ПОДХОД ЗА ПРИЛОЖЕНИЕ НА ОНТОЛОГИИ ПРИ ГЕНЕРИРАНЕТО НА ТЕСТОВИ ВЪПРОСИ НА ПРАВИЛЕН БЪЛГАРСКИ ЕЗИК

Мария Грънчарова- Христова[1], Тодорка Глушкова[2]
[1]Хуманитарна гимназия „Св. св. Кирил и Методий“, Пловдив
[2]Пловдивски университет „Паисий Хилендарски“, Пловдив

AN APPROACH FOR APPLYING ONTOLOGIES WHEN GENERATING TEST QUESTIONS IN CORRECT BULGARIAN LANGUAGE

Mariya Grancharova-Hristova[1], Todorka Glushkova[2]
[1]Humanitarian High School „St. St. Cyril and Methodius“, Plovdiv
[2]Plovdiv University ‘Paisii Hilendarski’, Plovdiv

**Авторите изказват благодарност към научен проект NextGenerationEU, финансиран от Европейския съюз, чрез Националния план за възстановяване и устойчивост на Република България, проект № BG-RRP-2.004-0001-C01, както и на научен проект FP23-FMI-002 „Интелигентни софтуерни инструменти и приложения в изследванията по математика, информатика и педагогика на обучението“ към ПУ „П. Хилендарски“. за частичното финансиране на настоящата работа.*

Abstract: Digitization and semantic processing of common sense knowledge is one of the most significant tasks related to the development and use of cyber-physical educational platforms. One of the good approaches for semantic modelling of this knowledge is the use of ontologies. Through appropriate structuring, the ontology provides opportunities for automatic generation of questions in the correct Bulgarian language, through which students' knowledge can be checked and evaluated. The article presents the results of the development and use of an ontology with knowledge of history and literature, which provides an opportunity to generate sentences and questions in grammatically correct Bulgarian. The authors' experience of using this approach in school education is also shared.

Keywords*:* Semantic modeling, ontologies, Artificial Intelligence in Education (AIEd)

Въведение. Дигитализирането и семантична обработката на знания с общ смисъл е една от най-значимите задачи, свързани с разработката и използването на съвременни образователни платформи [1]. Един от удачните подходи за семантично моделиране на тези знания е използването на онтологии [2]. Чрез подходящо структуриране онтологията предоставя възможности за автоматично генериране на въпроси на правилен български език, чрез които да могат да се проверяват и оценяват знанията на учениците.

В статията се представят резултатите от разработката и използване на онтология със знания по история и литература, в която се осигурява възможност за генериране на изречения и въпроси на граматически правилен български език. Споделен е и опитът на авторите от използването на този подход в училищното образование.

Семантично моделиране и онтологии. В днешно време, всеки потребител на глобалната мрежа се възползва от услугите, които предлага. Информацията е представена под формата на хипертекстови документи, които са съхранени и се предоставят от web сървъри, с възможност при поискване от конкретен потребител, тя да достига до него, като се разглежда чрез web браузъри. Семантичният уеб е разширение на сегашния (синтактичен уеб), при който информацията е представена с по-добре дефинирано значение, което позволява работата между хора и компютри да е по-успешна [3]. Чрез използването на формална семантика и подходящо прилагане на семантични технологии, системите за изкуствен интелект могат не само да „прочетат документи от компютри“, но и да „разберат“ семантичната (смислова) същност на информацията, предадена им от хора или други компютри и да могат да обработят информацията по начина, по който хората биха го направили. В това се състои същността на семантичното моделиране, а имено представяне на връзки, които съществуват между конкретни данни и те впоследствие да се отделят от разработчика и физически да бъдат поделяни между различни потребители и многократно използвани [4].

