ONTOLOGY-BASED GENERATION OF MULTILINGUAL QUESTIONS FOR ASSESSMENT IN MEDICAL EDUCATION

Maja Radović[1], Nenad Petrović[2], Milorad Tošić[2]

[1]University of Kragujevac, Faculty of Technical Sciences, Serbia
E-mail: maja.radovic@ftn.kg.ac.rs
[2]University of Niš, Faculty of Electronic Engineering, Serbia
E-mails: nenad.petrovic@elfak.ni.ac.rs, milorad.tosic@elfak.ni.ac.rs

Abstract. *The requirements of state-of-the-art curricula and teaching processes in medical education have brought both new and improved the existing assessment methods. Recently, several promising methods have emerged, among them the Comprehensive Integrative Puzzle (CIP), which shows great potential. However, the construction of such questions requires high efforts of a team of experts and is time-consuming. Furthermore, despite the fact that English language is accepted as an international language, for educational purposes there is also a need for representing data and knowledge in native language. In this paper, we present an approach for automatic generation of CIP assessment questions based on using ontologies for knowledge representation. In this way, it is possible to provide multilingual support in the teaching and learning process because the same ontological concept can be applied to corresponding language expressions in different languages. The proposed approach shows promising results indicated by dramatic speeding up of construction of CIP questions compared to manual methods. The presented results represent a strong indication that adoption of ontologies for knowledge representation may enable scalability in multilingual domain-specific education regardless of the language used. High level of automation in the assessment process proven on the CIP method in medical education as one of the most challenging domains, promises high potential for new innovative teaching methodologies in other educational domains as well.*

Key words: *medical education, assessment, multilingual teaching, ontology, artificial intelligence, comprehensive integrative puzzle*

1. INTRODUCTION

The main purpose of medical education (ME) is to train workforce to provide and improve patient treatment (Teodorczuk et al., 2017). In order to provide appropriate treatment medical physicians need to determine the correct diagnosis of the disease (Romero-Tris, Riaño & Real, 2011). The ability to communicate effectively with patients is a basic clinical skill that is essential for setting up right diagnosis and thus conducting all necessary treatments (Wette & Hawken (2016). Besides developing communication skills in their native language, medical students face the challenge of mastering English language as well, since it has become lingua franca of medical international communication.

In the context of newly advanced curricula and teaching processes, ME institutions are under high pressure to develop new and improve the existing assessment methods. Faculty

observation, oral examinations, and multiple-choice tests have traditionally been used as assessment methods in ME (Howley, 2004). Recently, promising new methods have emerged and of these, the Comprehensive Integrative Puzzle (CIP) shows great potential (Ber, 2003). This has been applied in various fields of medicine, and studies have shown positive attitudes towards it among students (Ber, 2003). It also has a strong discriminatory quality, due to the possibility of score differentiation (Van Bruggen et al., 2012). However, in the same way as for other assessment methods, the major obstacle to addressing higher cognitive skills lies in the construction of CIP. In order to reach its full potential, questions must be authored by highly skilled professionals. Such professionals are scarce for a number of reasons, including the difficulty of securing the right person for the job, the high costs of engaging specialists, and various organizational and management barriers. Accordingly, there is a strong need for the automatic generation of this type of assessment, as it will significantly reduce the workload of teachers.

One possible solution for CIP automatic generation is application of artificial intelligence (AI) that represents the ability of a machine to copy human intelligence processes, learn from experiences, adapt to new information and perform human-like activities. The main advantage of AI is precisely the automation of complicated and time-consuming tasks (Karsenti, 2019). For knowledge representation, we propose ontologies as a part of AI to achieve automatic generation of CIP assessment method. Ontology is defined in different areas of philosofy and science in somewhat different ways. The origine of the term lies in philosophy and relates to describing the world as seen by a group of people at certain time according to a school of thought that is based on a view of the world (El-Diraby &Osman). In informatic systems, an ontology is the means to formally model the structure of a system (Guarino 2009). Formal foundations of ontologies make them effective in using encoded knowledge in software development, such as in semantic (Shadbolt, Berners-Lee, & Hall, 2006) and big data technologies (Pop, Kołodziej, & Martino, 2016). Furthermore, their formal foundations allow for advanced domain modeling and algorithm development even in the most challenging applications in which sophisticated human expert reasoning has traditionally been the only choice.

Although English language is accepted as international language for publishing and research, in many applications and areas, there is a need for representing data and knowledge in native language as it enables content reuse and extends its usage to traditionally wider audience making it more effective (Ivanova, 2019). There are many approaches that rely on usage of ontologies and semantic web to improve learning and assessment but most of the developed ontologies target only English language. However, usage of multilingual ontologies for resource annotation could enable more efficient interlingual content delivery, as well as its reuse in e-learning systems, making the learning content itself adaptable for a much wider audience (Ivanova, 2019-2). In this paper, the multilingual ontologies based automatic generation of assessment in medical education is considered and evaluated in CIP case study.

