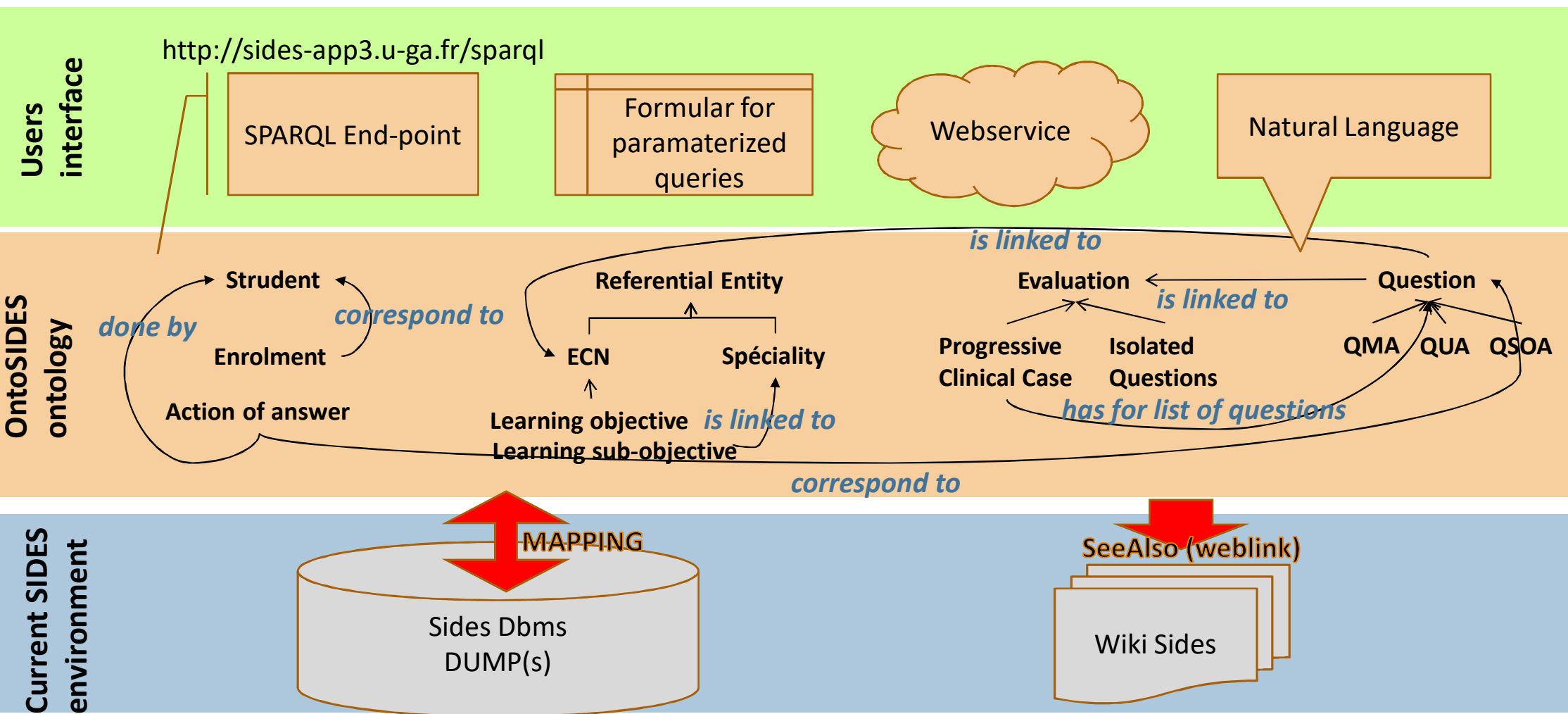




Ontology Mediated Query Rewriting challenged by aggregated queries in OntoSIDES

Fabrice Jouanot, Olivier Palombi, M-Christine Rousset, Adam Sanchez

OntoSIDES: the OBDA layer of SIDES 3.0



OntoSIDES ontology

67 classes, 42 properties, 1419 instances (medical specialties, official items of the ECN programme) + 8 rules

