
УНИВЕРСИТЕТ ПО БИБЛИОТЕКОЗНАНИЕ
И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ

РЕЦЕНЗИЯ

от проф. д.ик.н. Владимир Симов Йоцов – УниБИТ, София

на научните трудове
на

доц. д-р Владимир Борисов Периклиев,
представени за участие в конкурс за заемане на
академичната длъжност „ПРОФЕСОР“
в направление „Информатика“

УниБИТ, София - 2012 г.

I. Кандидати за конкурса

Относно обявения конкурс в ДВ от 2012 г. за заемане на академичната длъжност „професор“ по образователното направление „Информатика“ подадени са материали от един кандидат, доц. д-р Периклиев, чиито трудове се оценяват в тази рецензия.

II. Обща характеристика на трудовете

Кандидатът доц. д-р Периклиев има висше филологическо образование, английска филология. Той е един от изявените специалисти в областта на компютърната лингвистика и има голям практически опит в областите на статистиката и интелигентна обработка на данни за целите на лингвистиката и.

Дисертационния труд на доц. д-р Периклиев също е по темата на конкурса, но не се разглежда в настоящата рецензия. От представените по конкурса 32 научни труда на доц. д-р Периклиев, двете монографии и 25 доклада или статии се приемат за рецензиране в пълен обем. Пет от изданията (под номера 17, 18, 28, 29 и 30) не се разглеждат, защото са твърде кратки по обем, под 4 страници, и са публикувани преди написването на дисертацията. Рецензираните трудове с общ обем над 700 страници включват: две монографии в чужбина, 9 научни доклади от международни конференции в чужбина и 16 статии в чужди научни списания, от които 6 са с ИМПАКТ фактор, а почти всички останали са категоризирани според European Reference Index for the Humanities (ERIH 2011). Четири от докладите и четири статии са написани в съавторство, за което не е приложен разделителен протокол, подписан от всички съавтори. Дори ако те не се разглеждат, в комплекта остават повече от един монографичен труд и 15 публикации и изискванията на гл. 2, чл. 2, точка 5 на Правилника на ИМИ-БАН са изпълнени. От представените публикации поне 10 са в списания или в сборници от международни конференции: има

включени материали от списания с ИМПАКТ фактор. В документацията открих впечатляващи коментари на учени с висок международен престиж. Авторът представи доказателства за поне 20 цитирания (основно в чужбина и всъщност са над 100) и по *точка 5* не са представени само данни за успешно защитили поне двама докторанти.

Извод: Анализът на представените материали ми позволява да заявя, че приносите са оригинално дело на автора с достатъчен личен принос и при публикации в съавторство, които са само 8 от рецензираните 27 и представляват над 30% от общият обем на публикациите. Всички представени изследвания имат ясно очертана научна тематика, в която доц. д-р Периклиев е водещ изследовател.

III. Приноси в научен, научно-приложен и приложен план.

Оценка на педагогическата дейност

Приносите в разглежданите 27 труда са научни и научно-приложни и отразяват посоченият по-горе практически опит на доц. д-р Периклиев.

Основните приноси са в областта на компютърна лингвистика (обработка на естествен език, нееднозначност, свободен словоред), изкуствен интелект (интелигентна обработка на данни, откритие на нови знания), обща лингвистика (еволюция и история на езика, миграции, лингвистична типология), лингвистика на отделни или на групи езици (български, английски, семейството кайнганг (Южна Америка) и аустронезийското семейство (Тихия океан)). Посочените области напълно покриват тематиката на конкурсната област, а приносите са достатъчни за започване на процедурата за разглеждане на документите.

Много от работите са посветени на изследването на нееднозначността (ambiguity), която е основен проблем пред системите за разбиране и машинен превод. Това предполага както нейното изучаване (в отделни езици и сравнително), така и предлагане на ефективни методи за

нейното разрешаване. В тази област бих препоръчал да се направят и сравнения със съвременните изследвания по онтолозиите: проблемите на нееднозначността на онтолозиите в различен контекст е един от централните в направлението семантичен уеб и др.

Словоредът е друга важна тематика в системите за обработка на езика. За разлика от английски език, който има фиксиран словоред, повечето други езици, и в частност български, имат различни степени на свобода на наредбата, което изисква създаването на подходящи формализми за тяхното изразяване. Авторът представи неоспорими доказателства за извършени експерименти и успешно приключени проекти по автоматичната проверка на словореда, като част от проектната дейност в тази насока е наградена от MOMH.

Една от пионерските работи в областта е публикувана през 1999 г. от автора в *The Journal of Comparative Social Science*, където машинно се откриват значимите универсали от роднински модели (kin term patterns) в база от над 500 езика.

В книгата „Profiling language families by their kin term patterns: A computational approach“ доц. д-р Периклиев акцентира внимание върху т. нар. „компонентен анализ“ на термините за родство, тематика владяла лингвистиката и социалната антропология в продължение на десетилетия и много актуална и днес. Разполагайки с роднинските термини на един език и техните „описания“ (напр. в бълг. чичо = брат на баща ∨ брат на майка), програмата KINSHIP генерира техния „смисъл“ (чичо = мъжки пол & +1 поколение & родство по неправа линия), така че всеки термин се разграничава от всички останали по поне един признак, като при това анализът е максимално икономичен (използва минимален брой глобални признаци и най-кратки дефиниции на термините). KINSHIP е открила компонентните анализи на няколко езика, като този за английски дава най-много нови резултати.

