◈ 500 cases reported in the national CR database (ArisG5) among them 300 serious AE and 20 fatal AE
  - ◈ Notificators : doctors (400), chemists (50), patients (10)

Case report information

◈ <u>Notificator</u> identification,

◈ <u>Patient</u> characteristics
  ◈ (initials, age, sexe, background)

◈ ➢ All the <u>patient's drug(s)</u>
  ◈ (chronology, posology, drug route, indication)

◈ Produced a<u>dverse effect</u> : Gravity, description, evolution, results of the ethiology test;…

Adverse Effect Case report

Vigiterms project

◈ A French-founded (ANR) research project reusing TAO outputs

◈ The main objective is to **improve the process of detecting a new unforeseen signal between an adverse effect and a drug**.

→ **Easing case report documentation by Pharmacovigilants** based on available resources:
- the pharmacovigilance case report database,
- the medical literature abstract (MedLine),
- the medical terminologies and thesaurus,
- the drug characteristics,
- the external specialized databases or resources.

TAO

VigiTerms project

◈ **TAO components are used to:**

  ◈ Create a **Signal Detection Ontology** modelling adverse effects, drugs and the patient context

   ◈ based on existing medical terminologies (such as WHO-ART, SNOMED, MedDRA) and existing drug databases (Thesorimed, Vidal)

  ◈ **Annotate the medical and pharmaceutical literature** available into MedLine, DrugDex, etc. for the case report documentation

  ◈ **Transition the legacy application** (Aris G5) which collects Adverse Effect Drug Reaction reports with detail on patients, drugs, adverse reaction details, etc.

TAO

Transitioning medical resources

◈ existing drug databases such as the French Thesorimed with references to standard medical classifications
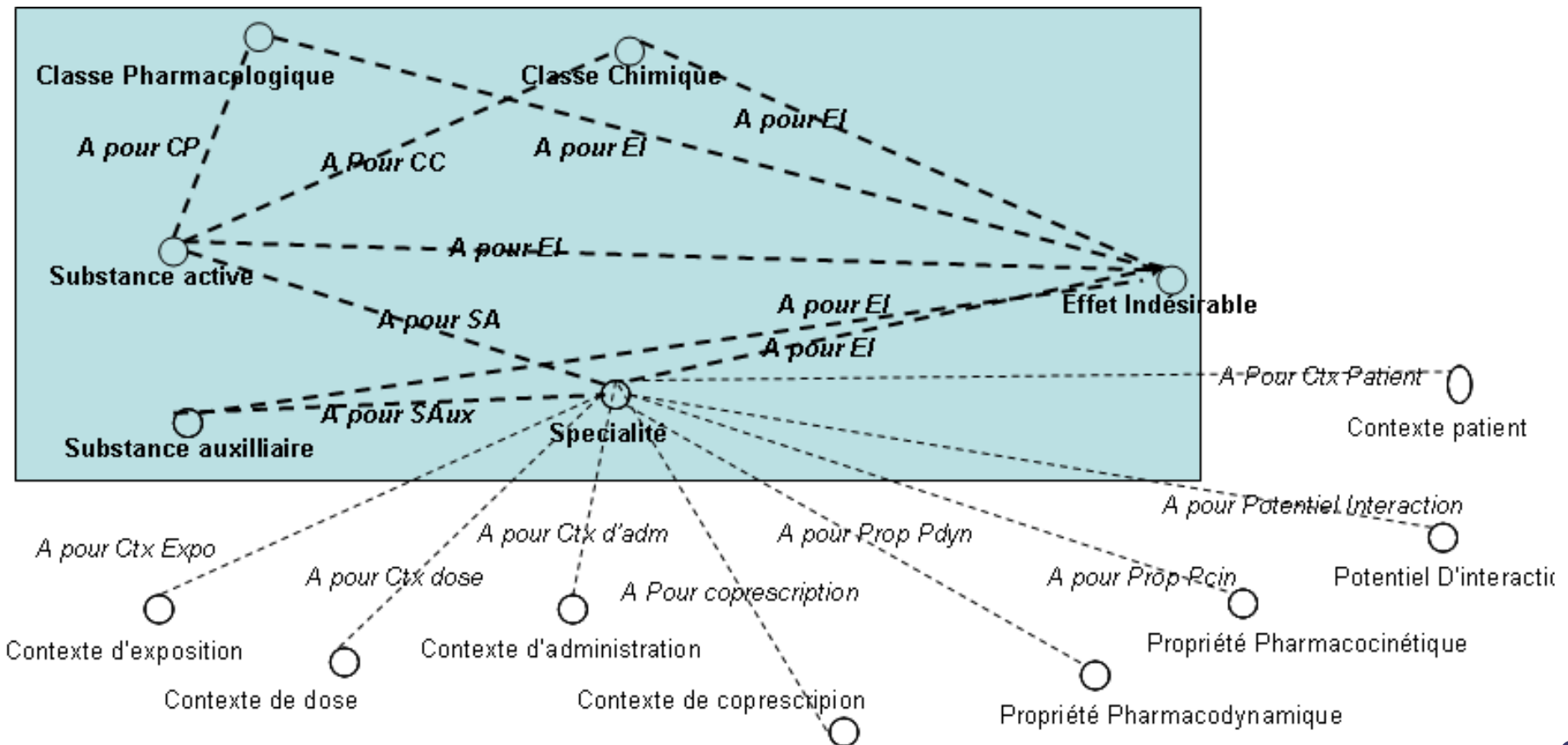
Transitioning medical resources

