



РЕЦЕНЗИЯ

по конкурс за професор с единствен кандидат
доц. Стоян Милков Михов, д.н.
Научна област: **4. Природни науки, математика и информатика**
Професионално направление: **4.6. Информатика и компютърни науки**
Обявен в "Държавен вестник" бр. 45/28.05.2021 г.

Настоящият конкурс за професор за нуждите на Института по информационни комуникационни технологии на БАН е обявен в бр. 45/28.05.2021 г. на "Държавен вестник". Документи е подал единствен кандидат:

доц. Стоян Милков Михов д.н.

Не забелязах допуснати нарушения по процедурата на конкурса. Ще анализирам последователно научната и педагогическата дейност на кандидата според материалите, представени за участие в конкурса.

1. Данни за кандидата

Стоян Михов е роден на 9. април 1968 г. в гр. София. През 1993 г. завършва Факултета по математика и информатика на СУ "Св. Кл. Охридски". с магистърска теза "Унификация на корегулярни множества". През 2000 г. защитава докторска дисертация в Института по информационни и комуникационни технологии на БАН с дисертационен труд на тема "Минимални ациклични автомати: конструкции, алгоритми, приложения" с научен ръководител проф. д-мн Димитър Скордев.

През годините от 1995 до 2006 Стоян Михов е последователно програмист, асистент и главен асистент в ИИКТ-БАН. През 2006 г. Стоян Михов е избран за доцент отново в ИИКТ, където работи и досега. През 2020 г. Стоян Михов защитава дисертационен труд за придобиване на научната степен доктор на науките на тема: "Крайни автомати, преобразуватели и бимашини – алгоритмични конструкции и имплементации".

От 2003 г. Стоян Михов работи и като хоноруван преподавател във Факултета по математика и информатика на СУ "Св. Климент Охридски", както и като изследовател в Рила Сълзушънс и Коммек ЕООД.

2. Описание на научните трудове

Кандидатът е представил за участие в конкурса хабилитационен труд, 17 научни статии, глава от книга и един патент. Научните му изследвания отговарят напълно на темата на конкурса и могат да бъдат отнесени към областта на теоретичната информатика – теория на крайните автомати, обработка на естествен език, разпознаване на реч, корекция и нормализация на текстове.

Като хабилитационен труд е представена монографията на кандидата и Клаус Шулц "Finite State Techniques: Automata, Transducers and Vimachines", издадена от престижното издателство Cambridge University Press. Тя е посветена на описание на методите, използващи машини с краен брой състояния, които са сред най-ефективните средства за анализ на текстови и символни данни. Тези методи дават решение на огромен брой практически задачи в изчислителната лингвистика и компютърните науки. Монографията покрива напълно тази област, започвайки на ниво изграждане на понятията и стигайки до по-сложни теми и приложения. Най-важните технологии са представени с нужната тематическа строгост. Специално внимание е отделено на възможностите за упростиране, трансформиране и комбиниране на устройства с краен брой състояния. Всички алгоритми са представени с доказателства за коректност и изпълним код на въведения за тази цел език C(M), който е фокусиран върху прозрачността на стъпките и простотата на кода. По мое мнение монографията напълно отговаря на очакванията от един хабилитационен труд и може да бъде приета за такъв.

Пет от статиите са публикувани в престижни научни списания, като четири от тях са с импакт-фактор и пет са със SJR. Единадесет работи са в сборници с доклади от конференции и симпозиуми като четири са в Lecture Notes in Computer Science. Една от работите е препринт, който е достъпен в arXiv. Статиите са отпечатани в следните издания:

- Natural Language Engineering - 1; (IF 1.065)
- Theoretical Computer Science - 1; (IF 1.231)
- Studies in Computational Intelligence - 1 (IF 1.052)
- Computational Linguistics - 1 (IF 1.800)
- Journal of Automata Languages and Combinatorics - 1 (SJR 0.325)
- Lecture Notes in Computer Science - 4; (SJR 0.293, 0.338, 0.427, 0.249)
- Proc. of the Int. Conf. Recent Advances in Natural Language Processing - 1
- Proc. of the 10th Int. Conf. on Language Resources and Evaluation - 1
- Proc. of the Int. Conf. Document Analysis and Recognition - 2

- Proc. of IJCAI 2007 Workshop on Analytics for Noisy Unstructured Text Data - 1
- ACM International Conference Proceeding Series - 1
- Proceedings - 11th IAPR International Workshop on Document Analysis Systems, DAS 2014 - 1
- препринт arXiv - 1

От представените работи пет статии са с един, седем статии са с двама, три – с трима, една с пет и една – с шест съавтори. Приемам участието на кандидата в съвместните работи за равностойно. За това свидетелства и приложената препоръка от проф. Клаус Шулц, който е съавтор в 10 от статиите, представени за участие в конкурса, както и в хабилитационния труд.

