



**Краеведските библиографско-информационни системи
като семантичен ресурс**

Калоян Здравков

**Local History Bibliographic Information Systems
as a Semantic Resource**

Kaloyan Zdravkov

***Abstract:** This article discusses the Local history bibliographic information systems in Bulgaria in the perspective of their transformation through semantic technologies into a new generation of information resources to facilitate and enrich the research activity.*

***Keywords:** local history, bibliographic information systems, classification systems, ontology.*

ВЪВЕДЕНИЕ

Бурното развитие на компютърните технологии създаде предпоставка за постоянно нарастване на количеството информация, ежедневно публикувана във всевъзможни медии. Улесненият достъп до глобални ресурси, породен от все по-широкото навлизане на Интернет сред всички слоеве на населението и възможността практически всеки да преработва и препубликува получената информация, поставя сериозни изпитания пред съвременните системи за управление на бази данни с оглед нейното адекватно съхраняване и управление. Не е за подценяване и нарастващата тенденция за създаване на т.нар. „фалшиви новини“ – преиначаване на факти с цел извличане на лична обществена, политическа или икономическа изгода.

Някак встрани от всеобщото развитие останаха библиотеките, или по-точно техните специализирани отдели, обработващи целия този информационен поток. Резултат от тяхната работа са библиографските бази от данни с краеведска информация, чието основно преимущество е изключително прецизният подбор на описваните материали, като се изключва явлението „компютърно генерирана информация“. Предимство, което се оказва и недостатък от гледна точка на ограничените възможности за извличане на данни в нестандартни сечения и детайлност. Ограничения, наложени от типа на използваните информационни системи.

КРАЕЗНАНИЕТО КАТО УНИВЕРСАЛНА НАУЧНА ДИСЦИПЛИНА

Най-елементарният, но и най-точен отговор на въпроса „Що е краезнание?“, е „знания за края“. По-формално краезнанието е „комплекс от научни дисциплини, различни по съдържание и по частни методи на изследване, но водещи в своята съвкупност към научното и всестранно изучаване на края“. [3, 13] Това е термин, който разглежда работата на широк кръг организации, научни институти и частни лица (основно местни научни и обществени дейци) по издирване, проучване, обработване и популяризиране на документална информация със значима научна, историческа, културна и т.н. стойност за съответния край като териториално-административна единица или район по етнически, етнографски, езиков или друг признак.

Краеведските¹ източници са универсални както по съдържание (история, география, изкуство, селско стопанство, медицина и др.), така и по вид. Могат да бъдат печатни (книги, статии в периодични

¹ Поради липса на производни на „краезнание“ се използват такива от остарялата форма „краеведение“ (сега съществуваща само в руския език) – краеведи, краеведски и т.н.

издания), архивни (писма, документи, фотографии), предметни (археологически обекти, паметници, предмети на бита и др.), а в последно време и електронни (изцяло цифрово създадени материали, онлайн ресурси).

Продукт на краеведската дейност са уникални по своя съдържателен обхват информационни масиви: библиографски (реферативни) бази от данни, в които се отразяват всички обработени документални източници, подходящо анотирани; фактографски бази, които предоставят значима обществена и социално-битова информация; картотеки „Бележити дати и събития“ и „Персоналия“, които са незаменим ресурс на генеалогичните изследвания [4, 62–66], [9, 67–70].

Начинът на съставяне на аналитично библиографско описание е регламентиран с Български държавен стандарт 17264-91 [3, 123].

Елементите, от които е изградено, представляват преди всичко метаданни за информационния източник и изданието, в което е публикуван. Най-голяма семантична значимост притежава анотацията. Въпреки формулираните методически указания относно нейното съдържание съставянето ѝ е на база субективната преценка на краеведите.

В България краеведските библиографско-информационни системи (БИС) се изграждат изолирано, почти изцяло от регионалните библиотеки. Всяка от тях дефинира собствен обхват на разглежданите източници и пълнотата на тяхното разкриване. Все още част от тези БИС са недостъпни онлайн. В резултат част от работата се дублира, предполага се липсата на важни връзки между обекти и субекти в разглежданите предметни области, което не дава възможност за създаване на пълноценна ИС с национален обхват, подходяща за научноизследователска дейност.

КРАЕЗНАНИЕТО КАТО СЕМАНТИЧЕН РЕСУРС

Краеведските БИС представляват ценен информационен ресурс с перспектива за широк кръг мултидисциплинарни научни изследвания.

