Towards an architecture for universities management

Assisting students in the choice of their specialization

Inaya Lahoud[1], Fatma Chamekh[2]

[1]Dept. of Computer Science, University of Galatasaray, Istanbul, Turkey
clahoud@gsu.edu.tr

[2]University of Lyon, Lyon, France
Fatma.chamekh@univ-lyon3.fr

ABSTRACT. The heterogeneity and the high volume of data on the Web are the main features that make it a promoter field of knowledge engineering for researcher. However, the user is getting lost towards the diversity of information on the Web. In this paper, our aim is to propose an approach to assist user. Our approach uses Semantic Web technologies. Our scenario is focused on the field of education and in particular higher education. This choice is motivated by the diversity of information sources where the student is dispersed.

Keywords: Ontologies; High education institution; universities management system

1 Introduction

Nowadays, the web offers advanced interactions between users and data providers. Since the open data initiative, the data on the Web has increased. Many institutions have published their data in heterogeneous and distributed way. The universities and educational institutions follow this way by providing a data concerning their training offers. Users retrieve and review educational training. The main idea is to share their experiences and get the best feedback about universities.

Our Objective is to match the user's needs and trainings offered by institutions. The management of the data from diverse resources is complex and tedious process, relying mainly on human-based and error-prone tasks. Globally, there is a lack of practical approach for converting and linking multi-origin data pieces into one coherent smart data set. More specifically, the following scientific locks make the transformation of data on a smart data is difficult task. For this end we present the main challenges behind our research work:

- Data usually comes in variable quality unorganized or not described, and not linked to other sources on the Web. There is a need to homogenize vocabularies by providing machine-readable explicit description of data semantics, and linking data sets to each other and to ontologies.

- The combination of heterogeneous data from different sources generates some issues. Those one could be classified on three levels: the syntactic level related to data format and syntactic, the semantic level when different knowledge presentations are used and the structural level due to the different data organizations.

We focused in this paper on assisting users in their information research by offering them a universities research system. This system seeks the most appropriate training for a student according to his criteria (university ranking, domain of training, geography of university and unemployment percent). In addition, this system offers to him trainings where he can benefit from social or excellence scholarships, grants from the state to encourage certain areas, or simply where the unemployment rate is lowest and therefore he has more chances to acquire employment at the end of his studies.

Our system uses the Open Data for higher education and unemployment available at data.gouv.fr.

The remainder of this paper is organized as follows: section 2 states our problem, section 3 presents some existing approach that use semantic web technologies for education. Section 4 depicts our framework. Future works and conclusion are presented in section 5.

2 Problem statement

We conducted our study in the field of education and more specifically on higher education institutions. Indeed, the number of higher education institutions is increasing every year. If we count the number of these institutions in a single country such as France (more than 3500[1]) and the United States (more than 4500 [1]), and the number of training courses that offer these institutions, we can understand why students are always lost in the choice of their studies.

The final year of high school is the year when students accrue the stress of the passage of the high school diploma, the daily difficulties and the choice of their future orientation.

Most students, who have managed to continue their studies, think that they were insufficiently or badly informed and advised and therefore they chose the training in which they can succeed. Not having a clear and precise idea of the job they wish to exercise, nor a long-term professional objective, these students eventually chose more often the curriculum that leaves most open doors. Nevertheless, the difficulties are not especially evacuated for those who are still looking for jobs, then they wonder about their chances to integrate quickly the job market and are asking questions about the appropriateness of their courses' choice [2].

In front of this situation, we need to propose to the student a system to manage the huge and heterogeneous amount of data by taking into account the student criteria. Our general challenge is summarized as follows: given the data provided by educational institutions, universities, government and companies, whose understandability remains difficult, semantic web technologies could provide an efficient solution.

[1] education.gouv.fr

3 Related work

The exploitation and the search for information on the Web become increasingly complicated task. How to find information that we seek quickly seen the diversity and the quantity of data on the Web? Where to search this information since there is no global database that stores all data of the whole world as wished Tim Berners Lee? What is the relevance of the information we found on the Web? Do they fully correspond to our research?

There are many researchers working on these issues and try to find solutions. As we explained in the introduction, we focused our study on the education and specifically on higher education institutions. Our goal is to consolidate information of higher education in one system where students can find what training is most appropriate for them basing on university ranking, percentage of employment in this field, geography, and available scholarships.

