Transitioning Applications to Ontologies

The VigiTerms Use Case

Florence Amardeilh
Mondeca

Funded by: European Commission – 6th Framework
Project Reference: IST-2004-026460
Talk Outline

- **Introduction**
  - VigiTerms Project

- **Ontology Learning**
  - 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’s Adverse 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

- **Notifier** identification,

- **Patient** characteristics
  - (initials, age, sexe, background)

- ➢ All the patient’s drug(s)
  - (chronology, posology, drug route, indication)

- Produced adverse effect: Gravity, description, evolution, results of the etiology test;...
Adverse Effect Case report
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.
**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.
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

<table>
<thead>
<tr>
<th>ID_CPH</th>
<th>ClassePharmacologique</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>APPAREIL CARDIOVASCUULAIRE</td>
</tr>
<tr>
<td>C1</td>
<td>ANTIHYPERTENSEURS</td>
</tr>
<tr>
<td>C11</td>
<td>ANTIHYPERTENSEUR</td>
</tr>
<tr>
<td>C12</td>
<td>ANTIHYPERTENSEUR CENTRAL</td>
</tr>
<tr>
<td>C13</td>
<td>ANTIHYPERTENSEUR INHIBITEUR CALCIQUE</td>
</tr>
<tr>
<td>C15</td>
<td>ANTIHYPERTENSEUR INHIB ENZYME CONVERSION</td>
</tr>
<tr>
<td>C16</td>
<td>ANTIHYPERTENSEUR ANTAG ANGIOTENSINE II</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>El_ID</th>
<th>EffetIndesirable</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>GASTROENTEROLOGIE HEPATOLOGIE</td>
</tr>
<tr>
<td>AB</td>
<td>TROUBLES GASTRO-DUODENAIRES</td>
</tr>
<tr>
<td>AB1</td>
<td>GASTRALGIE</td>
</tr>
<tr>
<td>AB2</td>
<td>EPIGASTRALGIE</td>
</tr>
<tr>
<td>AB3</td>
<td>NAUSEE VOMISSEMENT</td>
</tr>
<tr>
<td>AB4</td>
<td>ULCERATION GASTRIQUE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CodeATC</th>
<th>LibelléATC</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>SYSTEME CARDIOVASCUULAIRE</td>
</tr>
<tr>
<td>C01</td>
<td>MEDICAMENTS EN CARDIOLOGIE</td>
</tr>
<tr>
<td>C01A</td>
<td>GLUCOSIDES CARDIOTONIQUES</td>
</tr>
<tr>
<td>C01AA</td>
<td>GLUCOSIDES DE LA DIGITALE</td>
</tr>
<tr>
<td>C01AA01</td>
<td>ACETYLDIGITOXINE</td>
</tr>
<tr>
<td>C01AA02</td>
<td>ACETYLDIGOXINE</td>
</tr>
<tr>
<td>C01AA03</td>
<td>FEUILLES DE DIGITALE</td>
</tr>
<tr>
<td>C01AA04</td>
<td>DIGITOXINE</td>
</tr>
<tr>
<td>C01AA05</td>
<td>DIGOXINE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID_CCH</th>
<th>ClasseChimique</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>AMINES</td>
</tr>
<tr>
<td>C1</td>
<td>ALKYLAMINE</td>
</tr>
<tr>
<td>C11</td>
<td>ALKYLAMINE AUTRE</td>
</tr>
<tr>
<td>C12</td>
<td>ETHANOLAMINE</td>
</tr>
<tr>
<td>C13</td>
<td>ETHYLAMINE</td>
</tr>
<tr>
<td>C14</td>
<td>ETHYLENEDIAMINE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID_SA</th>
<th>SubstanceActive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LOMEFLOXACINE CHLORHYDRATE</td>
</tr>
<tr>
<td>2</td>
<td>OR ELEMENT MINERAL</td>
</tr>
<tr>
<td>3</td>
<td>SPARFLOXACINE</td>
</tr>
<tr>
<td>4</td>
<td>SUFENTANIL CITRATE</td>
</tr>
<tr>
<td>5</td>
<td>EMETINE CHLORHYDRATE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CodeMeSH</th>
<th>LibelleMeSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>D02.455.426.559.847.845</td>
<td>1-Naphthylamine</td>
</tr>
<tr>
<td>D04.615.636.845</td>
<td>1-Naphthylamine</td>
</tr>
<tr>
<td>D04.615.638</td>
<td>Naphthalenes</td>
</tr>
<tr>
<td>D02.455.426.559.847.638</td>
<td>Naphthalenes</td>
</tr>
<tr>
<td>D02.455.426.559.847</td>
<td>Polycyclic Hydrocarbons, Aromatic</td>
</tr>
<tr>
<td>D04.615</td>
<td>Polycyclic Hydrocarbons, Aromatic</td>
</tr>
<tr>
<td>D02.455.426.559</td>
<td>Hydrocarbons, Aromatic</td>
</tr>
<tr>
<td>D04</td>
<td>Polycyclic Compounds</td>
</tr>
<tr>
<td>D02</td>
<td>Organic Chemicals</td>
</tr>
<tr>
<td>D02.455</td>
<td>Hydrocarbons</td>
</tr>
</tbody>
</table>
Transitioning medical resources