The screenshot displays the OntoSIDES ontology editor interface. The left pane shows the 'Classes' hierarchy, with 'sides:speciality (31)' selected. The central pane shows the 'Class Form' for 'sides:speciality', including annotations for 'rdfs:label' in English ('Speciality') and French ('Spécialité'), and a 'Class Axiom' 'rdfs:subClassOf' 'sides:referential_entity'. The right pane lists 'Properties' such as 'sides:has_for_proposal_of_answer', 'sides:has_for_registration_place', 'sides:has_feedback', 'sides:has_for_comment', 'sides:has_for_context', 'sides:has_for_duration', 'sides:has_for_registration_date', 'sides:has_rating', 'sides:correspond_to', 'sides:correspond_to_training', 'sides:correspond_to_question', 'sides:correspond_to_student', 'sides:done_by', 'sides:done_during', and 'sides:ending_date_of_test'. The bottom pane shows a table of instances for 'sides:speciality'.

[Resource]	rdf:type	rdfs:label
◆ sides:addictology	sides:speciality	Addictologie
◆ sides:anesthesiology_resuscitati...	sides:speciality	Anesthesiologie - Reanimation - ...
◆ sides:cancerology_radiotherapy	sides:speciality	Cancerologie - Radiotherapie
◆ sides:cardiovascular	sides:speciality	Cardio-vasculaire
◆ sides:dermatology	sides:speciality	Dermatologie

Ontop: a framework for implementing OBDA

An open source system for querying relational data sources through an ontology using SPARQL

- a virtual approach based on query rewriting using mappings
- support SPARQL 1.0

D. Calvanese, B. Cogrel, E. G. Kalayci, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, and G. Xiao.

OBDA with the ontop framework.

In 23rd Italian Symposium on Advanced Database Systems, SEBD 2015, Gaeta, Italy, June 14-17, 2015., pages 296–303, 2015

Ontop mappings 1/2

The screenshot displays the Ontop Mapping Assistant interface. On the left, the 'Class hierarchy: 'Action to answer'' is shown, with 'Action to answer' selected. Below it, the 'Object property hierarchy:' is visible, listing various properties like 'graph name node', 'group by', 'has for author', etc. The main area on the right is the 'Mapping editor', which contains several mappings in SQL-like syntax. The mappings are:

- urn:test**
`sides:test{id} a sides:test .
select a.id as id, a.title, a.startdate, a.enddate from public.assessment a`
- urn:test_has_for_title**
`sides:test{id} sides:has_for_title {title}^^xsd:string .
select a.id as id, a.title as title, a.startdate, a.enddate from public.assessment a`
- urn:test_start_and_end_dates**
`sides:test{id} sides:starting_date_of_test {startdate}^^xsd:dateTime ; sides:ending_date_of_test {enddate}^^xsd:dateTime .
select a.id as id, a.title, a.startdate as startdate, a.enddate as enddate from public.assessment a`
- urn:evaluation_has_for_textual_content**
`sides:eval{id} sides:has_for_textual_content {eval_introduction}^^xsd:string .
SELECT di.id as id, string agg(di.introduction, '') as eval_introduction from ontosides.docimococontent_introduction di
inner join ontosides.evaluation ev on ev.pool_id = di.id
group by di.id`
- urn:relation_evaluation_type_question**
`sides:eval{pool_id} sides:has_for_question sides:q{question_id} .
select pg.pool_id as pool_id, pg.question_id as question_id, pg.position from pool_question pg
inner join ontosides.evaluation ev on pg.pool_id = ev.pool_id`
- urn:relation_test_evaluation_type**
`sides:test{assessment_id} sides:is_made_of sides:eval{docimococontent_id} .
select dca.assessment_id as assessment_id, dca.docimococontent_id as docimococontent_id from docimococontentassessment dca
inner join ontosides.evaluation ev on ev.pool_id = dca.docimococontent_id`
- urn:relation_test_student**
`sides:test{assessment_id} sides:has_for_registrant sides:stu{participant_id} .
select p.assessment_id as assessment_id, p.participant_id as participant_id from public.participant p`

At the bottom, the 'Mapping count: 189' is displayed, along with a search bar containing 'pred:action_to_answer' and an 'Enable filter' checkbox. A footer note says 'To use the reasoner click Reasoner > Start reasoner' and a 'Show inferences' checkbox is visible.

Ontop mappings 2/2

The screenshot displays the 'Mapping Assistant - BETA' window. The 'Mapping editor' tab is active, showing a list of mappings on the left and an 'Edit Mapping' dialog box on the right.