◈ Sub-part of Thesorimed database model to test the transitioning process with RDBToOnto

ID_CPH	ClassePharmacologique
C	APPAREIL CARDIOVASCULAIRE
C1	ANTIHYPERTENSEURS
C11	ANTIHYPERTENSEUR
C12	ANTIHYPERTENSEUR CENTRAL
C13	ANTIHYPERTENSEUR INHIBITEUR CALCIQUE
C15	ANTIHYPERTENSEUR INHIB ENZYME CONVERSION
C16	ANTIHYPERTENSEUR ANTAG ANGIOTENSINE II

EI_ID	EffetIndesirable
A	GASTROENTEROLOGIE HEPATOLOGIE
AB	TROUBLES GASTRO-DUODENAUX
AB1	GASTRALGIE
AB2	EPIGASTRALGIE
AB3	NAUSEE VOMISSEMENT
AB4	ULCERATION GASTRIQUE

CodeATC	LibelléATC
C	SYSTEME CARDIOVASCULAIRE
C01	MEDICAMENTS EN CARDIOLOGIE
C01A	GLUCOSIDES CARDIOTONIQUES
C01AA	GLUCOSIDES DE LA DIGITALE
C01AA01	ACETYLDIGITOXINE
C01AA02	ACETYLDIGOXINE
C01AA03	FEUILLES DE DIGITALE
C01AA04	DIGITOXINE
C01AA05	DIGOXINE

ID_CCH	ClasseChimique
C	AMINES
C1	ALKYLAMINE
C11	ALKYLAMINE AUTRE
C12	ETHANOLAMINE
C13	ETHYLAMINE
C14	ETHYLENEDIAMINE

ID_SA	SubstanceActive
1	LOMEFLOXACINE CHLORHYDRATE
2	OR ELEMENT MINERAL
3	SPARFLOXACINE
4	SUFENTANIL CITRATE
5	EMETINE CHLORHYDRATE

CodeMeSH	LibelleMeSH
D02.455.426.559.847.638.845	1-Naphthylamine
D04.615.638.845	1-Naphthylamine
D04.615.638	Naphthalenes
D02.455.426.559.847.638	Naphthalenes
D02.455.426.559.847	Polycyclic Hydrocarbons, Aromatic
D04.615	Polycyclic Hydrocarbons, Aromatic
D02.455.426.559	Hydrocarbons, Aromatic
D04	Polycyclic Compounds
D02	Organic Chemicals
D02.455	Hydrocarbons