Семантичното моделиране може да бъде реализирано посредством онтологии. Понятието онтология, от философска гледна точка, представлява определена система от категории, които са следствие от определена система от възгледи за света. Така реалният свят може да бъде разделен на части, да бъдат описани свойствата им, за да се представи тяхното развитие и поведение. В контекста на информатиката има различни дефиниции, но най-популярната е дадена от Томас Грюбер през 1993 г. и представлява яв-

на (експлицитна) спецификация на определена концептуализация [5]. Концептуализацията описва множество обекти и понятия, знания за тях и връзки между тях. Онтологиите изясняват структурата на знанията за дадена област и позволяват споделянето на тези знания. Независимо от езика за описание, обикновено те се състоят от концепти (понятия, класове, обекти), индивидуалности (основни компоненти от най-ниско ниво), свойства, зададени връзки между тях и правила за извод.

За формалното представяне на знания чрез онтологии са разработени множество езици. Основните базови езикови нива са:

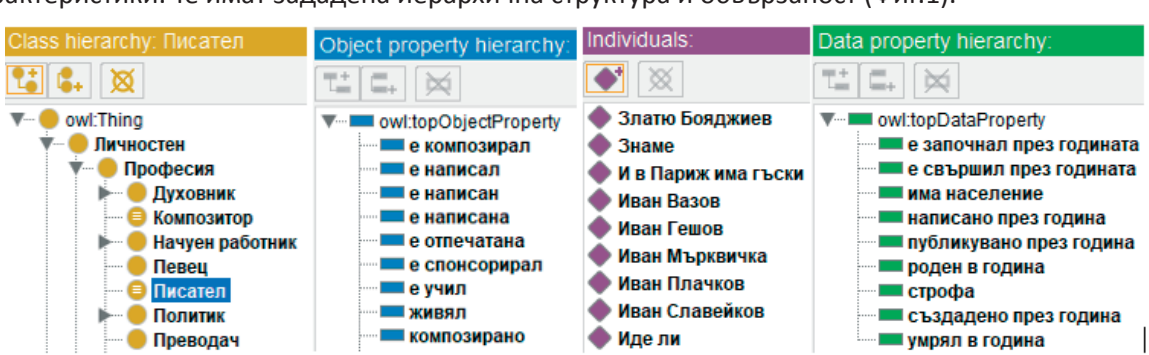
- **RDF** (Resource Description Framework) е модел на данни, съдържащ обекти ("ресурси") и отношения между обекти.
- **RDF scheme** (Resource Description Framework Schema) **е речник,** описващ свойствата и класовете на RDF ресурсите.
- **OWL** (Web Ontology Language) [6] – осигурява релации между различните обекти. Последната версия OWL 2 използва допускането за отворен свят – всичко, което не знаем, не е дефинирано. Допускането за затворен свят, въплъщава идеята, че всичко, което не знаем е грешно.

Онтологиите като модели за представяне на знания и извод намират широко приложение в различни области на обществения живот. Особено подходящи, намиращи място в областта на науката и образованието, когато са част от системи, които предлагат модул „Тестване" [7].

Разработка на онтология по хуманитарни науки. При разработката на една онтология е необходимо да се състави стратегия, за да се постигне търсения резултат. При създаване на онтология в хуманитарната област е необходимо да се конкретизира домейна. Той представлява база от знания, чрез която е описана богатата история на днешната Хуманитарна гимназия [8]. Тя е създадена в град Пловдив от Найден Геров през 1850 година като Епархийско училище и прераства в първата гимназия на територията на България. Много бележити личности, които оставят трайна диря в историята на България са били учители и възпитаници на училището. В онтологията се представят и структурират знания – за училището, за личностите и техните професии, периодите през които са живели и работили, населените места, които са обитавали, резултатите от тяхната работа, така че лесно да се обработва информацията и да се правят заключения, като се прави проверка на истинността на дадено твърдение. Концептите представляват понятията, които са зададени като йерархия от класове, които са свързани помежду си чрез различни свойства: обектни, анотационни и свойства данни. Тези връзки формират аксиоми, които оформят твърдения, които представляват триплети, подчинени на структурата: субект – предикат – обект и дават възможност да се правят изводи за истинността на твърдения в областта. За да бъдат описани тези знания се следват няколко важни задачи:

- Да се анализира областта и да се вземане решение кои знания е необходимо да бъдат описани в онтологията. Излишните и непотребни знания да бъдат изключени, за да не се утежнява обработваната информация;
- Определяне на основните понятия в областта и създаване на съответните класове. Изграждане на йерархия от тези класове;
- Определяне на необходимостта от конкретни свойства: обектни, свойства данни и анотационни;
- Задаване на ограничения на свойствата;
- Въвеждане на инстанции;
- Проверка на онтологията за съгласуваност и двусмислия.