Semantic web technologies are getting increasingly used in various contexts, higher education is no exception. Different approaches are designed to spread the services of the universities which are distributed across several departments to serve their substantial student base. In their survey, Dietze et al [3] highlight the growing use of linked data by various universities. Many platforms are made available for direct consumption and reuse. This includes for example the OU's linked open data platform (http://data.open.ac.uk), the University of Muenster (http://data.uni-muenster.de), the University of Southampton (http://data.southampton.ac.uk) among others. The data available through those platforms include: Courses information podcasts, Library catalogue, research publications, OpenLearn, reading experience database, the open arts archive events, information about university staff and buildings located across the UK. Besides that, the data of the platforms cited above are connected to the linked open data cloud. Dedicated graphs include links to the Dbpedia entities, Geonames entities and BBC entities. The external entities have the topics of media objects, web pages, courses.

The last few years, we have seen several websites style MOOCs such as Coursera[2] FUN[3]. These websites allow users a free access to online courses. However, the aim of our system is to search trainings offered by universities and not online courses as the MOOC case.

[4] [5] [6] have proposed methodologies and frameworks to transform the existing data sources into linked data. The main idea is to turn the available organizational data in a linked data cloud, using pre-programmed transformation pattern. To follow the success of social and knowledge graphs functionalities provided by facebook[4] and google[5], Health et al [7] proposed to create an education graph by processing courses information and learning material from various universities in UK. They rely on bibli-

[2] https://www.coursera.org/

[3] https://www.fun-mooc.fr/

[4] http://newsroom.fb.com/News/562/Introducing-Graph-Search-Beta

[5] http://www.google.com/insidesearch/features/search/knowledge.html

ographical data of material repositories to identify links to course resource. In this direction, the LinkedUpDataCatalog[6] or related community initiatives[7] are initial efforts to collect and catalog dataset have been made by universities or other institutions.

Zablith [8] created a semantic data layer to conceptually connect courses taught in a higher education program. The aim is to interlink courses within the same institution at the level of concepts covered in course topics. For this reason, the author proposed a courses information data model.

The limitations of approaches explained above are mainly related to the following aspects. First, the data provided by those platforms are mainly limited to the educational information such as courses program, publication. That means student can find the courses topics and courses material or web pages. Also, those dataset are related to external dataset such as Dbpedia or geonames but to enrich the educational data. Second, the first efforts to connect various datasets are proposed by [7] but it is limited to universities of UK.

In the other side we didn’t find researches who occur the subject of trainings’ classification to obtain scholarships except Rad [9] who presented in his paper a study in Iran on the classification of university courses using data mining techniques. These courses were classified according to different criteria to determine which training is the most promising in the future. The Iranian state uses the result of this classification then to invest in these trainings and pro-vide scholarships to encourage students to follow them.

In our context, the aim is to assist the student to select a training. This selection should be made following various criteria such age the university, the rank, scholarships existence, and the unemployment average. To our knowledge, there is no existing approach that covers these requirements. To achieve our objective, we have to connect the educational and government datasets. We present in the next section our approach of universities management system in order to assist students in their choice.

4 Global approach

The idea of our work is to deal with academic and government data and make them more accessible to users. Our methodology consists of four main steps: knowledge definition, linked open data, knowledge extraction and classification, and knowledge representation (Fig. 1). We use the word “knowledge” in our system because we enrich the data with semantic information.

The first step in our system is to backup data in ontological format. This allows us to represent them formally and semantically. The data used in our study are open data from the Website “data.gouv.fr”. Once the data are well represented, we take the two ontologies, in the second stage, and we try to link them with the online data as dbpedia, and others. All OWL files are saved in a knowledge base.

