

## РЕЦЕНЗИЯ

на дисертационен труд за придобиване на образователна и научна степен „доктор“ в област на висше образование 4. „Природни науки, математика и информатика“,  
Професионално направление: 4.6 „Информатика и компютърни науки“,  
Научна специалност: 01.01.12 „Информатика“

Автор на дисертационния труд: *Виктор Сендеров*

Тема: *„OpenBiodiv: отворена система за управление на знанието за биологичното разнообразие“*

Научен ръководител: *проф. д-р Любомир Пенев*

Научен консултант: *доц. д-р Кирил Симов*

Рецензент: *доц. д-р Светла Бойчева - ИИКТ-БАН*

Тази рецензия е написана и представена на основание на заповед 86/30.04.2019 г. на директора на ИИКТ-БАН във връзка с решение на Научния съвет на ИИКТ-БАН (протокол №5/24.04.2019 г.), както и на решението на научното жури по процедурата (Протокол №1/10.05.2019 г.). Тя е изготвена въз основа на ЗРАСРБ, Правилника за прилагане на ЗРАСРБ, Постановление № 26 от 13.02.2019 г. за изменение и допълнение на ППЗРАСРБ, касаещо Област 4. Природни науки, математика и информатика, проф. направления 4.1., 4.2., 4.3., 4.4., 4.5., 4.6., Правилник за условията и реда за придобиване на научни степени и за заемане на академични длъжности в БАН, Правилник за специфичните условия за придобиване на научни степени и за заемане на академични длъжности в ИИКТ-БАН.

### **1. Съдържателен анализ на научните и научно-приложните постижения в дисертационния труд. Характеризиране на основните постижения.**

Дисертацията е на английски език и е с обем 113 страници, в т.ч. 15 страници използвана литература, 2 страници авторска справка, 4 страници листинги, 1 приложение с обем 2 страници, както и 2 празни страници. Поради спецификата на избрания формат, във връзка с това, че дисертацията ще се отпечатва от издателство Pensoft, трябва да се отбележи, че има отклонения от стандарта за машинописна страница (38 реда, 66 удара) – 51 реда, 85 удара, което приравнено към стандарта се равнява на около 150 стандартни страници, без да се броят литературата, авторската справка и приложенията. Текстът е организиран в увод и 8 глави и заключение. Използваната литература се състои от 164 източника, в т.ч. и 9 публикации, на които е автор докторанта. Предимно цитираната литература е на английски език, като има и няколко източника на руски, френски, латински. В основната си част използваните източници са публикувани през последните 10 години. Използваната литература е цитирана по подходящ начин в текста на дисертацията.

### **Актуалност на проблема**

Темата на дисертацията е свързана с една много актуална и бързоразвиваща се тема – създаване и управление на свързани отворени данни (linked open data – LOD – <https://lod-cloud.net/>).

Представеният научен труд представя проектирането и създаването на система OpenBiodiv за работа с LOD за биологичното разнообразие. Тази тема е много важна, тъй като има сериозна необходимост от систематизиране на знанието в областта, както като релации/връзки между отделните понятия, организирани в таксономии и онтологии, така и за уеднаквяване и съгласуване на наименования на термините и създаване на едно цифрово хранилище от отворени свързани данни.

## **Познаване на състоянието на проблема**

От направените анализи на базите от знания и свързани отворени данни, както и обзора на знания и данни в биоразнообразието е видно познаване на проблема. На базата на това изследване е дефиниран и инициран проекта OpenBiodiv.. Използваните литературни източници също показват, че докторанта познава естеството на проблема.

## **Подход и решение на проблема**

Целта на дисертацията е създаването на отворена система за управление на знания и данни за биоразнообразието (OBKMS - Open Biodiversity Knowledge Management System), като се концентрира върху знание, извлечено от научната литература.