След като е определен домейнът е необходимо да се състави класовата организация, която описва основните обекти в онтологията. Класовете във всяка онтология представляват колекции от обекти с общи характеристики. Те имат зададена йерархична структура и обвързаност (Фиг.1).



Фигура 1. Класове, обектни свойства, индивидуалности и свойства данни

Обектните свойства (Фиг. 1) в една онтология позволяват да бъде създадена връзка между индивидуалностите на два класа създавайки Resourse Description Framework (RDF) тройка: субект-предикат-обект, където субектът се явява подлог в изречението, докато обектът играе ролята на допълнение. Чрез зада-

ване на област (domain) и обхват (range) на обектното свойство, се показва възможността за изграждане на връзка между индивидуалности от класа на областта с индивидуалности от класа на обхвата чрез даденото свойство.

Индивидуалностите (Фиг. 1) представляват конкретни обекти като хора, произведения, градове и т.н. и отговарят на определена дескриптивна логика. Те спадат към основните компоненти, които изграждат онтологията.

Свойства данни (Фиг. 1) описват данните на дадена индивидуалност с различни Resourse Description Framework Scheme RDF(S) типове данни. По този начин дадена индивидуалност може да се свърже с конкретна нейна стойност. Чрез задаване на област (domain) и обхват (range) на свойство данни, се дава възможност за изграждане на връзка между индивидуалности от класа на областта, със стойност принадлежаща на типа зададен като обхват на свойството.

Във всяка онтология могат да бъдат добавени мета знания. Те представляват знания за самите знания т. е. дават допълнителна информация за класовете, индивидуалностите, свойства данни, обектните свойства или за самите себе си. По този начин даден обект се свързва към определена стойност. Тази стойност може да бъде символен низ или сложен обект (сложно анатационно свойство). Говорим за сложно анатационно свойство, когато едно анотационно свойство притежава едно или множество поданотационни свойства.

Проблеми и решения с генерирането на въпроси на правилен български език. Съществуват много програми, с които могат да се създават онтологии. Една от тях е редакторът Protégé-OWL v.5.5.0 – свободно разпространявана програма, с която е направена онтологията BulgarianLiteratureAndHistoryOntology. Тя е предназначена за автоматично генериране на въпросителни и съобщителни изречения на български език, които да се използват за проверка на знанията на учениците по български език и българска история. Генерираните изречения са от вида:

- въпросително изречение с един верен и три грешни отговора (Фиг. 2);

questionLabel	individualLabel	correctAnswerA	wrongAnswerA	wrongAnswerB	wrongAnswerC
"През коя година е роден"@bg	"Димчо Дебелянов"@bg	"1887"	"1815"	"1866"	"1849"
"През коя година е роден"@bg	"Димчо Дебелянов"@bg	"1887"	"1815"	"1866"	"1903"
"През коя година е роден"@bg	"Димчо Дебелянов"@bg	"1887"	"1815"	"1866"	"1856"
"През коя година е роден"@bg	"Димчо Дебелянов"@bg	"1887"	"1815"	"1866"	"1898"
"През коя година е роден"@bg	"Димчо Дебелянов"@bg	"1887"	"1815"	"1866"	"1837"

