# l'IA au service de la formation en Médecine: retour d'expérience du projet SIDES 3.0

FINANCÉ PAR ANR
INVESTISSEMENTS D'AVENIR
ANR-16-DUNE-0002

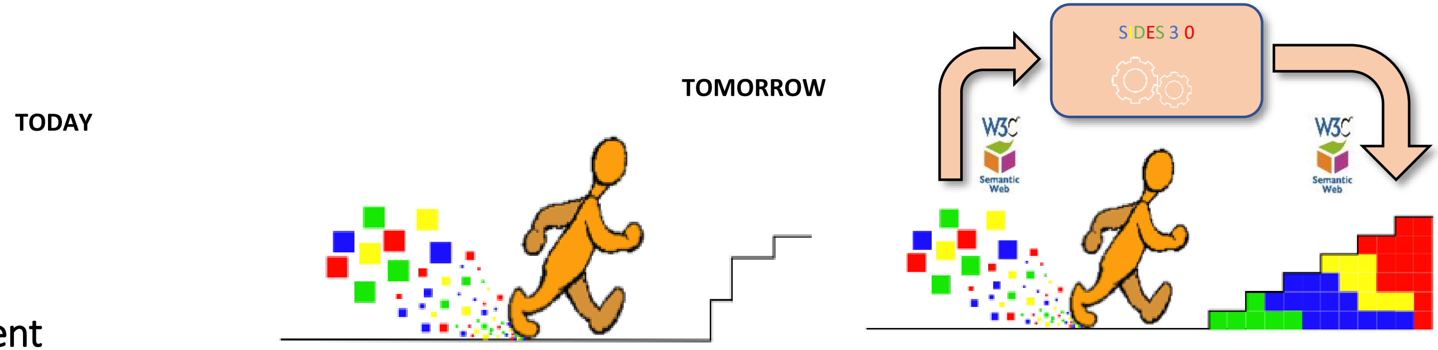
**Marie-Christine Rousset**
**LIG, Univ. Grenoble Alpes & Institut Universitaire de France**
**joint work with**
**Olivier Palombi (Univ. Grenoble Alpes, Inria-LJK, CHU, UNESS)**
**and**
**Fabrice Jouanot (LIG, Univ. Grenoble Alpes)**

UNIVERSITÉ Grenoble Alpes
LIG
LABORATOIRE JEAN KUNTZMANN MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE
GRICAD GRENOBLE ALPES RECHERCHE INFRASTRUCTURE DE CALCUL INTENSIF ET DE DONNÉES
UNESS.fr Université Numérique En Santé et Sport
Médecine
Pharmacie
Odontologie
Sport
Inria INVENTEURS DU MONDE NUMÉRIQUE
wimmics
privatics
ENS | PSL RESEARCH UNIVERSITY PARIS
samsei stratégie d'apprentissage des métiers de santé en environnement immersif
VISEO DIGITAL MAKERS

# SIDES 3.0 : DUNE project (August 2017-July 2021)

## Towards data-driven personalized self-assesment and training

A successful experience of online graduation, training and assessment for all medical students in France




**Produces a huge amount of low-level activity traces that are exploited by database administrators for predefined tasks**

**facilitates the empowerment of end-users in data analytics**

**by using Semantic Web technologies and Linked Data principles**

# Ontology-Based Data Access: OBDA

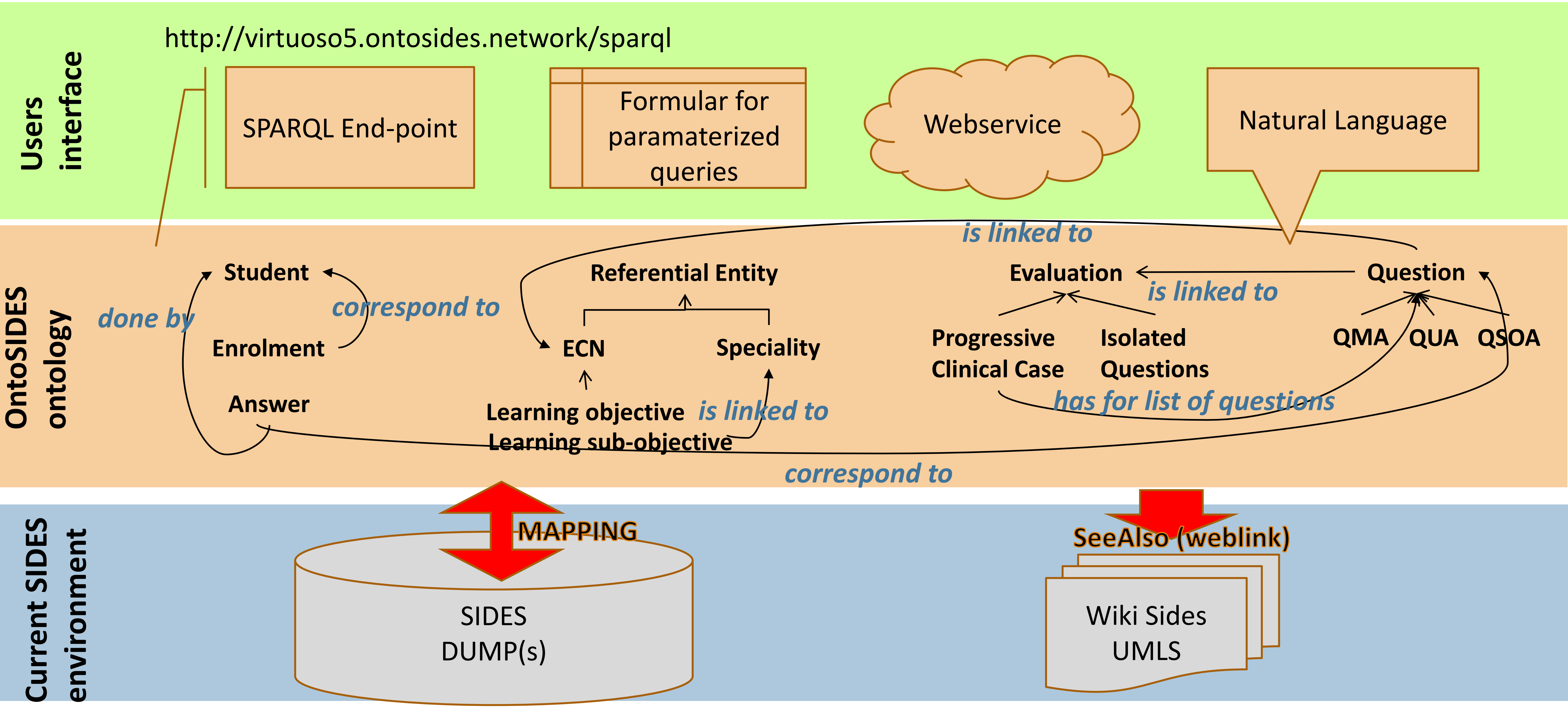
- A novel paradigm at the crossroad of Artificial Intelligence and Databases
  - a **domain ontology** serves as a **mediator** for expressing **users queries**