За изпълнение на целите на дисертацията са формулирани 6 задачи: (1) създаване на онтология; (2) дизайн на архитектура на системата, (3) създаване на свързани отворени данни на базата на публикуваните таксономични статии и създадената онтология, (4) разработване на методи за преобразуване на таксономични публикации в семантичния модел на онтологията от 1, (5) разработване на автоматизирано актуализиране на свързаните отворени данни, чрез преобразуване на таксономични данни в таксономични публикации, (6) Създаване на уеб портал с примерни приложения.

Като основен модел е използвано представяне на данните като семантичен граф и е използван GraphDB (<https://www.ontotext.com/products/graphdb/>), на базата на който е създадена OBKMS.

## **Анализ на научните и научно-приложните постижения в дисертационния труд**

В **уводът** е направен обзор на предметната област, идентифицирани са основни проблеми, които преставяват интерес за изследване от научна гледна точка, дефинирана е целта на дисертацията и са формулирани 6 основни задачи за постигането ѝ. Дисертационният труд описва проектът OpenBiodiv, чиято цел е да допринесе за OBKMS (Open Biodiversity Knowledge Management System) чрез създаването на отворена система за информация за биоразнообразието, основана на знание, извлечено от научната литература. Направена е обосновка за избраната методология за научното изследване. Разгледани са различни аспекти:

- за представяне на знанието са избрани семантични графи и представяне чрез RDF схема;
- обоснован е изборът на информационни източници -издателство Pensoft, което има бази от публикации с добавена семантична информация (enhanced publications); GBIF Backbone Taxonomy; и Plazi Treatment Bank.
- избран е гъвкав метод за софтуерна разработка (Agile).

Представена е и структурата на дисертацията.

**Първа глава** описва архитектурата на OpenBiodiv – (1) база от семантични графи – онтологии, отворени свързани данни; (2) back-end – методи за преобразуване на неструктурирана информация във вид, който позволява представяне на семантичните им връзки, (3) front-end част – уеб портал. При проектирането на системата е използван гъвкав метод за софтуерна разработка (Agile). Реализацията е базирана на GraphDB.

**Втора глава** представя онтологията OpenBiodiv-O. Подробно се изследва и описва процеса на създаване на таксономии, с цел този процес да може да се автоматизира при създаването на онтологията OpenBiodiv-O. Направен е обзор на предходни разработки в тази област. Представен е семантичен модел на Biodiversity Publishing Domain. Дефинирани са класове за тази област и са представени релациите между обектите от тези класове. Представено е и семантичното моделиране на биологичната номенклатура. Описана е комбинирана хибридна йерархия на класовете, която включва както традиционните таксономични наименования, така и използването на таксономични концептуални етикети и операционни таксономични единици. Представено е и семантичното моделиране на таксономичните понятия. Идентифицирани са два модела за съответствие между кодирането на латинизираните научни имена: модел „заместващо име“ и модел „свързани имена“.

**Трета глава** описва OpenBiodiv-LOD (отворени свързани данни), генерирани при обработка на три основни източника на информация: GBIF Backbone Taxonomy; публикациите от Pensoft; и Plazi Treatment Bank. OpenBiodiv-LOD е синтетично множество от данни, като то не съдържа нови данни, а само интегрира данни, които се съдържат в източниците на информация. Представен е и псевдокод на алгоритъм Екстрактор за трансформация от XML в RDF.

**Четвърта глава** описва библиотеката RDF4R, която е библиотека на езика R за работа с RDF. Описани са основните функции в пакета. Даден е пример за преобразуване на SPARQL заявка във функция на R.

**Пета глава** разглежда автоматизираният работен процес за обработка на потока данни, свързани с биоразнообразието. Представени са два метода: (1) за автоматично гвнасяне на записи за наблюдения на видове в ръкописи и (2) автоматично генериране на ръкописни материали от екологичните метаданни (Ecological Metadata Language - EML).