Фигура 2. Генерирани въпросителни изречения с един верен и три грешни отговора

- въпросително изречение с два верни и два грешни отговора (Фиг.3);

questionStart	partialMemberLabel	propertyLabel	individualLabel	correctAnswerALabel	correctAnswerBLabel	wrongAnswerALabel	wrongAnswerBLabel
"В кои"@bg	"Градове"@bg	"е живял"@bg	"Васил Левски"@bg	"Ловеч"@bg	"Сопот"@bg	"Габрово"@bg	"Милано"@bg
"В кои"@bg	"Градове"@bg	"е живял"@bg	"Васил Левски"@bg	"Ловеч"@bg	"Сопот"@bg	"Габрово"@bg	"Капофер"@bg
"В кои"@bg	"Градове"@bg	"е живял"@bg	"Васил Левски"@bg	"Ловеч"@bg	"Сопот"@bg	"Габрово"@bg	"Одеса"@bg
"В кои"@bg	"Градове"@bg	"е живял"@bg	"Васил Левски"@bg	"Ловеч"@bg	"Сопот"@bg	"Габрово"@bg	"Париж"@bg
"В кои"@bg	"Градове"@bg	"е живял"@bg	"Васил Левски"@bg	"Ловеч"@bg	"Сопот"@bg	"Габрово"@bg	"Котел"@bg

Фигура 3. Генерирани въпросителни изречения с два верни и два грешни отговора

- съобщителни изречения с дописване на верния отговор (Фиг.4);

individualLabel	propertyLabel	hiddenAnswer
"Димитър Агура"@bg	"е роден през годината"@bg	"1849"
"Леда Милева"@bg	"е родена през годината"@bg	"1920"

Фигура 4. Генерирани съобщителни изречения с дописване на верния отговор

- верни и неверни съобщителни изречения (Фиг.5);

fullMemberLabel	individualLabel	statementSentenceLabel	partialMemberLabel	validationLabel	
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателката"@bg	"Леда Милева"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Пейо Яворов"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Йордан Йовков"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Димчо Дебелянов"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Христо Смирненски"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Елин Пелин"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Евтим Евтимов"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Иван Вазов"@bg	"Иван Вазов"@bg
"Одата"@bg	"Опълченците на Шипка"@bg	"е написана от"@bg	"Писателя"@bg	"Петко Рачов Славейков"@bg	"Иван Вазов"@bg

Фигура 5. Генерирани верни и неверни съобщителни изречения

Граматичните норми на българския език са заложени в анотационните свойства. Алгоритмични заявки извличат заложената информация от тях. Поради тази причина изключително важна част от създадената онтологията са анотационните свойства, с които се постига генерирането на граматически правилни въпро-

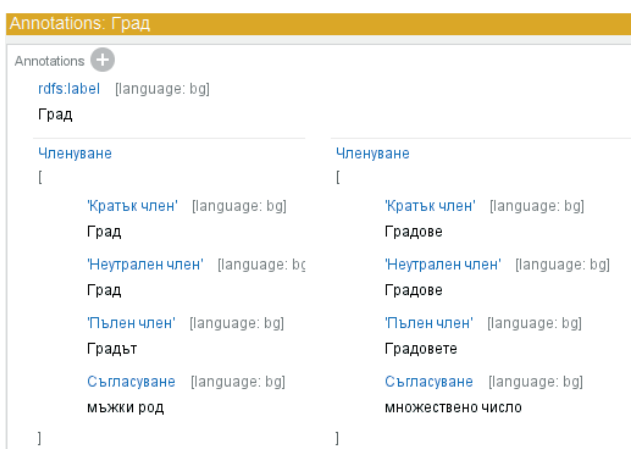
сителни и съобщителни изречения на български език. Ако няма допуснати грешки при въвеждането на информация в тях и въведената информация е подчинена на граматичните правила в българския език, то генерираните изречения са коректно съставени. При генерирането на част от въпросите се установи проблем при членуването на някои части на изречението, независимо че в анотационните свойства е заложено подлогът да се членува с пълен член, а второстепените части – с кратък член (Фиг.6). Граматичното правило има изключения, по-конкретно, че имената на населени места и титли не се членуват. Това наложи промяна в анотационните свойства за конкретните класове в онтологията, а имено добавяне на поданотационно свойство „Неутрален член“ (Фиг. 8), което да се използва, и проблемът беше разрешен (Фиг. 7).