[6] http://data.linkededucation.org/linkedup/catalog

[7] http://www.w3.org/community/opened

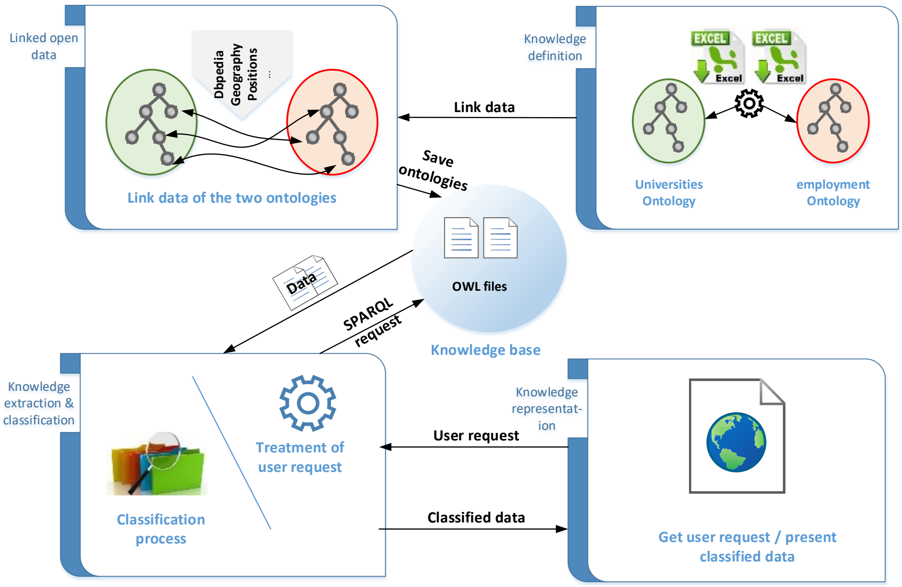


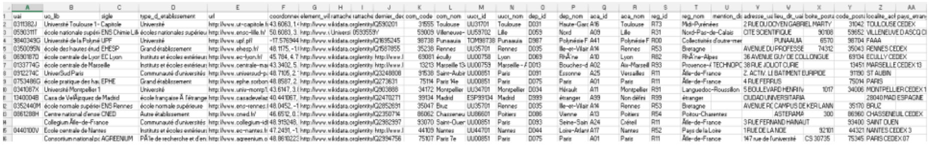
Fig. 1. Our global approach

The first three steps are in the systems level only while the fourth is in interaction with the human being. Indeed, the human being formulates his query in the search engine and sends it to the system. This request is received by the system in the third step of our methodology (knowledge extraction & classification). It treats the user's query and transforms it into SPARQL query in order to apply it on our OWL related files. The resulting data, from this extraction, will be sorted and sent back to the fourth step to present them to the user.

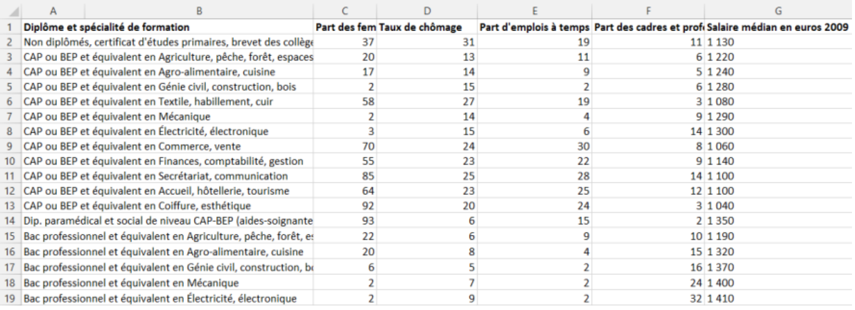
Currently the first two steps are performed only once. We do not manage the evolution of ontology over time. Whereas the last two steps are triggered for each request sent by user.

4.1 Knowledge definition

As we mentioned before, we use the data available on the government website. These data can be extracted from this website only in Excel format (Fig. 2). To represent the data in a more formal and semantic way, we have chosen to transform them into OWL language. This language allows to describe ontologies, that is to say, it defines terminology to describe specific areas. Ontologies have shown in recent years their ability to model a range of knowledge in a given field. The transformation of our data in OWL was done with an open source software RDBToOnto [10].



