Ontology-based learning analytics on the national e-learning platform of French Medical Schools

Marie-Christine Rousset
LIG, Univ. Grenoble Alpes & Institut Universitaire de France

Joint work with
Fabrice Jouanot (LIG, Univ. Grenoble Alpes), Olivier Palombi (Univ. Grenoble Alpes, Inria-LJK, CHU, UNESS) and Adam Sanchez (GRICAD), Mohannad Almasri, Shadi Baghernezhad
**SIDES 3.0**

AI for personalized students progress monitoring and training

**TODAY**

A successful experience of dematerialized exams in Medical Studies in France

**TOMORROW**

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

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb of students</td>
<td>80000</td>
</tr>
<tr>
<td>Nb of connections per month</td>
<td>250 000</td>
</tr>
<tr>
<td>Nb of validating exams</td>
<td>2600</td>
</tr>
<tr>
<td>Nb of clinical cases in exams</td>
<td>5 000</td>
</tr>
<tr>
<td>Nb of training clinical cases</td>
<td>1 000 000</td>
</tr>
</tbody>
</table>

facilitates the empowerment of end-users in data analytics by using Semantic Web technologies and Linked Data principles

**UNESS.fr**

**ANR-16-DUNE-0002**

AMEE 2019 symposium (August 27th 2019)
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: enriched and linked data

http://virtuoso5.ontosides.network/sparql

**OntoSIDES ontology**
- **Student**
- **Enrolment**
- **Answer**
- **Referential Entity**
  - **ECN**
  - **Speciality**
- **Learning objective**
- **Learning sub-objective**
- **Evaluation**
  - **Progressive Clinical Case**
  - **Isolated Questions**
- **Question**
  - **QMA**
  - **QUA**
  - **QSOA**

**Current SIDES environment**
- **SIDES DUMP(s)**
- **SeeAlso (weblink)**
  - **Wiki Sides**
  - **UMLS**

**Users interface**
- **SPARQL End-point**
- **Formular for paramaterized queries**
- **Webservice**
- **Natural Language**

**Mapping**

**AMEE 2019 SYMPOSIUM (AUGUST 27TH 2019)**
RDF: a unified model for data, metadata and constraints

- **RDF graph = a set of triples** `<subject, property, object/value>
  <ans1001, sides:done_by, stu100>` means that answer 1001 is done by student 100
  ✓ Extensible schema
  ✓ No strict separation between schema and instances

- **Ontological constraints on schema’s classes and properties** also expressed in RDF
  ✓ Typing constraints on subjects and objects of a given property
  ✓ Specialization constrains between classes and properties

**Triples describing instances**

```
sides:correspond_to_question
  rdf:type rdf:Property ;
  rdfs:domain sides:answer ;
  rdfs:range sides:question ;
.

sides:action11
  rdf:type sides:action_to_answer ;
  sides:is_part_of sides:answer1
  [sides:correspond_to_question sides:q148 ]
  sides:has_rightly_ticked
    sides:prop621 , sides:prop620 ,
    sides:prop623 , sides:prop622 ;
  sides:done_during
    sides:test861 ;
```
OntoSIDES ontology

52 classes, 50 properties, 1400+ instances (medical specialities, official items of the ECN programme) + 10 rules
Zoom on data linkage

- Student
- etu1001
- Medical Speciality
- Med_A5
- Correspond To Medical Speciality
- Correspond To Training
- Done by enrol10502
- act918
- Has For Registration Date: '1/09/2015'
- Has For timestamp: '28/01/2016'
- Correspond To Question

- Medical Speciality
- Haematology
- Is Linked To

- Learning Objective_208
- Correspond To Question

- Question
- q128
- Correspond To Question

- Question
- q128

- ECN
- (Hémogramme ... indications et interprétation)

- Wiki Sides

- Has for registration date: '28/01/2016'

- See also

- '1/09/2015'

AMEE 2019 symposium (August 27th 2019)
Actual status: ~ 1,5 billions of triples

- **64 957 students**

  ```sql
  select (count (distinct ?student)) where {?student a sides:student}
  ```

