

ПОДХОД ЗА ГЕНЕРИРАНЕ НА ТЕСТОВИ ВЪПРОСИ НА БЪЛГАРСКИ ЕЗИК ОТ ОНТОЛОГИИ

Ася Стоянова-Дойчева, Тодорка Глушкова,
Мария Грънчарова

Резюме. Статията представя подход за автоматично генериране на тестови въпроси на български език от онтологии. Описана е структурата на онтология, създадена с цел генериране на тестовите въпроси. Това включва елементи, свързани с българската граматика. Представени са примери за генерирани, граматически правилни, тестови въпроси от създадената онтология като за целта са разработени и използвани SPARQL алгоритми.

Ключови думи: автоматично генериране на тестови въпроси, онтологии, SPARQL.

Въведение

Семантичните модели са ефективен подход за структуриране и разширяване на знанията, който улеснява създаването на материали за преподаване, проверка на знания и повторна употреба. След изграждането на основната структура тя лесно може да се адаптира за различни учебни и изследователски процеси.

Статията разглежда автоматично генериране на въпроси на български език за оценка на знанията, което е особено важно при нарастващия обем на учебното съдържание. Автоматизацията би освободила преподавателите от създаването и проверката на тестове, позволявайки им да се фокусират върху по-добра подготовка и представяне на уроците.

Генерираните въпроси трябва да спазват синтактичните и граматичните норми на българския език, за да избегнат объркване или допълнителна работа за преподавателя, като целта е процесът да бъде оптимизиран, а не да създава нови предизвикателства.

Текущо състояние в областта

Автоматичното генериране на въпроси играе ключова роля в образователния процес. Затова много изследователи съсредоточават своите

усилия към автоматичното генериране на въпроси за образователни цели. Авторите на [1] правят преглед на съвременните подходи за разработване на образователни приложения за генериране на въпроси, позовавайки се на изследвания описани в [2]. В статията е представена система за автоматично генериране на въпроси (AQG), наречена G-Asks. Резултатите показват, че G-Asks може да генерира въпроси, които са изчистени от граматични и семантични грешки. Авторите на [3] представят няколко техники за автоматично генериране на въпроси и оценка на техните отговори, като се илюстрира цикълът на процеса на електронно оценяване в [4]. В [5] се прави проучване, като се използва базиран на онтологии подход, за автоматизиране на генерирането на въпрос. Информацията, свързана с онтологията, се разделя на информационни категории във формата на SPARQL заявки. След това заявките се преобразуват във въпроси. Експертите валидират генерираните въпроси и установяват, че формирането им е с голяма точност. Авторите на изследване [6] разглеждат и обобщават настоящите тенденции и напредъка в AQG, подчертават промените, които областта е претърпяла през последните години, и предлагат подходи за подобрене и бъдещи възможности за AQG, използвайки конструирани шаблони, с цел извличане на информация от онтологии.

Факултетът по математика и информатика на ПУ „Паисий Хилендарски“ работи по различни проекти за автоматично генериране на тестови въпроси [7, 8, 9]. В статия [10] се представя архитектурата на Virtual Physical Space (ViPS). Част от ViPS е средата за автоматично генериране на тестови въпроси. Архитектурата на средата за автоматично генериране на въпроси е подробно описана в [11]. Генерирането на тестове се поддържа от специализирани онтологии, които се обслужват от два интелигентни агента, известни като Questioner Operative (QO) и Assessment Operative (AO) [12]. QO е отговорен за генериране на въпросите, като за целта използва онтология, а AO – проверява отговорите. В [13] е представен формален модел Question Generation Model (QGM) за генериране на тестови въпроси, който е организиран в три нива: ниво 1 – Домейн, ниво 2 – Извличане, ниво 3 – Генератор. На базата на този модел е реализиран прототип на среда за провеждане на персонализирани тестове с автоматично генерирани и проверявани въпроси. Дефинирана е и концепция за изграждане на онтологии за целите на автоматично създаване на тестове. Тя е реализирана чрез изграждането на онтология в областта на езика UML (Unified Modelling Language) и е използвана за базата

знания на прототипната среда. Предложен е и вариант за обновяване на съществуващи онтологии, така че да могат да бъдат включени в базата знания на средата. В тази статия ще представим подход за генериране на тестови въпроси от онтология на български език. Основната цел е да представим структура на онтология за генериране на тестови въпроси и примери на генерирани тестови въпроси чрез SPARQL алгоритми.