WAO

Transitioning medical resources

◈ **Using RDBToOnto**
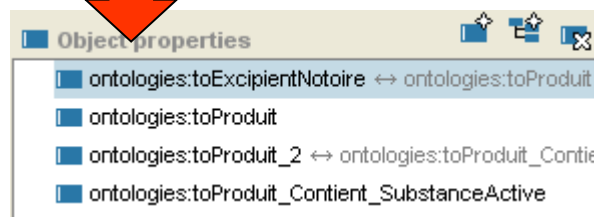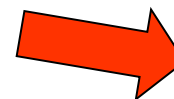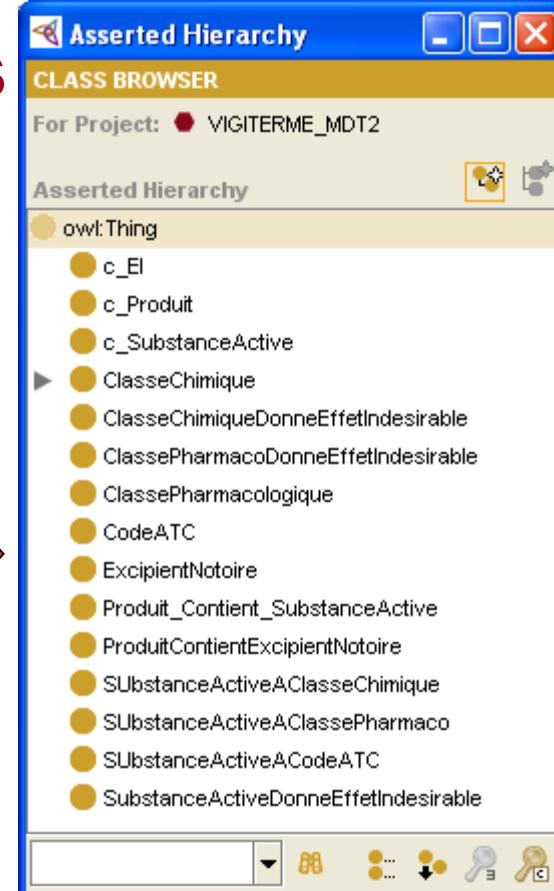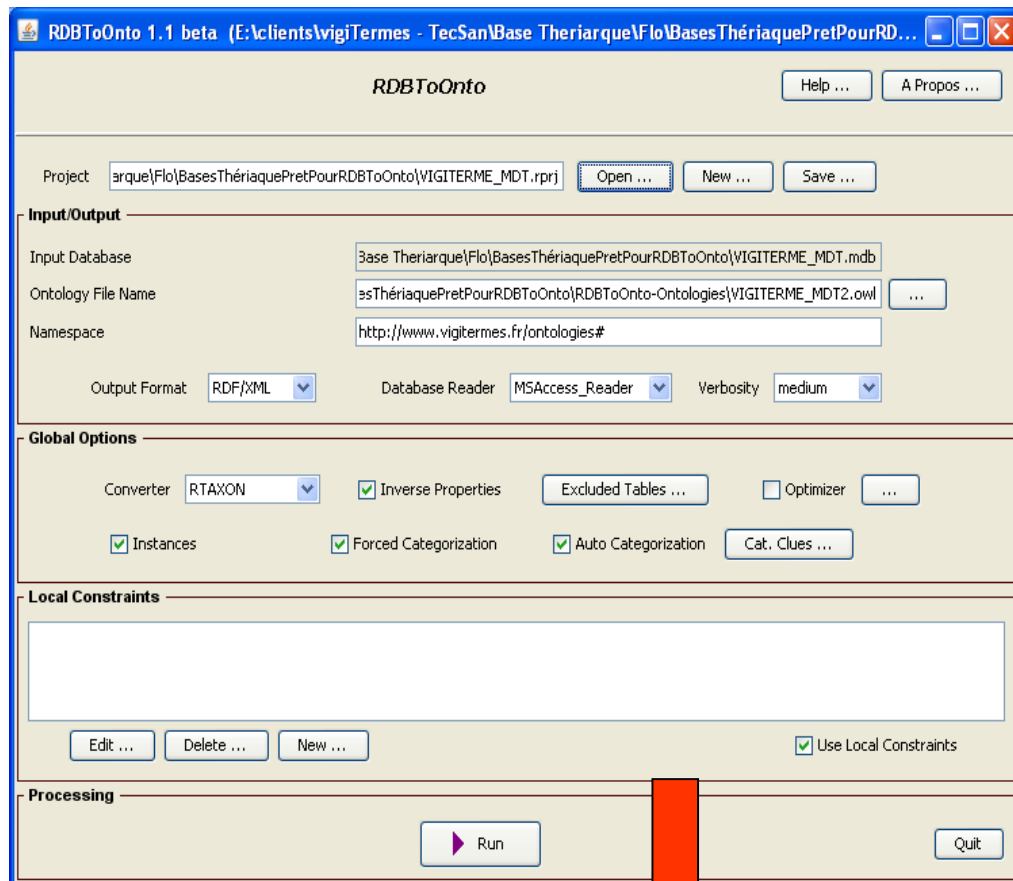
Improving the RDBToOnto algorithms for generating deeper taxonomies