Доц. д-р Периклиев разработи UNIVAUTO (UNIVersalsAuthoringTOol), една от първите програми за научно откритие, която е в състояние да представи своите резултати във формата на научна статия. Модулът за откритие UNIV открива универсали (абсолютни или статистически) от различен логически вид (типично, дву- или повече съставни импликации) и проверява тяхната статистическа значимост чрез използване на пермутационен тест или на разпределението χ -квадрат. След това модулът за генерация AUTO сравнява откритията на системата с тези на другия агент и генерира цялостна научна статия, съдържаща заглавие, уводна част, открити от програмата универсали, а може да съдържа и сравнение с тези на другия агент и заключение. По този начин са направени съществени промени в известната система на Гринберг. Експериментите на автора са довели до създаване на над 50 фонологични универсали, включени в известни световни системи и бази. При описанието на системата UNIVAUTO бих препоръчал сравнения с американските програми за автоматично съставяне на „научни статии“, чиито резултати за съжаление са доста многобройни в съвременното Интернет общество.

Втората представена от автора монография е публикувана от известното издателство, базирано в Англия: Equinox. Тя е от 2010 г.: Pericliev, V. Machine-Aided Linguistic Discovery: An Introduction and Some Examples. Equinox: London & Oakville. 330 pp. В нея е представен дългогодишния опит на автора в областта на машинното откритие в лингвистиката. И тук се използват термини от лексикологията (компонентен анализ на термините за родство), фонологията (анализ по дистинктивни признаци; профилиране на отделни езици), от езиковата патология и др. В седма глава се формулира задачата за намиране на статистически значими прилики и се илюстрира с откритието на лингвистична връзка между езиците в източната част на Южна Америка и

на Океания. Предложената компютъризация на теоретичната лингвистика ни дава интересен материал за изследване и на по-нататъшното развитие в тази бързо развиваща се област.

Седем от публикациите са в областта на компютърните модели и програми за изучаване на еволюцията на езика. И тук са извършени многобройни софтуерни експерименти във връзка с посочената тема.

При доц. д-р Периклиев педагогическите, изследователските, внедрителските, организаторските и популяризаторските качества неразривно се съчетават и няма смисъл изкуствено да се разделят неговите научно-приложни от практическите и от педагогическите качества. При разглеждане на неговите трудове и на подадените материали не може да се изчисли в явен вид никакъв икономически ефект, защото те касаят област, в която не личният принос, а работа екип води до успех. Сигурен съм, че иновативните технологии и елементи на неговото научно творчество ще доведат до успех и до подобряване на компютърната лингвистика на национално и международно ниво. Неговият стил на изложение, неговата ерудиция и организираност при провеждане на изследвания, както и при реализация на проектната дейност ми позволяват да направя извода, че той има необходимите качества за професура в посоченото направление.

Като цяло, давам положителна оценка на научноизследователската дейност на кандидата. Неговият дълбок практически опит в областта на компютърната лингвистика в съчетание с проектната, главно софтуерно насочена дейност и с неоспорими научно-приложни приноси към областта ми позволяват да дам висока оценка на неговите качества. Неговите трудове напълно отговарят на съвременните изисквания за академични публикации.

IV. Критични бележки и препоръки

Бих искал да посоча следните забележки от технически характер, които не касаят пряко същината на представените трудове.

На определени места е необходимо да се уточнят техническите детайли на изследването, например, да се дадат повече характеристики на софтуерните системи. По този начин ние, читателите-рецензенти или не, бихме могли по-добре да оценим експерименталната част на изследването на доц. д-р Периклиев и да му помогнем на базата на нашият опит да направи още по-интересни проекти с още по-значителни международни резултати. Проектната дейност би трябвало да се увеличи, особено в областта на международните проекти и за обучение на млади учени.

На места аз не съм съгласен с означаване чрез „машинно откривателство“ или с други думи *knowledge discovery*, откриване на нови знания на чисто статистически приложения в областта на лингвистиката. Известно е, че статистическите платформи за интелигентното обработване на данни не допринасят за откриване на логическите закономерности и затова логическата част управлява програмния комплекс като цяло (вж. *data mining*, *deep data mining*, *web mining* и др). Ако авторът има свой собствен подход към компоновката на подобен род системи, то все едно той трябва да сравни параметрите на своите системи с многобройните най-добрит представители от посочените области. Това би допринесло за повишаване ефективността на неговите приложения и намаляване на ръчната работа за откриване, потвърждаване и корекция на новите резултати.

Накрая, но не на последно място, бих искал да отбележа, че би било добре да се споменати и съвременните интелигентните средства за лингвистични приложения (*text mining*) – едно направление, което е най-близо до авторските проекти и публикации.

Всички публикации са написани на английски език, предимно в реномирани чуждестранни издания. Бих препоръчал доц. д-р Периклиев да

публикува не само в реномирани списания по компютърна лингвистика, но и по интелигентни системи, защото по този начин той ще получи ценни отзиви и коментари от една по-широка аудитория със сродни интереси.

V. Заключение

Критичните забележки не омаловажават ценността на представеното дисертационно изследване. Те служат за подобряване на перспективните изследвания на автора. Направеният анализ ми позволява да дам **положителна оценка** на кандидата в конкурса доц. д-р Периклиев за избор на професор по конкурс в професионално направление „Информатика“. Препоръчвам на уважаемото Научно жури да се направи предложение за избор на доц. д-р Периклиев за заемане на академичната длъжност **„ПРОФЕСОР“** в посоченото професионално направление.

Член на журито:

Гр. София,

3 септември 2012 г.

/проф. д.ик.н. Владимир Симов Йоцов/