**Шеста глава** описва уеб портала <http://openbiodiv.net/>. Представени са функционалностите на системата, които могат да се използват от трите типа потребители: (1) основно ниво: използва работна лента със средства, (2) ниво на специалист: използва приложения и (3) експерт: използва SPARQL или R. Все още не всички средства за търсене са разработени. Уеб порталът предлага основните функционалности за семантично търсене за трите типа потребители.

**Седма глава** съдържа листинг на сорс кода на някои SPARQL заявки, които са описани в предходните глави. Освен това е зададен и сорс кода на библиотеката RDF4R.

**Осма глава** описва представянето на уебинара iDigBio. Описан е профилът на слушателите.

**Заклучението** съдържа оценка на изпълнението на всяка от поставените задачи в дисертацията. Скицирани са насоки за бъдеща работа. Приложена е авторска справка, публикации, цитирания, апробация на резултатите - изнесени доклади. Общо са представени 8 публикации, забелязани са 20 цитирания. Една от публикациите представлява работен план за дисертационния труд. Резултати от дисертационния труд са представени на 3 вътрешни научни семинара в БАН и на 16 международни събития.

## Приноси в дисертацията

Дефинирани са 3 приноса на докторанта – един научен и два научно приложни:

**Научен принос:** Създаване на онтология OpenBiodiv-O и формален модел на областта на публикуване на знание за биологично разнообразие.

**Научно-приложни приноси:**

- Анализ на източниците на информация и създаване на OpenBiodiv-LOD
- Имплементация на софтуерните модули на OpenBiodiv

## Достоверност на получените резултати

Основен акцент на цялата работа е отвореност на данните и на програмния код. Приложени са линкове към всички данни, ресурси и програмни реализации. За достоверността на дисертацията говорят и осемте публикации, изнесените доклади по темата на дисертацията, прякото участие на докторанта в изследователския процес и описанието на направените проучвания.

## 2. Общо описание на публикациите, които отразяват дисертацията – монографии, статии, свидетелства и патенти, класифицирани по тематика или друг признак и редуцирани поради съвпадение или препокриване

От формална гледна точка, съгласно Правилника за прилагане на ЗРАСРБ, Постановление № 26 от 13.02.2019 г. за изменение и допълнение на ППЗРАСРБ, касаещо Област 4. Природни науки, математика и информатика, проф. направления 4.1., 4.2., 4.3., 4.4., 4.5., 4.6., и Правилник за специфичните условия за придобиване на научни степени и за заемане на академични длъжности в ИИКТ-БАН се изисква докторантът да има минимум 30 точки от група Г (показатели от 5 до 10).

Докторантът Виктор Сендеров е представил общо 8 статии, като всички са на английски език в рецензирани издания с отворен достъп, в т.ч.:

- 3 публикации (№3, 7, 8) в реферирани научни списания в SCOPUS, с импакт ранк SJR, освен това всичките 3 публикации са реферирани и в Web of Science, като две от тях (№7,8) имат и импакт фактор и попадат в квантил Q3. Публикации (№3,8) са индексирани и в PubMed
- 5 публикации (№1, 2, 4, 5, 6) в Research Ideas and Outcomes – електронно списание на издателство Pensoft. Трябва да се отбележи, че публикация №1 не е научна публикация, а работен план за докторантурата, ето защо няма да бъде включена при изчисление на точките.

Всички представени публикации са в съавторство, както с научните си ръководители, така и с български и чуждестранни учени. На три от публикациите е първи автор (№1,4,8). Докторантът не е представил нито една самостоятелна публикация.

Съгласно Постановление № 26 от 13.02.2019 г. за изменение и допълнение на ППЗРАСРБ се зачитат само научни публикации в издания, които са реферирани и индексирани в световноизвестни бази данни с научна информация (*Web of Science u Scopus, Zentralblatt, MathSciNet, ACM Digital Library, IEEE Xplore u AIS eLibrary*)

Така за точка Г (показатели 5 до 10) се отчитат само 3 от публикациите (№3, 7, 8) на Виктор Сендеров. Тогава сумарният брой точки по този показател е 80 т., което също покрива минималното изискване от 30 т.