- **224 919 141 actions of answer** for **457 873 questions** (66 531 questions with unique answer and 390 323 questions with multiple answers)

  ```sql
  select (count (?a)) where {?a sides:done_by ?student}
  select (count (distinct ?q)) where {?a sides:correspond_to_question ?q}
  select (count (distinct ?q)) where {?a sides:correspond_to_question ?q. ?q a sides:QUA}
  select (count (distinct ?q)) where {?a sides:correspond_to_question ?q. ?q a sides:QMA}
  ```
Some statistics about questions

- **Questions with multiple answers (QMA):** 467 498
  - with 77 175 without action of answer
    
    ```sql
    select (count (distinct ?q)) where { ?q a sides:QMA
    FILTER NOT EXISTS {?a sides:correspond_to_question ?q} }
    ```
  - with 50 550 related to a medical speciality, and 54 497 to an item of the ECN referential
    
    ```sql
    select (count (distinct ?q)) where { ?q a sides:QMA.
    ?q sides:is_linked_to_the_medical_speciality ?s}
    ```

- **Questions with unique answers (QUA):** 81 155
  - with 14 624 without action of answer
    
    ```sql
    ```
  - with 5181 related to a medical speciality, and 5912 to an item the ECN referential
Focus on a student average comparison of his/her results with average results of all students

```sql
WHERE {GRAPH <http://ontosides.en> { 
  { SELECT ?speciality ((AVG(?r1)) AS ?overallAvg) 
    ((COUNT(DISTINCT ?q1)) AS ?nbquestionspeciality) 
  WHERE { ?adr1 sides:has_for_result ?r1 . 
    ?adr1 sides:done_by ?etu . 
    ?adr1 sides:correspond_to_a_question ?q1 . 
    ?q1 sides:is_linked_to_ENC_referential_entity ?item . 
    ?item sides:is_linked_to_the_medical_speciality ?speciality . } 
  GROUP BY ?speciality . }
  { SELECT ?speciality ((AVG(?r)) AS ?studentAvg) 
    ((COUNT(DISTINCT ?q)) AS ?nbquestionItemEtu) 
  WHERE { ?adr sides:has_for_result ?r . 
    ?adr sides:correspond_to_a_question ?q . 
    ?q sides:is_linked_to_ENC_referential_entity ?item . 
    ?item sides:is_linked_to_the_medical_speciality ?speciality . 
    FILTER(?student = sides:etu12402) } 
  GROUP BY ?speciality . }
  OPTIONAL { SELECT ?speciality str(?label) as ?label 
    WHERE { ?speciality a sides:speciality. 
      ?speciality rdfs:label ?label . } 
  GROUP BY ?speciality . }}
```

AMEE 2019 symposium (August 27th 2019)

10
On-demand data analytics at different granularity levels and driven by end-users

- **Use case 1:** Define explainable and computable criteria from data about student questions and their related answers
  - For computing difficulty level, discriminatory factor, and quality level of questions (inside a test or a pool of questions, ....) or distractors (inside a question content)

- **Use case 2:** Define and compute a student *trajectory* (or for a group of students) related to items from the referential or to medical specialities
  - Temporality is central!
Ongoing work

• New SIDES dump covering 2010-2018 period
  • Parallel computing for mapping-based data materialization (+6 billions triples, + 160 GB)

• Enriching OntoSides with Standard Medical Ontologies
  • Discovery of links between learning objectives and MeSH and SNOMED thesaurus

• Automatic completion of links between questions and specialities or learning objectives
  • Comparison of several (supervised and unsupervised) Machine Learning methods

• Integration of an ontology for supporting and formalizing learning units for training by simulation in medical gestures
  • Ontology-based automatic generation of pre-filled forms for helping teachers to set up new simulation workshops

⇒ New architecture where OntoSIDES will replace SIDES database as data repository interfaced with Moodle or Mahara platforms as front-ends.
Thanks

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