

## СТАНОВИЩЕ

по конкурс за заемане на академичната длъжност „професор“,  
обявен за нуждите на секция „Изкуствен интелект и езикови технологии“  
към Института по информационни и комуникационни технологии при БАН  
в ДВ бр. 45 от 28.05.2021 г.

Област на висше образование: **4. Природни науки, математика и информатика**  
Професионално направление: **4.6. Информатика и компютърни науки**  
Специалност: **Информатика**

Изготвил становището: **акад. Веселин Стоянов Дренски**, д.м.н., пенсиониран професор  
в ИМИ-БАН, определен за член на Научното жури със заповед № 185 от 27.07.2021 г. на  
директора на ИИКТ.

Единствен кандидат: **дн Стоян Милков Михов**, доцент в ИИКТ-БАН.

За участие в конкурса кандидатът е представил всички изисквани от ЗРАСРБ и  
Правилника за специфичните условия за придобиване на научни степени и заемане на  
академични длъжности в ИИКТ. Освен това документацията съдържа и една препоръка  
(на немски и английски) от проф. Клаус Шулц от Университета в Мюнхен, който е един  
от основните съавтори на кандидата. Въпреки, че това не се изисква от Закона и  
Правилника, би било добре към документите да бъдат приложени извадка от обявата за  
конкурса в „Държавен вестник“ и списък на всички публикации на кандидата, което би  
дало цялостна представа за успехите на кандидата. (На страницата на ИИКТ има такъв  
списък, но той не е осъвременяван от 2019 г.)

**1. Биографични данни.** Доц. дн Стоян Михов е завършил като магистър по математика  
Факултета по математика и информатика на Софийския университет „Св. Климент  
Охридски“ (с дипломна работа „Унификация на корегулярни множества“ и научен  
ръководител доц. д-р Анатолий Буда). Защищава докторска дисертация по  
информатика в ИИКТ на тема „Минимални ациклични автомати: конструкции,  
алгоритми, приложения“ с научен ръководител проф. дмн Димитър Скордев. Отново  
в ИИКТ защитава дисертация за доктор на науките по информатика и компютърни  
науки на тема „Крайни автомати, преобразуватели и бимашини: алгоритмични  
конструкции и имплементации“. Цялата научна кариера на доц. дн Михов е свързана  
с ИИКТ и неговите предшественици. Последователно е програмист, асистент, главен  
асистент, а от 2006 г. е доцент. (В автобиографиите на страницата на ИИКТ и  
представените за участие в конкурса документи има известно разминаване в годините  
за това, кога кандидатът е започнал своята трудовата дейност в БАН.) От 2003 г. е  
хоноруван преподавател във ФМИ на СУ. Освен това е бил ръководител на група в  
Рила Солющън ЕАД и ръководител на изследователска дейност (с допълнителен  
трудов договор) в Коммек ЕООД.

**2. Обща характеристика на научно-изследователската дейност на кандидата.**  
Основните направления на изследванията на доц. дн Стоян Михов са свързани с  
теоретичната информатика (теория на крайните автомати – на границата с  
абстрактната теория на изчислимостта, приближеното търсене, синтез и разпознаване  
на реч, компютърната лингвистика) и практическо реализиране на много от  
получените резултати с конкретни приложения в обработката на текстове от  
естествен език, разпознаване на реч, корекция и нормализация на текстове. Повечето

от създадените техники са ефективни от теоретична гледна точка и са приложени за актуални от практическа гледна точка задачи. Кандидатът е автор (в повечето случаи съавтор) на над 60 научни публикации, включително 1 монография в авторитетно международно издание, а повече от половината статии са с импакт фактор или SJR. Публикациите са за периода от 1989 г. досега. Съгласно представените данни в автобиографията на кандидата неговите публикации са цитирани над 420 пъти в над 340 документа. В приложения документ за покриване на минималните изисквания на ЗРАСРБ и на правилника на ИИКТ е приложен списък от 213 цитата, като една от статиите (с Шулц от 2003 г.) е цитирана 93 пъти. Специално ще отбележа, че кандидатът е създател на програмния език C(M), който директно превръща математически конструкции в програми на програмния език C. Това позволява бърза компютърна реализация на сложни алгоритми и има широки възможности за приложения.

