Language and knowledge technologies to properly model textual medical data and support better reasoning

Dr Frédérique Segond,
Directrice du centre R&D, Viseo group
Professeure Associée, Institut National des Langues et Civilisations Orientales, ERTIM

With the contribution of SYNODOS partners.

inalco
Institut national des langues et civilisations orientales

VISEO
www.viseo.com

The context

Health goes to digital and becomes eHealth

Electronic medical record (EMR)

- support medical decision
- epidemiological surveillance
- data (semantic) mining
- Patients phenotyping
- Establishment of quality indicators, etc.

However, today, most data are still *unstructured* and very few are coded in the EMR

Hospital Acquired Infections (HAI):

some numbers

- 5% to 10% of Hospitalized patients will get an HAI
- In Europe, 3 Million cases / 50 000 death a year up to 30% considered as avoidable
- Related cost for France : 1 Billion €

Colon Cancer: some numbers

- One of the 5 most common cancers in Europe for men and women (the third). Less frequent in South America, Asia and Africa.
- Most cases, patients over 70.
- Worldwide standard rate of incidence between 35,5 and 43 per 100 0000.

(source www.e-cancer.fr)

VISEO

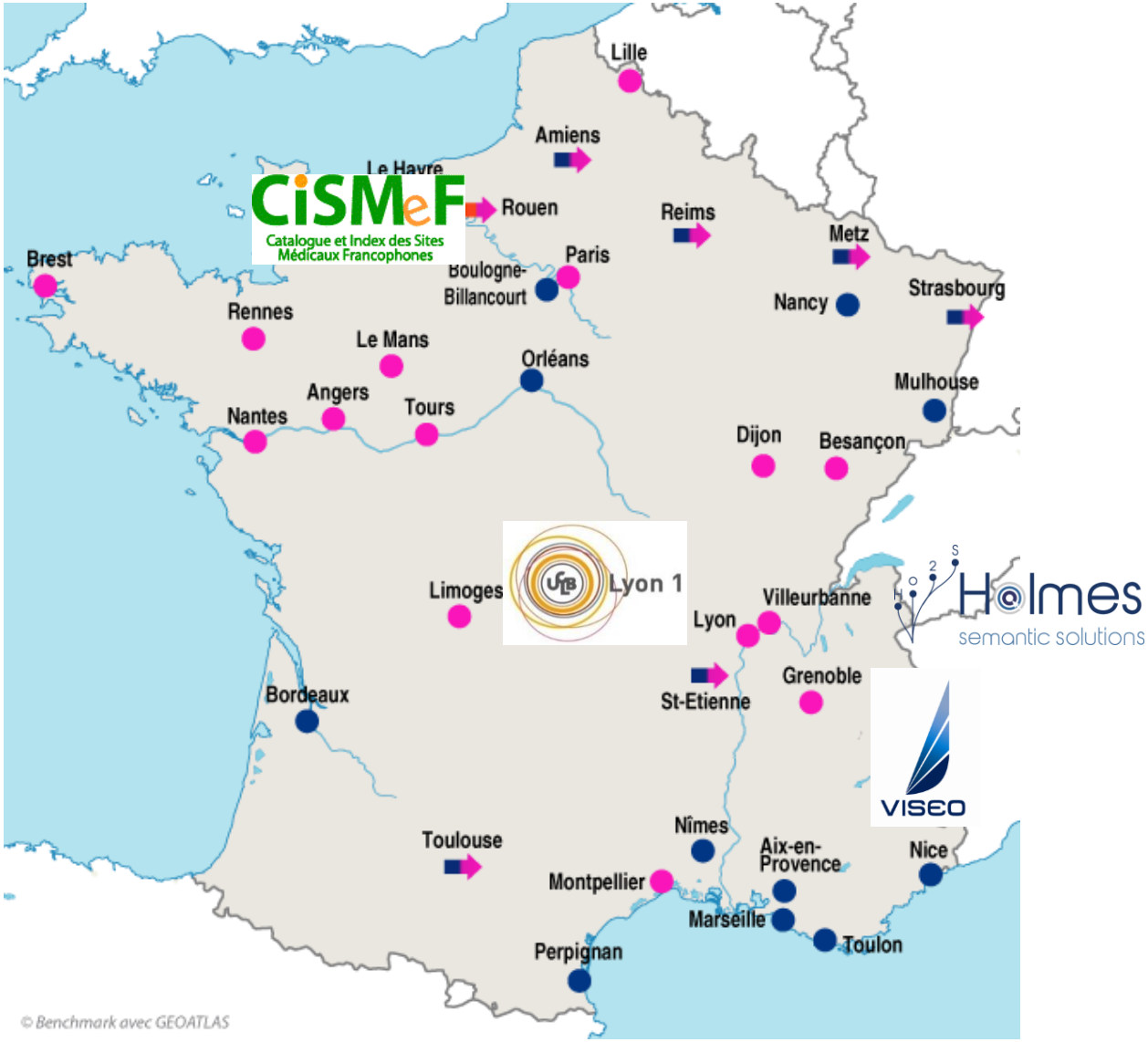
The Project : Synodos

Develop a generic solution to automatically make sense of textual data contained in EMRs

- Structure medical knowledge in order to use it in epidemiological studies or in medical decision support
- Combine natural language processing and knowledge engineering to extract “semantics” from medical data (EMRs)
- Give medical users the possibility to write their own expert rules

