

## АВТОРСКА СПРАВКА ЗА ПРИНОСНИЯ ХАРАКТЕР НА ТРУДОВЕТЕ

на гл.ас. д-р Валерия Николаева Симеонова,

представени за участие в конкурса за академична длъжност „Доцент“

по 4.б. Информатика и компютърни науки (Информатика),

обявен в Държавен вестник, бр. 33 от 17.04.2018 г.

Списъкът на представените за участие в конкурса за доцент трудове включва 17 публикации. Таблицата по долу дава разпределението на публикациите по години, брой, представяни публикации в други конкурси, брой статии с IF, брой статии без IF, брой учебно-методическа литература (УМЛ), брой участия с доклади на конференции:

Година	Представени в друг конкурс	с IF	без IF	УМЛ	Доклад от конф.	Общо
2018	0	3	2	2	2	9
2017	0	0	1	0	1	2
2015	1	2	0	0	0	2
2013	1	1	0	0	0	1
2012	2	1	1	0	0	2
2008	1	0	0	1	0	1
Общо	5	7	4	3	3	17
						11

- 7 са в списания с импакт фактор, 4 от които са приети за печат 2018
- 4 - пълнотекстови публикации без импакт фактор, от които 2 са приети за печат 2018
- 3 доклада от конференции, свързани с някои от другите публикации
- 1 публикуван на хартия учебник
- 2 електронно публикувани в сайта на ФМИ учебника, които са в процес на подготовка и оформление за печат.
- От представените публикации за участие в конкурса за заемане на длъжност „главен асистент“ са общо 5 публикации [7, 9,10,11,13,14].

От представените 17 публикации, 13 са в съавторство в резултат на съвместна работа с колеги от Факултета по математика и информатика (ФМИ) на Софийски университет (СУ) „Св. Климент Охридски“.

Публикациите, представените за участие в конкурса, могат да бъдат групирани в следните основни тематични направления:

- [1] Алгоритмични подходи в сферата на изкуствения интелект за работа с големи масиви от секвенционни данни;
- [2] Методи за анализ на данни в биоинформатиката;
- [3] Учебно-методическа литература

Представянето на приносния характер на публикациите е развито в съответствие с изложеното групиране, като отделните статии са цитирани с номерата им в списъка на трудовете по конкурса.

## 1. Алгоритмични подходи и софтуер за работа с големи масиви от секвенционни данни

---

В тази група попадат трудове [1,3,8,9,12,13]. Разработването и използването на различни алгоритмични подходи и софтуерни решения за анализ на големи масиви от данни в биоинформатиката заема особено място в преодоляването на непрекъснатото генериране на данни от паралелно геномно секвениране за целите на медицината, фармацевтиката и биологическите изследвания. Този кръг от задачи се свежда до намиране на най-бързите, по възможност най-точни и най-икономични способи за съхранение, търсене, съпоставяне и визуализация на такива данни.

Главните, обобщени приноси от тази група публикации биха могли да се разгледат в два аспекта – методически и приложен. В методическо отношение, основните приноси са разработване на модели за търсене на грешки и структурни варианти с използването на усъвършенствани аналитични подходи и подходи от машинно обучение. В приложно отношение приносите се изразяват в разработването на софтуер за анализ на данни от паралелно секвениране.

**Публикация 1** разглежда моделирането на подходящо място за внедряване на допълнителен протеинов участък на вируса на Инфлуенца във вируса на Хепатит Е. Биологическата цел на една такава операция е по-бързо разпознаване на вируса на Хепатит Е още в ранен стадий. Чрез използване на методи от областта на извличане на данни и изкуствен интелект се определят подходящите позиции в структурата за такова внедряване. Основният принос на публикацията е, че дава бърз алгоритъм за определяне на възможни позиции за внедряване, придружен с качествени оценки, позволяващи избор на позиция.

**Публикация 3** се фокусира върху оценката на качеството на различни модели за машинно самообучение приложени за откриване на грешките в данните от паралелно геномно секвениране. Работата е продължение на цикъл от публикации посветени на откриването на грешките в такъв тип данни чрез невронни мрежи и размита логика. Основният принос се изразява в това, че се установи, че резултатите получени чрез обучение по правила са подобни на тези при използване на случаен модел на дървета, като и в двата случая алгоритъмът на Hoefding създава точност, която е сравнима и понякога по-добра от модела на невронни мрежи. Това доказва, че последните два модела са подходящи за подобряване на точността и надеждността на данните за секвенциите за целите на метагеномни изследвания.