**3. Характеристика и оценка на преподавателската и проектната дейност на кандидата.** Доц. дн Стоян Михов има дългогодишна и успешна преподавателска дейност. Той е водил и продължава да води редица курсове във ФМИ и Факултета по славянски филологии на СУ „Св. Климент Охридски“. Има двама успешно защитили се докторанти (в ИИКТ и ФМИ на СУ), 8 защитили дипломанти във ФМИ на СУ и 2 във ФСФ на СУ. Към тези данни ще добавя изключително положителния отзив на проф. Шулц за педагогическата дейност на кандидата. Съгласно проф. Шулц в своите лекции и семинари доц. дн Стоян Михов включва множество интересни и дълбоки по научно съдържание теми, с което мотивира студентите. Като научен ръководител той предлага на своите докторанти и дипломанти задачи, които водят до интересни задачи и директно практическо използване в приложни проекти. Още от 1993 г. доц. дн Стоян Михов участва активно в редица научни и приложни проекти, в много от които прилага на практика своите разработки. От 1996 г. е ръководител на повечето от проектите, в които участва. В момента е ръководител на работен пакет в Националната научна програма „Електронно здравеопазване в България“.

**4. Общо описание на представените материали.** Кандидатът е представил за участие в конкурса 15 научни публикации, публикувани в периода 2006-2021 г., 1 публикация приета за печат, 1 препринт в популярната база от препринти arXiv.org и един патент, регистриран в САЩ. От статиите 4 са в списания (по 1 в Natural Language Engineering, Theoretical Computer Science, J. Automata, Languages and Combinatorics, Computational Linguistics), 5 са в поредици (Lect. Notes Comp. Sci - 4, Studies in Comp. Intelligence - 1) и 7 са в трудове на конференции. Всичките 16 публикувани или приети за печат публикации са в издания с SJR. Кандидатът не е отбелязал това, но според мене две от списанията имат импакт фактор в годината на публикуване. Би било добре, ако кандидатът беше представил данни и за точните стойности на SJR. Всички статии са съвместни (4 с един съавтор, 8 с двама съавтори, 3 с трима и по една с пет и шест съавтори). Сред съавторите са ученици и колеги на кандидата, 4 са учени от Германия и 1 е учен от Канада. Особено впечатляващо е сътрудничеството с Шулц (съавтор на 10 от представените в конкурса статии) и с успешно защитилите под ръководството на кандидата докторанти Митанкин и Герджиков (съавтори съответно на 8 и 5 статии). Кандидатът е декларирал, че във всички публикации съавторите имат равноправно участие. Считаю, че в конкретния случай съвместните статии са положителен факт, защото изследванията са с интердисциплинарен

характер и на границата на няколко области. Съвместната работа увеличава КПД-то на изследванията, позволява използването на методи от различни области на науката и показва способността за работа в екип. Лично аз високо ценя качествата на успешна работа в екип и изследвания на ръба на няколко области. **Съгласно сведенията, които имам, представените публикации не са използвани преди това за други процедури.** За конкурса кандидатът е представил списък от 213 цитата на 6 свои статии. От тях 23 цитата са на статия, представена за участие в конкурса. **От представената таблица се вижда, че кандидатът напълно удовлетворява, а по част от критериите значително надхвърля минималните изисквания на ИИКТ за участие в конкурс за „професор“.** **Не установих плагиатство в представените за конкурса трудове.**

5. **Основни научни и научно-приложни приноси.** Ще се спра накратко на основните резултати, съдържащи се в представените работи на кандидата, както и на оценката ми за тях. В справката за научните си приноси кандидатът е разделил публикациите си на три групи: (1) Теория на крайните автомати (статии [1-4] от списъка на публикации за участие в конкурса); (2) Обработка на естествен език и разпознаване на реч (статии [5-8] и патент [18]); (3) Приближено търсене, корекция и нормализация на текстове (статии [9-17]). Както отбелязва кандидатът, това деление е условно, защото много от резултатите попадат в повече от едно направление.