- Ontology: a formal specification of a domain of expertise
  - a structured vocabulary (classes and properties) meaningful for domain experts
  - a conceptual yet computational model of a domain

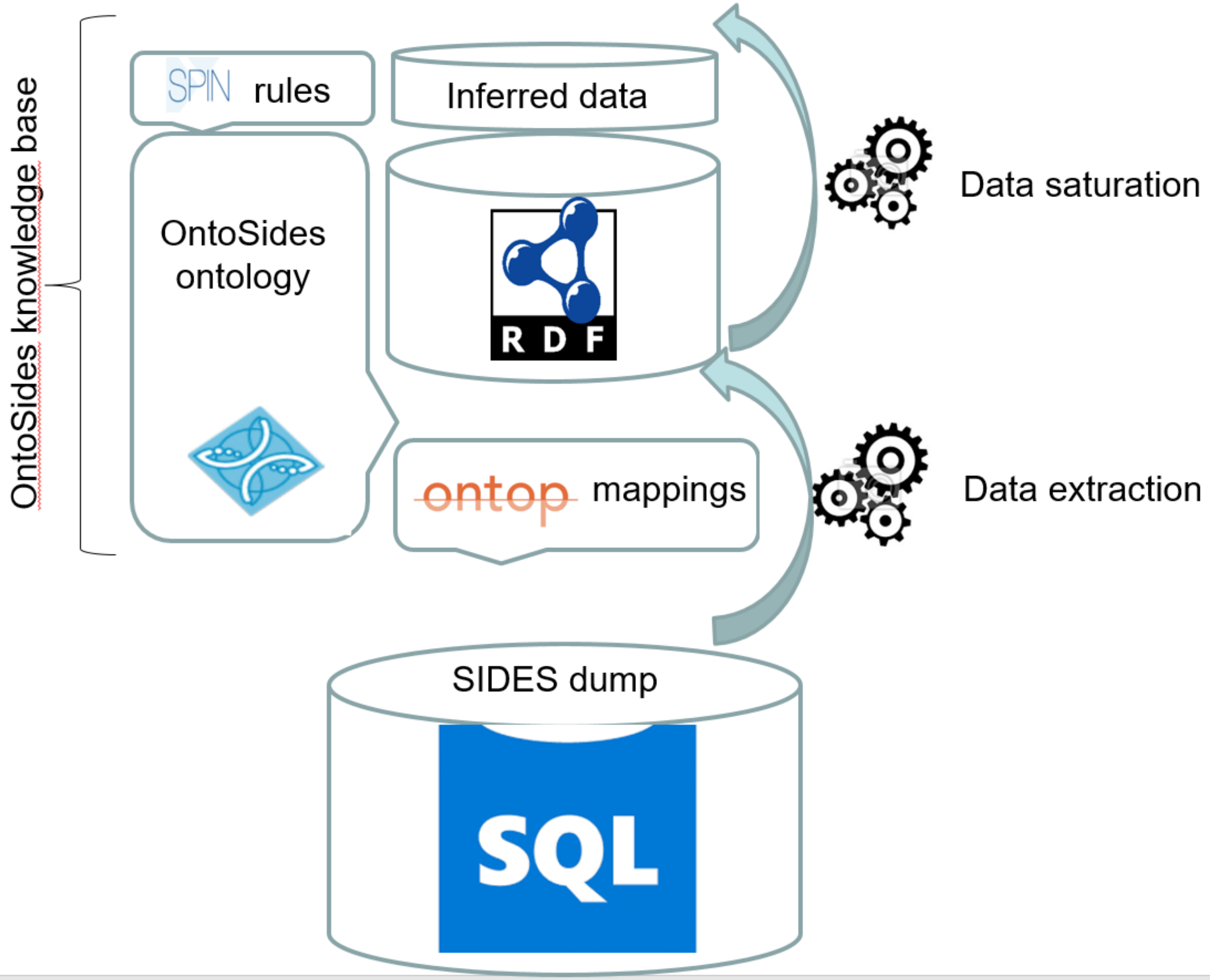
$\Rightarrow$ **humans** can express their data analysis needs using terms of a **shared vocabulary in their domain of interest or of expertise**