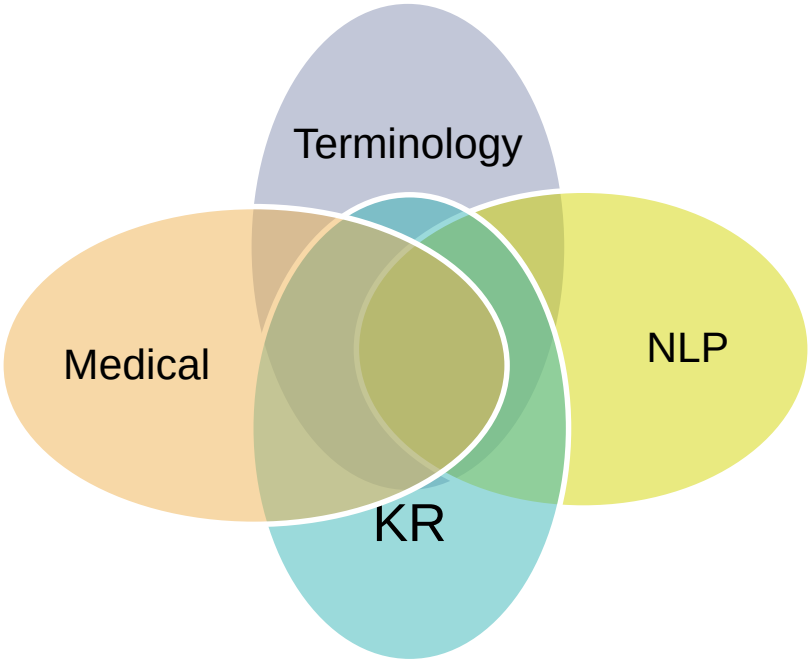
VISEO

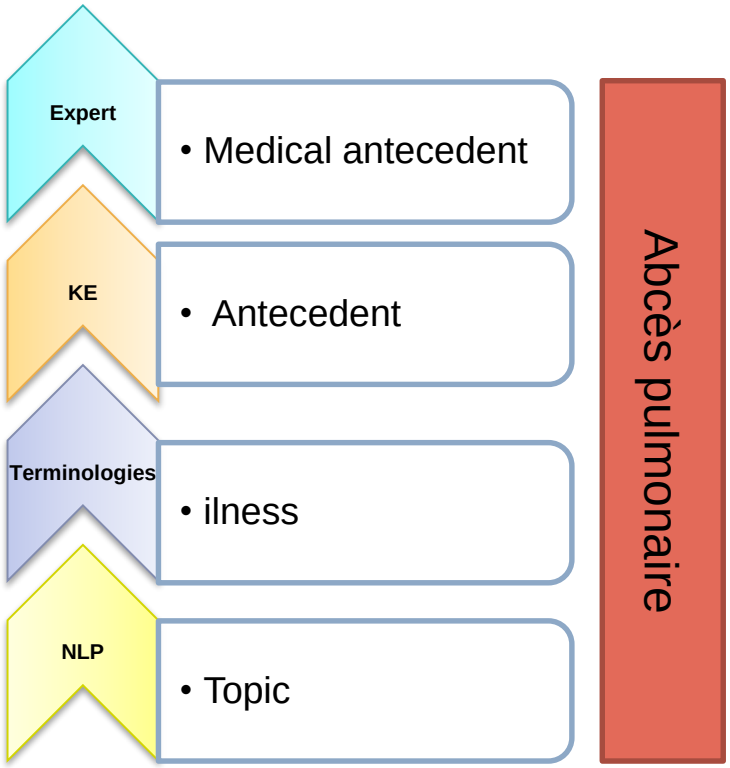
Synodos partners



VISEO

Combination of scientific fields





Nous avons découvert un abcès pulmonaire chez le patient en [T-10A]

(We discovered a lung abscess in this patient in [T-10A])

The data

SYNODOS Data

Reports

Initial Operative Report

Cette patiente est connue du service depuis près de 24 ans. On sait qu'elle est porteuse de gastrinome récidivant dont on n'a jamais trouvé de manière certaine la lésion primitive..... Elle a subi plusieurs interventions pour ablation de masses ganglionnaires sous et para pancréatiques, pour curage

Consultation Report

Antécédents chirurgicaux :
- Prothèse d'épaule bilatérale en [T-5A] et [T-4A]
- arthro GG en [T+5M]
- résection endo urétrale de prostate en [T-15A]

Exit report

Des EFR réalisées en [T+0M] retrouvent un trouble ventila
réversible, avec un VEMS à 1,2 litre, diminution de la CV
hypoxémie artérielle de repos, une DLCO normale. A noter
abcès pulmonaire en [T-12A]
- Crise de goutte
- Prostatisme avec nycturie x 3
- Glaucome chronique à angle ouvert

Other documents

suite discussion téléphonique avec Dr [PERSONNE] :
poursuite des ATB pendant encore 3mois.

Structured EMR

I. Parcours de soins
Identifiant patient
Code postal (ou département)
Age
Sexe

Episode de soins 1:
N° de venue:
Service:
Date d'entrée:
Type de document:

II. Antécédents
Antécédents médicaux
Type
Date de diagnostic
Examens biologiques
Examens anatomo-pathologiques
Examens d'imagerie

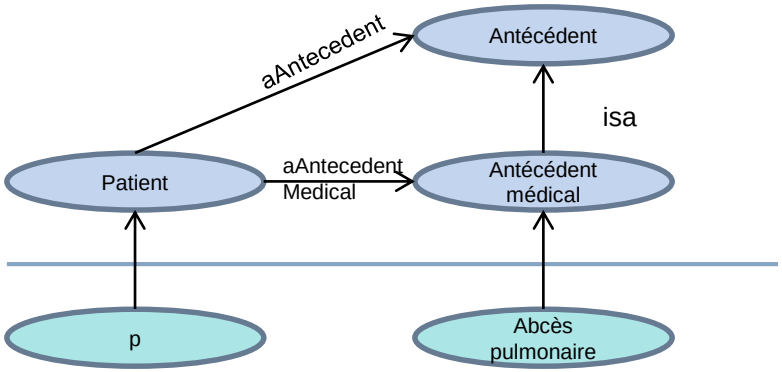
Episode de soins
Motif du contact
Examen clinique à l'entrée
Prise en charge à l'entrée
Actes thérapeutiques
Actes diagnostiques
Evolution
Symptômes
Examens
Actes thérapeutiques
Actes diagnostiques
Résultat de l'épisode de soins