С автоматизирането на дейността в библиотеките са правени различни опити да се създаде крос-платформен каталог, обединяващ ресурси от различни организации („Регина“², „Регина-ПИИ“³, „Споделена библиотечна база за аналитично описание на статии от вестници, списания, рецензии и части от книги“⁴). Проектите се оказаха неуспешни и бяха оставени без развитие поради неспазването на единна методология за формиране на библиографско описание, нуждата от голям обем ръчна обработка на данни при всяко синхронизиране на създадения масив, затворения принцип за формиране на общата база данни и ключовата зависимост от решенията на компанията – инициатор на изграждането на каталога.

Към настоящия момент всички краеведски центрове в регионални библиотеки използват релационни бази данни⁵.

През последните години се наблюдава сериозен подем в разработката на бази данни, които силно се различават от установения като традиционен релационен модел на структуриране и достъп до данните.

Релационният модел страда от някои ограничения, които се превръщат във все по-значим фактор с нарастване обемите на съхраняваните и обработваните данни, като скалируемост – възможност за справяне с нарастващ обем данни и с нараснало натоварване, както и възможността за адаптиране и измяна при работещо приложение.

Нерелационните бази данни са изначално проектирани да се справят с тези проблеми. Приложенията могат да използват набор от сравнително евтини машини, за да поддържат устойчива работа на огромни по обем хранилища от данни. Липсата на фиксирана схема за съхранение на данните прави добавянето на концептуално различна информация в базата тривиално и се отнася само до новите данни.

В момента съществуват достатъчно сериозни алтернативи на RDBMS, някои от които и с отворен код – GraphDB, Stardog, Oracle RDF Graph [1, 134–143].

² <http://ruc.ilib.primasoft.bg/> (31.03.2019)

³ <http://reginapi.ilib.libsofia.bg/> (31.03.2019)

⁴ Понастоящем прекратен, няма онлайн достъп.

⁵ <https://www.pc-tm.eu/ab-2/потребители/#2>; <http://softlib.bg/klienti>. И двете фирми, производители на софтуер за библиотечна автоматизация, предоставят релационни бази данни на своите клиенти (D-Base, Access).

Основното предимство на семантично ориентираните приложения е ефективното управление на информацията. Увеличаващото се разнообразие на човешкото познание генерира все по-бързо и все по-разнородни данни и поставя с цялата му тежест проблема за тяхното адекватно съхранение, намиране, анализ и последваща употреба.

Трансформирането на съществуващите релационни краеведски БИС е възможно посредством използването на подходяща технология за разбиването им на атомарни частици от данни и обвързването им със система от правила.

В научната литература с термина „онтология“ се означават формално представени знания на основата на определена концептуализация. Концептуализацията предполага описание на множество от обекти и понятия, знания за тях и връзки между тях.

Според най-често използваната дефиниция онтология се нарича експлицитната спецификация на концептуализацията [11]. Формално онтологията е съвкупност от термини, организирани в таксономия, техните определения и атрибути, както и свързаните с тях правила за извод.

Онтологиите играят ключова роля в постигането на целите на семантичния Web. Те дефинират семантиката на засегнатата предметна област и са основен способ за натрупване, извличане и повторно използване на знания в информационното общество.

За създаването на една онтологията най-важното е правилното дефиниране на обхвата на предметната област, типовете запитвания, на които трябва да може да отговаря онтологията, както и потребителите, които ще я използват. Необходимо е да се прецизират детайлно всички понятия и свързаните с тях класове и подкласове (концепти), както и свойствата, които представляват връзките между данните в създаваната информационна система.

В Глобалната мрежа съществуват множество свободно достъпни разработки на спецификации в различни тематични области, ранни версии на които датират още от 2000 г. За съжаление, по различни причини след време част от проектите прекратяват своето развитие, напр. онтологичната библиотека на DAML (<http://www.daml.org/ontologies/category.html>) – голяма част от връзките вече са невалидни, *Ontolingua* (<http://www.ksl.stanford.edu/software/ontolingua/>) и др.

В резултат на продължително проучване за целите на настоящото изследване са избрани три онтологии, които заедно в максимална степен се доближават до необходимото терминологично покритие на дисциплината „Краезнание“. И трите са в достатъчно зряла фаза на разработване, което дава основание да се предположи положителната перспектива за тяхното развитие.