$\Rightarrow$ **computer systems** can base decisions on **reasoning on domain knowledge**

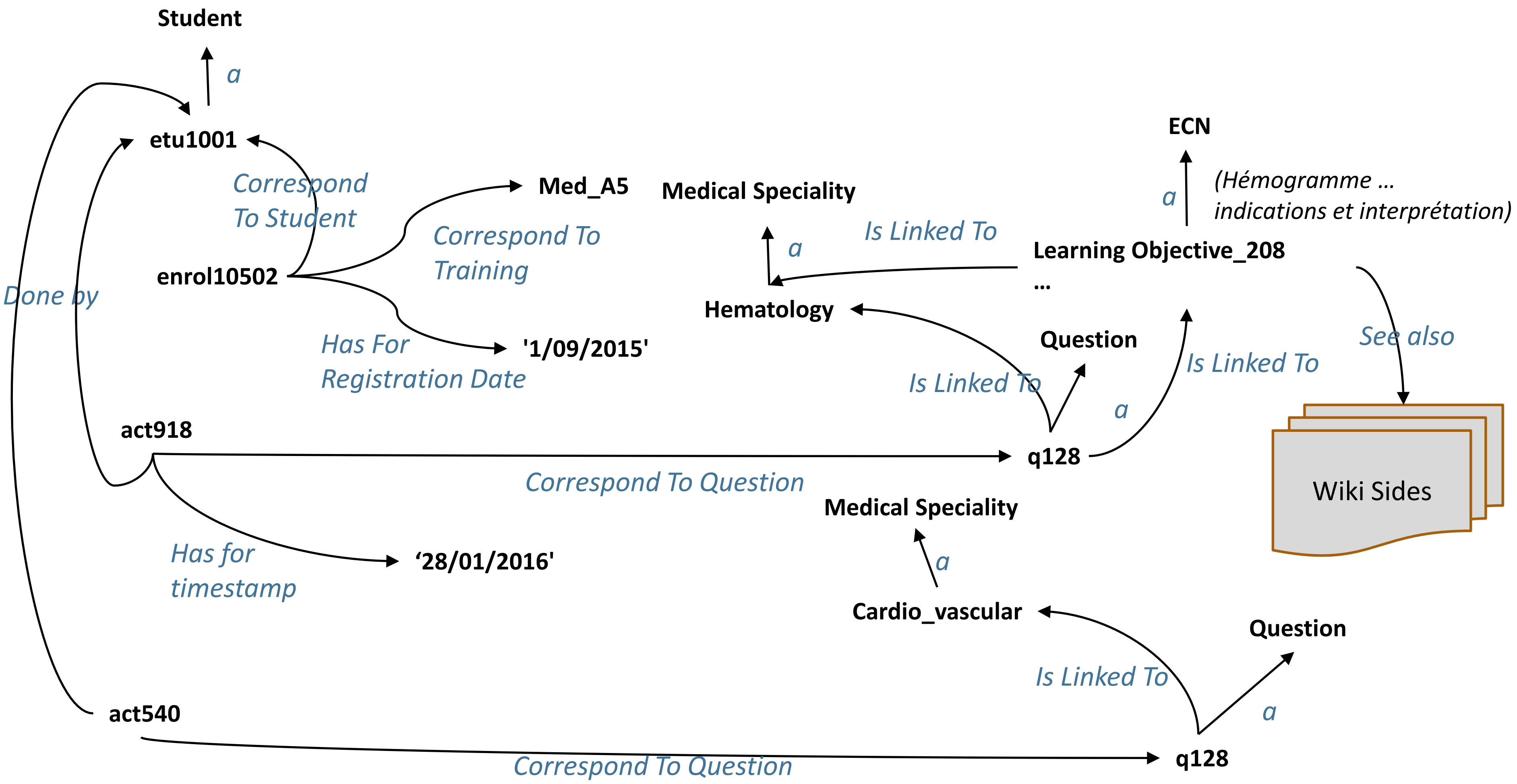
# OntoSIDES ontology: interface for data access and analytics



# Mapping-based automatic data extraction guided by the ontology



# Zoom on data linkage



# OntoSIDES knowledge graph

- the linked data layer of SIDES 3.0
  - describes **training** and **assessments activities** performed by more than **145,000 students** in Medicine **over almost 6 years**
    - exams and training tests are made of **multiple choices questions**
    - students **answers** are described at the granularity of **time-stamped clicks of answers** done by students for choosing among the proposals of answers (correct or distractors) associated to questions

⇒**7,8 billions triples** with almost **400 millions clicks** coming from the answers of students to almost **1,4 million questions.**

# Knowledge Graphs

- Modern knowledge representation formalism based on **RDF data model**
  - more flexible than the relational model
  - adapted to data/knowledge sharing between distributed data sources over the Web
- a set of triples <subject, property, object/value>
  - **subject, property** and **object** are URIs (http Uniform Resource Identifiers)
  - **dereferencable URIs** (pointers to Web pages) versus **local URIs**
  - **value** is a literal (string, integer, date, boolean)
- Tractable reasoning
  - Simple knowledge (OntoSides ontology: 52 classes, 50 properties, 1400 instances, 18 rules)
  - Big data associated with a powerful query language (SPARQL)