Difficulties

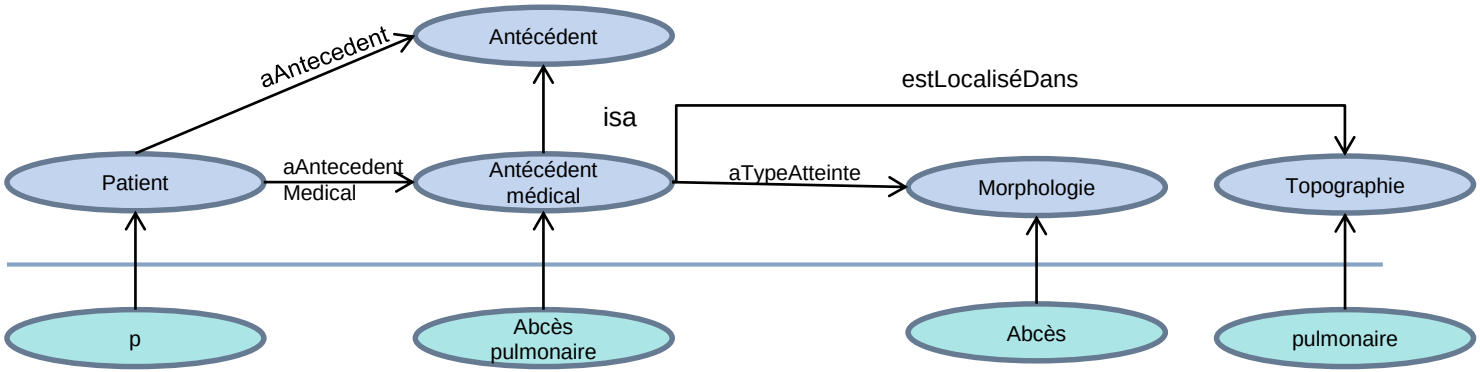
- Definition of a « stable » conceptual model general enough
- Definition and Identification of an episode of care:
  - An EMR can be composed of different reports.
  - An episode of care can be attached to one or more reports
  - How to detect an episode of care? No linguistic nor terminological information, solution on the expert side. Experts have trouble themselves to formally define an episode of care

```
Identification du patient : 22212

EPISODE DE SOINS1
Service : chirurgie digestive
Type de document : CRO initial, CR sortie

EPISODE DE SOINS2
Service :
Type de document : CR consultation
```

Difficulties



- What is the best model/formalism to represent these data?
- Degree of expressivity / granularty
- Separate linguistic processing, knowledge representtion and expert knowledge



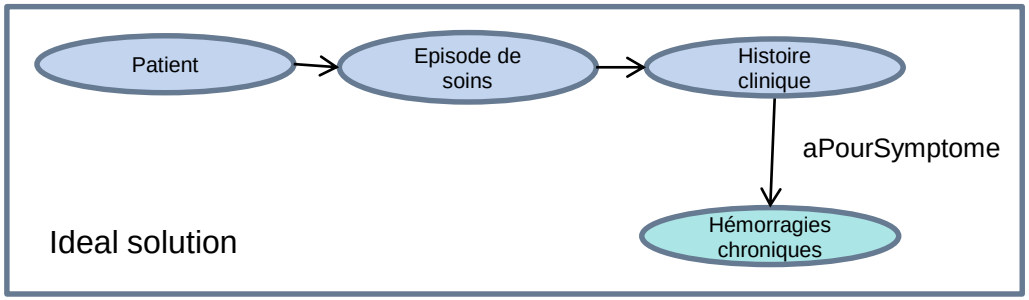
Difficulties

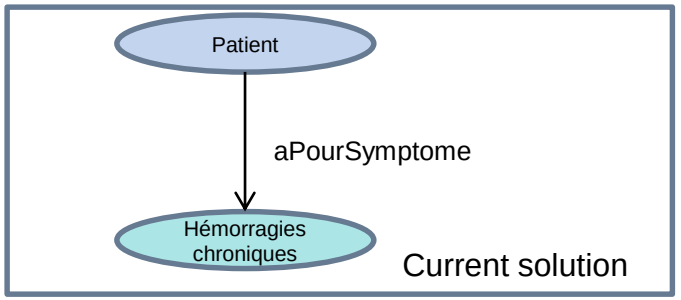
Categorisation of medical terms

How to determine that a symptom belongs to « patient clinical history « or to « reason for consultation »

- Information from terminologies: *hémorragies chroniques* = symptom
- Need medical expert to categories this symptom

On sait qu'il existe une masse sous hépatique dans la région du 2ème duodénum correspondant probablement à une récidive du gastrinome. Cette masse fait saillie à l'intérieur du 2ème duodénum sous la forme d'un champignon bien visible sur le TOGD et l'IRM. Elle est ulcérée et est responsable sans doute des hémorragies chroniques.