The rest of this paper is structured into 7 sections. In Section 2, background and related work provide an overview of concepts and technologies used to propose our approach, such as medical education and the English for Medical Purposes ontologies, mapping between relational database and ontologies, multilingual mapping, etc. In Section 3, the problem of automated generation of CIP assessment method is defined. In Section 4, the OntoCIP ontology which semantically presents CIP assessment method is described. Section 5 covers the process of determining candidate answer in question generation. In section 6

architecture and implementation overview of the proposed approach for automated question generation is explained. Evaluation and results are given in Section 7. Finally, Section 8 covers conclusion and future work.

2. BACKGROUND AND RELATED WORK

2.1. English for Medical Purposes education

Proficient knowledge of English is fundamental for medical students, especially for those wishing to acquire and develop their skills by following the changes and developments in their field (Çelik, 2017). Although medical education institutions in many countries where English in not official language (Bakić-Mirić, 2013) (Yurchuk et. all, 2015) (Çelik, 2017) have adopted it as a medium of instruction, it has been noticed that students in such environmentss are most likely facing problems in dealing with their medical subjects. Consequently, designing appropriate courses to assist students in coping with the language challenges is becoming ever more relevant (Faraj 2015).

Since internationalization of higher education has been increasing, many medical students now study in English speaking institutions. Conducted researches (Hoekje, 2007) (Dahm, 2011) (Wette & Hawken, 2016) shown that language, communication and cultural issues can still persist for foreign students during and after their education and residency training. Accordingly, English speaking counties seek to develop ESP courses that will meet the needs of foreign students and residents regarding acculturation, the language and culture of the patient community, the language of the hospital, and intelligibility in performing key medical texts (Hoekje, 2007).

2.2. Assessment in medical education

Although medical education and training varies considerably across the world (Flores-Mateo & Argimon, 2007) one thing that is common for all countries and regions, is assessment, which is necessary at all stages of a medical student's education. From the moment they enroll into medical school and during all phases of medical education and practice, medical students and physicians are constantly being assessed for multiple purposes.

The use of a variety of different assessment methods has been characteristic of every field of education, and medical education is no exception. Since 1950s, the medical knowledge and clinical skills of students and doctors were often assessed using oral and written examinations. With the advent of new technologies, the way assessment is conducted in medical education changed rapidly. New methods, such as clinical simulations or multisource assessment, have focused on clinical skills (taking history from a patient and performing a physical examination), communication skills, procedural skills, and professionalism (Norcini & MyKinley, 2007).

2.2.1. The Comprehensive Integrative Puzzle Assessment Method

The preparation of a CIP is work that needs a team of experienced teachers or experts in different fields of medicine: clinician, pathologist, microbiologist, pharmacologist, biochemist, radiologist, etc. Each team determines clinical scenarios and prepares a pool of relevant material, such as ECG and EEG strips or interpretations, X-ray, CT, MRI or their interpretations, pathology pictures and/or slides or their descriptions and endoscopy

photographs or descriptions (Ber, 2003). Considering all these facts, it can be noticed that preparation of CIP is a time-consuming task even for an experienced teacher. According to Van Bruggen, preparation of one CIP question requires 1 to 2 hours for an experienced teacher (Van Bruggen et al., 2012). The CIP assessment takes the format of an extended matching crossword puzzle with an appropriate pool of options (Figure 1). The CIP answer sheet takes the form of a puzzle grid composed of between four and seven rows and columns (Ber, 2003). The left-hand column contains possible diagnoses or brief clinical vignettes, while the rest of the columns contain information about patient's story (medical history), physical exam, chest X-ray and electrocardiogram, laboratory, pathology and treatment. Beneath the puzzle grid, several parts marked Sections I, Section II, etc., correspond to the columns in the grid. Each section contains several options indicated by letters (for example a–f). In order to complete the puzzle, students choose one option from each section and place it in the cell in the appropriate column.

Matching columns						
Diagnosis	I: Medical history	II: Physical examination	III: Chest X-ray and ECG	IV: Laboratory and other tests	V: Treatment and follow-up	VI: Pathology
Chronic Obstructive Pulmonary Disease	B	F	B	B	A	C
Hemorrhagic Stroke	C	C	D	F	E	A
Type 1 Diabetes	D	B	E	C	B	E
Acute Renal Failure	E	D	C	D	C	F
Type 2 Diabetes	A	E	F	E	F	B
Myocardial Infarction	F	A	A	A	D	D

Section I: Patient's presenting stories
(a) Patient complains of recurrent moderate tingling of the feet over the past month. The pt reports moderate weak urinary stream, moderate foamy urine. Patient is a 57-year m. Patient is a heavy smoker of cigarettes. Patient reports that he never drinks alcohol.
(b) pt c/o 3 weeks h/o critical cough (not checked), critical dyspnea, critical shortness of breath. NKDA. Patient is a female landlord aged 34 yrs.
...
(d) Patient presents with 3 years history of severe weight loss. Pt also reports increased frequency of severe frequent urination, severe increased thirst. Patient is a 35 yr old female receptionist.
...
Section IV: Laboratory and other tests
(a) Transthoracic echocardiography: shows Ventricular Septal Rupture
(b) FEV1=35 %
(c) HCO3=17 meq/L
...
Section V: Treatment and follow-up
(a) OMS 50, via nasal cannula (contin)
(b) Potassium Chl po qd
(c) AVAPRO, by mouth daily
(d) nitroglycerine 2xday
(e) mannitol IV
(f) glyburide daily
...