# Illustration : RDF modeling multiple choice questions in OntoSides

**Q30986** has_for_textual_content **"Concernant la péritonite appendiculaire, donnez la ou les propositions exactes :"** ;

is_linked_to_the_medical_speciality **digestive_surgery**

has_for_proposal_of_answer_ **prop98552 [** has_for_textual_content **"les signes infectieux sont présents d'emblée » ;**

has_for_correction **« true »]**

**prop98553** [ has_for_textual_content

**"il n'y a pas de défense abdominale ou de contracture" ;**

has_for_correction **« false »]**

**prop98604[** has_for_textual_content

**"elle peut se présenter comme une occlusion fébrile" ;**

has_for_correction **« true»]**

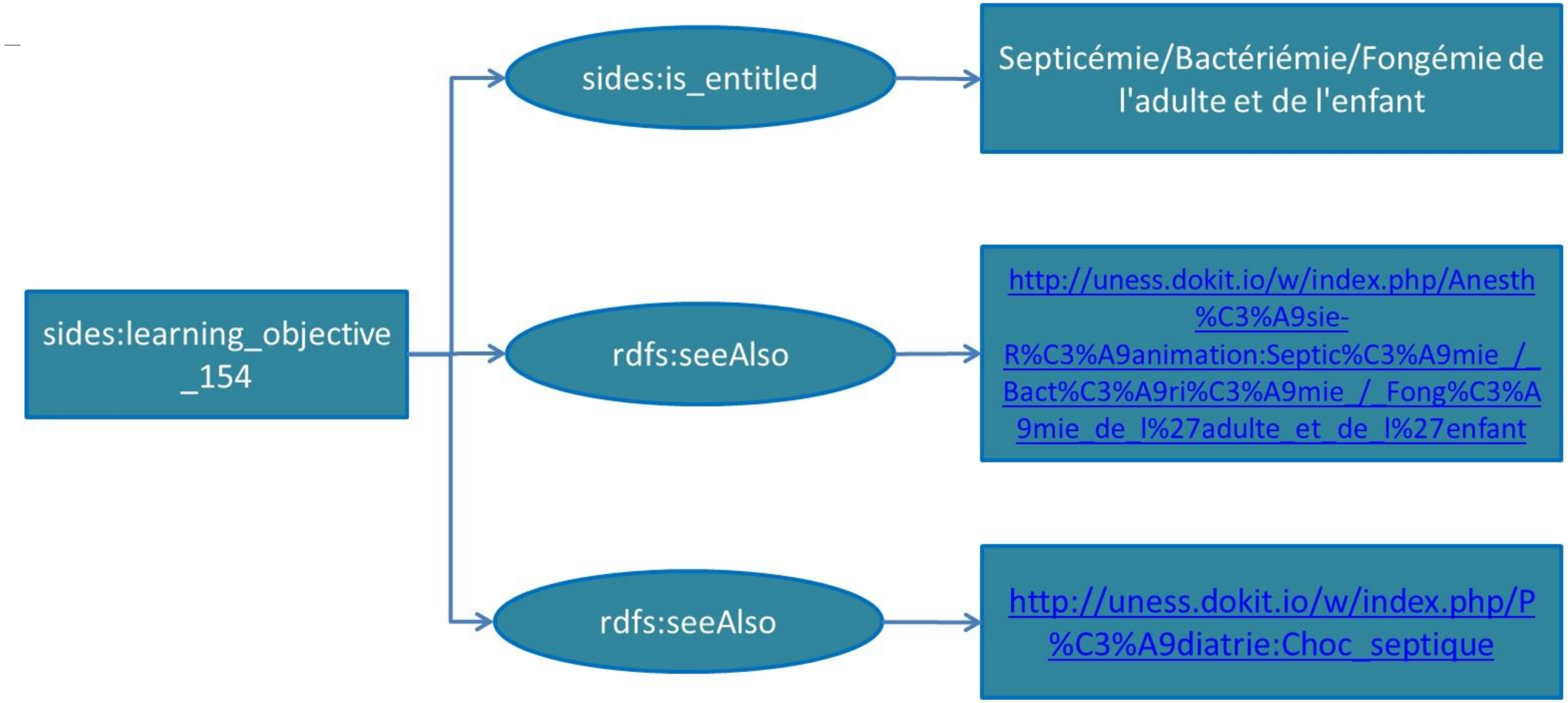
**prop98605[** has_for_textual_content **"il n'y a pas de pneumopéritoine" ;**