FOAF (Friend Of A Friend) е проект, посветен на свързването на хора и информация чрез уеб. Макар и привидно да изглежда като аналог на връзките в социалните мрежи, спецификацията се простира отвъд тези граници. Разширява се бавно и постепенно се формира като устойчива във времето структура [7].

VIBO (Bibliographic Ontology). Спецификацията на библиографската онтология предоставя основни понятия и свойства за описване на цитати и библиографски справки (т.е. цитати, книги, статии и т.н.) в семантичния уеб в RDF. Тази онтология може да се използва като цитираща, като онтология на класификацията на документи, или просто като начин да се опише всеки вид документ в RDF. Той е вдъхновен от много съществуващи формати на метаданни за описание на документи и може да се използва като обща основа за конвертиране на други източници на библиографски данни [5].

Географската база данни GeoNames съдържа над 25 милиона географски имена и се състои от над 11 милиона уникални характеристики, от които 4,8 милиона са за населени места и 13 милиона са алтернативни имена. GeoNames интегрира географски данни, като имена на места на различни езици, надморска височина, население и други от различни източници [8].

Мощността на трите онтологии, взети заедно, е достатъчна за представяне на основните връзки между личности, географски понятия и документи. За обогатяване обхвата на една базирана преди всичко на личности информационна система бе взето решение да се добави функционалност за изграждане на генеалогични връзки. Тя би била от особена полза за откриване на далечни родословни зависимости – ценен факт при историческите изследвания. Подходяща спецификация – FamilyTree [6] – предоставя проектът CO-ODE на университета в Манчестър, Великобритания. Разработено е подробно

представяне на всички кръвни и сродни връзки до трета степен по възходяща и сребрена линия. Макар и приключил през 2009 г., проектът предоставя свободно онтологията в OWL формат⁶.

За пълноценното функциониране на една краеведска информационна система е необходимо подходящо представяне на публичната обществена роля на личностите, засегнати в нея. Такава възможност не бе реализирана в никоя от проучените спецификации. След анализ и сравняване описанията на известни българи, взели дейно участие в различни исторически, културни и образователни събития, открити в библиографските бази данни на по-големи библиотеки в страната, Wikipedia и др. източници, бе взето решение да се изготви нова схема от концепти и свойства, която да разкрие и този информационен аспект.

Създаден е набор от 39 класа, като понятията от най-високо ниво (Общественик, Учен, Политик, Приложна професия, Занаятчия) представят обща концептуализация на обществените роли, актуални през разгледания времеви период – предимно периода на Възраждането. Те, както и следващите ги конкретизации от следващите нива, са подбрани консервативно, с оглед гъвкавост и устойчивост при бъдещо разширение на терминологичния речник. По-долу е изложен йерархичен списък на класовете, представен и графично на фиг. 1.

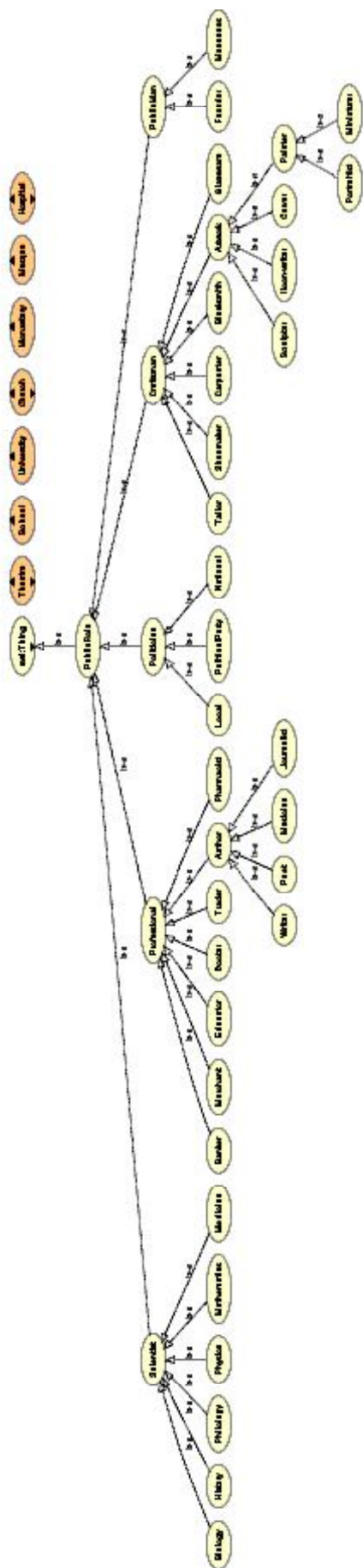
PublicMan	Educator
Maecenas	Doctor
Founder	Pharmacist
Scientist	Banker
History	Trader
Medicine	Merchant
Mathematics	Craftsman
Biology	Carpenter
Physics	Blacksmith
Philology	Shoemaker
Politician	Glassware
National	Tailor
Local	Artwork
PoliticalParty	Painter
Professional	Miniaturer
Author	Portraitist
Writer	Icon-writer
Poet	Sculptor
Journalist	Carver
Musician	