Структура на онтология за генериране на тестови въпроси

Онтологията създадена за генериране на тестови въпроси на български език е в областта на българската литература и история. Създадената онтология съдържа знания за много бележити личности и техните дела. Онтологията може лесно да се променя и обновява, като се разширява и допълва с нови знания и да става все по-ясен и изчерпателен източник на информация. За създаване на онтологията е използван редакторът Protégé-OWL v.5.5.0.

Основните елементи на онтологията са представени в клас йерархия с подходящи концепции от домейна (писатели, композитори, исторически личности и политици), object properties, представящи семантичните връзки между концепциите, data properties, които свързват основните концепции и индивидуалности с конкретни стойности и индивидуалности, които представят конкретни инстанции на класовете. На Фигура 1 е представена част от структурата на онтологията.

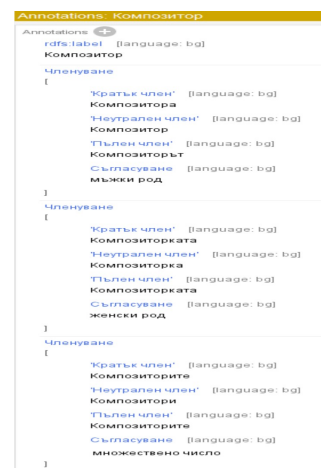
За да включим граматичните норми на българския език в онтологията, използвахме annotation properties, които добавят метазнания към класове, свойства и индивидуалности. Тези свойства съдържат информация като:

- Род на съществителните имена (мъжки, женски, среден),
- Число (единствено или множествено),
- Съгласуване на глаголите по време, лице и число.

Това осигурява граматически правилно генериране на тестови въпроси. Например, класът „Град“ е маркиран като мъжки род, а свойството „е разположен в“ съгласува израза граматически правилно. Тези анотации улесняват автоматичното създаване на синтактично и морфологично коректни изречения. На Фигура 2 са представени анотационните свойства за класа „Композитор“.

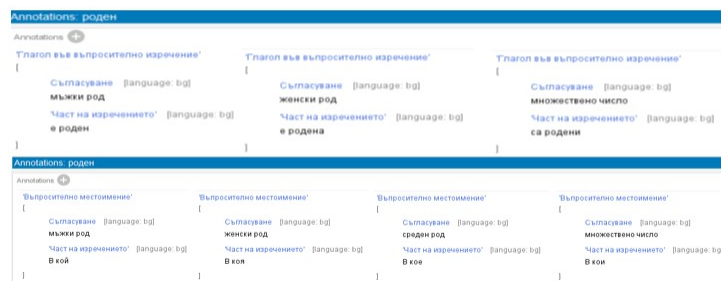


Фигура 1. Структура на онтологията



Фигура 2. Анотационни свойства за клас „Композитор“

На Фигура 3 са представени анотационните свойства за обект property, които представят глаголите. Има спецификация за спрежение на глагола за мъжки и женски род и за множествено число. С цел да е ясно как да се сформират изреченията са добавени анотационни свойства за въпросително местоимение, което също има род и число.



Фигура 3. Анотационни свойства за глагол (object property)

Анотационните свойства са ключов инструмент за добавяне на мета-знания към елементите на онтологията. Те позволяват описание на допълнителни характеристики като граматичен род, число и съгласуване, което осигурява: контекстуалност: по-добро адаптиране на информацията към конкретни задачи, като генериране на граматически правилни въпроси; гъвкавост: лесно разширяване с нови правила и характеристики; семантична коректност: устойчивост на синтактични и морфологични грешки.

Анотационните свойства увеличават ефективността и приложимостта на онтологията, особено в хуманитарни и езикови домейни.

Резултат от SPARQL алгоритми

Алгоритмите разработени за извличане на информация от онтологията под формата на въпросителни и съобщителни изречения са написани на езика за заявки към онтологии SPARQL (SPARQL Protocol and RDF Query Language). При изпълнението на даден алгоритъм, върнатата информация е структурирана в таблица съдържаща редове и колони. Всеки един ред от таблицата представлява уникално изречение и отговорите, които се предлагат за него. Колоните на таблицата са отделните компоненти, които като се свържат, създават лексикално-логически и смислени въпросителни или съобщителни изречения на български език и възможни отговори към тях.