	uai	uo_lib	sigle	type_d_etablissement	url	coordonne	element_wi	rattache	rattache	dernier_dec	com_code	com_nom	uucr_id	uucr_nom	dep_id	dep_nom	aca_id	aca_nom	reg_id	reg_nom	mention_di	adresse_ua	lieu_dit_uai	boite_posta	code_posta	localite_ach	pays_etran
2	0311382J	Université Toulouse 1 -	Capitole	Université	http://www.ut-capitole.f	43.6063, 1.4	http://www.wikidata.org/entity/Q590201				31555	Toulouse	UU31701	Toulouse	D031	Haute-Garc	A16	Toulouse	R73	Midi-Pyrénées		2 RUE DU DOYEN GABRIEL MARTY			31042	TOULOUSE CEDEX	
3	0593311T	école nationale supéri	ENS Chimie Lill	écoles nationales supérieu	http://www.ensc-lille.fr/	50.6083, 3.	http://www.(	Universi	0593559Y'		59009	Villeneuve-	UU59702	Lille	D059	Nord	A09	Lille	R31	Nord-Pas-de-Calais		CITE SCIENTIFIQUE		90108	59652	VILLENEUVE D ASCQ C	
4	9840349G	Université de la Polyné	UPF	Université	http://www.upf.pf/	-17.576944	http://www.wikidata.org/entity/Q1695245				98738	Punaauia	TOM98738	Punaauia	D987	Polynésie F	A41	Polynésie F	R00	Collectivités d'outre-mer			PUNAAUIA	6570	98704	FAAA	
5	0350095N	école des hautes étud	EHESP	Grand établissement	http://www.ehesp.fr/	48.1175, -1.	http://www.wikidata.org/entity/Q1587855				35238	Rennes	UU35701	Rennes	D035	Ille-et-Vilair	A14	Rennes	R53	Bretagne		AVENUE DU PROFESSE		74312	35043	RENNES CEDEX	
6	0690187D	école centrale de Lyon	EC Lyon	Instituts et écoles extérieur	http://www.ec-lyon.fr/	45.784, 4.7	http://www.wikidata.org/entity			http://www.l	69081	écully	UU00758	Lyon	D069	RhÃ´ne	A10	Lyon	R82	RhÃ´ne-Alpes		36 AVENUE GUY DE COLLONGUE			69134	ECULLY CEDEX	
7	0133774G	école centrale de Marseille		Instituts et écoles extérieur	http://www.centrale-ma	43.3402, 5.	http://www.wikidata.org/entity			http://www.l	13213	Marseille 13	UU00759	Marseille - A	D013	Bouches-d	A02	Aix-Marseill	R93	Provence-A	TECHNOPC	38 RUE JOLIOT CURIE			13451	MARSEILLE CEDEX 13	
8	0912274C	UniverSud Paris		Communauté d'universités	http://www.universud-p	48.7105, 2.1	http://www.wikidata.org/entity/Q3248808				91538	Saint-Aubir	UU00851	Paris	D091	Essonne	A25	Versailles	R11	Ãle-de-France		Z. ACTIV. LE BATIMENT EURIPIDE			91190	ST AUBIN	
9	0753486G	école pratique des hau	EPHE	Grand établissement	http://www.ephe.sorbon	48.8587, 2.	http://www.wikidata.org/entity/Q273631				75114	Paris 14e	UU00851	Paris	D075	Paris	A01	Paris	R11	Ãle-de-France		4 RUE FERRUS			75014	PARIS	
10	0341087X	Université Montpellier 1		Université	http://www.univ-montp1.	43.6147, 3.(	http://www.wikidata.org/entity/Q903888				34172	Montpellier	UU34701	Montpellier	D034	Hérault	A11	Montpellier	R91	Languedoc-Roussillon		5 BOULEVARD HENRI IV		1017	34006	MONTPELLIER CEDEX 1	
11	1340004B	Casa de VelÃ¡zquez de	Madrid	école française Ã l'étrange	http://www.casadevelaz	40.441067,	http://www.wikidata.org/entity/Q2470271				99134	Madrid	ESP99134	Madrid	D999	étranger	A99	Non défini	R99	étranger		CIUDAD UNIVERSITARIA				28040 MAD	ESPAGNE
12	0352440M	école normale supérie	ENS Rennes	école normale supérieure	http://www.ens-rennes.f	48.0452, -1	http://www.wikidata.org/entity/Q2852691				35047	Bruz	UU35701	Rennes	D035	Ille-et-Vilair	A14	Rennes	R53	Bretagne		AVENUE RC	CAMPUS DE KER LANN		35170	BRUZ	
13	0861288H	Centre national d'ense	CNED	Autre établissement	http://www.cned.fr/	46.6512, 0.3	http://www.wikidata.org/entity/Q2350714				86062	Chasseneu	UU86601	Poitiers	D086	Vienne	A13	Poitiers	R54	Poitou-Charentes			ASTERAMA	300	86960	CHASSENEUIL CEDEX	
14		Collegium Ãle-de-France		Communauté d'universités	http://www.collegium-idf	48.919246,	http://www.wikidata.org/entity/Q2982997				93070	Saint-Ouer	UU00851	Paris	D093	Seine-Sain	A24	Créteil	R11	Ãle-de-France		3 RUE FERNAND HAINAUT			93400	SAINT OUEN	
15	0440100V	Ecole centrale de Nantes		Instituts et écoles extérieur	http://www.ec-nantes.fr	47.2491, -1.	http://www.wikidata.org/entity			http://www.l	44109	Nantes	UU44701	Nantes	D044	Loire-Atlant	A17	Nantes	R52	Pays de la Loire		1 RUE DE LA NOE		92101	44321	NANTES CEDEX 3	
16		Consortium national pc	AGREENIUM	PÃ´le de recherche et d'en	http://www.agreenium.o	48.8610223	http://www.wikidata.org/entity/Q2994756				75107	Paris 7e	UU00851	Paris	D075	Paris	A01	Paris	R11	Ãle-de-France		147 rue de l'université		CS 30735	75345	PARIS CEDEX 07	