Figure 1 CIP answer sheet, generated from the used data source

It should be noted that CIP allows for adjustments to different degrees of difficulty. For example, at the lowest level of difficulty, each option may be used only once; that is, one description of a medical history (or physical, treatment, etc.) can be assigned to only one diagnosis. CIPs at higher degrees of difficulty may offer more than six sections (columns), more than six options/distracters per section, and may include instructions that 'each option may be used once, more than once, or not at all' (Ber, 2003). In addition, not all columns need to be displayed. If a diagnosis does not need a chest X-ray or ECG, the teacher can choose not to display this column in the assessment.

2.3. Ontologies

Over the years, ontologies have matured into a cornerstone technology in conjunction with semantic web services and the semantic grid (Gladun, Rogushina, García-Sanchez, Martínez-Béjar, & Fernández-Breis, 2009). Advances in the semantic web (Tim, Lee, Hendler, & Lassila, 2001) and linked data (Bizer, Heath, & Berners-Lee, 2009) have facilitated the emergence of a new trend in question generation that is based on the use of ontologies. The adoption of ontologies is particularly promising when scalability is required, either due to the massive amounts of data used or the large number of users involved, and when the task at hand requires an expert level of knowledge that is specific to the target application domain. In the literature, various approaches involving the successful application of ontologies in the generation of assessment methods can be found (Cubric & Tosic, 2011)(Jelenkovic & Tosic, 2013)(Vinu & Kumar, 2015)(Vinu, Alsubait, & Kumar, 2016)(Radovic et al., 2017)(Radovic et al., 2018). Ontologies can be grouped into four types: (i) top-level ontologies describing the most general concepts (space, time, emotions, etc.) that are independent of a particular problem or domain; (2) domain ontologies; and (3) task ontologies describing generic tasks or activities in general domains (medicine, cars, biology) such as diagnosing (for medicine) and selling (for cars); these types of ontologies refine the concepts introduced in the top-level ontology; (4) application ontologies that describe concepts depending on a particular domain and task. For the purpose of automated assessment methods task ontologies should be applied.

2.3.1. Multilingual Approach in Ontologies

Multilingual ontology mapping is defined as a type of ontology mapping where the matcher is capable of dealing with ontologies expressed (or labelled) in multiple languages. Multilingual ontology mapping requires modifications of some of the existing mapping algorithms in order to make them work in a multilingual environment. However, most of the approaches to multilingual ontology mapping rely on general-purpose machine translation services (Ivanova, 2019) (Ivanova, 2019-2). Furthermore, multilingual labels are supported using the language tagging facility of RDF literals, which makes RDF suitable for development of multilingual semantic knowledge bases. In the existing literature, several solutions have been proposed. In (Alatrish, E. S. et al., 2014), various software tools available on the Internet were used (most notably DODDLE-OWL, WordNet, Protégé and XSLT transformations) to implement a procedure to construct domain ontology for any natural language. On the other side, in (Ivanova, 2019), a framework leveraging a set of domain ontologies for representation of pedagogical knowledge was introduced. In that paper, Bilingual Linguistic Domain Ontology (BLDO) was used to presents the structure of the course knowledge (or learning content) using labels in both Bulgarian and English.

Moreover, in (Trojahn et al., 2008), there is presented a framework for automatic mapping of multilingual Description Logics (DL) ontologies. Their focus was on the process of translating ontologies, where the source ontology is translated to the target ontology language, using an appropriate dictionary by translation agent, where translation agent refers to a software service, component or platform which performs the translation of given input text originally written in one language to some other target language.

3. THE ONTOCIP ONTOLOGY

For our case related to assessment in medical sciences based on Comprehensive Integrative Puzzle Assessment method, we rely on our OntoCIP ontology (Radovic et al., 2018). In Figure 2, the main classes and properties of OntoCIP ontology are shown.
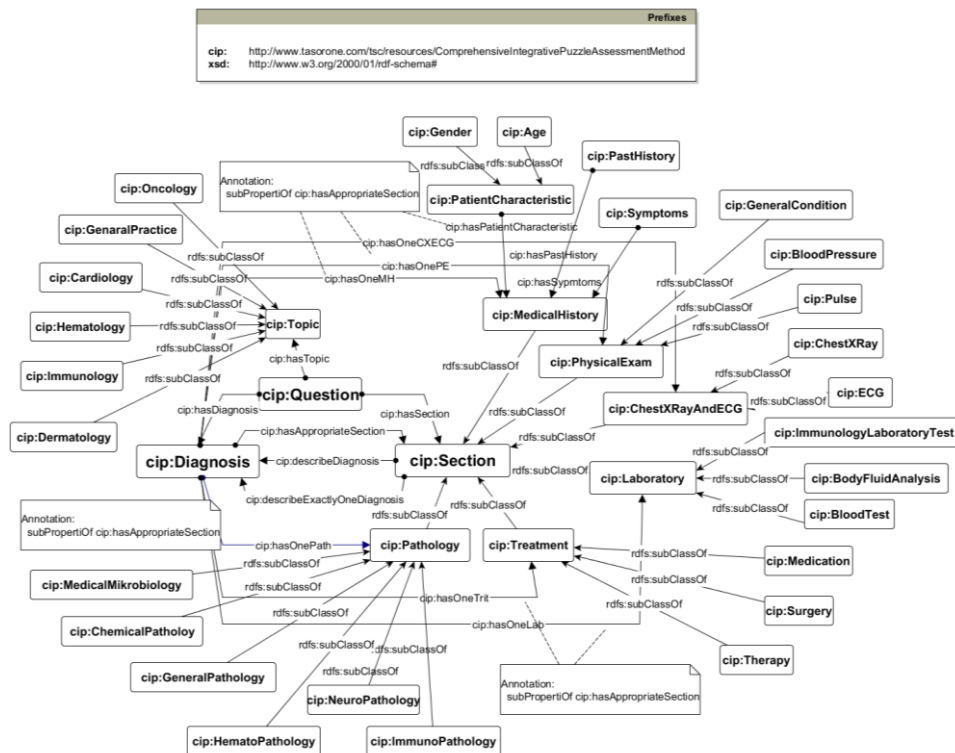