Difficulties

Desambiguisation:

Both « diagnostic » and « symptom » are Associated to *fever* in terminologies

Anaphora resolution

L'IRM a été concluant. Il a décelé une anomalie dans le cortex.
On sait qu'il existe une masse sous hépatique. Cette masse fait saillie à l'intérieur du 2é duodénum

Temporal expressions

- Granularity: day
- Reference date coverage
- Notion of interval
- Approximate date (e.g. about ten days ago)
- Special cases (Such medical examination that happened in [T 2 J] will be repeated 15 days later
- Several dates in the same sentence

Coordination

Masses sous et para pancréatiques

– **Abbreviations, spelling errors**

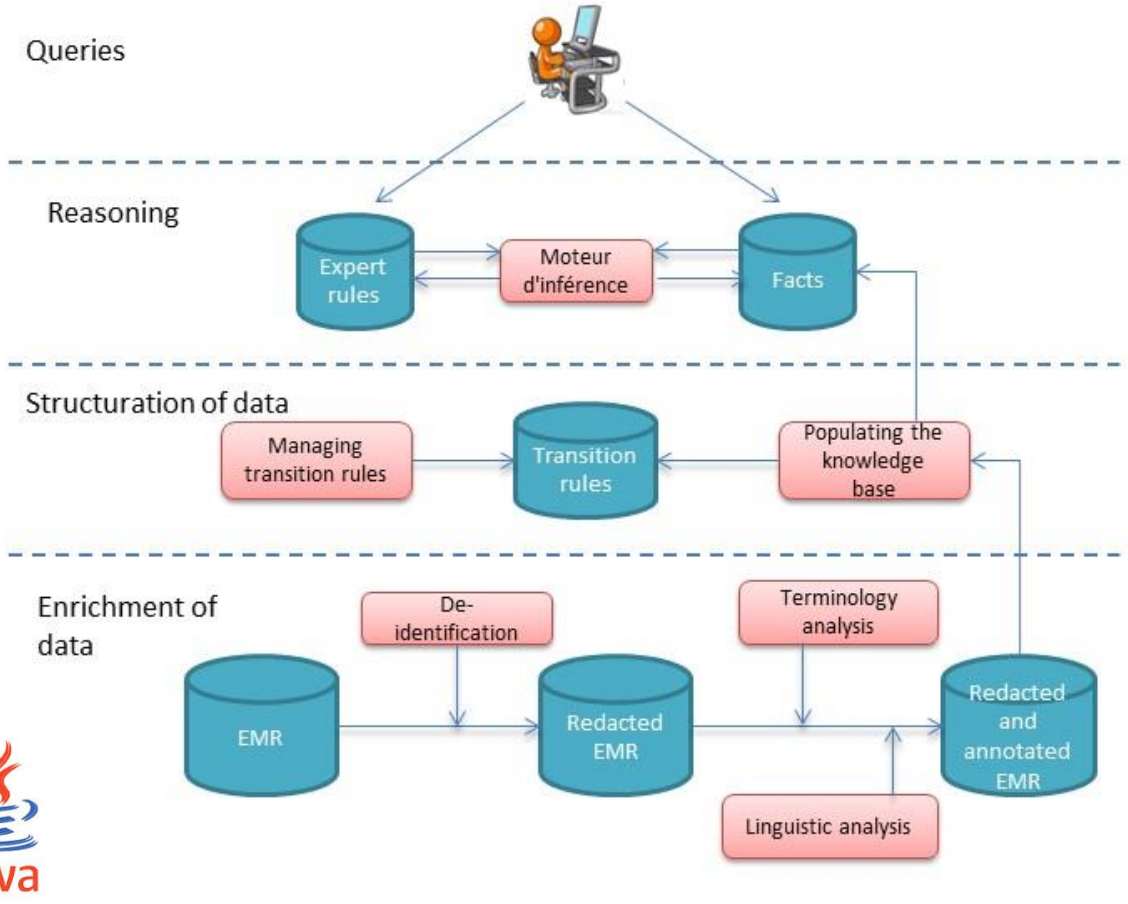
Elle a opté pour une AG+PCA

The biggest difficulty?

Who does what or where are the borders between disciplines?