	Diplôme et spécialité de formation	Part des fem	Taux de chômage	Part d'emplois à temps	Part des cadres et profe	Salaire médian en euros 2009
2	Non diplômés, certificat d'études primaires, brevet des collège	37	31	19	11	1 130
3	CAP ou BEP et équivalent en Agriculture, pêche, forêt, espaces	20	13	11	6	1 220
4	CAP ou BEP et équivalent en Agro-alimentaire, cuisine	17	14	9	5	1 240
5	CAP ou BEP et équivalent en Génie civil, construction, bois	2	15	2	6	1 280
6	CAP ou BEP et équivalent en Textile, habillement, cuir	58	27	19	3	1 080
7	CAP ou BEP et équivalent en Mécanique	2	14	4	9	1 290
8	CAP ou BEP et équivalent en Électricité, électronique	3	15	6	14	1 300
9	CAP ou BEP et équivalent en Commerce, vente	70	24	30	8	1 060
10	CAP ou BEP et équivalent en Finances, comptabilité, gestion	55	23	22	9	1 140
11	CAP ou BEP et équivalent en Secrétariat, communication	85	25	28	14	1 100
12	CAP ou BEP et équivalent en Accueil, hôtellerie, tourisme	64	23	25	12	1 100
13	CAP ou BEP et équivalent en Coiffure, esthétique	92	20	24	3	1 040
14	Dip. paramédical et social de niveau CAP-BEP (aides-soignante	93	6	15	2	1 350
15	Bac professionnel et équivalent en Agriculture, pêche, forêt, e:	22	6	9	10	1 190
16	Bac professionnel et équivalent en Agro-alimentaire, cuisine	20	8	4	15	1 320
17	Bac professionnel et équivalent en Génie civil, construction, bo	6	5	2	16	1 370
18	Bac professionnel et équivalent en Mécanique	2	7	2	24	1 400
19	Bac professionnel et équivalent en Électricité, électronique	2	9	2	32	1 410

Fig. 2. Excerpt of data from data.gouv.fr website

Before the data is transformed into ontologies we clean them to optimize their quality whether that of higher education institutions or employment. Our data cleaning process will affect incomplete, noisy and inconsistent data. As the result of RDBToOnto is unsatisfactory, the cleaning process is done manually by a human expert.

4.2 Linked open data

This step consists of taking the OWL files from the knowledge base and link them with open data available online such as the geographical data from INSEE. Indeed, the two OWL files on which we work contain geographical data. Therefore, we have linked these data with the geographical data of INSEE as in the following example.

Fig. 3 shows the geographic data (municipality, department, and region) of a university and the rate of employment / unemployment by region and how we related them to the open data on the Web.

```
<Universite rdf:ID="Autre_établissement_12">
	<owl:sameAs rdf:resource="http://fr.dbpedia.org/page/Centre_national_d'enseignement_%C3%A0_distance"/>
	<statut_juridique_long>établissement public à caractère administratif</statut_juridique_long>
	<uo_lib>Centre national d&apos;enseignement à  distance</uo_lib>
	<identifiant_freebase>http://www.freebase.com/m/0d98ld</identifiant_freebase>
	<code_postal_uai>86960</code_postal_uai>
		<aCommeFormation rdf:resource="#Licence_professionnelle_-_Droit,_économie,_gestion_-_hôtellerie_et_tourisme_-_spécialité_chef_de_projet_et_créateur_d'entreprises_touristiques" />
```

```
                        <aCommeFormation                    rdf:resource="#Licence_professionnelle_-
_Sciences,_technologies,_santé_-_mécanique_-_spécialité_coordinateur_technique_des_mé-
thodes_d'industrialisation" />
                        <aCommeFormation      rdf:resource="#Diplôme_d'ingénieur_du_Conserva-
toire_national_des_arts_et_métiers_spécialité_aéronautique_et_spatial_en_conven-
tion_avec_l'ISAE-ENSMA,_en_partenariat_avec_AEROTEAM" />
    …….
    …….
    ……..
          <libelle_mission_chef_de_file>Enseignement scolaire</libelle_mission_chef_de_file>
          <uai>0861288H</uai>
          <lieu_dit_uai>ASTERAMA 2 AV DU TELEPOR</lieu_dit_uai>
          <localite_acheminement_uai>CHASSENEUIL CEDEX</localite_acheminement_uai>
          <uucr_id>UU86601</uucr_id>
          <element_wikidata>http://www.wikidata.org/entity/Q2350714</element_wikidata>
          <boite_postale_uai>300</boite_postale_uai>
            <url>http://www.cned.fr/</url>
          <identifiant_programme_lolf_chef_de_file>214</identifiant_pro-
gramme_lolf_chef_de_file>
          <statut_juridique_court>EPA</statut_juridique_court>
          <statut_operateur_lolf>Opérateurs LOLF Hors MIRES 2014</statut_operateur_lolf>
          <uucr_nom>Poitiers</uucr_nom>
          <geo:Departement rdf:about="DEP_86">
                        <geo:code_departement>86</geo:code_departement>
                        <geo:nom xml:lang="fr">Vienne</geo:nom>
          </geo:Departement>
          <coordonnees>
                        <gs:long>46.6512</gs:long>
                        <gs:lat>0.372263</gs:lat>
         </coordonnees>
          <aca_id>A13</aca_id>
          <flux_rss>http://www.cned.fr/rss/communiques-de-presse.xml</flux_rss>
          <numero_telephone_uai>0549493400</numero_telephone_uai>
       </Universite>
```