individualLabel	propertyLabel	partialMember
"Леда Милева"@bg	"е родена в"@bg	"Града @bg
"Анастасия Димитрова"@bg	"е родена в"@bg	"Града @bg
"Димчо Дебелянов"@bg	"е роден в"@bg	"Града @bg

Фигура 6. Генерирани изречения преди корекция на анотационното свойство за класа „Град“

individualLabel	propertyLabel	partialMember
"Леда Милева"@bg	"е родена в"@bg	"Град "@bg
"Анастасия Димитрова"@bg	"е родена в"@bg	"Град "@bg
"Димчо Дебелянов"@bg	"е роден в"@bg	"Град "@bg

Фигура 7. Генерирани изречения след корекция на анотационното свойство за класа „Град“



Фигура 8. Анотационното свойство за класа „Град“

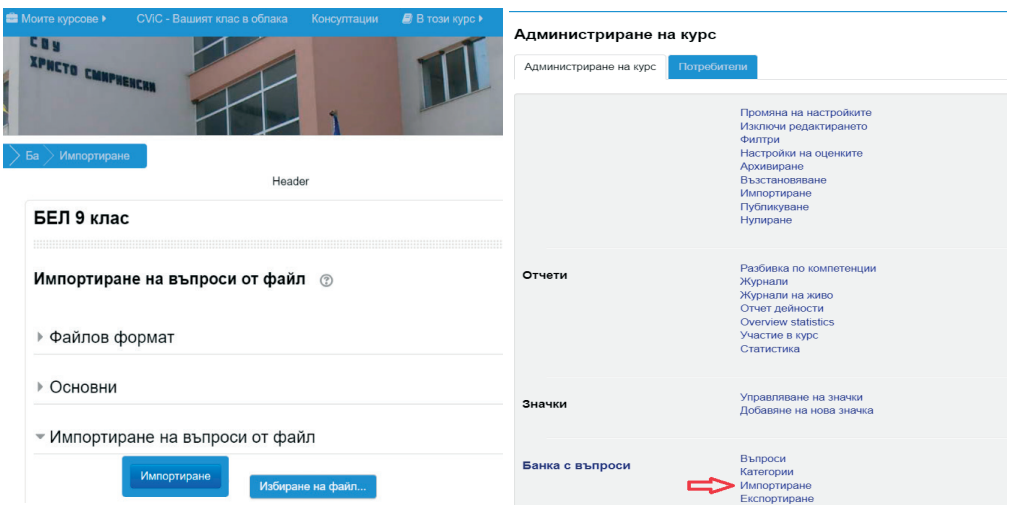
В резултат на описания подход от създадената онтология могат да се генерират различни видове тестови въпроси по различни хуманитарни учебни дисциплини – история, литература, философия и др. Генерираните тестови въпроси се записват във файлове в JSON формат и могат да бъдат използвани в различни тестови системи или учебни платформи.

Форматът на данните JSON (JavaScript Object Notation) е отворен стандартен файл (.json) и формат на данни, използвани за обмен на данни, представени чрез различни форми [9]. JSON се дефинира като файлов формат, използван в обектно-ориентираното програмиране, който използва четим от човека език, текст и синтаксис за съхраняване и комуникация на обекти с данни между различни приложения. Тези характеристики на формата позволяват лесно имплементиране на генерираните от онтологията тестови въпроси в различни тестови системи и създаването на подходящи тестове за различни групи ученици.

През последните години и особено в периода на световната Covid’19 пандемия в средното училище се наложиха редица образователни среди и платформи. Moodle (Modular Object-Oriented Dynamic Learning Environment) е една от най-популярните системи за дистанционно обучение, която широко се използва както в университетското, така и в училищното образование. Moodle, въпреки по-високото и ниво на сложност в сравнение с популярните у нас Google Classroom и MS Teams се налага с разнообразните възможности, които предлага [10] – различни видове и форми на уроци и учебни ресурси, интерактивни инструменти, учебни игри, както и различни форми на бързи, тренировъчни и изпитващи тестове. За настоящото изследване е съществен факта, че Moodle използва специфичен JSON формат за импортиране и експортиране на тестове [11].