Fig. 2 Main concepts in the OntoCIP ontology

The main classes in OntoCIP are cip:Diagnosis, cip:Section and cip:Question. Diagnosis is defined in medical terminology as the identification of a disease or condition through evaluation and examination[1]. It is the process of determining which disease or condition

---

[1] https://tinyurl.com/82v7u5d

explains person's symptoms and signs. It can also be presented as a short patient vignette. Each individual instance of the cip:Question class has a containment relation with several (between four and seven) instances of the cip:Diagnosis class and several (between four and seven) instances of the cip:Section class. In other words, each question consists of several diagnoses and sections. This is modeled using the properties cip:hasDiagnosis and cip:hasSection.

The cip:Section class is specialised into the cip:MedicalHistory, cip:PhysicalExam, cip:ChestXRayAndECG, cip:Laboratory, cip:Treatment and cip:Pathology subclasses. Regardless of the level of difficulty, each instance of the cip:Diagnosis class has a containment relation with exactly one instance of the subclasses cip:MedicalHistory, cip:PhysicalExam, cip:ChestXRayAndECG, cip:Laboratory, cip:Treatment and cip:Pathology. This is modelled using the properties cip:hasOneMH, cip:hasOnePE, cip:hasOneCXRECG, cip:hasOneTrit, cip:hasOnePath and cip:hasOneLab. All of these properties are subproperties of the cip:hasAppropriateSection property. Looking at this from the opposite direction, if the lowest difficulty level of the CIP assessment is targeted, then cip:Section will have instances that have a relationship with exactly one appropriate instance of the cip:Diagnosis. In this case, the property cip:describeExactlyOneDiagnosis is used for implementation.

The cip:MedicalHistory has a relationship with the classes cip:Gender, cip:Age, cip:Symptoms and cip:PastHistory. This is modelled due to the fact that the patient's generalities, symptoms or patient subject feelings with a past history of illness are important in order for a physician to establish a diagnosis. The cip:PhysicalExam class contains the subclasses cip:GeneralCondition, cip:BloodPreasure and cip:Pulse. These subclasses represent the main actions that physician carries out during an initial patient exam. The cip:ChestXRayAndECG class is further divided into the cip:ChestXRay and cip:ECG classes. These represent ECG strips and X-ray films, or their interpretation, respectively. The cip:Laboratory represents laboratory tests and is further specialised into cip:ImmunologyLaboratoryTest, cip:BodyFluidAnalysis, cip:BloodTest. Depending on the patient's symptoms, the physician will order certain laboratory tests. The cip:Treatment class is divided into several subclasses: cip:Medication, cip:Surgery and cip:Therapy. Each of these is further specialised into several subclasses. The cip:Pathology class is also divided into several subclasses such as cip:ImmunoPathology, cip:NeuroPathology, etc.

4. CANDIDATE ANSWER GENERATION

Based on the difficulty level, the cip:Section class may have instances that have a relationship with one or more instances of the cip:Diagnosis class. This is implemented using the property cip:describeDiagnosis. The cip:Section class can also have instances that have no relationship with instances of the cip:Diagnosis class that represents the distractors in the question. In order to generate distractors, a semantic similarity strategy was used. Semantic similarity can be defined in several different ways, but for the purposes of this paper, we define it as a function of the distance between the concepts in a graph corresponding to the hierarchical structure of the underlying ontology (Gan, Dou, & Jiang, 2013). The distance between two concepts (Lee, Shah, Sundlass, & Musen, 2008) is a numerical representation of how far apart two concepts are from one another in some geometric space. It can be considered the inverse of semantic similarity. Figure 3