Mapping editor:

Mapping manager: Create Remove Copy Select all Select none

urn:test
sides:test{id} a sides:test .
select a.id as id, a.title, a.startdate, a.

urn:test_has_for_title
sides:test{id} sides:has_for_title {title}^^xsd:string .
select a.id as id, a.title as title, a.startdate as startdate

urn:test_start_and_end_dates
sides:test{id} sides:starting_date_of_test {startdate}^^xsd:date .
select a.id as id, a.title, a.startdate as startdate

urn:evaluation_has_for_textual_content
sides:eval{id} sides:has_for_textual_content {eval}^^xsd:string .
SELECT di.id as id, string_agg(di.introduction, ' ' inner join ontosides.evaluation ev group by di.id

urn:relation_evaluation_type_question
sides:eval{pool_id} sides:has_for_question sides:eval{question_id}^^xsd:string .
select pq.pool_id as pool_id, pq.question_id as question_id inner join ontosides.evaluation ev

urn:relation_test_evaluation_type
sides:test{assessment_id} sides:is_made_of sides:eval{evaluation_id}^^xsd:string .
select dca.assessment_id as assessment_id, dca.evaluation_id as evaluation_id inner join ontosides.evaluation ev

urn:relation_test_student
sides:test{assessment_id} sides:has_for_registrant sides:participant{participant_id}^^xsd:string .
select p.assessment_id as assessment_id, p.participant_id as participant_id from public.participant p

Edit Mapping

Mapping ID: urn:test_has_for_title

Target (Triples Template): sides:test{id} sides:has_for_title {title}^^xsd:string .

Source (SQL Query): select a.id as id, a.title as title from public.assessment a limit 10

id	title
1	CONF-2015-2016-HEMATO n°1- M2-Cpe F-DM
13894	PHARMA 3A EVAL UE 2014-2015 - 3A S05 - 1er déce...
11	C6-ORL-02/09/2015-DP1
15567	SDP
16	CONFBDI/CHIRVISC/DFASM3/01
17	CONFBDI/MAL INF/DFASM3/01

Test SQL Query

Update Cancel

Mapping count: 189 Search: pred:action_to_answer Enable filter

To use the reasoner click Reasoner > Start reasoner Show Inferences

Limitations of Ontop

- Mappings do not enable to capture RDF containers
 - needed for relating a test of type « progressive_clinical_case » to its ordered list of questions
 - ⇒ we extended the class `ModelIOManager` to generate such mappings
- Query rewriting does not support SPARQL 1.1 queries
 - Aggregated queries are needed
 - ⇒ we used mappings to materialize data
 - ⇒ accessible using a sparql endpoint: <http://virtuoso.ontosides.network/sparql>

Actual status: ~ 1,5 billions of triplets

■ **64 957 students** `select (count (distinct ?student)) where {?student a sides:student}`

■ **947 students from Grenoble**

`select (count (distinct ?student)) where {?student a sides:student. ?e sides:correspond_to_student ?student. ?e sides:has_for_registration_place sides:grenoble_alpes_university }`

■ **224 919 141 action of answer** (68 126 for Grenoble Students)
for **457 873 questions** (66 531 questions with unique answer and 390 323 questions with multiple answers)

`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 }`

Number of students by University

```
select (count (distinct ?student)) as ?NbEtu ?univ where {?student a sides:student.  
?e sides:correspond_to_student ?student. ?e sides:has_for_registration_place ?univ } group by ?univ
```

Sample of answers:

NbEtu	univ
456	http://www.side-sante.fr/sides#paris_xi_university
824	http://www.side-sante.fr/sides#caen_university
1419	http://www.side-sante.fr/sides#bordeaux_university
537	http://www.side-sante.fr/sides#paris-ile_de_france_ouest_university
1127	http://www.side-sante.fr/sides#lille_university
39	http://www.side-sante.fr/sides#paris_xii_university
451	http://www.side-sante.fr/sides#paris_xiii_university
3842	http://www.side-sante.fr/sides#lyon_est_university
83	http://www.side-sante.fr/sides#la_reunion_university
757	http://www.side-sante.fr/sides#lyon_sud_university
732	http://www.side-sante.fr/sides#tours_university
947	http://www.side-sante.fr/sides#grenoble_alpes_university
743	http://www.side-sante.fr/sides#nantes_university
554	http://www.side-sante.fr/sides#brest_university