Our approach

General architecture

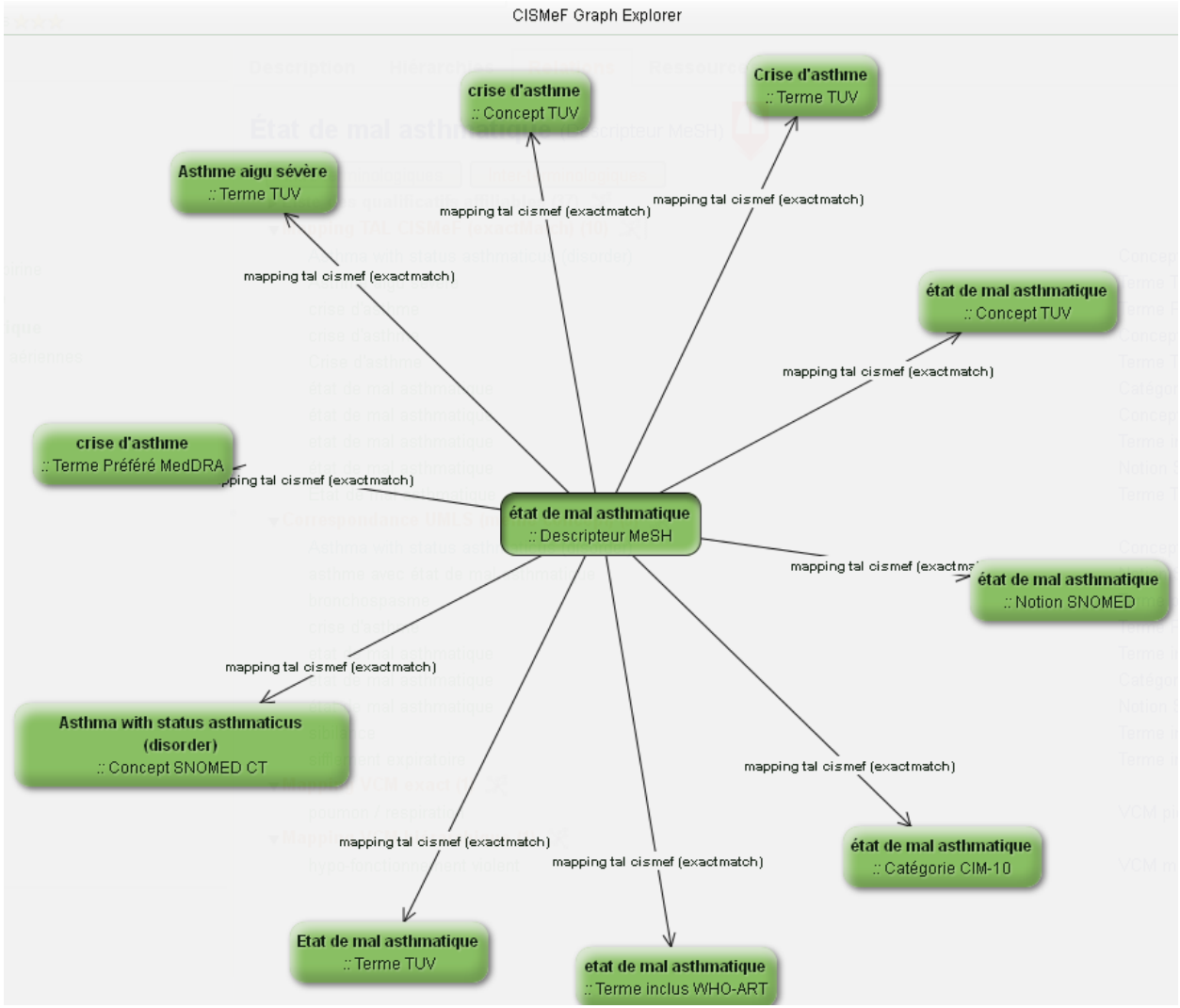


Different steps

- De-Identification and re identification of patient records
- Linguistic processing and terminology
  - Adaptation of general linguistic analyzer to the medical field
  - Temporal processing
- Knowledge modelization
- Development of transition rules
- Development of expert rules
- Evaluation

Terminology retrieval and Linguistic processing

Multi-terminology server



Terminology retreival

- Medical coding
- Preferred term
- Terminology Source
- Semantic type and group (UMLS)
- Super concepts Cismef

La situation actuelle est la suivante résultant d'un bilan très complet fait dans le service du XXX, avec endoscopie, échoendoscopie, scanner et IRM.
Il a été réalisé aussi une hépatectomie droite puis une hépatectomie de la pointe du foie gauche et un traitement par chémoembolisation des masses hépatiques récidivantes.

- MRI
  => Exam
  => Type of exam: imagery
  => magnetic resonance imaging
- Liver
  => Organ

www.hetop.eu

Syntactic Dependency parsing : an example

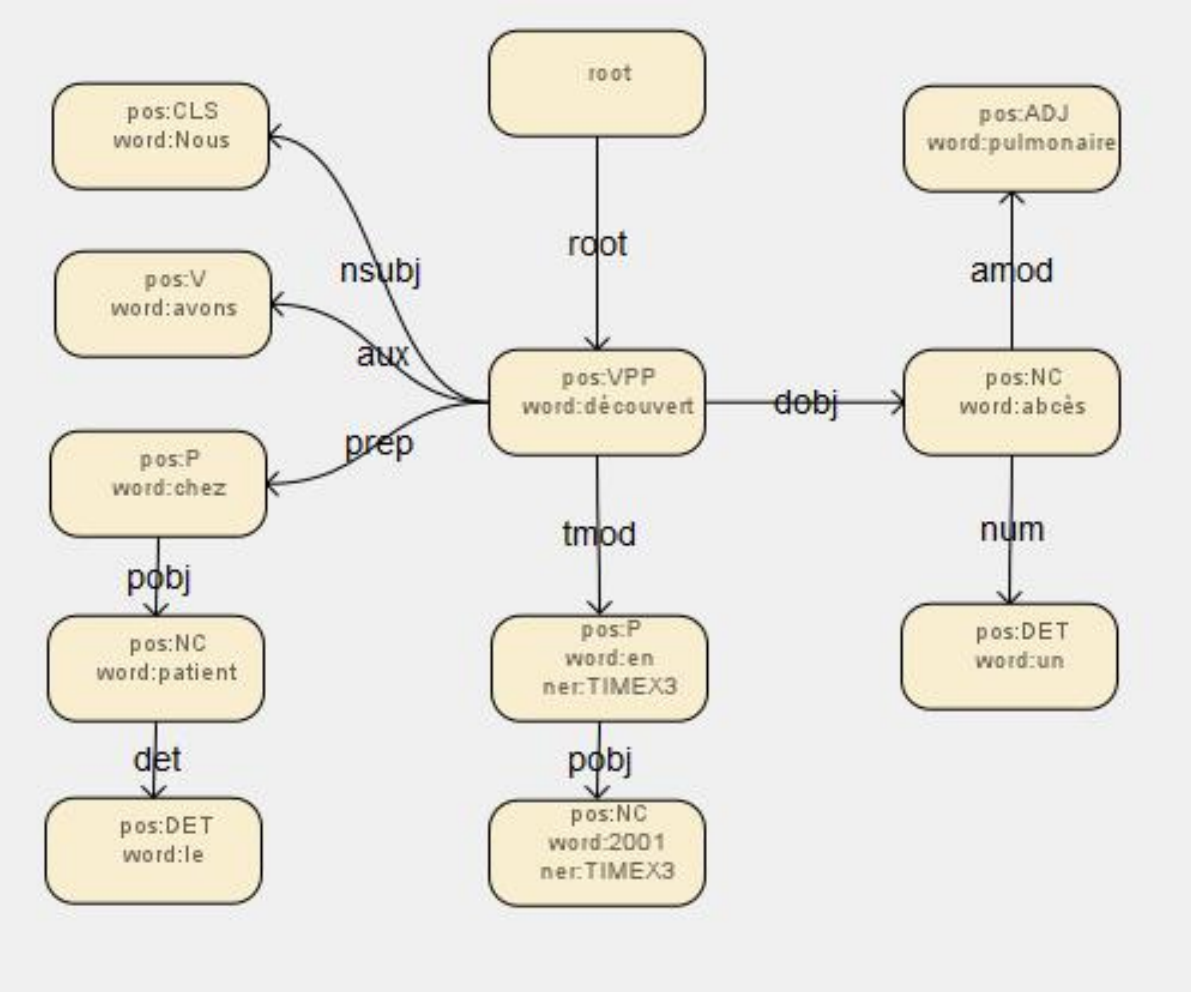
Nous avons découvert un abcès pulmonaire chez le patient en 2001.