gives an example of this distance in the classification (diagnosis) of chronic diseases. The dashed line indicates the large distance between *Chronic ischemic colitis* and *Intermittent exotropia of the right eye*; these are not very similar, and putting them in the same question will produce an easy choice for students. On the other hand, *Chronic ischemic colitis* and *Chronic bronchiolitis* have a short distance (marked with a dashed/dotted line), and are very similar (they are both in the same level in the hierarchy); hence, putting them into a given question would require a higher level of knowledge of diagnosis and appropriate medical history, treatment, etc. to give the right answer.
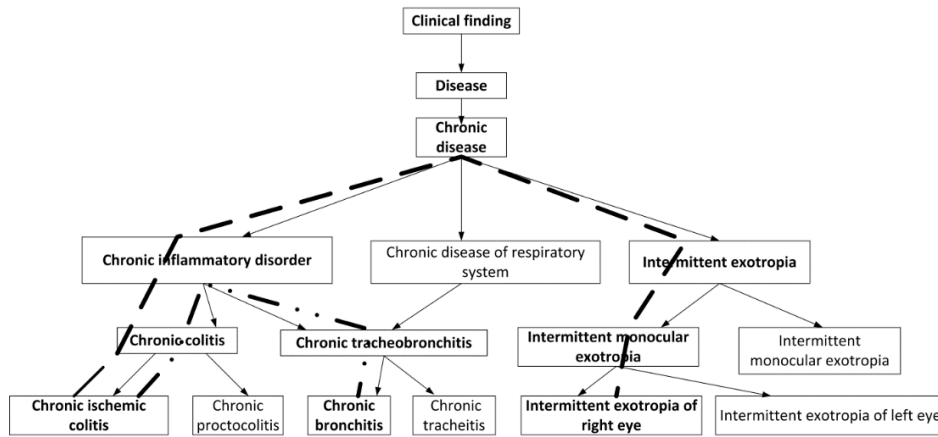

Fig. 3 Illustration of how distance in knowledge taxonomy influences
the difficulty of the question

5. ARCHITECTURE AND IMPLEMENTATION OVERVIEW

The main components, their interactions and basic principles of the proposed approach are illustrated in Figure 4.
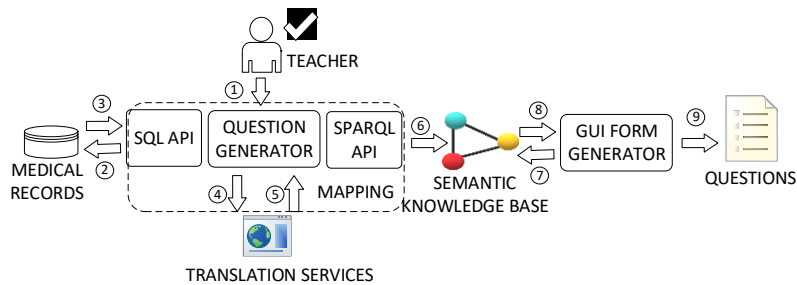

Fig. 4 Solution overview: 1-Difficulty and language settings 2-SQL queries 3-SQL query results 4-Translation service call 5-Translation results 6-RDF triplets 7-SPARQL queries 8-SPARQL query results 9-Question GUI form

First, teacher selects the question difficulty level, inputs the diagnosis name and sets the target language. After that, for each of the ontology properties, question generator executes queries against the relational medical record database using SQL API in order to retrieve the necessary question property values. The medical record database used in this paper contains lab results, diagnostic and medication details of a large number of patients collected during ten years. It is available online as a public, free-to-use dataset on data.world2. It contains data about 3063 medical encounters, 7509 lab results and 1176 medical fulfillments.

In Table 1, full mapping between the medical record database and ontology property values is described. The first table column denotes the name of column(s) from the medical database used for mapping. The second column shows the filtering criteria against the previous database columns used for mapping. The third table is the name of ontology property that is being mapped. To each of entries from Figure 1, SQL query similar to the previously given example is formed. These mapping queries are executed when knowledge about given input diagnosis has to be constructed.

Table 1 Mapping between medical record database and OntoCIP ontology

Medical record database column	Filtering criteria	Ontology property
encounter_dx.description	encounter_dx.description= *'input diagnosis'*	hasDiagnosis
encounter.soap_note	encounter_dx.description= *'input diagnosis'*	hasAppropriateME
lab_results.result_description	(lab_results.result_name LIKE '%X-Ray%' or lab_results.result_name LIKE '%ECG%') and encounter_dx.description=*'input diagnosis'*	hasAppropriateCHECG
lab_results.result_name lab_results.result_description	encounter_dx.description=*'input diagnosis'*	hasAppropriateLAB
medication_fulfillment.drug_name medication_fulfillment.sig	encounter_dx.description=*'input diagnosis'*	hasAppropriateTR

6. EVALUATION AND RESULTS

In this section, the performance evaluation of the proposed framework for automated, ontology-based generation of questions for multilingual assessment in medical education is given and compared to the time needed for manual creation of the same questions. The evaluation was performed on a laptop equipped with Intel i7 7700HQ CPU, 16GB DDR4 RAM and 1TB HDD, running on Windows 10. The backend of the platform for code generation was entirely written in Java, while the ontology management, triplet insertion and querying were done using TaaSOR online service and its Java API (Tosic et al., 2012).

---

[2] https://data.world/arvin6/medical-records-10-yrs

In Table 2, the achieved results are given. The first column shows the number of diagnoses present within the puzzle. The second column contains the time necessary for execution of SPARQL queries against the medical records database during the process of mapping. Moreover, the third column is the time spent for generation of candidate diagnoses according to the pre-selected difficulty level based on similarity. The fourth column presents the time needed for insertion of corresponding triplets into the semantic knowledge base with respect to given question ontology. Furthermore, the fifth column is the time needed for generation of corresponding question's visual representation (HTML web page containing table for puzzle representation in our case). The sixth columns is the question's target language. Finally, the last column is the total time needed for automated question generation, obtained as a sum of previous four times columns.