Some statistics about questions

■ Questions with multiple answers (QMA): 467 498

- ✓ with 77 175 without action of answer

```
select (count (distinct ?q)) where { ?q a sides:QM  
                                     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

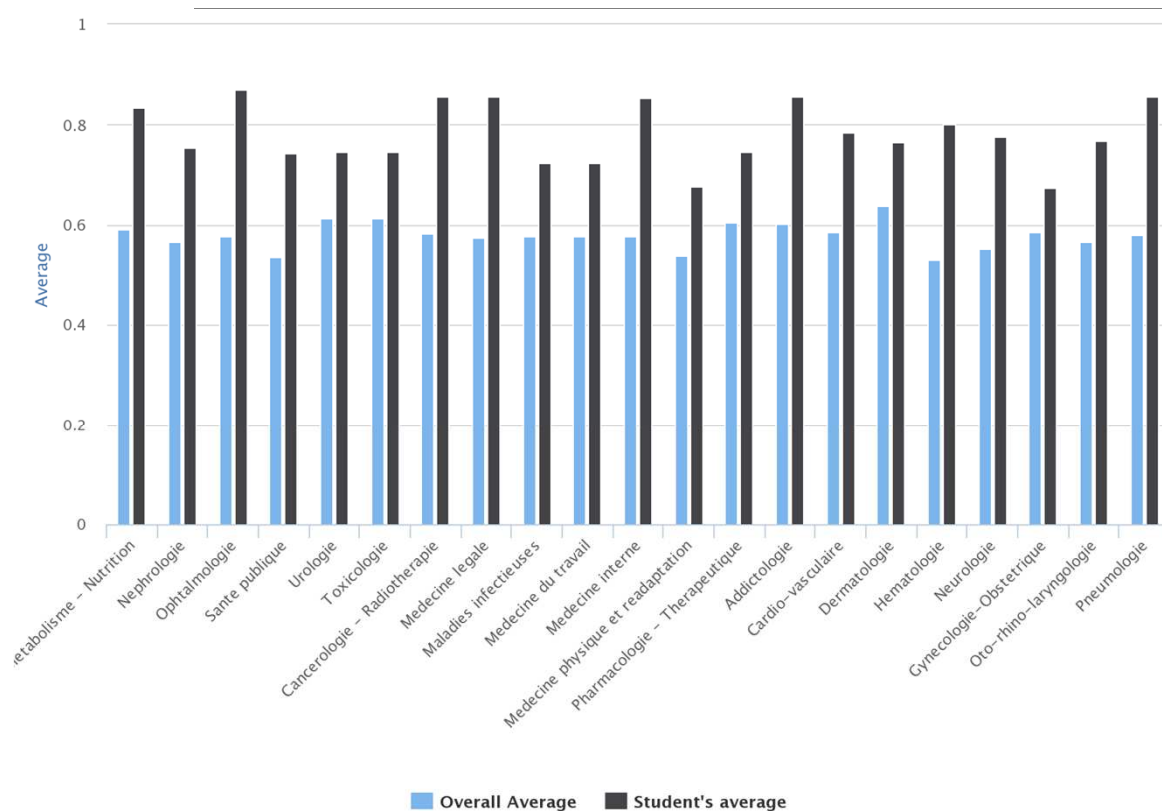
```
select (count (distinct ?q)) where { ?q a sides:QMA.  
                                     ?q sides:is_linked_to_the_medical_speciality ?s }
```

■ Questions with unique answers (QUA): 81 155

- ✓ with 14 624 without action of answer
- ✓ with 5181 related to a medical speciality, and 5912 to an item the ECN referential

Focus on a student: stu12402

average comparison of his results with average results of all students