Fig. 3. Extract from higher education ontology

Then we have linked other data such as areas or trainings. As the employment rate is also classified by domain (agriculture, commerce, computer, construction ...), and institutions offer trainings (IT, networks, mechanical, civil Engineering, political science, medicine ...) that are applied in specific areas, we can well link these two fields (Fig. 4). However, these data are not available in open data until now so we created our own links to connect them. TrainingLevel corresponds to Master, Bachelor, and Engineer degree.

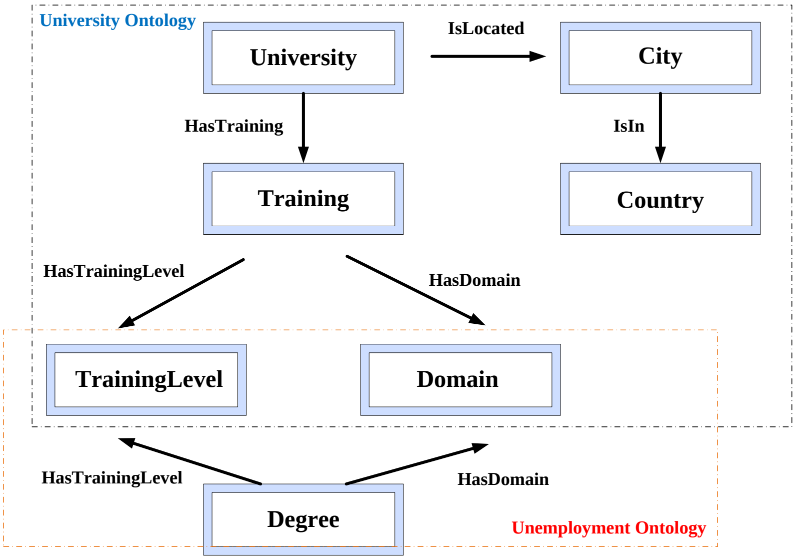


Fig. 4. Linking university and unemployment ontologies

4.3 Knowledge extraction & classification

This step is divided into two distinct parts: the extraction and classification.

In the first part, the system retrieves the request submitted by the user as shown in (Fig. 6) and reformulates it in SPARQL query language to execute it on OWL files. The result of this query will then be classified by the most relevant to user (Fig. 7). Indeed, given a user query, traditional search engines output a list of results which are ranked according to their relevance to the query. However, the ranking is independent of document topic. Therefore, the results of different topics are not grouped together within the result output from a search engine. This can be problematic, as the user must scroll though many irrelevant results until his desired information need is found. This might arise when the user is a novice or has superficial knowledge about the domain of interest, but more typically, it is due to the query being short and ambiguous. Therefore, to avoid this problem we will present in the future a classification method for results search to satisfy the request of user.