has_for_correction **« true»]**

**prop98606[** has_for_textual_content **« le traitement est chirurgical" ;**

has_for_correction **« true»]**

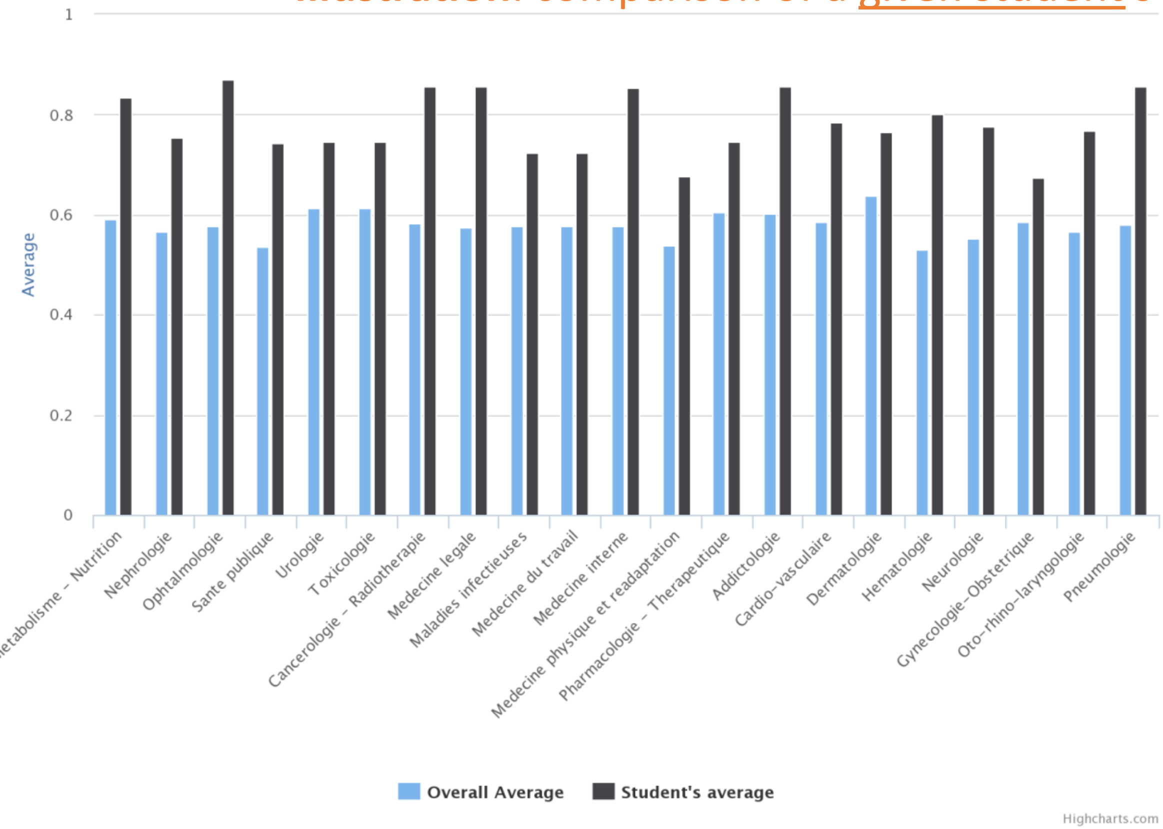
# Illustration: Learning objectives linked to web pages of the wiki SIDES (*)




(*) official educational content provided by the association of French Medical colleges covering the French pre-residency program examination

# Parametrized queries:
# a step towards personalized and explainable data analytics

**Illustration**: comparison of a given student's results with results of all students by medical specialty




```
SELECT ?specialty  ?globalAverage  ?studentAverage
WHERE  {
  {  SELECT ?specialty ( AVG(?result) AS ?globalAverage)
   WHERE { ?answer sides:has_for_result ?result .
              ?answer sides:done_by ?student .
              ?answer sides:correspond_to_a_question ?q .
              ?q sides:is_linked_to_the_medical_speciality ?specialty . }
     GROUP BY ?specialty } .
 { SELECT ?specialty (AVG(?result) AS ?studentAverage)
   WHERE { ?answer sides:has_for_result ?result .
               ?answer sides:done_by sides:etu12402 .
               ?answer sides:correspond_to_a_question ?q .
               ?q sides:is_linked_to_the_medical_speciality ?specialty .}
   GROUP BY ?specialty} .
}
```

## Aggregate queries (SPARQL 1.1)

- not supported by query rewriting approaches
- **requires data completeness**

# Data incompleteness

- Problematic for conducting **well-grounded learning analytics**
  - partial answers for basic Select From Where queries
  - **wrong results for aggregate or counting queries**

- This may occur on some specific properties likely to be involved in aggregate queries to define dimensions
  - is_linked_to_medical_specialty (from questions to medical specialties)
    - **13% questions** have been explicitly **linked** by their authors **to medical specialties**
  - is_linked_to_ECN_item (from questions to learning objectives)
    - **12% questions linked to learning objectives**

# Knowledge graph completion and enrichment

- Knowledge graph completion
  - automatically inferring missing facts from existing ones
    - between **questions** and **medical specialties** or **learning objectives**

- Knowledge graph enrichment
  - automatically discovering links with external reference knowledge graphs or standard ontologies
    - Standard **UMLS (Unified Medical Language System)** medical terminologies like **MeSH (Medical Subject Headings)** and **SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms)**

=> Can be modeled as **classification** or **matching** problems

- depending on the availability of textual description of the target entities and of training data

# Inferring links between **questions** and **medical specialties**

- can be solved as a **multi-label** and **multi-class** classification problem
  - **multi-class**: 31 possible classes (the different medical specialties)
  - **multi-label**:  question can be linked to more than one medical specialty
  - **training set:** the 149,000 (13%) questions (with their textual description) for which the property is linked to the medical specialty is valued
- using several classifiers
  - Naive Bayes, Maximum Entropy, CNN (Convolutional Neural Network)

# Inferring links between **questions** and **learning objectives**

- a **multi-label** and **multi-class classification** problem
  - **multi-class:** 362 possible classes (the different learning objectives) and **multi-label**
  - **training set** : 144,000 (12%) questions for which the corresponding property is valued
- can be also solved as a **matching** problem between the textual descriptions of the questions and of the learning objectives
  - only for the **236 learning objectives** that have a textual description

using several variants of TF-IDF ranking function used in **Information Retrieval** to return the top-k learning objectives for each question