№	Издание	Издание	Издание	Издание
3	Biodiversity Data Journal	Pensoft	WoS SCOPUS, SJR 0.465	30
7	ZooKeys	Pensoft	WoS IF 1.079, Q3 SCOPUS, SJR 0.533	30
8	Journal of Biomedical Semantics	Springer Nature	WoS IF 1.6, Q3 SCOPUS, SJR 0.952,	20
			<b>Total</b>	<b>80</b>

### **3. Отражение на резултатите от дисертацията в трудове на други автори. Числови показатели – цитати (без автоцитатите), импакт-фактор и др.**

Забелязани са 18 цитирания на публикации на докторанта, представени в дисертационния труд. Публикациите в реномирани международни научни издания, както и множеството им цитирания от световни учени, показват значимостта на получените резултати от научните изследвания за областта.

### **4. При колективни публикации да се отрази приносът на кандидата**

Според мен приносът на кандидата при колективните публикации е ясен и съществен. Той личи от засегнатата тематика. На три от публикациите докторанта е първи автор.

### **5. Критични бележки и препоръки на рецензента**

Бих направила някои технически бележки, които не намаляват стойността на представения дисертационен труд от гледна точка на постигнатите резултати от кандидата, но биха способствали за подобряването на изложението:

1. Добре би било да се отделят печатните от електронните литературни източници.
2. Текстът на дисертационния труд се нуждае от корекции от стилистична и терминологична гледна точка. Има известни пунктуационни и правописни грешки, чието отстраняване само би спомогнало за подобряване на общото впечатление от представените научни резултати. Добре би било да се избягва употребата на чуждици и жаргодни думи.
3. Като цяло структурата на текста на дисертацията не отговаря на традиционно приетото оформление. Липсва глава обзор – има само 3 страници преглед на литературата в уводната част, но в тях е включено и описание на структурата на дисертацията. Листингите на код (глава 7) трябва да бъдат изнесени като приложение извън основния текст на дисертацията. Приложението (глава 8) също не трябва да бъде част от основния текст, а да бъде след използваните литературни източници. Липсва списък с фигури и списък с таблици, каквито обикновено се включват в описанието.
4. Според мен целта на дисертационния труд не е много добре дефинирана – нуждае се от известно прецизиране в посока акцент върху научния проблем, който се решава.

5. Бих искала да препоръчам на докторанта да публикува самостоятелни публикации в научно направление 4.6 Информатика и компютърни науки, с цел популяризиране на резултатите и получаване на обективна оценка на постиженията му в тази област от международни рецензенти.

## **6. Качества на автореферата, включително доколко правилно отразява приносите на дисертацията**

Беше ми предоставен автореферат на български език. Той съдържа 54 страници и отразяват основните глави и резултати от дисертацията.

## **7. Заключение**

В заключение считам, че така представените публикации и описаните резултати в дисертацията, показват изградени качества у докторанта Виктор Сендеров за извършване самостоятелно на цялостни научни и научно-приложни изследвания в съвременна интердисциплинарна област. Представеният за рецензиране дисертационен труд отговаря на изискванията на Закона за РАСРБ и на съответните Правилници на МОНМ, БАН и ИИКТ-БАН. Покрити минималните изисквания за присъждане на ОНС „Доктор“ според Правилниците за приложение на ЗРАСРБ, БАН и ИИКТ-БАН за професионално направление 4.6 Информатика и компютърни. Предвид горното и поради научните приноси на кандидата в дисертационния труд, давам **положителна оценка и предлагам на уважаемото жури да присъди на Виктор Сендеров образователната и научна степен “доктор”** в област на висше образование 4. „Природни науки, математика и информатика“, Професионално направление: 4.6 „Информатика и компютърни науки“, Научна специалност: 01.01.12 „Информатика“.

София, 7 юни 2019 г.

Рецензент.

**NOT FOR  
PUBLIC RELEASE**

/доц. д-р Светла Бойчева/