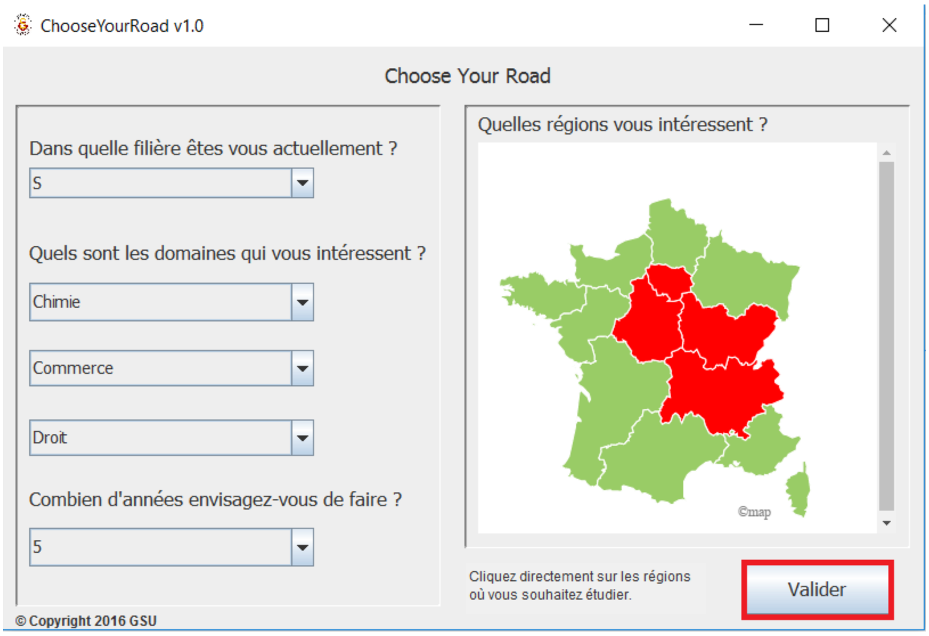
```
String sparqlQuery = "PREFIX ontologie1:<http://www.uni-
versities-project.fr/ontologies/universites#>\n" +
"PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-
ns#>\n" +
"PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>\n"+
                  "SELECT ?nom_univ ?nom_formations \n" +
                  "WHERE {\n" ;
```

```
             if(!domaine1.equals("Choisissez_un_do-
maine"))
             {
             sparqlQuery += " ?formations ontologie1:has-
Domaine ontologie1:"+ domaine1 +"  .\n" ;
             }
             else // Si l'utilisateur demande tous les 
domaines
             {
                     lblNomdomaine.setText("Tous les do-
maines");
               panel_res_3.setBounds(0, 0, 785, 460);
             }
…
…
{sparqlQuery += "?formations ontologie1:hasNiveauForma-
tion ontologie1:"+ ite +" .\n";}
…
…
sparqlQuery += "?universites ontologie1:aCommeFormation 
?formations .\n";
…
…
sparqlQuery += "{ ?universites ontologie1:new_reg_nom 
\""+ region + "\" } UNION ";
…
…
}
```

Fig. 5. Excerpt of SPARQL query

4.4 Knowledge representation

This step allows us to offer a website to the user where he can submit his request for a search and receive the corresponding information. The system then retrieves the query when the user submits it and sends it to the step "Knowledge Extraction & classification" to treat it. The system receives by return the query result already classified and displayed it to the user by order of relevance.



Choose Your Road

Dans quelle filière êtes vous actuellement ?
S

Quels sont les domaines qui vous intéressent ?
Chimie
Commerce
Droit

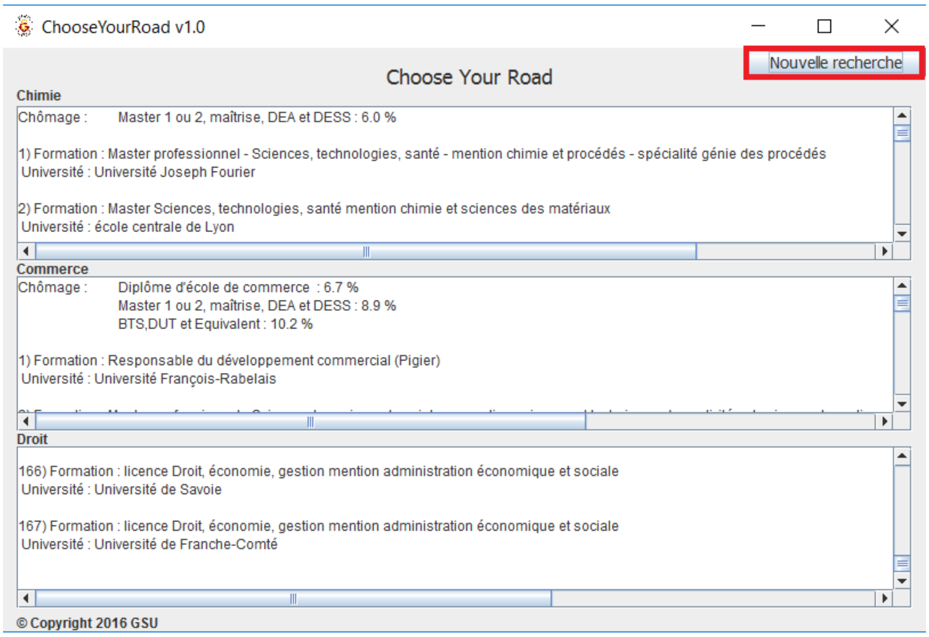
Combien d'années envisagez-vous de faire ?
5

Quelles régions vous intéressent ?