Дефинирани са свойства, позволяващи оптимално да се обвържат обектите със субектите в различните им роли:

hasAuthor (*инверсия: isAuthorOf*)
 hasEducator (*инверсия: isEducatorOf*)
 hasMaker (*инверсия: isMakerOf*)
 hasMember (*инверсия: isMemberOf*)
 hasFoundedBy (*инверсия: isFounderOf*)
 hasSponsoredBy (*инверсия: isSponsorerOf*)

За последните две свойства като домейн са вписани понятията "Church", "Monastery", "Mosque", "Hospital", "School", "University", "Theatre" и "PoliticalParty" (от терминологичния набор на GeoNames), а като диапазон съответно: "Founder" и "Maecenas". Ограничения от този вид спомагат за предотвратяване на фактологически грешки при програмна реализация на онтологията.

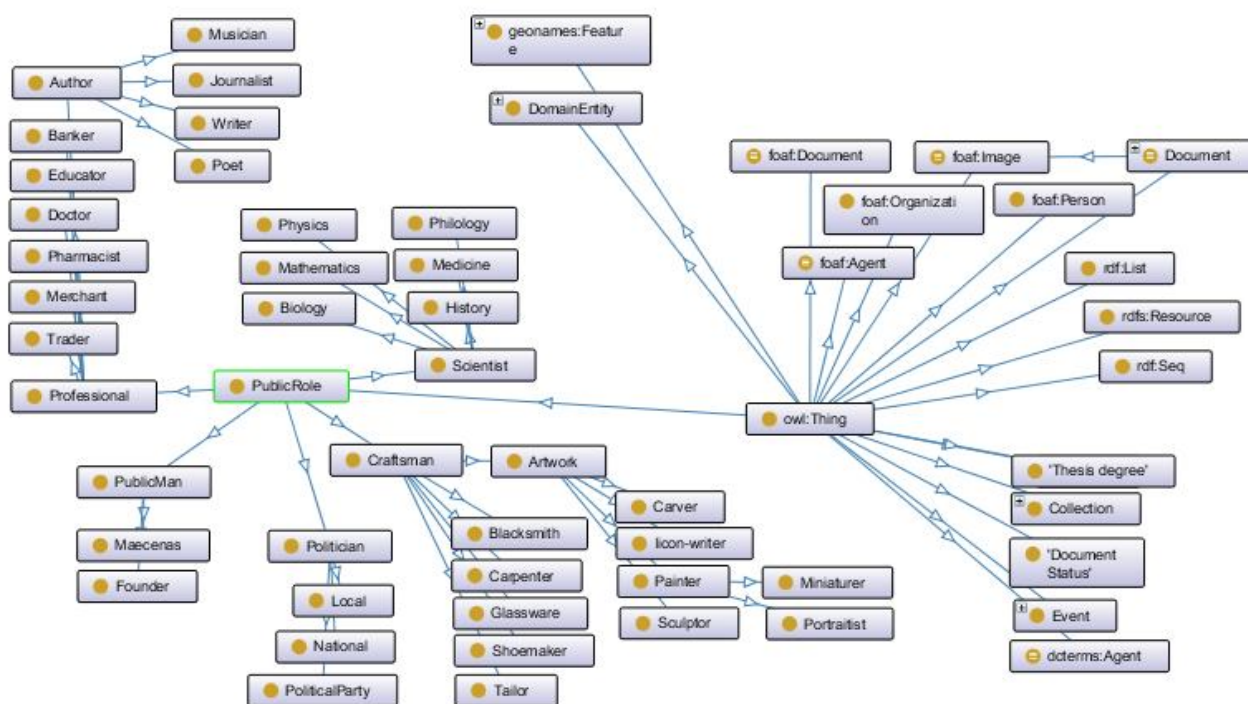
⁶ <http://www.co-ode.org/roberts/family-tree.owl> (31.03.2019)