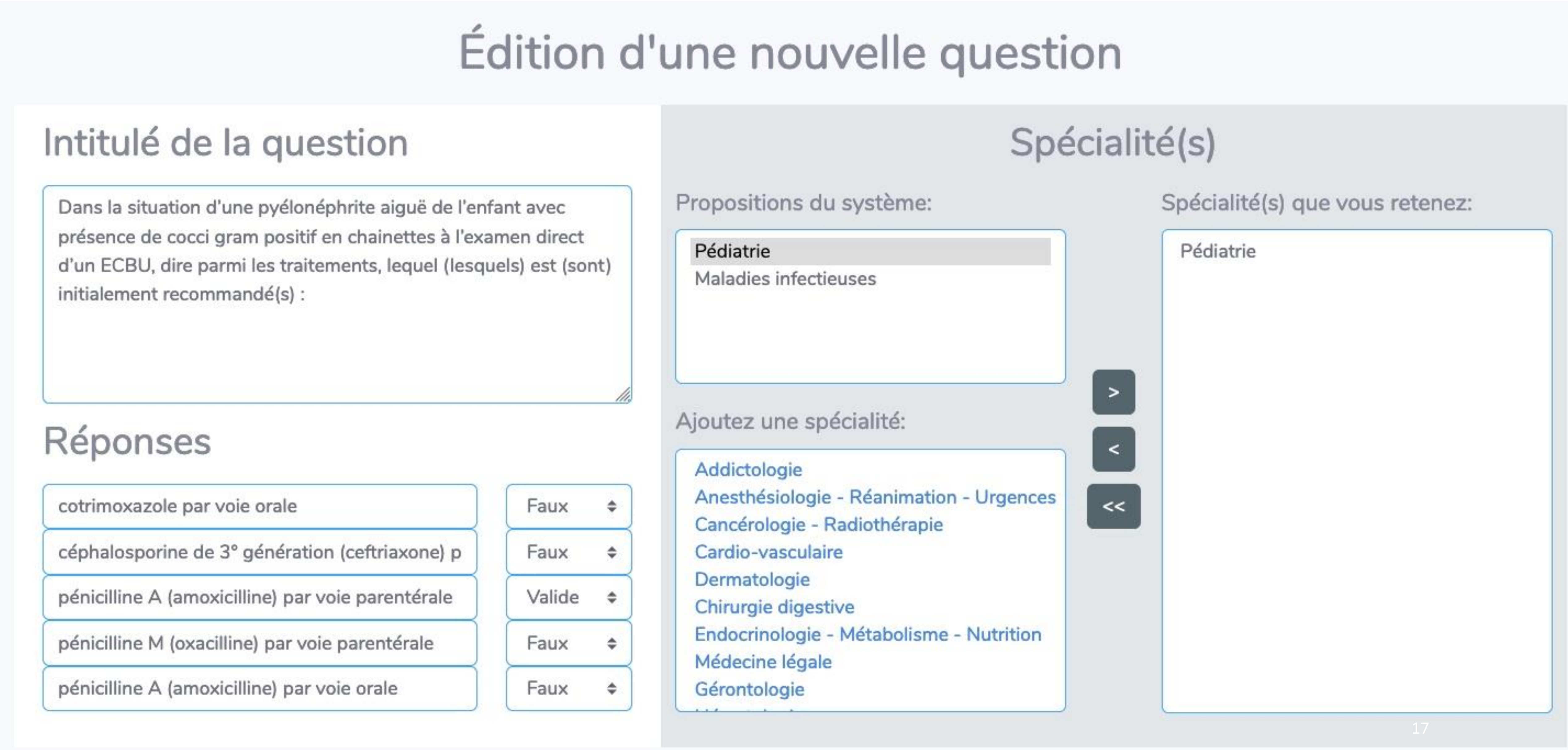
# Experimental results for classification

| Dataset | Classifier | Hits@1 | Hits@2 | Hits@5 | Hits@10 | MRR |
|---|---|---|---|---|---|---|
| Dataset1 | Naive Bayes classifier | 73.8% | 83.1% | 84.2% | 84.3% | 79.9% |
| | Maximum Entropy classifier | 75.1% | 88.9% | 95.4% | 96.8% | 84% |
| | CNN classifier | 76.4% | 89.4% | 96.3% | 98.5% | 85.2% |
| Dataset2 | Naive Bayes classifier | 56.4% | 64.8% | 67.8% | 67.9% | 61.5% |
| | Maximum Entropy classifier | 68% | 81.7% | 90.6% | 93.6% | 78.2% |
| | CNN classifier | 66.4% | 78.9% | 88.8% | 93.4% | 76% |

**Dataset1**: 149145 questions -> 31 medical specialties
**Dataset2**: 144708 questions -> 362 learning objectives
**Hits@k (Precision at k):** average number of times a correct result appears in the top-k answers
**MRR (Mean Reciprocal Rank)**: average of the rank inverses of the first correct answer

- **All the classifiers perform better on Dataset1 than on Dataset2**
  - the number of classes for Dataset2 is more than 10 times the number of classes for Dataset1 for almost the same number of items to classify
- **Naive Bayes outperformed by Maximum Entropy and CNN**
- Maximum Entropy gives slightly better results than CNN classifier on Dataset2
- **In more than 96% (93%) of the cases, the correct medical specialties (learning objectives) are returned in the top-10 answers**

# Application: suggestions to teachers while editing questions



# Application to OntoSides completion

- Focus on Hits@1
  - set up a threshold to obtain the desired balance between high precision and an acceptable recall.