Алгоритмите могат да бъдат изпълнени един единствен път и информацията върната от тях да бъде запазена в JSON (JavaScript Object Notation) формат или друг и в последствие да се използва от системи за тестване предназначени за обучение. Докато структурата или информацията в онтологията не бъде променяна, не е нужно алгоритмите да се изпълняват отново, защото информацията, която връщат винаги ще е една и съща. Приложените алгоритми са тестови и не предоставят всички възможности за генериране на въпросително изречение от текущата структура на онтологията. Те служат, за да се покажат възможностите на предложената архитектура. Чрез алгоритмите могат да бъдат генерирани два типа изречения – съобщителни (Фигура 4) и въпросителни (Фигура 5). Всеки един алгоритъм създава изречения във формата на един от двата споменати типа.

questionStart	answerType	questionLabel	fullMember	individual	answerCorrectA	answerWrongA	answerWrongB	answerWrongC
"В коя @bg	Стихосбирка	"е отпечатана"@bg	"Поемата"@bg	Кино	Моторни песни	Детство мое	Подир сенките на облаците	Сън за щастие
"В коя @bg	Стихосбирка	"е отпечатана"@bg	"Поемата"@bg	Кино	Моторни песни	Подир сенките на облаците	Подир сенките на облаците	Хубава си, татковино
"В коя @bg	Стихосбирка	"е отпечатана"@bg	"Поемата"@bg	Кино	Моторни песни	Весели балони	Подир сенките на облаците	Хубава си моя горо

Фигура 4. Генерирани въпросителни изречения с един верен и три грешни отговора

fullMember	individualLabel	propertyLabel	partialMember	answerLabel
"Мъжът"@bg	"Атанас Беляев"@bg	"е работил в"@bg	"Гимназията"@bg	"Пловдивска мъжка гимназия"@bg
"Мъжът"@bg	"Атанас Беляев"@bg	"е ръководил"@bg	"Гимназията"@bg	"Пловдивска мъжка гимназия"@bg

Фигура 5. Генериране на съобщителни изречения с дописване на верния отговор

При генерирането на изречения са разгледани три подхода, базирани на RDF триплети. Първият включва изречения, съставени от връзка между две индивидуалности чрез обектно свойство. Вторият разглежда

връзки между индивидуалност и нейна стойност чрез свойство за данни. Третият се фокусира върху връзки между индивидуалност и снимка чрез анотационно свойство. Тези подходи осигуряват създаването на раз-нообразни синтактично коректни изречения.

Заклучение

Разработената онтология и алгоритмите за генериране на тестови въпроси илюстрират практическото използване на семантичното моделиране в образованието. Чрез включване на правилата на българската граматика и използване на заявки на SPARQL, системата създава граматически точни въпроси, които оптимизират оценката на знанията на учениците. Този подход подчертава полезността на онтологиите в образователен план и изтъква тяхната адаптивност към други области и езици, демонстрирайки потенциала им за усъвършенстване на процесите на преподаване и учене. Като бъдещо развитие планираме автоматизиране на процеса на добавяне на мета знания за граматиката на българския език към онтологиите като използваме речници като WordNet [14] и BulNet [15] и интелигентни компоненти.

Благодарности

Това научно изследване е финансирано по проект FP23-FMI-002 „Интелигентни софтуерни инструменти и приложения в изследванията по математика, информатика и педагогика на обучението“ в Пловдивски университет „Паисий Хилендарски“.

Литература

- [1] N. Le, T. Kojiri, N. Pinkwart, Automatic question generation for educational applications—the state of art, *Advanced Computational Methods for Knowledge Engineering: Proc. of the 2nd International Conference on Computer Science, Applied Mathematics and Applications (ICCSAMA 2014)*, Springer International Publishing, 2014, 325–338, DOI: 10.1007/978-3-319-06569-4_24.
- [2] M. Liu, R. Calvo, V. Rus, G-Asks: An intelligent automatic question generation system for academic writing support, *Dialogue & Discourse*, 3 (2), 2012, 101–124, DOI: <https://doi.org/10.5087/dad.2012.205>.
- [3] B. Das, M. Majumder, S. Phadikar, A. Sekh, Automatic question generation and answer assessment: a survey, *Research and Practice in*