(We discovered a lung abscess in this patient in 2001.)

Token	Lemma	POS
Nous	nous	CLS
avons	avoir	V
découvert	découvrir	VPP
un	un	DET
abcès	abcès	NC
pulmonaire	pulmonaire	ADJ
chez	chez	P
le	le	DET
patient	patient	NC
en	en	P
2001	2001	NC
.	.	PONCT

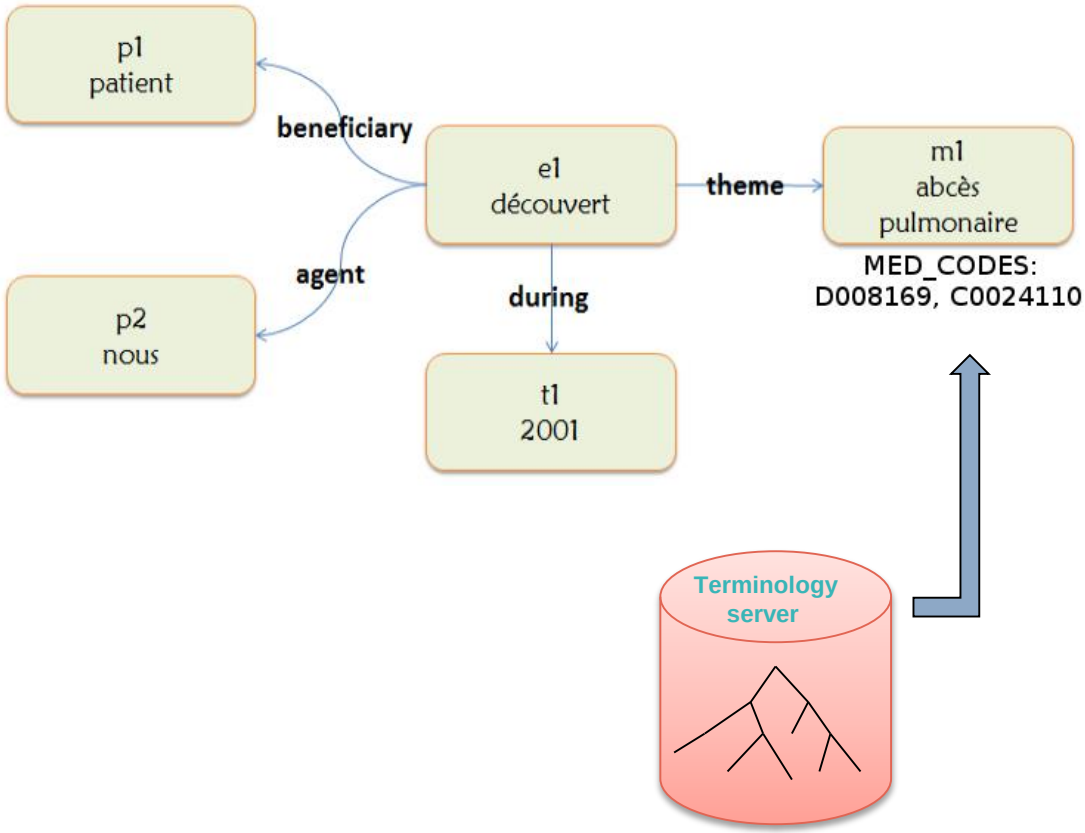
Syntactic Dependency parsing : an example

Syntactic parsing yields a graph structure of syntactic dependencies



Semantic Parsing : an example

Semantic parsing yields a graph structure of semantic dependencies built on both the syntactic dependency graph and access to the medical terminology