Table 2 Evaluation results for automated generation of questions using OntoCIP ontology

Number of diagnoses	SQL queries [s]	Candidate generation [s]	Triplet insertion [s]	Page generation [s]	Target language	Total time [s]
4	0.98	0.089	0.163 1.624	0.038	English Italian	1.270 2.731
5	0.96	0.091	0.208 1.881	0.040	English Italian	1.299 2.972
6	0.99	0.092	0.208 2.015	0.044	English French	1.334 3.305
7	1.02	0.094	0.211 2.427	0.048	English French	1.373 3.589

In Figure 5, a screenshot of the web page containing the generated question for evaluation purposes in English is given. According to the achieved results, it can be noticed that the total question generation time increases with the number of diagnoses considered, but it does not exceed order of magnitude of 1 in our experiments (from 4 to 7 diagnoses per question). Furthermore, it can be noticed that triplet insertion is much longer (up to more than 10 times) when target language is not English. It can be explained by the fact that in cases when other target language is selected, an online translation service is used, which introduces additional delay. In this study, we have used translation service Yandex instead of widely adopted Google Translate because the usage of its programmatic interface in third party applications is still entirely free unlike Google Translate API. Moreover, it was shown that Yandex Translate outperforms Google Translate in some cases (Bülbül et al., 2020).

Compared to manual procedure of CIP question construction, the proposed approach obviously speeds up the process which usually takes 1-2 hours for puzzle with 7 diagnoses (van Bruggen et al., 2012). Moreover, it also has huge potential to reduce costs and time, as manual procedure usually requires several experts (around 4-7 if not the whole team for each of the Sections), while multilingual assessment is enabled, which requires much more efforts for manual construction.

Diagnosis	Answers		
	Medical history	Laboratory and other tests	Treatment and follow-up
Chronic Obstructive Pulmonary Disease	B	B	A
Hemorrhagic Stroke	C	F	E
Type 1 Diabetes	D	C	B
Acute Renal Failure	E	D	C
Type 2 Diabetes	A	E	F
Myocardial Infarction	F	A ✕	D

Medical history

- A. Patient complains of recurrent moderate tingling of the feet over the past month. The pt reports moderate weak urinary stream, moderate foamy urine. patient is a 57 year male patient is a heavy smoker of cigarettes. Patient reports that he never drinks alcohol.
- B. Patient complains of 3 weeks history of critical cough (not checked), critical dyspnea, critical shortness of breath. NKDA. patient is a female landlord aged 34 years.
- C. Patient presents with chronic Hemorrhagic Stroke. He is a 45 years old male. He states he has never smoked a cigarette in his life.
- D. Patient presents with 3 years history of severe weight loss. pt also reports increased frequency of severe frequent urination, severe increased thirst. patient is a 35 yr old female receptionist.
- E. A 44 years old male complains of 3 weeks history of Acute Renal Failure. NKDA. he smokes one pack daily for 4 years. denies any alcohol use.
- F. Patient complains of Myocardial Infarction and other medical problems. NKDA. she is a female flight attendant aged 40 years. Denies any smoking. Patient is a moderate drinker.

Laboratory and other tests

- A. Transthoracic echocardiography: shows Ventricular Septal Rupture
- B. FEV1=35 %
- C. HCO3=17 meq/L
- D. Na=43 meq/L
- E. HbA1c=9%
- F. CT Head:findings of acute hemorrhage of hyperdense focal collection

Treatment and follow-up

- A. OMS 50, via nasal cannula (contin)
- B. Potassium Chl po qd
- C. AVAPRO, by mouth daily
- D. nitroglycerine 2xday
- E. mannitol IV
- F. glyburide daily

Fig. 5 Screenshot of generated CIP question in English

7. CONCLUSION AND FUTURE WORK

The medical students enrolled in medical schools all around the world are going to become the doctors of tomorrow, carrying values, skills and hopes for the profession into the future. Medical education today represents the cornerstone of the overall medicine in future. As a Lingua Franca, knowledge of English language has become imperative for students of medicine particularly in a highly connected and globalized world where medical crises travel around the world in an unprecedented speed. Therefore, multilingual medical education, where different native languages are used together with English language, represents the key enabler for addressing the urgent need for new generation of medical personnel.

As one of the crucial parts of medical education, assessment methods must be able to address the new challenging circumstances. Artificial Intelligence represents a promising approach, particularly adoption of ontologies that have powerful knowledge representation and reasoning capabilities. However, in addition to new technologies, new advanced assessment methodologies are required as well. CIP represents one such assessment method with a strong discriminatory quality. In other words, CIP items exhibit minimal false-

positives (cases when students answer the item correctly even though they do not have necessary knowledge or competence) and false-negatives (cases when students give wrong answers even though they have relevant knowledge or competence). The novelty and widespread adoption of this assessment method highlight the need for automatic generation because time consuming manual construction is not a practical option.