Pharmacovigilant Knowledge Store

◈ **Storing and querying the Knowledge Store (ITM)**

Case report documentation use case

13
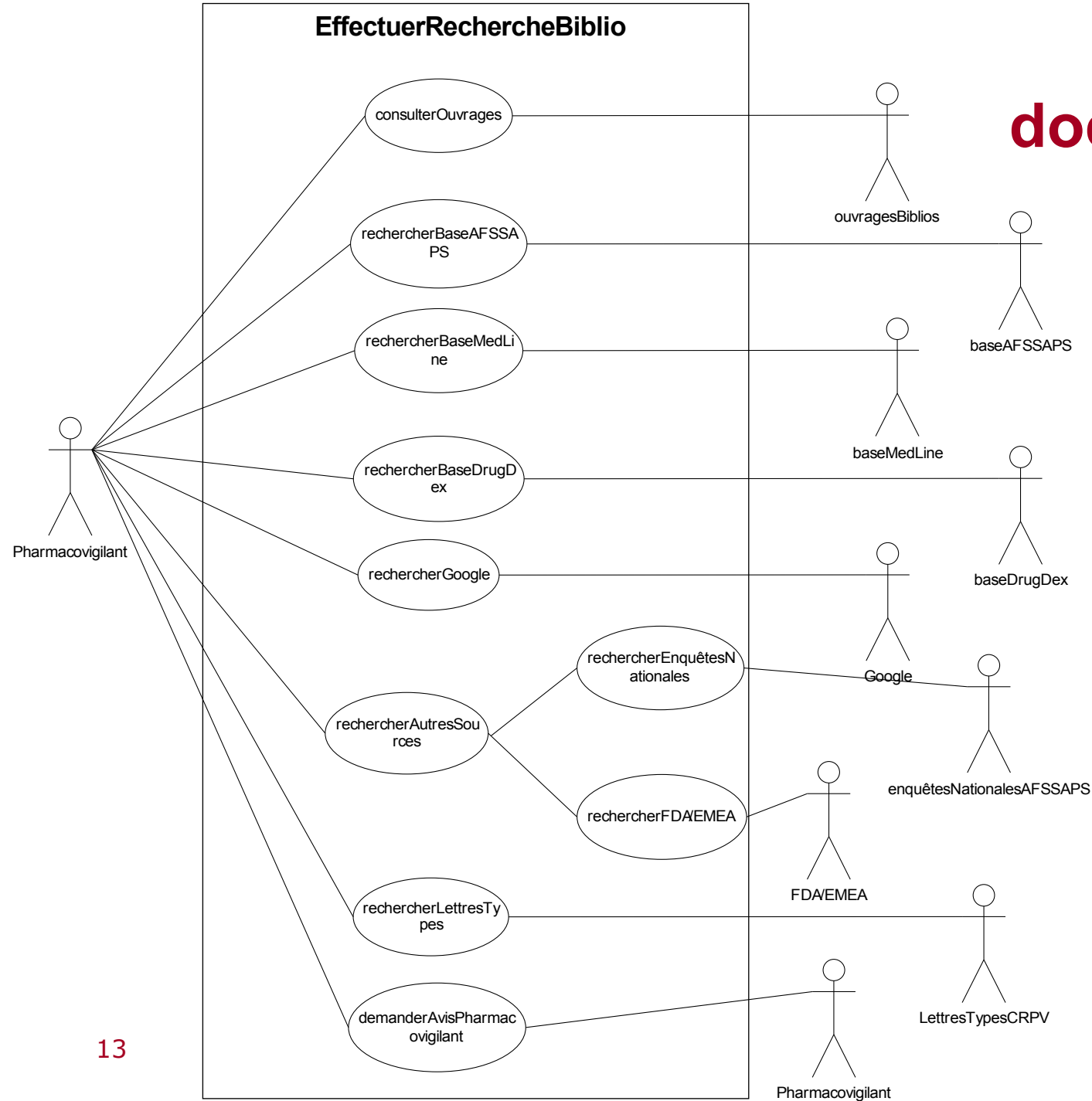
Ontology Population & Semantic annotation Paradigms

Text Mining

Audio Mining

Structured information

Ontology and vocabulary control

Reasoning engines

Input documents

Format transformation

Metadata generation

User validation

Knowledge population (knowledge store integration)