- Technology Enhanced Learning*, 16 (1), 2021, 1–15, <https://doi.org/10.1186/s41039-021-00151-1>.
- [4] N. Alruwais, G. Wills, M. Wald, Advantages and challenges of using e-assessment, *International Journal of Information and Education Technology*, 8 (1), 2018, 34–37, DOI: 10.18178/ijiet.2018.8.1.1008.
- [5] S. Kusuma, D. Siahaan, C. Fatchah, Automatic Question Generation in Education Domain Based on Ontology, *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, IEEE, November 2020, pp. 251–256, DOI: 10.1109/CENIM51130.2020.929799.
- [6] A. Malinova, V. Ivanova, A. Rahnev, Computer algebra aided generation of English language tests, *The 11th Annual International Conference on Computer Science and Education in Computer Science CSECS 2015*, June 04–07 2015, Boston, MA, USA, Computer Science Education & Computer Science Research Journal, Vol. 11, 2015, pp. 66–74, ISSN: 1313-8624.
- [7] K. Gramatova, S. Stoyanov, E. Doychev, V. Valkanov, Integration of eTesting in an IoT eLearning ecosystem – Virtual eLearning Space, *BCI'15*, September 02–4 2015, Craiova, Romania, 2015 ACM, Art. 14, ISBN: 978-1-4503-3335-1/15/09, DOI: <http://dx.doi.org/10.1145/2801081.2801086>.
- [8] O. Rahneva, N. Pavlov, *Distributed systems and applications in learning*, Plovdiv University Press, 2021, Plovdiv, ISBN: 978-619-7663-06-8.
- [9] G. Kurdi, J. Leo, B. Parsia, U. Sattler, S. Al-Emari, A systematic review of automatic question generation for educational purposes, *International Journal of Artificial Intelligence in Education*, 30, 2020, 121–204, DOI: <https://doi.org/10.1007/s40593-019-00186-y>.
- [10] S. Stoyanov, T. Glushkova, A. Stoyanova-Doycheva, E. Doychev, Virtual physical space—an architecture supporting internet of things applications, *2018 20th International Symposium on Electrical Apparatus and Technologies (SIELA)*, 2018, June, pp. 1–3, DOI: 10.1109/SIELA.2018.8447156.
- [11] N. Stancheva, A. Stoyanova-Doycheva, S. Stoyanov, I. Popchev, V. Ivanova, An environment for automatic test generation, *Cybernetics and information Technologies*, 17 (2), 2017, 183–196, DOI: 10.1515/cait-2017-0025.
- [12] N. Stancheva, I. Popchev, A. Stoyanova-Doycheva, S. Stoyanov,

- Automatic generation of test questions by software agents using ontologies, *2016 IEEE 8th International Conference on Intelligent Systems (IS)*, 2016, September, pp. 741–746, DOI: 10.1109/IS.2016.7737395.
- [13] N. Stancheva, A. Stoyanova-Doycheva, S. Stoyanov, I. Popchev, V. Ivanova, A model for generation of test questions, *Compt. Rend. Acad. bulg. Sci.*, 70 (5), 2017, 619–630, Print ISSN: 1311-9702, Online ISSN: 1314-408, DOI: 10.2478/cait-2020-0032.
- [14] WordNet – https://dcl.bas.bg/wordnet_bg.html.
- [15] BulNet – https://dcl.bas.bg/completed_projects_bg.html.

Ася Стоянова-Дойчева¹, Тодорка Глушкова², Мария Грънчарова³,
^{1,2,3} Пловдивски университет „Паисий Хилендарски“,
Факултет по математика и информатика,
бул. „България“ № 236, 4027 Пловдив, България
Автор за кореспонденция: astoyanova@uni-plovdiv.bg

APPROACH FOR GENERATING TEST QUESTIONS IN BULGARIAN FROM ONTOLOGIES

Asya Stoyanova-Doycheva, Todorka Glushkova, Maria
Grancharova

Abstract. *The article introduces an approach for automatically generating test questions using ontologies in the Bulgarian language. It details the structure of the ontology designed specifically for this purpose, incorporating elements of Bulgarian grammar. Examples of grammatically correct test questions generated from the ontology are provided, demonstrating the use of SPARQL algorithms developed to achieve this.*

Key words: Automatic Test Generation, Ontology, SPARQL Algorithms.