Cliquez directement sur les régions où vous souhaitez étudier.

© Copyright 2016 GSU

Fig. 6. GUI of our system



Choose Your Road

Chimie

```
Chômage :      Master 1 ou 2, maîtrise, DEA et DESS : 6.0 %

1) Formation : Master professionnel - Sciences, technologies, santé - mention chimie et procédés - spécialité génie des procédés
 Université : Université Joseph Fourier

2) Formation : Master Sciences, technologies, santé mention chimie et sciences des matériaux
 Université : école centrale de Lyon
```

Commerce

```
Chômage :      Diplôme d'école de commerce  : 6.7 %
               Master 1 ou 2, maîtrise, DEA et DESS : 8.9 %
               BTS,DUT et Equivalent : 10.2 %

1) Formation : Responsable du développement commercial (Pigier)
 Université : Université François-Rabelais
```

Droit

```
166) Formation : licence Droit, économie, gestion mention administration économique et sociale
 Université : Université de Savoie

167) Formation : licence Droit, économie, gestion mention administration économique et sociale
 Université : Université de Franche-Comté
```

© Copyright 2016 GSU

Fig. 7. Results of research

In this example, the student search for a training in chemistry, commerce or law domain in four regions in France.

Fig. 7 shows the result of his search. It is classified by domain. In each one, student can find the name of training and the university that offers it. The system shows also the percent of unemployment by domain and by training level depending on the information we extracted from the data.gouv.fr website.

We didn't implement yet the classification system to classify results in each domain by the most relevant to the student.

5 Conclusion and future work

We presented in this paper an idea for universities management system. The aim of this system is to assist students in the choice of their domain of specialization. This choice can be done by selecting different criteria: university ranking, location, scholarships, and the average of unemployment. Our system collect information from government open data, clean them and represent them formally in ontological format.

Our work is still in progress. We worked now on the phase of linking our ontologies to available open data or create our own open data when necessary.

We presented in section 4 our first results. We will improve this system by working on different points such as creating new vocabularies, implementing the classification system and allow students to submit their queries in natural language.

References

1. National Center for Education Statistics: Table 5 Number of educational institutions, by level and control of institution: Selected years, 1980-81 through 2010-11. Department of Education, U.S. (2014).
2. Bresfelean, V.P.: Data Mining Applications in Higher Education and Academic Intelligence Management. Presented at the (2009).
3. Stefan Dietze, Salvador Sanchez-Alonso, Hannes Ebner, Hong Qing Yu, Daniela Giordano, Ivana Marenzi, Bernardo Pereira Nunes: Interlinking educational resources and the web of data: A survey of challenges and approaches. Program. 47, 60–91 (2013).
4. Daquin, M.: Linked data for open and distance learning. Commonweathof learning report. (2014).
5. Zablith, F., Fernandez, M., Rowe, M.: Production and consumption of university Linked Data. Interact. Learn. Environ. 23, 55–78 (2015).
6. Zablith, F., d'Aquin, M., Brown, S., Green-Hughes, L.: Consuming Linked Data within a Large Educational Organization. Presented at the Second International Workshop on Consuming Linked Data (COLD) at 10th International Semantic Web Conference (ISWC 2011) , Bonn, Germany (2011).
7. Heath, T., Clarke, C., Singer, R., Leavesley, J., Shabir, N.: Assembling and applying an education graph based on learning resources in universities. In: In: Linked Learning (LILE) Workshop (2012).

8. Zablith, F.: Interconnecting and Enriching Higher Education Programs Using Linked Data. In: Proceedings of the 24th International Conference on World Wide Web. pp. 711–716. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2015).
9. Rad, A., Naderi, B., Soltani, M.: Clustering and ranking university majors using data mining and AHP algorithms: A case study in Iran. Expert Syst. Appl. 38, 755–763 (2011).
10. Cerbah, F.: RDBToOnto: un logiciel dédié à l'apprentissage d'ontologies à partir de bases de données relationnelles. In: In proceeding of: Extraction et gestion des connaissances (EGC'2009). p. 495. Cépaduès, Strasbourg (2009).