ITM repository
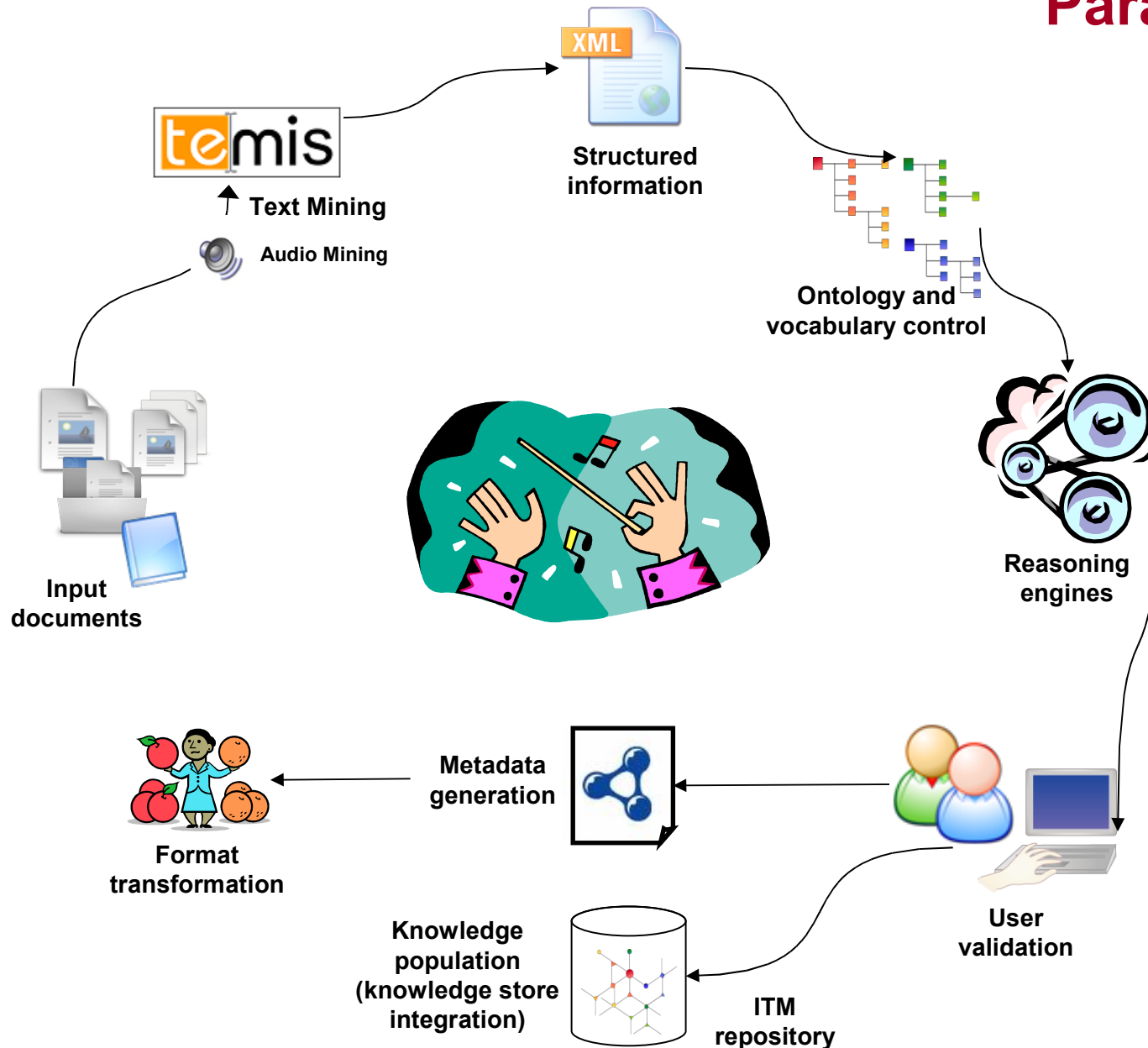
Case report Documentation Workflow