Всички работи са написани след представянето на дисертационния труд за присъждане на степента “доктор” и не са използвани в процедурите за доцент и доктор на науките.

Кандидатът е класифицирал работите, представени за участие в конкурса в няколко направления: теория на крайните автомати; обработка на естествен език и разпознаване на реч: приближено търсене, корекция и нормализация на текстове.

(1) Теория на крайните автомати.

Към тази група се отнасят работи [1–4]. В работа [1] е представен ефективен метод, който по зададен речник за презапис конструира подпоследователен преобразувател, който приема входен текст и извежда пре-записан резултат по стратегията “най-ляво, най-дълго срещане”. Някои недостатъци на конструкцията от [1] са отстранени в [4]. В работа [3] е представена нова методология, която позволява композирането на вероятностни преобразуватели с преходи при неуспех с условно вероятностни преобразуватели. Описани са конкретни конструкции за създаване на преобразувател за автоматично разпознаване на реч. В работа [2] е представен принцип за конструиране на бимашини, при който стъпките на бимашината отчитат алтернативните паралелни пътища на преобразувателя. Представен е клас от функционални преобразуватели, работещи в реално време, с $n+2$ състояния, за които горният принцип води до 2^n+n+3 състояния. Демонстрирано е, че сложността по памет на конструкцията е близка до оптималната.

(2) Обработка на естествен език и разпознаване на реч.

Тук могат да бъдат отнесени работи [5–8], които са посветени на обработката на естествен език и машинното разпознаване на реч. В статия

[5] е описана реализация на система за разпознаване на българска реч при наличие на голям речник. Целта е разработване на система за автоматично разпознаване на българска реч за въвеждането на юридически документи. Един от приносите тук е разработването на фонетична система и специфична конструкция за биграмен езиков модел при използване на голям речник. В работа [7] са описани принципите при създаване на специален корпус на реч, наречен BulPhonC, който е използван за трениране на системата от [5]. С това е намалена грешката при юридически текстове под тази, докладвана в [5].

Статия [6] е посветена на развитие на метод за генериране на първите n най-добри хипотези при разпознаване на реч. Той се свежда до конструиране на детерминиран краен автомат, който в рамките на дадено време представя по-голям брой хипотези. В [8] е представена методология, използвана за създаване на нов корпус от българска реч, подходящ за обучение и оценка на модерни системи за разпознаване на реч. Той е получен от записите на пленарните сесии на българското Народно събрание.

(3) Приближено търсене, корекция и нормализация на текстове.

Към това направление спадат статии [9–17]. Много задачи с огромна практическа значимост могат да бъдат формулирани като задачи за приближено търсене. В работа [9] - методика за коригиране на орфографични грешки в Интернет-текстове. Представена е методология за автоматично създаване на речници с грешни изписвания, както и статистика на грешките от определен тип. Работа [10] е представен метод за извличане на мярка за близост, която е по-релевантна от класическото разстояние на Левенштейн. По-нататък тази мярка е използвана за конструиране на абстрактен автомат на Левенштейн. Направените експерименти показват, че разработените методи са с висока ефективност. Статия [11] е посветена на методология, аналогична на тази от [10], но която работи без наличието на корпус от документи с орфографични грешки и техни коригирани копия. В [12] е описан биграмен езиков модел за ранкиране на кандидатите за корекция от предните работи, основан на вериги на Марков. В работа [13] е изследвана задачата за нормализация на исторически текст, разгледана като задача за приближено търсене. Тук е предложено решение на базата на специален вид изчисление, разпространено в компютърната лингвистика. В работа [14] е разработен ефективен алгоритъм за приближено търсене, а конкретна негова имплементация на ANSI C се съдържа в [15]. В работа [16] е представена нова методика за извличане на вариации в правописа на исторически текстове, а в [17] е реализирана цялостна

система за автоматично нормализиране на исторически текстове.

