

## СТАНОВИЩЕ

относно

дисертационния труд на **Виктор Ернестов Сендеров**

за придобиване на образователната и научна степен „доктор” по докторска програма „Информатика” в професионално направление 4.6 „Информатика и компютърни науки”

по тема

**„The Open Biodiversity Knowledge Management System in Scholarly Publishing”  
(OpenBiodiv: отворена система за управление на данни за биологичното разнообразие”)**

от

**доц. д-р Геннадий Павлович Агре**

Институт по информационни и комуникационни технологии – БАН

Със заповед № 86/30.04.2019 г. на директора на Института по информационни и комуникационни технологии – БАН (ИИКТ-БАН) съм утвърден за член на Научното жури във връзка с процедура за придобиване на образователната и научна степен „доктор” по докторска програма „Информатика” в професионално направление 4.6 „Информатика и компютърни науки” по тема „The Open Biodiversity Knowledge Management System in Scholarly Publishing” (OpenBiodiv: отворена система за управление на данни за биологичното разнообразие”) с научен ръководител проф. д-р Любомир Пенев (Издателство „Пенсофт”) и научен консултант доц. д-р Кирил Симов (ИИКТ-БАН)

Като член на журито получих:

1. Дисертация за придобиване на образователната и научна степен „доктор” по докторска програма „Информатика” в професионално направление 4.6 „Информатика и компютърни науки” (на английски език)
2. Автореферат (на български език)
3. Професионалната биография (на английски език)
4. Копия от публикации на автора, свързани с дисертацията.

Дисертацията съдържа 113 страници, структурирани в Увод, 8 глави, заключение, списък на използваните съкращения и библиография.

### **1. Актуалност на тематика на дисертационния труд и целесъобразност на поставените цели и задачи**

Дисертацията е свързана с една актуална тема от областта на семантичните технологии – създаване и управление на отворени свързани данни (linked open data – LOD) с приложение в една конкретна предметна област – публикуване на данни за биологичното разнообразие. Целта е да бъде предложен формален семантичен модел на областта, който след това да бъде приложен за решаване на задача за разработване на отворена система за създаване и управление на свързани отворени данни от тази област. Поставените цели и задачи са много важни за избраната проблемна област, тъй като водят до систематизиране на знанието в областта чрез онтологично описание на използваните понятия, уеднаквяване и съгласуване на наименованията на термините, както и чрез създаване на цифрово хранилище от свързани отворени данни.

### **2. Познаване на състоянието на тематичната област**

Дисертантът показва много добро познаване както на съвременните тенденции и постижения от областта на семантичните технологии, свързани с разработка на бази от

знания и свързани отворени данни, така и на важни специфики от избраната проблемна област – публикуване на знания и данни за биоразнообразие. Библиографията съдържа 163 източника, от които само 3 са на български език.

### **3. Методика на изследването**

Изборът на методология на изследването е обусловен от формулираните от дисертанта 6 конкретни задачи, реализиращи поставените в дисертацията цели. Методологията включва мотивиран избор на технология за създаване на база от данни за биоразнообразието, избор на основни източници на информация за проблемната област, избор на подходящата средата за програмиране, както и решение да се използва отворен изходен код, за осигуряване на проверимостта и възпроизводимостта на получените резултати.

### **4. Кратка аналитична характеристика на дисертационния труд**

Уводната част започва с мотивацията на изследването и продължава с аналитичен обзор на литература в избраната проблемна област. След анализ на използваните подходи и постигнатите резултати дисертантът формулира основните цели на дисертацията, задачи, решаването на които е необходимо за постигане на тези цели, както и методология за решаване на поставените задачи. Първата глава описва архитектурата на системата, която е разбита на три функционални блока - база от семантични графи (онтологии и отворени свързани данни), набор от функции за управление и поддръжка на базата и динамичната страница (уеб портал) за достъп към базата. Втората глава (която, по моето мнение, е централна за дисертацията) описва създадения формален модел на областта на публикуване на данни за биоразнообразие – онтология OpenBiodiv-O. Онтологията предоставя един концептуален модел на публикации в областта на биоразнообразие, като въвежда класове и релации от тази проблемна област и осъществява семантично моделиране на биологична номенклатура и таксонометрични понятия. Третата глава е посветена на отворени свързани данни, създадени от системата чрез извличане на данни за биологично разнообразие от статии, публикувани от списания на издателствата Пенсофт и Плаци чрез прилагане на разработената онтология OpenBiodiv-O. Представени са и резултати от тестове за поведение на системата и времето на обработка на различни обеми данни. Четвъртата глава представя функционалността на разработения от дисертанта пакет за работа с RDF в програмната среда R – RDF4R, като се прави и сравнение с подобни съществуващи пакети. Глава 5 разглежда два метода за автоматизиране на процеса на обработка на данни за биоразнообразие, разработени като част от системата - автоматично внасяне на записи за наблюдение на видове в таксономични статии и автоматично генериране на статии от екологични метаданни. Глава 6 представя интерфейса на уеб портала, който дава достъп до разработената система. Глави 7 и 8 представят разпечатка на част от изходния код на системата, както и детайли от уебинар, споменат в Глава 5. Смятам, че тези глави не са от съществено значение за дисертацията и би трябвало да бъдат оформени като приложения. Заключението обобщава резултатите в дисертацията, представя приносите на дисертанта, неговата публикационна активност, както и информацията за апробация на резултатите.