**Публикация 8** е посветена на използването на различни алгоритмични подходи за намаляване нивото на грешките в силно вариращи големи масиви метагеномни секвенционни данни. Оригиналният методически принос на работата е представянето на т.н. размит индикатор на надеждност за обезшумяване на данни от паралелно секвениране. Други приноси на работата са използването на различни дистанционни метрики, основани на модифицирани алгоритми на Хеминг и Левенщайн, и

приложението им при анализ на данни от паралелно секвениране. Разработените подходи дават подобрение при откриването на трудно разпознаваеми грешки.

**Публикация 9** показва рационалното използване на различни модели на машинно обучение за повишаване точността на анализите за търсене на грешки в големи масиви данни от паралелно секвениране. Основните приноси на работата се заключават в усъвършенстване на стандартните аналитични алгоритми, базирани на метрики за подобие и съпоставяне на низове. Целта е намаляване на шума (грешките) поради високото ниво на грешки от тип 1 (*false positives*) в области от данни с голямо вариране. В публикацията се използват модели на машинно обучение (невронни мрежи и случайна гора) за филтриране на някои от грешките. Допълнителното използване на тези алгоритми дава повишаване на точността до над 99%. Работата демонстрира, че обучените модели предлагат добър филтър срещу неправилно открити грешки, без да увеличават грешките от тип 2 (*false negatives*) в големи масиви от секвенционни данни.

**Публикация 12** Филогенетичните дървета могат да се разглеждат като форма на йерархичен клъстер, който визуализира отношенията между гени, протеини или цели видове, които имат общ произход. Това изследване е съсредоточено върху растителни генотипи и описва обработката на публично достъпни филогенетични данни със собствени Python скриптове на авторите. Основният принос е в разработването на методологии за намаляване, прегрупиране, манипулиране и идентифициране на хомоложните връзки и топологичните модели в набора от данни, като по този начин се осигуряват по-добри познания за еволюцията на основните филогенетични черти в растенията. Описаните методи и скриптове са успешно изпълнени при разпознаване на шаблони за специфични черти и изследвания на функционалните анотации в растенията.

**Публикация 13** има обзорно-методичен характер и е писана предимно от доц. Д-р Димитър Василев. Основният ѝ принос е дискутирането и сравняването на различни алгоритмични подходи за анализ на данни от секвениране следващо поколение – по-специално алгоритмични подходи за асемблиране на такива данни, които по принцип са NP сложни задачи. Разглежданите подходи са основани и на паралелни изчисления, и на работа в *cloud* структури. Сред приносите на работата може да се посочи и това, че тя представлява обзорна статия, предназначена за докторанти и дипломанти, както и за колеги, работещи в областта на модели за анализ на данни от паралелно секвениране.

## 2. Методи за анализ на данни в биоинформатиката

---

Данните, генерирани от биоинформатиката и по-специално данните, генерирани от технологиите за паралелно секвениране, са предмет на много и различни проекти. Тези проекти са фокусирани върху използването на различни изчислителни методи от информатиката, основани на класически аналитични подходи, както и на съвременни подходи на статистически анализ, и особено в последно време - на модели на машинно обучение. Хетерогенността на данните в големите масиви от секвенционни данни, както и потребността тези данни да бъдат анализирани бързо и сравнително евтино, изисква използването и развитието на по-широк кръг от методи и модели за техния анализ, съхранение и визуализация.

От публикациите, представени за конкурса, към тази група спадат работите с номера [2,4а, 4b,10,11,14,15,16].

Главните, обобщени приноси от тази група публикации са както методически, така и приложни. В методическо отношение това са създаването и използването на модели за анализ на различни видове биоинформатични данни, основани на статистика, както и на машинно обучение. Основен принос е интегрирането на тези подходи за анализ на данни и приложението на разработените подходи върху реални данни.

Тук, публикациите е правилно да се разгледат в 3 подгрупи:

1. **Публикации [2,15, 16]** разглеждат един често срещан проблем на реалните лабораторни данни, който масово се игнорира. Резултатите от такова игнориране в повечето случаи води до непубликуване. Става въпрос за анализ и обработка на компрометирани данни. Обичайният резултат от анализа им показва, че не се наблюдават значими резултати, което от биологическа гледна точка означава отхвърляне на дадена теория. Основният принос на публикациите е, че се предлага нов подход за последователността на анализа и обработката на биологически данни. В основата му е залегнала Булевата алгебра и съставянето на булеви функции въз основа на статическите резултати за значимост. Подробен статистически анализ на суровите данни валидира получените резултати от прилагането на подхода.
2. **Публикации [4a,4b,10,14]** разглеждат нов подход за извличане на закономерности между т.нар. хранителни среди и получените външни белези при лечебното растение Златен корен. Подобен тип закономерности се използват при лекарствения дизайн, като оригиналното наименование е QSAR. Това е и една от причините нашият модел да се нарича „адаптиран QSAR“. Адаптацията е предимно при определянето на входните и изходните данни. Използват се невронни мрежи за установяване на закономерностите. По този начин се постига лаборатория за изкуствено тестване на хранителни среди и тяхното влияние върху биологическите процеси вкореняване, растеж, размножаване. Предложеният подход реализира икономии и ефективност при определяне на теоретичните кандидат-хранителни среди. Разработките са свързани с конкретен проект в тази област.
3. **Публикация 11** решава предимно приложна задача, която е свързана от една страна с намиране на бързо решение, а от друга страна с извличане на дескриптивни характеристики за определен геном. Основният принос е свързан с ефективното програмиране, т.к. обемът на данните е значителен, а задачата е решавана със стандартни технологични средства. В публикацията се описва изследване, обхващащо данни от паралелно геномно секвениране и се отнася към работите, свързани с различни методи за откриване на грешки в големи масиви от секвенционни данни. Основните приноси на тази публикация се отнасят до разработването на подходи за откриване на грешки в секвенционни данни, представени в различни публични, свързани бази от данни. Примерът включва съпоставянето на секвенционни данни за ориз в две глобално използвани бази от данни – *NCBI* и *PlantGDB*. Сравняването е извършено по зададени параметри в определени локации в геномните данни, а разработеният подход представлява друг методически принос на изследването.

### 3. Учебно-методическа литература

---

**Публикация 5** представлява набор от лекциите по дисциплината „Математика и Информатика“, модул „Информатика“. Материалът е достъпен за студентите в електронната среда за обучение Moodle, както и на сайта на ФМИ, в личната ми страница. В момента се подготвя и дооформя, за да бъде предложен за официално отпечатване под заглавието „Програмиране с R за начинаещи“. По същият курс лекции е проведено обучение за служители на Теленор АД, като курсът предизвика голям интерес и имам покана за изнасяне на продължаващо обучение за работа с езика за програмиране R.

**Публикация 6** представлява набор от лекциите по дисциплината „Информационни системи и технологии“. Материалът е достъпен за студентите в електронната среда за обучение Moodle, както и на сайта на ФМИ, в личната ми страница. В момента се подготвя и дооформя, за да бъде предложен за официално отпечатване под заглавието „Информационни системи и технологии“. Поднася един материал, който има изключително практическа насоченост в ежедневните задачи при работа техника и документи.

**Публикация 7** е ръководство за работа с MS Office. То разглежда в дълбочина възможностите на офис пакета чрез подробно описани примери и цветен материал за по-добро възприятие. Т.к. е писан в стил 2 в 1, т.е. разглежда едновременно офис пакетите преди и след версия 2007, въпреки времето остава актуална публикация. В последните три години ръководството е предоставено в електронен вид чрез Мудъл на студентите, слушащи дисциплината „Информационни системи и технологии“, а така също и на магистрите слушащи „Методи за планиране, провеждане и документиране на научни проучвания и проекти“. То е в основата на упражненията по дисциплината „Информационни системи и технологии“. Ръководството се оценява високо и се възприема много добре от студентите.

---

СПРАВКА във връзка с чл.108 от ПУРПНСЗД в СУ "Св. Климент Охридски"  
и чл.5 (2) и чл. 6 от ПУРПНСЗД във ФМИ на СУ "Св. Климент Охридски"

## I. Преподавателска дейност за периода 2015-2018 година

---

### *Основен титуляр за следните дисциплини в бакалавърски програми на СУ:*

- „Информационни системи и технологии“, задължителен за студенти от Биологически факултет
- „Практикум информационни системи и технологии“, задължителен за студенти от Биологически факултет
- „Математика и Информатика“, модул „Информатика, задължителен за студенти от Биологически факултет

### *Основен титуляр за следните дисциплини в магистърски програми на СУ:*

- „Методи за планиране, провеждане и документиране на научни проучвания и проекти“, задължителен за студенти от МП: „Електронно обучение“ и избираем за „Био-медицинска информатика“, ФМИ
- „Обработка и анализ на данни“, избираем за студенти от МП: „Електронно обучение“ и „Био-медицинска информатика“, ФМИ
- „Невронни мрежи и генетични алгоритми“, избираем за студенти от МП: „Изкуствен интелект“, ФМИ