| Dataset | Threshold | Hits@1 | % Questions |
|---|---|---|---|
| Dataset1 | 0 | 75.1% | 100% |
| | 0.3 | 78.2% | 96.5% |
| | 0.5 | 82.1% | 84.5% |
| | 0.7 | 86.6% | 66% |
| | 0.9 | 91% | 43% |
| | 0.95 | 92.6% | 33.3% |
| | 0.99 | 94.5% | 18.3% |

| | Threshold | Hits@1 | % Questions |
|---|---|---|---|
| Dataset2 | 0 | 67.5% | 100% |
| | 0.1 | 75.2% | 85% |
| | 0.3 | 81.3% | 61% |
| | 0.5 | 86% | 43% |
| | 0.7 | 89% | 29% |
| | 0.8 | 91% | 23% |
| | 0.9 | 92.4% | 16% |

- if we consider that **a precision more than 90% is acceptable** for adding reliable links from questions to medical specialties (Dataset1), and to learning objectives (Dataset2), it suffices to **fix the threshold to 0.9 for Dataset1 and 0.8 for Dataset2**.
- it decreases the percentage of questions for which such links can be added (**43% of links from questions to medical specialties**, and **23% of links from questions to learning objectives**).

# Comparative results of classification and matching

**Dataset3:** 108818 questions -> 236 learning objectives with textual description

| Method | | Hits@1 | Hits@2 | Hits@5 | Hits@10 | MRR |
|---|---|---|---|---|---|---|
| unsupervised | Jelinek-Mercer applied to bags of words | 44.5% | 58.2% | 72.2% | 80.9% | 57% |
| | BM25 applied to bags of semantic terms | 51.5% | 64% | 75.7% | 81.9% | 62.4% |
| supervised | Naive Bayes classifier | 56.2% | 64.6% | 67.5% | 67.7% | 61.3% |
| | Maximum Entropy classifier | 68.2% | 81.4% | 90.4% | 93.6% | 78% |
| | CNN classifier | 66.2% | 78.9% | 88.6% | 93% | 75.9% |

- **The "bag of semantic terms" representation** leads to more accurate results than **the "bag of words" representation**
  - **Semantic terms** are **medical concepts** that are **automatically extracted** from the textual descriptions **by** using the **SIFR BioPortal Annotator** (LIRMM, Clément Jonquet) applied to the French versions of the reference biomedical terminologies MESH and SNOMED
- Not surprisingly, **supervised classification methods outperform the unsupervised ones** (except Naïve Bayes at precision 5 and 10)
- However, **unsupervised methods provide good results (above 80%) at precision 10**

# Automatic Discovery of Links with External Ontologies

- From the 236 **learning objectives** with a textual description to standard **medical concepts** described in biomedical ontologies
  - UMLS concepts in MeSH and SNOMED CT (French and English version)
- Method overview
  - applied to the set of learning objectives, **seen as a corpus**, each learning objective being seen as a document described by a bag of medical concepts
    - Computation of the term frequency (TF) and the inverse document frequency (IDF) for each medical concept present in the corpus
    - Filtering out the medical concepts with a low IDF (below a certain threshold fixed experimentally )
    - For each learning objective, return the top-k medical concepts (ordered by decreasing TF): k is also fixed experimentally

# Two-step validation: method and results

No training dataset available

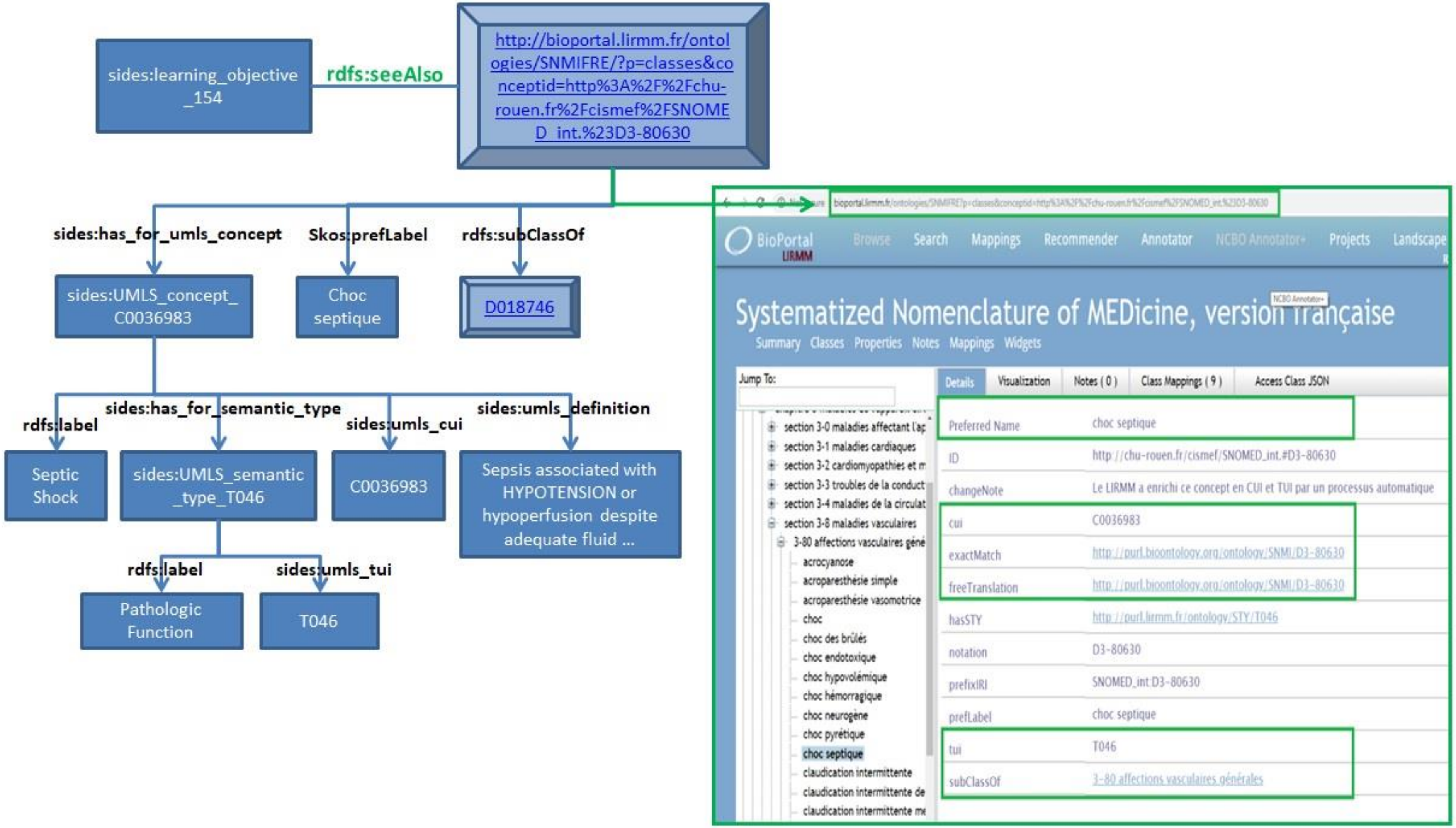
- First step of validation on 15 learning objectives with a domain expert (O.Palombi)
  - calibration of the parameters to get the best precision at precision k
- Second step of validation with medical experts through an online validation interface
  - answers of experts on 96 learning objectives