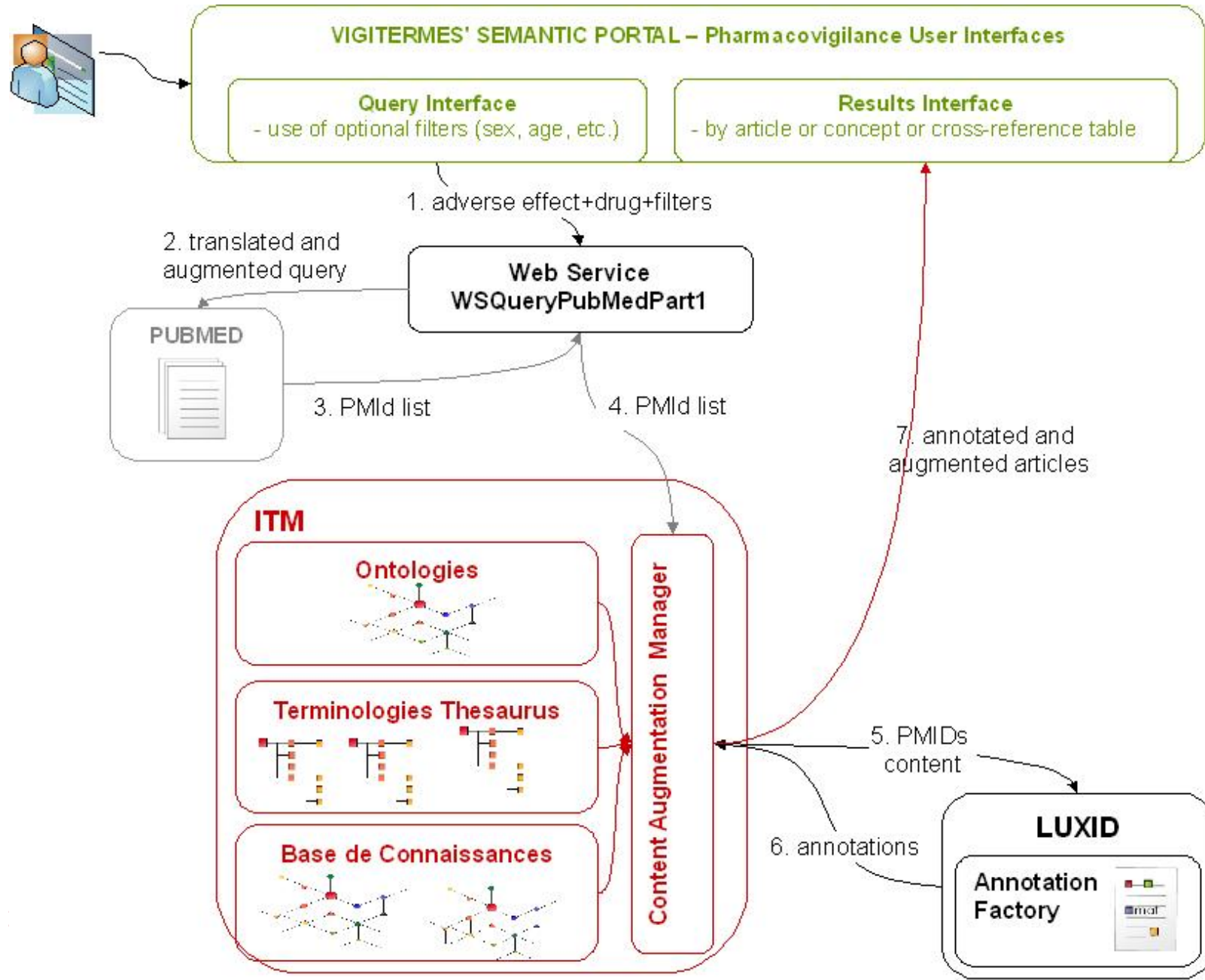
Content Augmentation Process