### *Участник в екипи по други дисциплини в бакалавърски програми на СУ:*

- „Компютърни системи и технологии“, задължителен за студенти на Факултет по математика и информатика
- „Увод в програмирането“, задължителен за студенти на Факултет по математика и информатика
- „Проектиране на човеко-машинен интерфейс“, задължителен за студенти на Факултет по математика и информатика

### *Дисциплини с изцяло обновено съдържание:*

- **Информационни системи и технологии:** преподава се на студенти от Биологически факултет, бакалавърска степен; лекционният материал включва както основи на информационните системи, така и изключително практически познания, свързани с ежедневното търсене, достъп и обработка на данни и файлове; материалът за упражнения покрива офис технологии за документна обработка, като се набляга на придобиване на умения за професионално и полу-автоматизирано форматиране на текст, обработка на таблични данни и създаване на презентации. Освен разработените специално за курса материала, д-р Симеонова има издадено ръководство „ Възможностите на MS OFFICE 2007 в теоретично обяснени и решени задачи с приложени адаптации към MS OFFICE 2003“, който се предоставя на студентите под формата на pdf файл в средата за обучение Moodle.
- **Практикум ИСТ:** преподава се на студенти от Биологически факултет, бакалавърска степен; дисциплината включва запознаване и работа на студентите със специфични за тяхната сфера информационни системи в различните под-области на биологията свързани с достъпа и

обработката на данни от сферите на геномиката, протеомиката, метагеномиката, епигенетиката; различни браузъри и реактоми и други.

- **Методи за планиране, провеждане и документиране на научни проучвания и проекти:** преподава се на студенти от ФМИ, магистърска степен; дисциплината има за цел да подготви студентите за планирането и написването на дипломен проект, като се разглеждат структурите на различните видове научни проучвания. Студентите придобиват знания „къде и как“ да търсят и намират ефективно качествена научна литература, как да цитират правилни. Придобиват навики за използването на различни информационни системи за цитиране. Научават се правилно да дефинират целите и задачите и да структурират даден научен текст, в зависимост от неговото предназначение.
- **Обработка и анализ на данни:** преподава се на студенти от ФМИ, магистърска степен; Включва теми като: Предварителна обработка на данните; Видове графични представяния на данните и разчитането им; Техники за сравняване на алгоритми за обработка и анализ на данни
- **Невронни мрежи и генетични алгоритми:** преподава се на студенти от ФМИ, магистърска степен; Включва теми и за съвременните методи и алгоритми за невронни мрежи от областта на дълбокото обучение

#### *Нови дисциплини:*

- **Математика и Информатика (модул Информатика):** преподава се на студенти от Биологически факултет, бакалавърска степен; лекционният материал представя основите на програмирането в средата R Project. Има подготвено учебно помагало, което е написано на достъпен за студентите биолози език с подходящи от тяхната сфера примери.

За всички посочени дисциплини, по които съм основен титуляр са разработени авторски електронно-учебни материали, които ежегодно се актуализират и публикуват в СУО Moodle: <https://learn.fmi.uni-sofia.bg/>

## II. Самостоятелно научно ръководство на защитили дипломанти

---

1. *Антония Ставрева* – „Модернизирани на реален международен образователен портал“, защитила 2016 г., МП РСМТ
2. *Елена Миронова* – „Сравнителен анализ на избрани методи за клъстеризация на сигнали“, защитила 2016 г., МП Изкуствен Интелект
3. *Ивайло Бараков* – „Метод за подготовка, мониторинг и отдалечена настройка на клиентски стационарни и мобилни компютри“, защитил 2016 г., МП РСМТ
4. *Ивайло Иванов* – „Изграждане на графичен потребителски интерфейс за използване на API услуга“, защитил 2017 г., МП РСМТ
5. *Кристиян Авкатев* – „Графичен потребителски интерфейс за използване на API услуги – надграждане с нови функционалности“, защитил 2018 г., МП РСМТ

## III. Изследователска дейност

---

**Общ списък публикации** – съобразно правилника за конкурса е представен отделно общ списък от 57 публикации, включително представените за конкурса. В списъка са включени 28 статии, 10 доклада в пълен текст, резюмета от доклади и постери от конференции, една глава от книга, както и копие от автореферата на дисертацията ми.