Using RDBToOnto

**Output Format**
- RDF/XML

**Database Reader**
- MSAccess Reader

**Ontology File Name**
- c_TherapiePrepOntoOntologies/VGITERME_MDT1.owl

**Namespace**
- http://www.vgitermes.fr/ontologies/

**Global Options**
- Converter: RITATION
- Inverse Properties
- Excluded Tables
- Optimizer

**Local Constraints**
- Instances
- Forced Categorization
- Auto Categorization

**Processing**
- Run

**Class Browser**
- owl:Thing
- c_El
- c_Produit
- c_SubstanceActive
- ClasseChimique
- ClasseChimiqueDonneEffetIndesirable
- ClassePharmacoDonneEffetIndesirable
- ClassePharmacoLOGique
- CodeATC
- ExcipientNotaire
- Produit_Content_SubstanceActive
- ProduitContentExcipientNotaire
- SubstanceActiveA ClasseChimique
- SubstanceActiveAClaassePharmaco
- SubstanceActiveACodeATC
- SubstanceActiveDonneEffetIndesirable
Improving the RDBToOnto algorithms for generating deeper taxonomies.
Pharmacovigilant Knowledge Store

- Storing and querying the Knowledge Store (ITM)
EffectuerRechercheBiblio

- consulterOuvrages
- rechercherBaseAFSSAPS
- rechercherBaseMedLine
- rechercherBaseDrugDex
- rechercherGoogle
- rechercherEnquêtesNationales
- rechercherAutresSources
- rechercherFDA/EMEA
- rechercherLettresTypes
- demanderAvisPharmacovigilant

ouvragesBiblos
- baseAFSSAPS
- baseMedLine
- baseDrugDex
- Google
- enquêtesNationalesAFSSAPS
- FDA/EMEA

rechercherLettresTypesCRPV

Pharmacovigilant
Ontology Population & Semantic annotation Paradigms

Text Mining

Audio Mining

Input documents

Structured information

Ontology and vocabulary control

Reasoning engines

Metadata generation

Format transformation

Knowledge population (knowledge store integration)

ITM repository

User validation
Case report Documentation Workflow

VIGITERMES' SEMANTIC PORTAL – Pharmacovigilance User Interfaces

Query Interface
- use of optional filters (sex, age, etc.)

Results Interface
- by article or concept or cross-reference table

1. adverse effect+drug+filters

Web Service
WSQueryPubMedPart1

2. translated and augmented query

PUBMED

3. PMID list

4. PMID list

ITM

Ontologies

Terminologies Thesaurus

Base de Connaissances

Content Augmentation Manager

5. PMIDs content

6. annotations

5. PMID list

7. annotated and augmented articles

TAO
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)
1. Information Extraction

- Split
- Extract
- Luxid

2. Information Consolidation

- Clean + Merge
- Control
- Infer
- ITM

3. Information Storage

- Serialize
- Store
- RDF
- OWL
- 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
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
Achieved in second year: CA-Manager architecture

Content Augmentation Service Type System

UIMA

CA-Manager Service

UIMA (WebService)

TS traductor

AE

Type System XYZ

UIMA

Client API

CLIENT APP

Web service

EJB

TS traductor

AE

Type System XYZ

UIMA

Content Augmentation Service Type System

UIMA

Regexp

Gate

OWLIM

HKS

RDF

Split

Extract

Merge

Match

Control

Infer

Store

Serialize

Transitoning text mining to semantic and ontology based tools
Achieved in second year: CA-Manager architecture

- The architecture
  - Is a specialized UIMA pipeline
  - Focuses on transitioning 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