In this paper we have presented one possible solution for automated question generation based on reasoning about domain knowledge represented by ontologies. Ontologies are used for domain knowledge as well as questions type conceptualization so that assessment is done at the conceptualization and reasoning level. Because ontologies allow annotation of concepts in different languages, domain questions outgrow domain knowledge assessment, that is traditionally their primary purpose, and become a new advanced tool for language for specific purposes teaching. In this paper, we primarily have been focused on teaching English for specific purposes, but the proposed approach can be equally effectively applied to other languages as well. In this way, domain specific problem solving is integrated into the language for specific purposes teaching.

First, OntoCIP ontology is developed to semantically represent structure and methodology of the CIP assessment method. After multilingual ontology mapping from medical relational database to OntoCIP ontology, automatic generation of question's visual representation is performed by populating the question template. The main components of the proposed architecture as well as implementation overview are explained. Conducted evaluation shows that proposed approach significantly reduces teachers' time for multilingual CIP generation. In our future work, we plan to develop an open access online service that would provide the proposed concept to a global community as a knowledge sharing platform. We envision a heterogeneous community that would include language experts, domain experts, practitioners, and students, but eventually institutions and organizations as well.

REFERENCES

Abbas, M. A., Berio, G. (2013). Creating Ontologies Using Ontology Mappings: Compatible and Incompatible Ontology Mappings. 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), pp. 143-146. https://doi.org/10.1109/wi-iat.2013.169

Alatrish, E. S., Tošić D., Milenković, N. (2014). Building ontologies for different natural languages Computer Science and Information Systems 11(2), pp. 623-644.

Bakić-Mirić N. (2013). The importance of teaching cultural competence in Serbian medical schools – a conceptual framework the journal of teaching English for specific and academic purposes vol. 1, no 1, 2013, pp. 45-52

Barrasa, J., Corcjo, O., Gomez-Perez, A. (2004). R2O, an Extensible and Semantically based Database-to-Ontology Mapping Language, Proceedings of the 2nd Workshop on Semantic Web and Databases (SWDB2004), pp. 1-17.

Basim M. Abubaker Faraj. (2015). English for medical education in EFL context. The journal of teaching English for specific and academic purposes Vol. 3, No 1, 2015, pp. 121–148

Ber, R. (2003). The CIP (comprehensive integrative puzzle) assessment method. Medical Teacher, 25(2), 171–176. https://doi.org/10.1080/0142159031000092571

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web will enable machines to comprehend semantic documents, Scientific American: Feature Article: The Semantic Web), 1–5.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data - The story so far. International Journal on Semantic Web & Information Systems - Special Issue on Linked Data, 1–22.

Bülbül, Z., Nur Çetinkaya, A., Buse Arıcı, F. (2020). Google Translate and Yandex Translate's Differences in Naturalness, Clarity, and Accuracy: A Comparison Study on Machine Translation (preprint), available at: https://www.researchgate.net/publication/339029502_Google_Translate_and_Yandex_Translate%27s_Differences_in_Naturalness_Clarity_and_Accuracy_A_Comparison_Study_on_Machine_Translation.

Capaldi V.F, Durning S.J, Pangaro L.N., Ber R. (2015). The Clinical Integrative Puzzle for Teaching and Assessing Clinical Reasoning: Preliminary Feasibility, Reliability, and Validity Evidence, Military Medicine, Volume 180, Issue suppl_4, 1 pp. 54–60, https://doi.org/10.7205/MILMED-D-14-00564

Çelik H. (2017). Perceived language needs of undergraduate medical students enrolled in a consecutively running ESP course. The journal of teaching English for specific and academic purposes vol. 5, no 2, pp. 209–218

Cubric, M., & Tosic, M. (2011). Towards automatic generation of e-assessment using semantic web technologies. International Journal of E-Assessment,7(1)

Dahm M. (2011). Exploring perception and use of everyday language and medical terminology among international medical graduates in a medical ESP course in Australia. English for Specific Purposes 30. pp: 186–197 http://dx.doi.org/10.1016/j.esp.2011.02.004

El-Diraby, T. E., & Osman, H. (2011). A domain ontology for construction concepts in urban infrastructure products. Automation in Construction, 20(8), 1120–1132. doi:10.1016/j.autcon.2011.04.014

Farouk M., Ishizuka M. (2012). An extensible approach for mapping relational DB to RDF. 2012 Japan-Egypt Conference on Electronics, Communications and Computers, pp. 163-167. https://doi.org/10.1109/jec-ecc.2012.6186976

Flores-Mateo G; Argimon JM (2007). "Evidence based practice in postgraduate healthcare education: A systematic review". BMC Health Serv Res 7: 119. doi:10.1186/1472-6963-7-119.

Gan, M., Dou, X., & Jiang, R. (2013). From ontology to semantic similarity: Calculation of ontology-based semantic similarity. The Scientific World Journal, 2013, 793091. https://doi.org/10.1155/2013/793091

Gladun, A., Rogushina, J., García-Sanchez, F., Martínez-Béjar, R., & Fernández-Breis, J.T. (2009). An application of intelligent techniques and semantic web technologies in e-learning environments. Expert Systems with Applications, 36(2 PART 1), 1922–1931. https://doi.org/10.1016/j.eswa.2007.12.019

Guarino, N. (1998). Formal ontology and information systems. International Journal of Human-Computer Studies - Special Issue: The Role of Formal Ontology in the Information Technology, (June), 3–15.