### **5. Приноси на дисертационния труд**

Дисертантът посочва един научен и два научно-приложни приноса. Като научен е посочено създаване на онтология OpenBiodiv-O предоставяща формален модел на публикации в областта на биоразнообразие. Научно-приложните приноси са свързани със създаване на свързани отворени данни за изследвана проблемна област и с реализация на програмните модули на системата. Смятам, че тези научни и научно-приложни приноси са посочени коректно.

## **6. Преценка на публикациите по дисертационния труд**

Дисертантът представи 8 статии по темата на дисертацията. Смятам, че публикациите отразяват коректно съдържанието на дисертацията и представляват оригинални постижения на дисертанта.

Съгласно действащият в момента Правилник за специфичните условия за придобиване на научни степени и за заемане на академични длъжности в ИИКТ-БАН, дисертантът трябва да има най-малко 30 точки от неговите публикации, свързани с дисертацията, като на оценяване подлежат само публикациите, индексирани в световно известни бази данни. Изчисляването на точките става по начина, приет с решение от 20.05.2019 г. на четиридесет и осмото заседание на VII-то Общо събрание на БАН.

От представените от дисертанта 8 статии само 3 (№8, №7 и №3) са публикувани в списания, индексирани в световноизвестни бази данни Web of Science (WoS) и/или Scopus, като статии №8 и №7 са публикувани в списания с импакт фактор (IF) (и двете попадат в трети квартал – Q3), а статия №3 – в списание с импакт ранг, индексирано в SCOPUS. Съгласно приетия начин за оценяване на публикации в научно направление 4.6. „Информатика и компютърни науки”, указаните три публикации събират  $2 \times 30 + 1 \times 20 = 80$  т., което значително превишава както националните минимални изисквания, така и специфичните изисквания на ИИКТ-БАН за придобиване на научна степен „доктор”.

Освен това, трябва да се отбележи, че до момента посочените 3 публикации имат 15 независими цитирания в Scopus.

## **7. Лично участие на докторанта**

Макар че докторантът до момента няма самостоятелни публикации по темата на дисертацията, в една от двете публикации с импакт фактор (тази с най-висок IF - №8) той е първият автор. Това показва, че личното участие на дисертанта е съществено.

Анализът на текста на дисертацията, както и на всички представени от дисертанта публикации позволява ми да направя заключение за липса на плагиатство в представената на мен за рецензиране дисертация.

## **8. Автореферат**

Авторефератът отразява коректно целите, задачите и резултатите от дисертационния труд.

## **9. Забележки**

Имам определени забележки към структурата на дисертацията – смятам, че материалът, представен в Глави 7 и 8, трябва да бъде оформен не като глави, а като приложения. Освен това имам определени резерви към превода на някои термини, като, например „графични” бази данни (graph databases), „буквални” стойности (literal values) и др.

Обаче тези забележки не намаляват научната и научно-приложна стойност на предложения дисертационен труд.

## 10. Заключение

От всичко приведено по-горе следва, че всичките изисквания на Закона за развитие на академичния състав, Правилника за неговото прилагане и Правилника за специфичните условия за придобиване на научни степени и за заемане на академични длъжности в Институт по информационни и комуникационни технологии – БАН са напълно изпълнени. Смятам, че дисертацията на Виктор Ернестов Сендеров има всички качества, изисквани от един дисертационен труд. Получените от дисертанта резултати са значителни и допринасят за по-широко прилагане на семантични технологии в различни проблемни области и особено в областта на публикуване на данни за биологично разнообразие.

Всичко това ми дава основание за положителната оценка на представения дисертационен труд и убедено препоръчвам на почитаемото Научно жури да присъди образователната и научна степен „доктор” в научно направление 4.6 „Информатика и компютърни науки” на Виктор Ернестов Сендеров.

10 юни 2019 г.

София

**NOT FOR  
PUBLIC RELEASE**