| #Evaluated Learning Objectives | #MSHFRE and SNMIFRE Semantic Terms | P@5 |
|---|---|---|
| 96 | 510 | 94.5% |

=> OntoSIDES enrichment by adding useful triples in addition to the discovered links

- 15371 triples added

# Example



sides:learning_objective_154
rdfs:seeAlso
http://bioportal.lirmm.fr/ontologies/SNMIFRE/?p=classes&conceptid=http%3A%2F%2Fchu-rouen.fr%2Fcismef%2FSNOMED_int.%23D3-80630
sides:has_for_umls_concept
Skos:prefLabel
rdfs:subClassOf
sides:UMLS_concept_C0036983
Choc septique
D018746
rdfs:label
sides:has_for_semantic_type
sides:umls_cui
sides:umls_definition
Septic Shock
sides:UMLS_semantic_type_T046
C0036983
Sepsis associated with HYPOTENSION or hypoperfusion despite adequate fluid ...
rdfs:label
sides:umls_tui
Pathologic Function
T046
BioPortal
LIRMM
Browse
Search
Mappings
Recommender
Annotator
NCBO Annotator+
Projects
Landscape
Systematized Nomenclature of MEDicine, version française
Summary
Classes
Properties
Notes
Mappings
Widgets
Jump To:
Details
Visualization
Notes ( 0 )
Class Mappings ( 9 )
Access Class JSON
section 3-0 maladies affectant l'ap
section 3-1 maladies cardiaques
section 3-2 cardiomyopathies et m
section 3-3 troubles de la conduct
section 3-4 maladies de la circulat
section 3-8 maladies vasculaires
3-80 affections vasculaires géné
acrocyanose
acroparesthésie simple
acroparesthésie vasomotrice
choc
choc des brûlés
choc endotoxique
choc hypovolémique
choc hémorragique
choc neurogène
choc pyrétique
choc septique
claudication intermittente
claudication intermittente de
claudication intermittente me
Preferred Name
choc septique
ID
http://chu-rouen.fr/cismef/SNOMED_int.#D3-80630
changeNote
Le LIRMM a enrichi ce concept en CUI et TUI par un processus automatique
cui
C0036983
exactMatch
http://purl.bioontology.org/ontology/SNMI/D3-80630
freeTranslation
http://purl.bioontology.org/ontology/SNMI/D3-80630
hasSTY
http://purl.lirmm.fr/ontology/STY/T046
notation
D3-80630
prefixIRI
SNOMED_int:D3-80630
prefLabel
choc septique
tui
T046
subClassOf
3-80 affections vasculaires générales

# In summary

## Specific completion and enrichment problems

- targeting property  of interest guided by the needs in data analytics of domain experts.

## Generic methodology

- exploiting textual information found in knowledge graphs through datatype properties or rdfs:seeAlso links to web pages.

## Experimental results

- demonstrated that it can effectively perform big knowledge graph completion and enrichment with a precision up to 95%

# Ongoing and future work

- Automatic generation of personalized Quizz
  - based on models that evaluate jointly the level of students and the difficulty of questions

    Inria Nice WIMMICS and ENS Ulm Cognitive Science partner

- Full deployment of SIDES NG as a moodle front-end on top of OntoSIDES

    UNESS

- SIDES LAB : new ANR project (starting in March 2022)
  - Large-scale in situ experimentation of several evidence-based strategies to enhance learning

=> continuously improve the learning platform to offer medical students **individually optimised learning paths based on the newly obtained results**.

# https://sides3.uness.fr/

**Artificial Intelligence In Medicine, Volume 96, May 2019, Pages 59-67**

OntoSIDES: Ontology-based student progress monitoring on the national evaluation system of French Medical Schools.

*Authors:* Olivier Palombi, Fabrice Jouanot, Nafissetou Nziengam, Behrooz Omidvar-Tehrani, Marie-Christine Rousset, Adam Sanchez.

https://doi.org/10.1016/j.artmed.2019.03.006

**AMEE 2019 Symposium: Understanding student behaviour: the role of digital data**

E-poster : https://my.ltb.io/#/viewStack/BHKJS