Hoekje B. (2007). Medical discourse and ESP courses for international medical graduates (IMGs). English for Specific Purposes 26 pp. 327–343 https://doi:10.1016/j.esp. 2006.09.002

Howley, L. D. (2004). Performance assessment in medical education: Where we've been and where we're going. Evaluation and the Health Professions, 27(3), 285–303. https://doi.org/10.1177/0163278704267044

Ivanova, T. (2019). E-Learning resource reuse, based on bilingual ontology annotation and ontology mapping. International Journal of Advanced Computer Research, Vol 9(45) ISSN (Print): 2249-7277 ISSN (Online): 2277-7970 pp. 351-364. http://dx.doi.org/ 10.19101/IJACR.2019.940101

Ivanova, T. (2019). A bilingual ontology mapping and enrichment approach for domain ontologies in e-learning. Proceedings of the 20th International Conference on Computer Systems and Technologies - CompSysTech '19, pp. 284-291. https://doi.org/10.1145/3345252.3345257

Jain, V., Prasad, S. V. A. V. (2014). Mapping Between RDBMS And Ontology: A Review, International Journal of Scientific & Technology Research, vol. 3 issue 11, pp. 307-313.

Jalali, V., Bagheri, A. (2007). Semi-Automated Mapping from RDB to Ontology. Proceedings of the 3rd International Conference on Information & Knowledge Technology, pp. 1-4.

Jelenkovic, F., & Tosic, M. (2013, May). Semantic multiple-choice question generation and concept-based assessment. The First International Conference on Teaching English for Specific Purposes, Nis, Serbia (pp.1-11)

Karsenti, T. (2019). Artificial intelligence in education: The urgent need to prepare teachers for tomorrow's schools. Formation et profession, 27(1), 105-111. http://dx.doi.org/ 10.18162/fp.2019.a166

Lee, W., Shah, N., Sundlass, K., & Musen, M. (2008, November). Comparison of ontology-based semantic-similarity measures. AMIA Annual Symposium Proceedings, 384–388.

Norcini J., McKinley D. (2007). Assessment methods in medical education. Teaching and Teacher Education, Volume 23, Issue 3, pp. 239–250. ISSN-0742-051X

Radovic, M., Tosic, M., & Milosevic, D. (2017, September). An ontology based approach to assessment in medical education, Eighth International Conference on eLearning 2017, Belgrade, Serbia (pp. 28–29).

Radovic, M., Tosic, M., Milosevic, D., & Jankovic, D. (2018). OntoCIP - An ontology of comprehensive integrative puzzle assessment method suitable for automatic question generation. Advances in Intelligent Systems and Computing, 716(September 2017), 394–399. https://doi.org/10.1007/978-3-319-73204-6_44

Romero-Tris, C., Riaño, D., & Real, F. (2011). Ontology-Based Retrospective and Prospective Diagnosis and Medical Knowledge Personalization. Lecture Notes in Computer Science, 1–15. doi:10.1007/978-3-642-18050-7_1

Teodorczuk, A., Yardley, S., Patel, R., Rogers, G. D., Billett, S., & Worley, P. (2017). Medical education research should extend further into clinical practice, Med Educ, 51(11), 1098-1100. https://doi.org/10.1111/medu.13459

Tosic, M., Seskar, I., Jelenkoivc, F. (2012). TaaSOR – Testbed-as-a-Service Ontology Repository. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 44, pp. 419-420.

Trojahn C., Quaresma, P., Vieira, R. (2008). A Framework for Multilingual Ontology Mapping. Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, pp. 1034-1037.

Van Bruggen, L., Manrique-van Woudenbergh, M., Spierenburg, E., & Vos, J. (2012). Preferred question types for computer-based assessment of clinical reasoning: A literature study. Perspectives on Medical Education, 1(4), 162–171. https://doi.org/10.1007/s40037-012-0024-1

Vinu, E. V., Alsubait, T., & Kumar, P. S. (2016). Modeling of item-difficulty for ontology-based MCQs. CoRR, abs/1607.00869. Retrieved from http://arxiv.org/abs/1607.00869

Vinu, E. V., & Kumar, P. S. (2015). Automated generation of assessment tests from domain ontologies. Semantic Web, 8(6), 1023–1047. https://doi.org/10.3233/SW-170252

Wette, R., Hawken, S. J. (2016). Measuring gains in an EMP course and the perspectives of language and medical educators as assessors. English for Specific Purposes, 42, 38–49. doi:10.1016/j.esp.2015.11.002

Yeoh K. G. (2019). The future of medical education. Singapore medical journal, 60(1), 3–8. https://doi.org/10.11622/smedj.2019003

Yurchuk V., Lakhno A., Yurchuk V., Petrova E., Nosova L. (2015). Foreign language teaching based on organizing practice-oriented communication in the classroom (from the experience of teaching medical English). The journal of teaching English for specific and academic purposes vol. 3, no 1, pp. 77–83