Нашата цел е да използваме тази възможност за използване на Moodle платформата в процеса на приложение на тестовото изпитване с генерираните от онтологията тестови въпроси. Използваният подход включва следните стъпки:

- Стъпка 1. Създаване на онтология с подходящи анотационни и поданотационни свойства по описаната в предходната секция методика
- Стъпка 2. Генериране на различни видове въпроси от онтологията на коректен български език
- Стъпка 3. Съхраняване на генерираните въпроси от различни видове в JSON формат
- Стъпка 4. Имплементиране на тези въпроси към банката с въпроси в Moodle и използване на тестовата система на платформата за структуриране, провеждане и оценяване на тестовата проверка. Тъй като този формат не е стандартно достъпен за импортиране в банката с въпроси на Moodle, е необходимо или да конвентираме файла в подходящ формат, или да инсталираме JSON плъгин (Фиг. 9).



Фигура 9. Импортиране на генерирания файл в Банката с въпроси в Moodle

Заключение. Персонализацията и адаптивността в обучението са основни характеристики на образованието в дигиталната ера. За реализацията на обучение, което отговаря на тези изисквания в голяма степен може да се използват възможностите на семантичното моделиране и в частност на онтологиите. В статията представяме един подход за подходящо структуриране на онтологии чрез добавяне на анотационни свойства, чрез които става възможно генерирането на изречения и въпроси на коректен български език – проблем, който не е еднозначно решен от научната общност у нас. След генерирането на множество от различни видове въпроси върху знанията в онтологията, те могат да се съхраняват в различни файлови формати и да се имплементират в съществуващи обучителни платформи. Един удачен вариант за това е използването на обучителната среда Moodle.

References:

1. Glushkova, T., Stoyanov, S., Ivanova, V., Krasteva, I., AN IDEA FOR CONDUCTING THE EDUCATIONAL PROCESS IN A VIRTUAL-PHYSICAL SPACE, AIP Conference Proceedings 2333, 050006 (2021); https://doi.org/10.1063/5.0041727 AIP
2. Hammer, M., & McLeod, D. (1981, September). Database Description with SDM: A Semantic Database Model. 6, 351-386. ACM
3. T. Berners Lee, J. Handler, O. Lassila, The Semantic Web, Scientific American, vol. 284, pp. 34-43, May 2001.
4. Allemang, D., & Hendler, J. (2011). Semantic web for the working ontologist: effective modeling in RDFS and OWL. Elsevier.
5. Gruber T. R. A translation approach to portable ontologies / Knowledge Acquisition, 1993, V. 5(2), P. 199-220
6. OWL Web Ontology Language https://www.w3.org/TR/2004/REC-owl-features-20040210/
7. Stancheva, N. S., Popchev, I., Stoyanova-Doycheva, A., & Stoyanov, S. (2016, September). Automatic generation of test questions by software agents using ontologies. In 2016 IEEE 8th International Conference on Intelligent Systems (IS) (pp. 741-746). IEEE.
8. Dimov, D., (2000). Istoricheskiyat podvig na Plovdivskata gimnaziya Sv. Kiril i Metodiy. Izdatelska kashta “Zhanet-45“, ISBN 954-491-006-2.
9. Bourhis, P., Reutter, J. L., Suárez, F., & Vrgoč, D. (2017, May). JSON: data model, query languages and schema specification. In Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI symposium on principles of database systems (pp. 123-135).
10. Gamage, S. H., Ayres, J. R., & Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. International journal of STEM education, 9(1), 9.
11. Zampirolli, F. D. A., Batista, V. R., Pimentel, E., & Braga, J. (2021, October). Facilitating the Generation of Parametric Questions and their export to Moodle. In 2021 IEEE Frontiers in Education Conference (FIE) (pp. 1-8). IEEE.