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

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)

ANR-16-DUNE-0002
SIDES 3.0 : DUNE project (August 2017-July 2021)

Towards data-driven personalized self-assessment and training

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

Produce 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
    ⇒ humans can express their data analysis needs using terms of a shared vocabulary in their domain of interest or of expertise
    ⇒ computer systems can base decisions on reasoning on domain knowledge
OntoSIDES ontology: interface for data access and analytics

http://virtuoso5.ontosides.network/sparql

SPARQL End-point

Formular for paramaterized queries

Webservice

Natural Language

Users interface

OntoSIDES ontology

done by

Student

Enrolment

Answer

correspond to

Referential Entity

Learning objective

Speciality

Learning sub-objective

is linked to

Evaluation

is linked to

Progressive Clinical Case

Isolated Questions

has for list of questions

Question

Answer

Progressive Clinical Case

Isolated Questions

QMA

QUA

QSOA

correspond to

MAPPING

SIDES DUMP(s)

SeeAlso (weblink)

Wiki Sides

UMLS

Current SIDES environment
Mapping-based automatic data extraction guided by the ontology
Zoom on data linkage

- **etu1001**
- **enrol10502**
- **act918**
- **act540**

**Medical Speciality**
- **Hematology**
- **Cardiovascular**

**Learning Objective_208**

- **ECN**

**Question**: q128

**Has For Registration Date**: 1/09/2015

**Has for timestamp**: 28/01/2016

- **Correspond To Student**: enrol10502
- **Done by**: act918

- **Correspond To Question**: Med_A5
- **Medical Speciality**

See also '1/09/2015' Has For Registration Date '28/01/2016' Has for timestamp

(Hémogramme ... indications et interprétation)
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**.
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)
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

```sparql
SELECT ?specialty ?globalAverage ?studentAverage
WHERE {
  { SELECT ?specialty ( AVG(?result) AS ?globalAverage )
    WHERE { ?answer sides:has_for_result ?result .
      ?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: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

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

**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**
Édition d'une nouvelle question

**Intitulé de la question**
Dans la situation d'une pyélonéphrite aiguë de l'enfant avec présence de cocci gram positif en chainettes à l'examen direct d'un ECBU, dire parmi les traitements, lequel (lesquels) est (sont) initialement recommandé(s) :

**Spécialité(s)**

<table>
<thead>
<tr>
<th>Propositions du système:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pédiatrie</td>
</tr>
<tr>
<td>Maladies infectieuses</td>
</tr>
</tbody>
</table>

**Spécialité(s) que vous retenez:**
Pédiatrie

**Réponses**

<table>
<thead>
<tr>
<th>Réponse</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>cotrimoxazole par voie orale</td>
<td>Faux</td>
</tr>
<tr>
<td>céphalosporine de 3e génération (ceftriaxone) p</td>
<td>Faux</td>
</tr>
<tr>
<td>pénicilline A (amoxicilline) par voie parentérale</td>
<td>Valide</td>
</tr>
<tr>
<td>pénicilline M (oxacilline) par voie parentérale</td>
<td>Faux</td>
</tr>
<tr>
<td>pénicilline A (amoxicilline) par voie orale</td>
<td>Faux</td>
</tr>
</tbody>
</table>

**Ajoutez une spécialité:**
- Addictologie
- Anesthésiologie - Réanimation - Urgences
- Cancérologie - Radiothérapie
- Cardio-vasculaire
- Dermatologie
- Chirurgie digestive
- Endocrinologie - Métabolisme - Nutrition
- Médecine légale
- Gérontologie

Application to OntoSides completion

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Threshold</th>
<th>Hits@1</th>
<th>% Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset1</td>
<td>0</td>
<td>75.1%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>78.2%</td>
<td>96.5%</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>82.1%</td>
<td>84.5%</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>86.6%</td>
<td>66%</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>91%</td>
<td>43%</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>92.6%</td>
<td>33.3%</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>94.5%</td>
<td>18.3%</td>
</tr>
</tbody>
</table>

- 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

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

- 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

=> OntoSIDES enrichment by adding useful triples in addition to the discovered links
  - 15371 triples added

<table>
<thead>
<tr>
<th>#Evaluated Learning Objectives</th>
<th>#MSHFRE and SNMIFRE Semantic Terms</th>
<th>P@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>510</td>
<td>94.5%</td>
</tr>
</tbody>
</table>
Example
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.
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