Фиг. 1. Визуализация от тип граф на новосъздадения набор от класове, реализирана в Prolog чрез плагина OWL Viz

Представената структура влиза в състава на една нова онтология (LocHis, Local History = Краезнание), в която са добавени споменатите вече спецификации Geonames, BIBO и FamilyTree. FOAF не е директно импортирана, тъй като BIBO включва достатъчен за целите на предметната област диапазон класове от нея, а редакторът на онтологии Protégé, с който е реализиран практическият пример, я добавя по подразбиране в дефиниционната област:

```
<rdf:RDF xmlns="http://www.semanticweb.org/demo/ontologies/2019/3/LocHis#"
  xml:base="http://www.semanticweb.org/demo/ontologies/2019/3/LocHis"
  xmlns:ns="http://www.w3.org/2003/06/sw-vocab-status/ns#"
  xmlns:ns1="http://creativecommons.org/ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:terms="http://purl.org/dc/terms/">
  <owl:Ontology rdf:about="http://www.semanticweb.org/demo/ontologies/2019/3/LocHis">
    <owl:versionIRI rdf:resource="http://www.semanticweb.org/demo/ontologies/2019/LocHis"/>
    <owl:imports rdf:resource="http://purl.org/ontology/bibo/">
    <owl:imports rdf:resource="http://www.co-ode.org/roberts/family-tree.owl"/>
    <owl:imports rdf:resource="http://www.geonames.org/ontology/mappings"/>
  </owl:Ontology>
```



Фиг. 2. Визуализация от тип граф на обединената онтология LocHis, реализирана в Protégé чрез плъгина OntoGraf

За формиране на пълноценна онтологична рамка за целите на „Краезнание“ е необходимо съществуващата схема да бъде надградена с връзки към предметните области на научните дисциплини, към които имат отношение обработваните документи.

За представяне на многообразието и обхвата на различните видове човешко знание се използват множество класификационни схеми. За страните от Европа (вкл. България) това е схемата за Десетична класификация на Дюи⁷, като се използва развитието ѝ до Универсална десетична класификация⁸ [12]. Всички области на знанието са сведени на най-горно ниво до 10 основни класа (от 0 до 9), като всяка детайлизация добавя по една цифра отдясно на съществуващия индекс. Получената дървовидна структура е изключително прегледна и устойчива и позволява без особени трудности да бъде интерпретирана в алтернативна класификационна схема – изграждане на онтология за система за библиотечен мениджмънт [10].

С отчитане на допълнителните специфики на краеведската област у нас за краезнание е разработена различна класификационна схема, която представлява вариация на стандартната УДК [2, 165–170]. За база отново се използва същия модел на десетично индексирание, като основните разлики са в добавените класификатори и определители.

ЗАКЛЮЧЕНИЕ И БЪДЕЩА РАБОТА

Направеният обзор на краеведските информационни системи в България и тяхната пригодност за моделиране посредством семантични технологии очертава перспектива за бъдещи научно-практически разработки, свързани с детайлен анализ на съществуващата библиографска онтология и разработването на нейни разширения, очертаване на пътна карта и разработване на специализирана онтология на краеведската десетична класификационна система, както и реализиране на информационни търсения в разрези и дълбочина, каквито досега не са били възможни.

ЛИТЕРАТУРА

[1] **Здравков**, Калоян. 2017. Сравнителен анализ на семантичните бази от данни GraphDB, Stardog, Oracle RDF Graph. // *Сборник доклади от Научна конференция на факултет Математика и информатика на младите изследователи*, Велико Търново, 134–143. // **Zdravkov**, Kaloyan. 2017. Sravnitelnen analiz na semantichnite bazi ot dannii GraphDB, Stardog, Oracle RDF Graph. // *Sbornik dokladi ot Nauchna konferentsia na fakultet Matematika i informatika na mladite izsledovateli*, Veliko Tarnovo, 134–143.