**(1) Теория на крайните автомати.** Теорията на крайните автомати и свързаната с нея теория на формалните езици са важни области на съвременните математика и теоретична информатика с многобройни приложения в други математически дисциплини (например в алгебрата, математическата логика и комбинаториката) и в други области на знанието, включително с приложения за решаване на практически задачи. В частност, тези теории се прилагат широко в лингвистиката за решаване на трудни задачи в обработката на текстове и естествени езици, обработката на реч, търсене на шаблони, приближено търсене, корекции на текстове и други. Статия [1] от 2007 г. се разглежда задачата за пренаписване на текст по даден речник за презапис. Авторите предлагат ефективен метод за конструиране на подпоследователен преобразувател, който при даден текст на входа дава изходен текст за време, зависещо линейно от дължината на входния и изходния текст. Методът има две предимства – оптимална ефективност и стандартни възможности за комбиниране с други преобразуватели за решаване на по-сложни задачи само с една стъпка. Останалите три статии са от последните две години. В статия [4] се усъвършенства конструкцията в [1], с което се разширяват възможностите за приложение, допълнително се подобрява ефективността и позволява работа с огромни по обем речници. В работата е докладван експеримент с построяването на преобразувател, който заменя ключовите думи от Wikipedia с връзка към съответната страница. Статия [2] е посветена на нова конструкция на бимашини и е продължение на предишна работа, която не е включена за участие в конкурса. Бимашините са клас от детерминирани машини с краен брой състояния, които представят класа на регулярните функции между думи. Съществува стандартна конструкция на бимашина, започвайки от функционален преобразувател. В предходни изследвания на кандидата се предлага нова конструкция, която превръща преобразувателя директно в бимашина. Сега тази конструкция е усъвършенствана, като се налага да се преодолеят редица трудности, включително от алгебричен характер. В частност, за

нуждите на конструкцията се въвежда клас от моноиди, който обхваща свободните моноиди, групите и други алгебрични обекти. Новата конструкция е близка до оптималната. В работа [3] се предлага оригинална методология за конструиране на вероятностни преобразуватели. Достойнство на метода е, че в процеса на работа коректно се запазват вероятностните величини, а при неуспех при изпълнението на някоя стъпка се запазва и детерминизма на преобразувателя. Представени са конкретни конструкции и експериментални резултати при разпознаване на реч.

**(2) Обработка на естествен език и разпознаване на реч.** Статии [5, 7, 8] са посветени на конкретни реализации на системи за обработка на естествен език и разпознаването на реч. Разработките са дело на колектив, ръководен от доц. дн Стоян Михов. В [5] е описана реализацията на първата система за разпознаване на непрекъсната българска реч при голям речник за целите на въвеждане на юридически документи. Експериментите показват, че допуснатите грешки са <12% за юридически текстове и <16% за общи текстове. В [7] са описани принципите, положени в основата на корпуса на реч BulPhonC. За направата на корпуса са използвани записи на 147 диктори с около 22 000 изкази с обща продължителност около 32 часа. Приложена към системата от [5], разработката дава грешка под 7% на ниво думи за юридически текстове. Оказва се, че корпусът BulPhonC е твърде малък за нуждите на машинното обучение на дълбоки невронни мрежи. В [8] е представен корпусът от реч BG-PARLAMA, който преодолява този недостатък. Записите на корпуса са извлечени от пленарните сесии на Народното събрание за период от 10 години. Впечатляващо е количеството на използваните данни – 250 часа записи на 572 диктори. Ще отбележим, че са използвани технологии, чрез които дейностите се извършват почти напълно автоматизирано при минимални човешки ресурси. Статията [6] е от съвсем друг, по-теоретичен характер. Използва се теорията на детерминирани крайни автомати. В резултат, освен намирането на няколко най-добри кандидата за правилна транскрипция на базата на извършените експерименти, се дават и значително по-голям брой кандидати без използване на допълнително време. Направените експерименти показват под 4% грешка. В патента [18] е регистрирана методология за анализ на влиянието на отделните субекти в дадено медийно събитие. Използва се теория на графите, като се прилагат разнообразни техники за анализ на графи. Това води до практически методи за оценка и подобряване на ефективността на медийната комуникация.

**(3) Приближено търсене, корекция и нормализация на текстове.** Статиите от тази група са публикувани в периода 2006-2014 г., а втората версия на препринта [14] е от 2015 г. Изследванията са в три поднаправления: алгоритми за приближено търсене (статии [13-15]), корекция на текстове (статии [9-12]) и нормализация и модернизация на исторически текстове (статии [16-17]).

**3.1. Алгоритми за приближено търсене.** В статия [13] са представени приложения в обработката на естествен език, като търсенето на близост се разглежда като специален вид изчисление. В статия [14] е представен нов ефективен метод за приближено търсене в електронен речник, като се постига ефективност, по-добра от тази на съществуващите техники и се предлагат интересни възможности за търсене в колекции от дълги низове (например изречения). В статия [15] се описва системата за търсене на близост WallBreaker, с която е участвано в състезание, организирано от Хумболтовия университет в Берлин. Преодолени са редица принципни и технически

трудности. Например, налага се създаване на ново представяне на речника, заложен в системата.