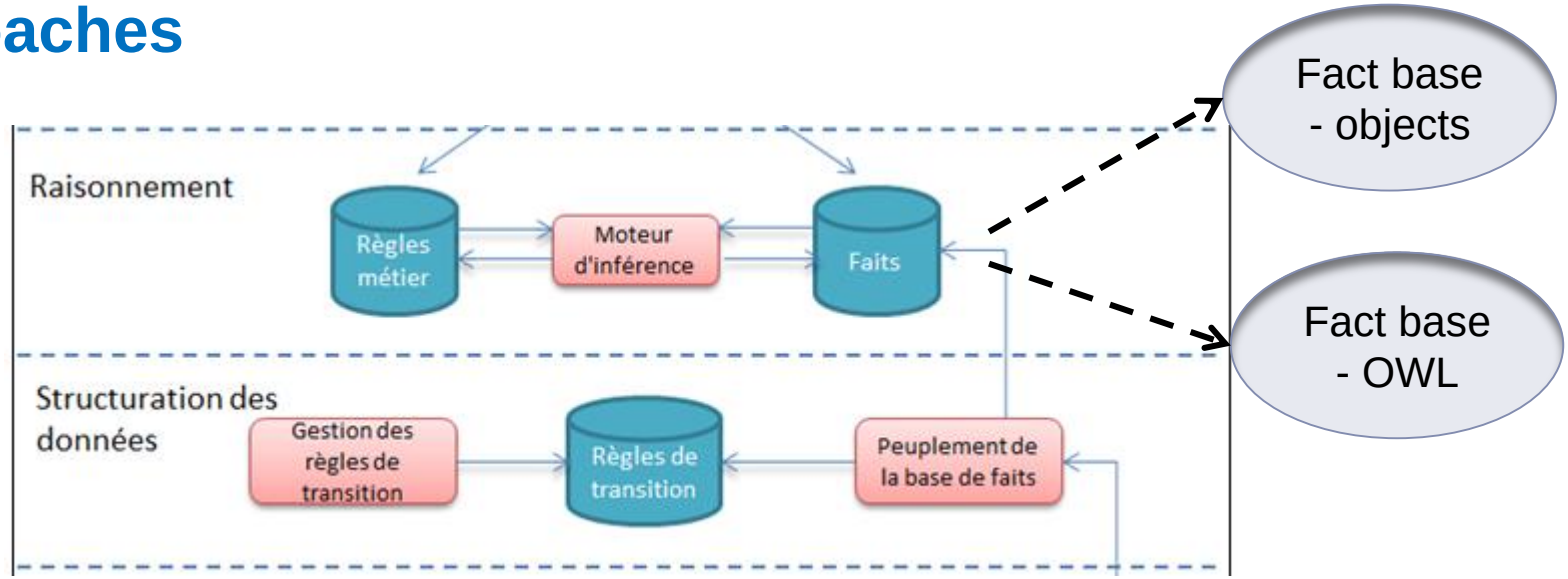


VISEO

Representing Knowledge

Knowledge bases

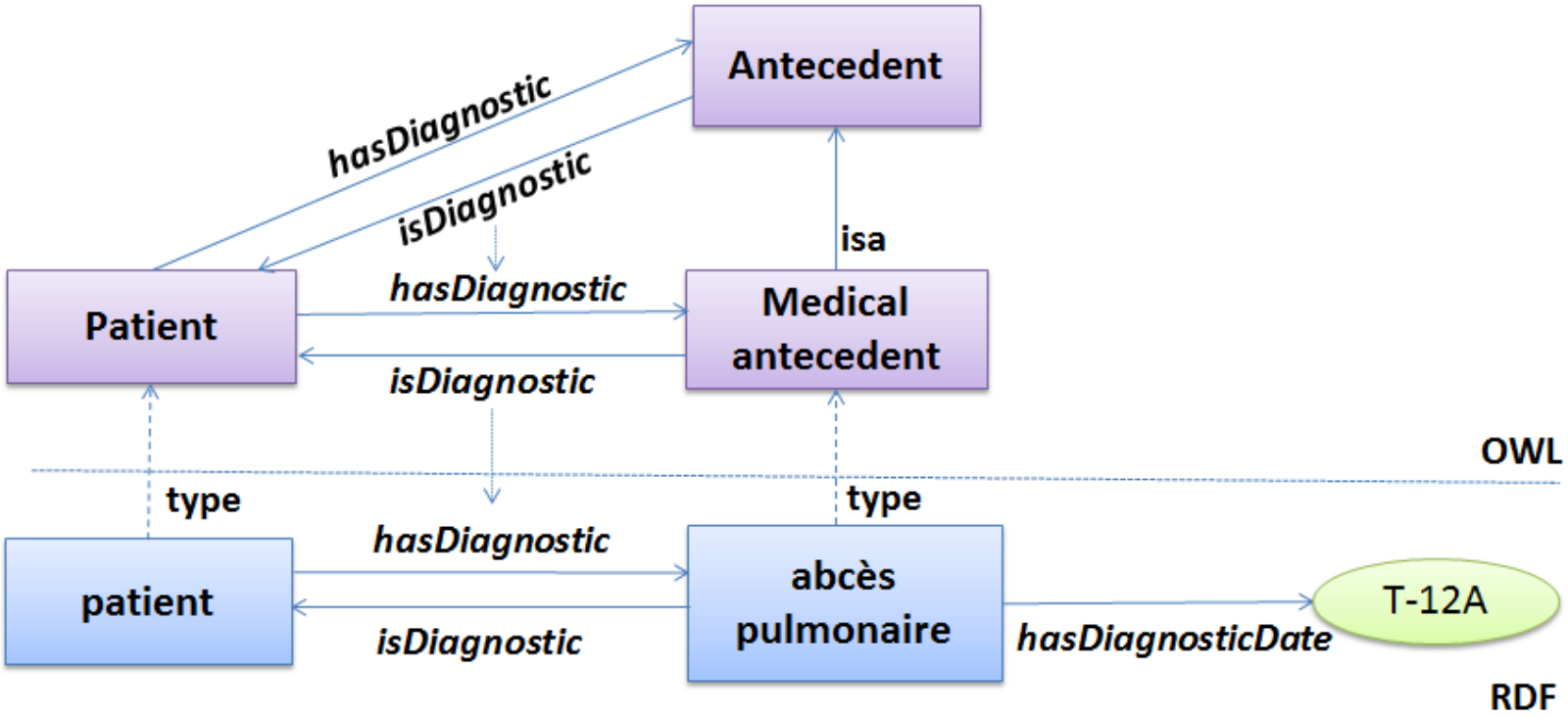
- **2 approaches**



- Classical approach
  - **Relational database**
  - **BRMS (Business Rules Management System)**
- Research Approach
- OWL-DL
  SWRL