[2] **Ковачев**, Марин. 1994. *Класификационни схеми и класифициране на краеведски информационни източници в България*. Велико Търново: Сириус 4. // **Kovachev**, Marin. 1994. *Klasifikatsionni shemi i klasifitsirane na kraevedski informatsionni iztochnitsi v Bulgariya*. Veliko Tarnovo: Sirius 4.

[3] **Ковачев**, Марин. 2004. *Краезнание*. Велико Търново: Астартал // **Kovachev**, Marin. 2004. *Kraeznanie*. Veliko Tarnovo: Astarta.

[4] **Панчева**, Цветанка. 1998. Новите технологии и развитието на краеведската дейност. // *Национална конференция Общодостъпните библиотеки и предизвикателството на промените*, Благоевград, 62–66. // **Pancheva**, Tsvetanka. 1998. *Novite tehnologii i razvitiето na kraevedskata deynost*. // *Natsionalna konferentsia Obshtodostapnrite biblioteki i predizvikelstvoto na promenite*, Blagoevgrad, 62–66.

[5] Спецификация на онтология BIBO. <<http://purl.org/ontology/bibo/>> (31.03.2019). // Spetsifikatsia na ontologiya BIBO. Retrieved from <<http://purl.org/ontology/bibo/>> (31.03.2019)

[6] Спецификация на онтология FamilyTree. Retrieved from <<http://www.co-ode.org/roberts/family-tree.owl>> (31.03.2019) // Spetsifikatsia na ontologia FamilyTree. Retrieved from <<http://www.co-ode.org/roberts/family-tree.owl>> (31.03.2019)

[7] Спецификация на онтология FOAF. Retrieved from <<http://xmlns.com/foaf/spec/>> (31.03.2019) // Spetsifikatsia na ontologia FOAF. Retrieved from <<http://xmlns.com/foaf/spec/>> (31.03.2019)

[8] Спецификация на онтология GeoNames. Retrieved from <http://www.geonames.org/ontology/ontology_v3.1.rdf> (31.03.2019) // Spetsifikatsia na ontologia GeoNames. Retrieved from <http://www.geonames.org/ontology/ontology_v3.1.rdf> (31.03.2019)

[9] **Томова**, Вили. 1998. Краеведската фактографска и фирмена информация в регионалната библиотека – проблеми и перспективи (опита на Регионална библиотека, Плевен). // *Национална конференция Общодостъпните библиотеки и предизвикателството на промените*, Благоевград, 67–70. // **Tomova**, Vili. 1998. *Kraevedskata faktografaska i firmena informatsia v regionalnata biblioteka – problemi i perspektivi (opita na Regionalna biblioteka*,

⁷ Разработена от американеца Мелвил Дюи през 1876 година. (Източник: Wikipedia)

⁸ Пълната версия на УДК включва над 220 000 раздела. (Източник: Wikipedia)

Pleven). // *Natsionalna konferentsia Obshtodostapnite biblioteki i predizvikatelstvoto na promenite*, Blagoevgrad, 67–70.

[10] **Dar AR, Razzaq S.** 2018. Building Ontology for Library Management System Using Dewey Decimal Classification Scheme. *J Comput Eng Inf Technol* 7:1. doi: 10.4172/2324-9307.1000194 <https://www.scitechnol.com/peer-review/building-ontology-for-library-management-system-using-dewey-decimal-classification-scheme-D0kv.php?article_id=7584> (31.03.2019)

[11] **Gruber, T. R.** 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *International Journal Human-Computer Studies* 43 (Aug. 1993), 907–928. Retrieved from <<http://tomgruber.org/writing/onto-design.pdf>> (31.03.2019)

[12] **Slavic, Aida.** 2008. Use of the Universal Decimal Classification. *Journal of Documentation*. 64. 211–228. 10.1108/00220410810858029. Retrieved from <https://www.researchgate.net/publication/241707135_Use_of_the_Universal_Decimal_Classification> (31.03.2019)

ИНФОРМАЦИЯ ЗА АВТОР(И)

Калоян Здравков, докторант, Факултет „Математика и информатика“, Великотърновски университет, „Св. св. Кирил и Методий“, E-mail: kzdravkov@gmail.com.

ABOUT THE AUTHOR(S)

Kaloyan Zdravkov, doctorate, Faculty of Mathematics and Informatics, Department of Information Technologies, “St. Cyril and St. Methodius” University of VelikoTarnovo, E-mail: kzdravkov@gmail.com.