**3.2. Корекция на текстове.** В статия [9] се анализират ортографичните (правописните) грешки и тяхната корекция в Интернет и се прави статистика на грешките от различен тип. Представена е методология за съставяне на речници с грешни изписвания и с правилно изписани думи. Методите могат да намерят приложение в корпусната лингвистика. В статии [10-11] се дава вариант на разстоянието на Левенщайн, който може да бъде ефективно реализиран чрез универсален автомат на Левенщайн. Докато методите от статия [10] изискват наличието на корпус от документи с ортографични грешки и ръчно коригирани техни копия, в [11] този проблем се решава чрез създаването на методология за автоматично създаване на речници с грешни изписвания. В статия [9] се показва, че точността на корекцията се подобрява при добавяне на честотата на биграмите на думите от интернет корпусите като нов фактор за оценка на близостта и униграмната честота на думите. Експериментите показват, че езиковите модели от интернет корпуси дават по-добри резултати от тези, получени от стандартните корпуси.

**3.3. Нормализация и модернизация на исторически текстове.** В статии [16-17] в рамките на проекта REBELS е разработен нов подход, който е продължение на оригиналната методология за нормализация и модернизация на исторически текстове, създадена в проекта CULTURA. Отново се използват техники от теорията на крайните автомати. Приведени са експериментални резултати, които са по-добри в сравнение на тези, получени със съществуващи системи за нормализация на английски текстове и за машинен превод.

В заключение на коментарите си по научните и научно-приложните приноси на кандидата ще отбележа, че кандидатът е запознат много добре с основните задачи в областта, със съществуващите теоретични и практически средства за тяхното решаване и използва богат арсенал от методи. Достоверността на аргументите и на резултатите от направените експерименти не буди съмнение. Не съм забелязал и съществени неточности.

Авторската справка и резюметата на представените трудове правилно отразяват основните приноси на трудовете, представени за участие в конкурса.

6. **Значимост на приносите за науката и практиката.** Получените резултати в научно-изследователските и научно-приложните работи на кандидата са интересни и съдържателни. Резултатите и създадените методи могат директно да се използват, а някои от тях вече се използват успешно в компютърната лингвистика.
7. **Критични бележки и препоръки.** Вече отбелязах някои пожелания относно изготвянето на документацията по конкурса. Допълнително ще препоръчам кандидатът да се стреми да публикува в още по-авторитетни издания, особено от наукометрична гледна точка. Това би повишило както неговия международен авторитет, така и би донесло непосредствена полза (научна и финансова) за ИИКТ и за БАН като цяло. Аз съм убеден, че представените материали са напълно достатъчни за покриване на формалните и научните критерии за академичната длъжност професор, но смятам, че като доктор на науките, кандидатът трябва да защити „честта на пагона“ и да представи за участие в конкурса по-авторитетна (по отношение на цитиранията и наукометрията) колекция от свои резултати. Считам, че моите забележки имат доброжелателен характер.

8. **Лични впечатления като член на Научното жури.** Познавам доц. дн Стоян Михов главно по работи, както и по работи на неговия ученик доц. д-р Стефан Герджиков. Имам много добри впечатления от него като учен и научен ръководител и високо го ценя.

### ЗАКЛЮЧЕНИЕ

В представените научни трудове доц. дн Стоян Милков Михов е получил интересни резултати в актуални области и има съществени приноси към теорията и нейните приложения. Повечето от резултатите вече са използвани или могат да бъдат използвани при подобен род изследвания от други автори. Те имат и практическо приложение в редица важни области на информатиката и лингвистиката. Голяма част от резултатите са публикувани в авторитетни издания. Към това ще добавя неговата успешна дейност като лектор и научен ръководител, както и успешното му ръководство или участие във важни научни и научно-приложни проекти. Въпреки критичните забележки, които направих в становището си, имам всички основания убедено да предложа доц. дн Стоян Милков Михов да заеме академичната длъжност „професор” в област на висше образование 4. Природни науки, математика и информатика, професионално направление 4.6. Информатика и компютърни науки, специалност: Информатика.

София, 27 септември 2021 г.

Подпис:

**NOT FOR  
PUBLIC RELEASE**  
(акад. д-р М. Н. В. Дунчев)