- Defining the CA process as based on 3 main steps :
  - Information Extraction
  - Information Consolidation
  - Information Storage

- Having a unique exchange format between each step

- Having a set of consolidation services according to the objectives and tools at disposal of the target application

- Being open-source and thus, each component must be independent from any existing platform or tool

- Relying on a Service Oriented Architecture (SOA)

CA manager pipeline

1. Information Extraction

2. Information Consolidation

3. Information Storage

Split | Extract | Luxid

Clean + Merge | Control | ITM | Infer

Serialize | RDF | OWL | Store | ITM | Sesame
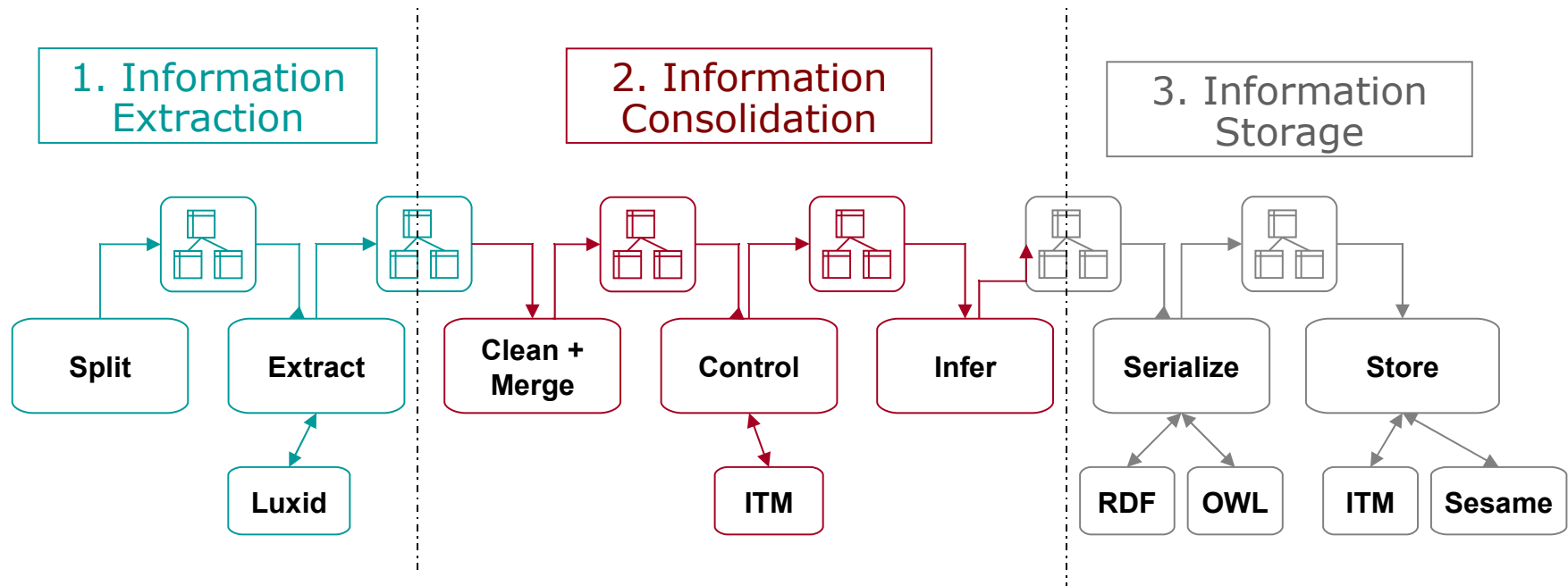
◈ Processing pipeline

◈ Common Analysis Structure (CAS) (cf. UIMA)

CA-Manager applied to Case Report Documentation

◈ Demo of CA-Manager

Conclusion & Future work

◈ Purpose

 ◈ Support case report documentation
 ◈ Detect new signals

◈ On going work

 ◈ Integrating components together
 ◈ Defining text mining patterns
 ◈ Integrating existing ontologies and terminologies together in an ontology model for pharmacovigilance
 ◈ User interfaces for reporting and interrogating case reports

◈ Future plans

 ◈ Implement data mining component based on LCA techniques to detect automatically new signals based on the knowledge base

19

Questions? Discussion welcome!

Achieved in second year : CA-Manager architecture

- ◈ Choice of UIMA as CA-Manager backbone.
  - ◈ IBM open-source framework
  - ◈ « UIMA supports the development, discovery, composition and deployment of multi-modal analytics for the analysis of unstructured information and its integration with search technologies »
  - ◈ Benefits :
    - ◈ Ability to define flexible processing flows
    - ◈ distributed & pluggable components
    - ◈ XML configuration of components
    - ◈ A common data structure between components
    - ◈ Sharing & reusing open-source components
- ◈ Other solution : 100% webservices
  - ◈ No need to have webservices inside internal components of TAO
  - ◈ Potentially slow & hard to maintain
  - ◈ Lots of things to reimplement

TAO

Achieved in second year : CA-Manager architecture

Transtioning text mining to
semantic and ontology based tools

Regexp | Gate | OWLIM | HKS | RDF

Split → Extract → Merge → Match → Control → Infer → Store → Serialize → ...

CA-Manager Service

UIMA (WebService)