```

SELECT ?specialite ?label ((?moyennespecialite) AS ?moy) ?moyenneetu
WHERE {GRAPH <http://ontosides.en> {
  { SELECT ?specialite ((AVG(?note1)) AS ?moyennespecialite)
    ((COUNT(DISTINCT ?q1)) AS ?nbquestionspecialite)
    WHERE { ?adr1 sides:has_for_result ?note1 .
      ?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 ?specialite . }
    GROUP BY ?specialite } .
  { SELECT ?specialite ((AVG(?note)) AS ?moyenneetu)
    ((COUNT(DISTINCT ?q)) AS ?nbquestionItemEtu)
    WHERE { ?adr sides:has_for_result ?note .
      ?adr sides:done_by ?etudiant .
      ?adr sides:correspond_to_a_question ?q .
      ?q sides:is_linked_to_ENC_referential_entity ?item .
      ?item sides:is_linked_to_the_medical_speciality ?specialite .
      FILTER(?etudiant = sides:etu12402) }
    GROUP BY ?specialite } .
  OPTIONAL { SELECT ?specialite str(?label) as ?label
    WHERE { ?specialite a sides:speciality.
      ?specialite rdfs:label ?label. }
    GROUP BY ?specialite } . }
  
```

On-demand data extraction at different granularity level and driven by analysts

- **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 !

For each QUA question linked to the speciality about infectious disease, with at least 100 action of answers and at least one correct answer, give the level of difficulty (**number of correct answers**/ number of action of answer)

```
SELECT ?question ?NbReponsesCorrectes ?NbReponses
      |(1-(xsd:decimal(?NbReponsesCorrectes)/xsd:decimal(?NbReponses)) AS ?NiveauDifficulte)
FROM <http://ontosides>
WHERE {
  {SELECT ?question (COUNT(?action) AS ?NbReponses)
  WHERE
  {?question a sides:QUA.
  ?question sides:is_linked_to_the_medical_speciality sides:speciality_infectious_diseases.
  ?action sides:correspond_to_question ?question.
  FILTER EXISTS {?actionCorrect sides:correspond_to_question ?question.
                  ?actionCorrect sides:has_rightly_ticked ?rep.}}
  GROUP BY ?question
  HAVING(COUNT (?action) > 100)
  }
  {SELECT ?question (COUNT(?actionCorrect) AS ?NbReponsesCorrectes)
  WHERE
  {?question a sides:QUA.
  ?question sides:is_linked_to_the_medical_speciality sides:speciality_infectious_diseases.
  ?actionCorrect sides:correspond_to_question ?question.
  ?actionCorrect sides:has_rightly_ticked ?rep.)
  GROUP BY ?question
  }
}
ORDER BY DESC(?NiveauDifficulte)
```


For each QUA question linked to the speciality about Infectious disease, with at least 100 action of answers and at least one correct answer, give the level of difficulty

question	NbReponsesCorrectes	NbReponses	NiveauDifficulte
http://www.side-sante.fr/sides#q67019	1	397	0.997481108312343
http://www.side-sante.fr/sides#q67008	1	396	0.997474747474747
http://www.side-sante.fr/sides#q40779	4	169	0.976331360946746
http://www.side-sante.fr/sides#q49540	5	108	0.953703703703704
http://www.side-sante.fr/sides#q49541	9	111	0.918918918918919
http://www.side-sante.fr/sides#q34192	11	131	0.916030534351145
http://www.side-sante.fr/sides#q39026	29	276	0.894927536231884
http://www.side-sante.fr/sides#q37511	31	268	0.884328358208955
http://www.side-sante.fr/sides#q107641	20	142	0.859154929577465
http://www.side-sante.fr/sides#q144749	83	541	0.846580406654344
http://www.side-sante.fr/sides#q90832	48	270	0.822222222222222
http://www.side-sante.fr/sides#q84795	34	187	0.818181818181818
http://www.side-sante.fr/sides#q695556	25	137	0.817518248175182
http://www.side-sante.fr/sides#q82564	34	185	0.816216216216216
http://www.side-sante.fr/sides#q85874	49	254	0.807086614173228
http://www.side-sante.fr/sides#q52985	35	171	0.795321637426901
http://www.side-sante.fr/sides#q82030	66	303	0.782178217821782
http://www.side-sante.fr/sides#q39085	106	447	0.762863534675615

Challenges

■ Technical Challenges

- ✓ Tracking student history across SIDES DB dumps
- ✓ Developing an efficient method for scaling forward reasoning on big data (triple store saturation)

■ Methodological Challenges

- ✓ Managing the whole workflow for an analyst use case: from data extraction to recommendation