## IV. Научни проекти

---

### Участие в научни проекти, финансирани от ЕС

1. **Университети за Наука, Информатика и Технологии в е-обществото (УНИТе)**, Член, Европейски фонд за регионално развитие по ОП НОИР, Номер на договора: BG05M2OP001-1.001-0004-C01, длъжност: „Главен секретар“, начало 2018, текущ

### Участие в научни проекти, финансирани от български източници

1. Фонд Научни Изследвания на Р. България. Проект/Контракт Ю2/7/2014 “Методи за анализ на данни и откриване на знания в големи масиви секвенционни данни” *Ръководител: Проф. Мария Нишева, Факултет по Математика и Информатика, Софийски Университет „Св. Климент Охридски“*, начало 2018, текущ
2. Фонд Научни Изследвания на Р. България. Проект/Контракт ТК18-010 “Модели за семантично интегриране на биомедицински данни” *Ръководител: доц. Димитър Василев, Факултет по Математика и Информатика, Софийски Университет „Св. Климент Охридски“*, начало: 2014, приключил: 2018
3. Фонд Научни Изследвания на Р. България. Проект/Контракт ДМУ 03/55 „Приложение на биотехнологични и биоинформатични методи за получаване in vitro на биомаса от застрашения вид *Rhodiola Rosea* (златен корен) и за повишаване продукцията на биологично активни вещества“, *ръководител: д-р Красимира Ташева, ИФРГ БАН*, начало: 2011, приключил

## V. Научно-организационна дейност

---

### Участие в органи на управление

1. Член на Факултетен съвет на Факултет по математика и информатика от 06.07.2016 г.
2. Член на Комисията по качество към Факултет по математика и информатика от Февруари 2016 г.



## VI. Експертна дейност

---

### Рецензии на дипломни работи

1. Яна Марчева Николова, „Интегриране на софтуерни приложения за съпоставяне на нуклеотидни секвенции в разпределена среда с помощта на Hadoop“, научни ръководители: проф. д-р Владимир Димитров и доц. д-р Димитър Василев, катедра “Компютърна Информатика”
2. Цветан Красимиров Цонков, „Извличане на данни от голям масив от данни от сензори на превозни средства“, научен ръководител: доц. д-р Десислава Георгиева Петрова-Антонова, катедра „Софтуерни технологии“
3. Павел Костадинов Асенов, “Проектиране и разработка на информационна система за фактуриране на телекомуникационна компания“, научен ръководител: доц. д-р Евгений Кръстев
4. Милена Ненкова Соколова, “Система за оценка достоверността на резултати от паралелно секвениране“, научен ръководител: доц. д-р Евгений Кръстев, доц. д-р Димитър Василев, научен консултант: д-р Милко Крачунов, катедра „Информационни технологии“
5. Стоян Димитров, "Дизайн на високонадеждна ресурсно-ефективна бизнес инфраструктура", научен ръководител: доц. д-р Елиза Стефанова, катедра „Информационни технологии“
6. Йордан Лазаров, "Синхронизиране на локална активна директория с облачна активна директория с цел интеграция с облачни услуги", научен ръководител: доц. д-р Елиза Стефанова, катедра „Информационни технологии“
7. Христо Малеев, "Създаване на система за следене на физическите параметри на сървъри реализирана чрез IoT модули", научен ръководител: проф. д-р Красен Стефанов, катедра „Информационни технологии“
8. Венелин Венелинов, "Моделиране и разработка на информационна система за отчитане на свършена работа по Европейски проекти", научен ръководител: проф. д-р Евгений Кръстев, катедра „Информационни технологии“

### Рецензии по вътрешни проекти към ФМИ

1. „Систематичен подход за изследване и оценка на интелигентни градове SystEmatic approach for eXPloRation and assEssment of Smart cities (EXPRESS)“, ръководител на проекта: доц. д-р Десислава Георгиева Петрова-Антонова, 2018
2. „Големите обеми от данни като инструмент за подпомагане взимането на решение в областта на образованието“, ръководител на проекта: доц. д-р Десислава Георгиева Петрова-Антонова, 2017
3. „Откриване на знания при анализ на комплексна сложност на гени Анотации“, ръководител на проекта: доц. д-р Димитър Иванов Василев, 2017
4. „Интеграционна рамка за комуникация на устройства с ограничен обхват и изграждане на „умен“ дом“, ръководител на проекта доц. д-р Десислава Георгиева Петрова-Антонова, 2016

14.06.2018 г.  
София

Подпис: \_\_\_\_\_  
(Валерия Николаева Симеонова)