В приложения патент е регистрирана методология за анализ на влиянието на отделните субекти в дадено медийно покритие. Тя се основава на създаването на специален граф с върхове – имена на журналисти, експерти, организации и пр. и ребра – специални релации като “цитира”, “отразява”, “говори за” и др. Така полученият граф се анализира, от което се получава обективна оценка за влиянието на отделните субекти. Продуктът е инструмент за подпомагане на организации, за измерване, анализ и планиране на медийната политика.

3. Научни приноси

По мое мнение по-важните приноси на кандидата се свеждат до следното:

- (1) Конструирани са подпоследователен преобразувател за презапис на текст, както и негови различни подобрения.
- (2) Създаден е ефективен алгоритъм за приближено търсене.
- (3) Създадени са вероятностни модели чрез подпоследователни преобразуватели с преходи при неуспех.
- (4) Реализирани са системи за разпознаване на българска реч.
- (5) Създадена е методология за автоматично създаване на корпуси от българска реч за целите на машинното обучение на системи за разпознаване на естествена реч.
- (6) Реализирана е система за модернизация на исторически текстове.

4. Преподавателска дейност

Макар преподавателската дейност да не е водеща в институтите на БАН, кандидатът има около двадесетгодишен преподавателски опит във водене на лекции и упражнения в Софийския Университет. Под негово ръководство докторска степен са получили двама докторанти – един в ИИКТ и един във ФМИ на СУ. Под негово ръководство защитени и 10 дипломни работи.

5. Проектна дейност, участия в конференции и др.

В периода след 2000 г. кандидатът е ръководител и участник в много национални и международни проекти, от които следва да се изтъкнат няколко проекта по Седма Рамкова Програма, както и Националната научна програма “Електронно здравеопазване в България” (е-Здраве).

6. Числови показатели

Съгласно представените материали работите на доц. Стоян Михов могат да бъдат класифицирани както следва:

- научни списания с IF:	4
- сборници с доклади от конференции със SJR :	5
- сборници с доклади от конференции без SJR :	7
- патенти:	1
- препринти:	1

Общият импакт-фактор на представените статии е 5.148, а индексът SJR е 1.632. За участие в конкурса кандидатът е представил 213 цитирания. Те се отнасят само до шест работи на кандидата като една от работите му е цитирана 93 пъти. Това е убедително доказателство, че изследванията на кандидата са значими и се приемат добре от научната общност.

Приемам без забелжки справката за регистриране в НАЦИД, представена от кандидата, от която е видимо и изпълнението на минималните национални изисквания. Става ясно, че кандидатът не просто удовлетворява, но и далеч надхвърля както минималните законови, така и специфичните изисквания от Правилника на ИИКТ-БАН, за заемане на длъжността "професор".

7. Критични бележки

Нямам критични бележки по същество.

8. Лични впечатления

Познавам лично кандидата от около 5 години. Присъствал съм на негови доклади, изнасяни на наши научни форуми. Впечатленията ми са, че той е сериозен изследовател със задълбочени познания в областта на информатиката, който се радва на добро име в научната общност у нас и щ чужбина. За мен е извън всяко съмнение, че той удовлетворява изискванията за заемане на длъжността "професор" на Института по информационни и комуникационни технологии на БАН.

9. Оценка на кандидата

Считам, че в своята научно-изследователска работа доц. Стоян Михов е получил значими научни резултати, които съответстват на съвременните постижения и представляват оригинален принос в математиката и информатиката. С работите си кандидатът показва задълбочени теоретични познания в областта на теоретичната информатика. Заедно с това той взема активно участие в проектната дейност, както и в живота на математическата общност. Горезиложеното ми дава основание да дам **положителна** оценка

на кандидатурата на доц. Стоян Милков Михов в конкурса за “професор” в професионално направление: 4.6. Информатика за нуждите на Института по информационни и комуникационни технологии на БАН.

Заклучение

Всичко гореизложено ми да основание убедено да препоръчам кандидатурата на доц. **Стоян Милков Михов, д.н.**, за професор на Института по информационни и комуникационни технологии на БАН в научна област: 4. Природни науки, математика и информатика, професионално направление: 4.6. Информатика и компютърни науки.

София, 07.09.2021 г.



(проф. д.м.н. Иван Ланджев)