PostgreSQL Drools protégé VISEO

OWL representation model example



Transition rules

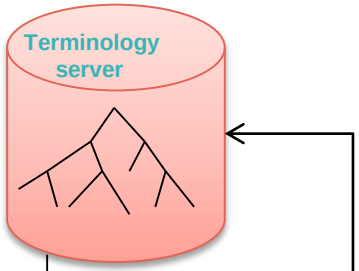
Transition Rules

To enrich information extracted by the linguistic Rules

```
CR initial

...
Nous avons découvert un abcès pulmonaire chez le patient en 2001.
...
```

RT1: if x.lemma="patient" and x.male et x.depRel="suj" then sex.patient=male.
RT2: If date(x) < date(today) then x=antecedent.
RT3: If x.semantic_type=T061 then x=therapeutic_act.

Transition Rules



m1-isa?

Linguistic server

(m1, abcès pulmonaire)
(t1, 2001)
(e1, découvert)
THEME (e1, m1)
TIME (t1, T-12A)
DURING (e1, t1)

If THEME(e1,m1) and DURING(e1,t1) and t1 > d
Then (m1-isa-Antecedent)

d = T

If THEME(e1,m1) and DURING(e1,t1) and t1 > d and (m1-isa-DISEASE)
then (m1-isa-MedicalAntecedent)

d = T

For the same rule, if (m1-isa-SURGICAL PROCEDURE)
Then (m1-isa-SurgicalAntecedent)

- **Knowledge server**
  - (abcès pulmonaire, *isa*, Antecedent)
  - (abcès pulmonaire, *isa*, MedicalAntecedent)
  - (abcès pulmonaire, *isa*, SurgicalAntecedent)

VISEO

Expert Rules

Expert Rules

Defined by medical experts after transition rules using a metalanguage

- Intensive Care:

If there is a sentence mentioning explicitely an HAI

If in « near by :close» sentences there is an entity refereing to an infection 5e;g; germ) , an antibiotic in a 2 days time frame after T0 (entry date in Intence care unit). In addition, the patient should be alive and no infection should have been detected before T0;

If patient has an infection at T0 or if he/she died during his/her stay at the hospital and if at least two occurrences of the following type of events are detected within a time frame greater than 2 days after T0: infection, antibiotic prescribed, temperature, use of invasive equipment.

- Surgery:

If there is an sentence mentioning a surgical site infection.

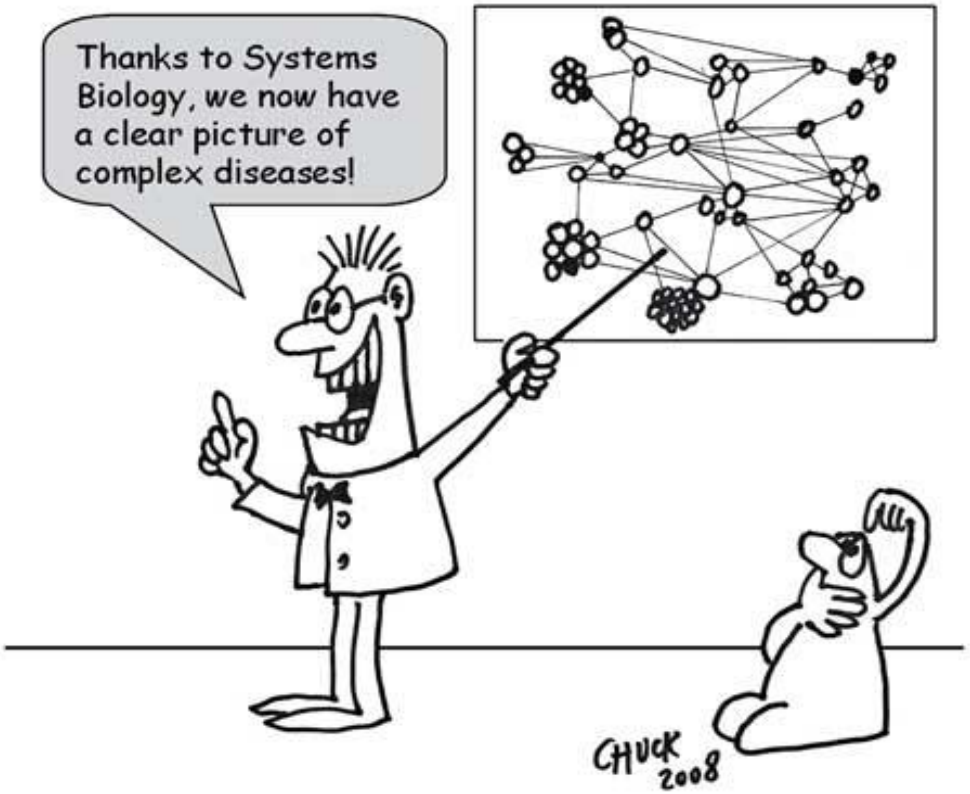
If one of the following events can be detected within a timeframe greater than T0: infections, antibiotic prescribing, use of antiseptic, germ detected, bacteriological exam.

VISEO

Other challenging issues we are facing

- Metalanguage to write expert rules: impact to the rest of the model? Templates are not enough but they are safe enough
- Metalanguage to query the database: from templates to NLP? Combined information coming from different analysis
- How to visualize the search results so that make sense? So that they are understandable? How much information? From how many patient? Level of granularity?
- How to integrate different data in the Knowledge base and reason on them: biomedical data, images etc.
- How to integrate this system into hospital information systems?

Many thanks!


Thanks to Systems Biology, we now have a clear picture of complex diseases!
CHUCK 2008