TS traductor — CAS — AE

Type System XYZ

UIMA

AE → CAS → AE → CAS → ... → CAS

Content Augmentation Service Type System

UIMA

EJB

Web service

Client API

CLIENT APP
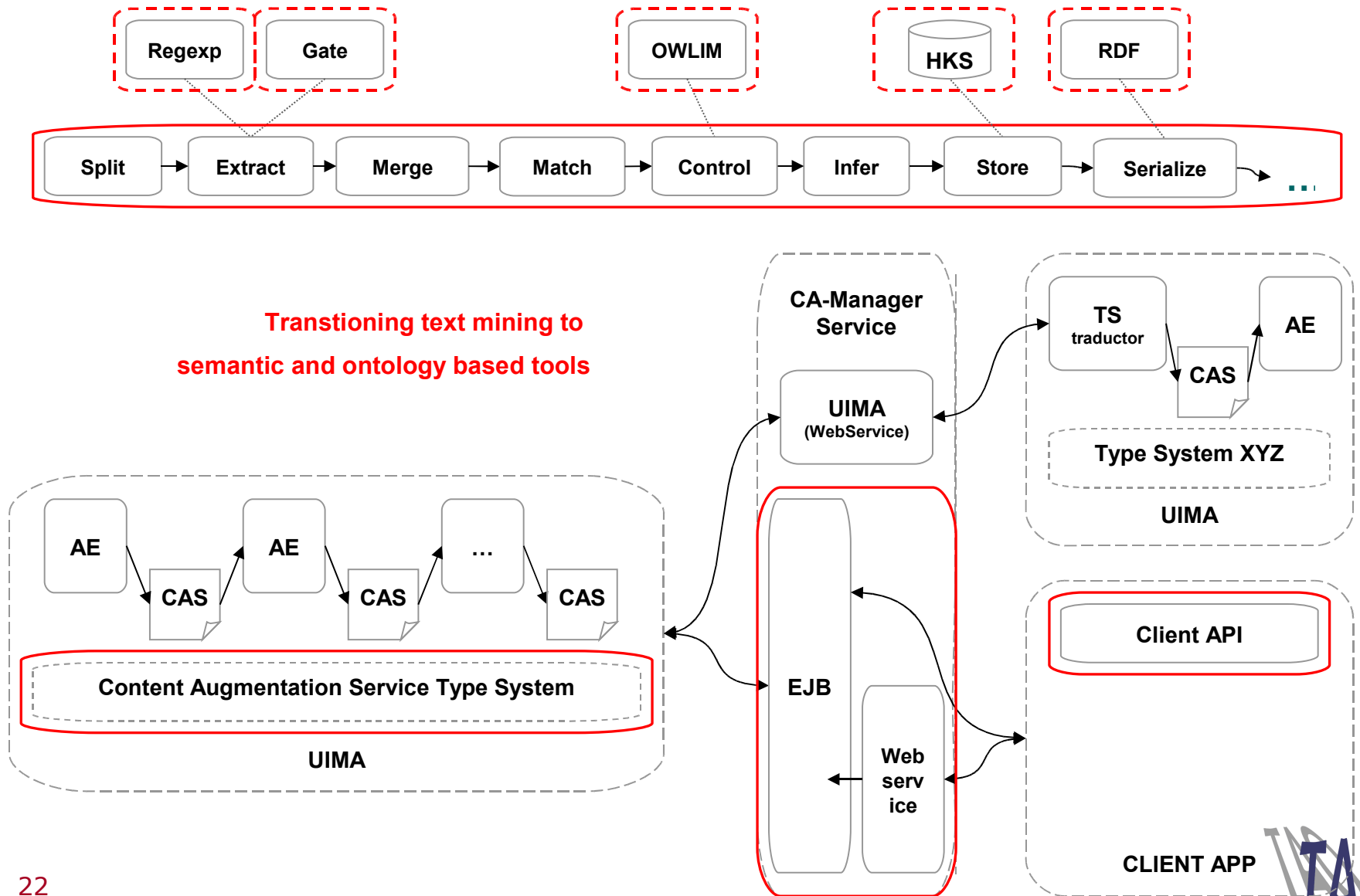
Achieved in second year : CA-Manager architecture

◈ The architecture
- ◈ Is a specialized UIMA pipeline
- ◈ Focuses on transioning text-mining results to semantic tools
- ◈ Is meant to hide UIMA complexity if not required
- ◈ Provides a _service_ (simple, remote)
- ◈ Benefits from UIMA framework (process configuration, distributed architecture, connection to other UIMA based tools, etc.)
- ◈ Is pluggable (possibility to write a connector to a new external tool)

◈ Added value (how is it better than just UIMA ?)
- ◈ Definition of a generic Common Analysis Structure
- ◈ Mapping language to map IE results to CAS
- ◈ Specialization of UIMA pipeline
- ◈ Connectors to various external tools
- ◈ Hide UIMA complexity

Achieved in second year: CA-Manager

◈ Demo of CA-Manager

CA-Manager : Coming next

- OK so far :
  - Specifications and architecture
  - KCIT integration
  - Working prototype
- Coming next :
  - HKS integration
  - Ontology-based controls
  - SA-WSDL generation
  - Integration with annotation GUI
  - Open-sourcing

TAO