



Страница на
ГАЛЯ
МЛАДЕНОВА

От 4 до 9 август 51. годишна среща на Асоциацията по компютърна лингвистика ще събере в София над 800 специалисти от цял свят. На форума ще бъдат представени най-новите научни и приложни разработки в областта на компютърната лингвистика и обработката на естествения език. Сред тях са програми за извличане на информация от текст, автоматично отговаряне на въпроси и резюмиране на документи, разпознаването и генерирането на реч, автоматичен превод.

Програмата на събитието включва две конференции, 11 семинара, уъркшоп за студенти и докторанти, обучения и демонстрации. Ще бъдат представени над 200 доклада. Годишната среща в София ще се проведе под егидата на президента Росен Плевнелиев и се организира от Секцията по компютърна лингвистика при Института за български език на БАН.

Проф. Хинрих Шутце, ръководител на Центъра за обработка на информация и език към Университета в Мюнхен:

И малките езици имат бъдеще

Компютърната лингвистика е навсякъде – в търсачките, програмите за превод и мобилните приложения

– Проф. Шутце, какви ще бъдат акцентите на предстоящата 51. годишна среща на Асоциацията за компютърна лингвистика в София?

– Един от акцентите на предстоящия форум е автоматичният превод от един език на друг. Работата ни в тази област има голямо теоретично и практическо значение. Ще обърнем специално внимание на автоматичното определяне на отношението на автора на текста. То има голямо комерсиално приложение в интернет-търсачките и контент-анализа, като отзиви за филми, за ресторанти. В програмата са включени и семинари, които са особено подходящи за студентите и докторантите. Позволяват им да се запознаят с това, какво представлява компютърната лингвистика, какви са най-новите изследвания и да се срещнат със специалисти в тази област. Ще имаме уъркшоп и обучения само за студенти.

– Кои са най-популярните приложения на компютърната лингвистика в живота на хората?

– На първо място бих поставил компютърните търсачки. В началото те бяха много слаби – просто търсеха съвпадения с думите, въведени от потребителите. Днес те са на много по-високо ниво, защото голяма част от тях всъщност са компютърна лингвистика. Друг пример са приложенията за мобилни телефони, които приемат гласови команди.

Всекидневно много хора използват и платформи за автоматичен превод. Признавам, те все още не работят перфектно. На

Проф. Хинрих Шутце е един от най-изявените учени в световната компютърна лингвистика. Той е ръководител на Центъра за обработка на информация и език към Университета в Мюнхен и председател на организационния комитет на 51. годишна среща на Асоциацията за компютърна лингвистика. Проф. Шутце е съавтор на едни от най-влиятелните книги за обработка на естествения език – „Основи на статистическата обработка на естествения език“, „Въведение в извличането на информация“. Има множество разработки в областта на търсенето и извличането на информация, машинното обучение и създаването на езикови модели.

този етап тяхната цел е не точен превод, а да покажат смисъла на текста. Но нека припомним, че те се появиха едва през последните десетина години и качеството им тепърва ще се подобрява.

– Може ли компютърната лингвистика да помогне на хора в неравностойно положение?

– Имаме много добри разработки на софтуер, който разпознава глас. Той може да бъде изключително полезен за хора с увредено зрение. На срещата в София ще имаме и уъркшоп за компютърна система, която да помага на напълно парализирани хора. Тя ще приема команди от движенията на околото или от звуци.

– А каква е Вашата оценка за българската школа по компютърна лингвистика?

– Нивото на работата в региона е много високо, особено в Българската академия на науките. Смятам, че страната трябва да инвестира повече в компютърната лингвистика. Не е тайна, че финансирането в България е по-ниско в сравнение с останалите

Ще продължаваме да усъвършенстваме автоматичния превод, но винаги ще имаме нужда от преводачи.

държави. Голяма част от парите са по линия на европейски програми, което ви прави зависими. При забавяне на средствата няма достатъчно национално финансиране, което да поддържа изследванията. Избрахме София за домакин на конференцията, за да стимулираме развитието на компютърната лингвистика в региона. Надяваме се да спомогнем финансирането на изследванията в тази област да се повиши. Всеки език има свои специфични проблеми, затова българският трябва да бъде изследван повече.

– При голяма част от „малките“ езици, включително и българския, езиковите технологии се развиват по-бавно. Заплашени ли са те от „дигитална смърт“?

– Безспорно най-много изслед-

вания са съсредоточени върху английския език. За да могат останалите езици да навакнат, са нужни инвестиции. За изследванията е необходимо изграждане на ресурси като електронни речници и бази данни, а това е дълъг и скъп процес. Резултатите са осезаеми в дългосрочен план – за учените и за икономиката, така че няма как да го спестим. Лично аз не съм съгласен с песимизма на прогнозите за „дигитална смърт“ на малките езици. Вярвам, че ни очаква голям напредък в областта на базовите проучвания в компютърната лингвистика и ще можем да прилагаме по-гъвкаво ресурсите, с които разполагаме. В момента проблемът е, че при превод програмите не могат да се адаптират добре в контекст и не могат да научават нови неща. Ако успеем да разберем по-добре умението на хората да се адаптират, да разбират значението на езика в контекст и след това да го прилагат при други условия, можем да приложим методите в

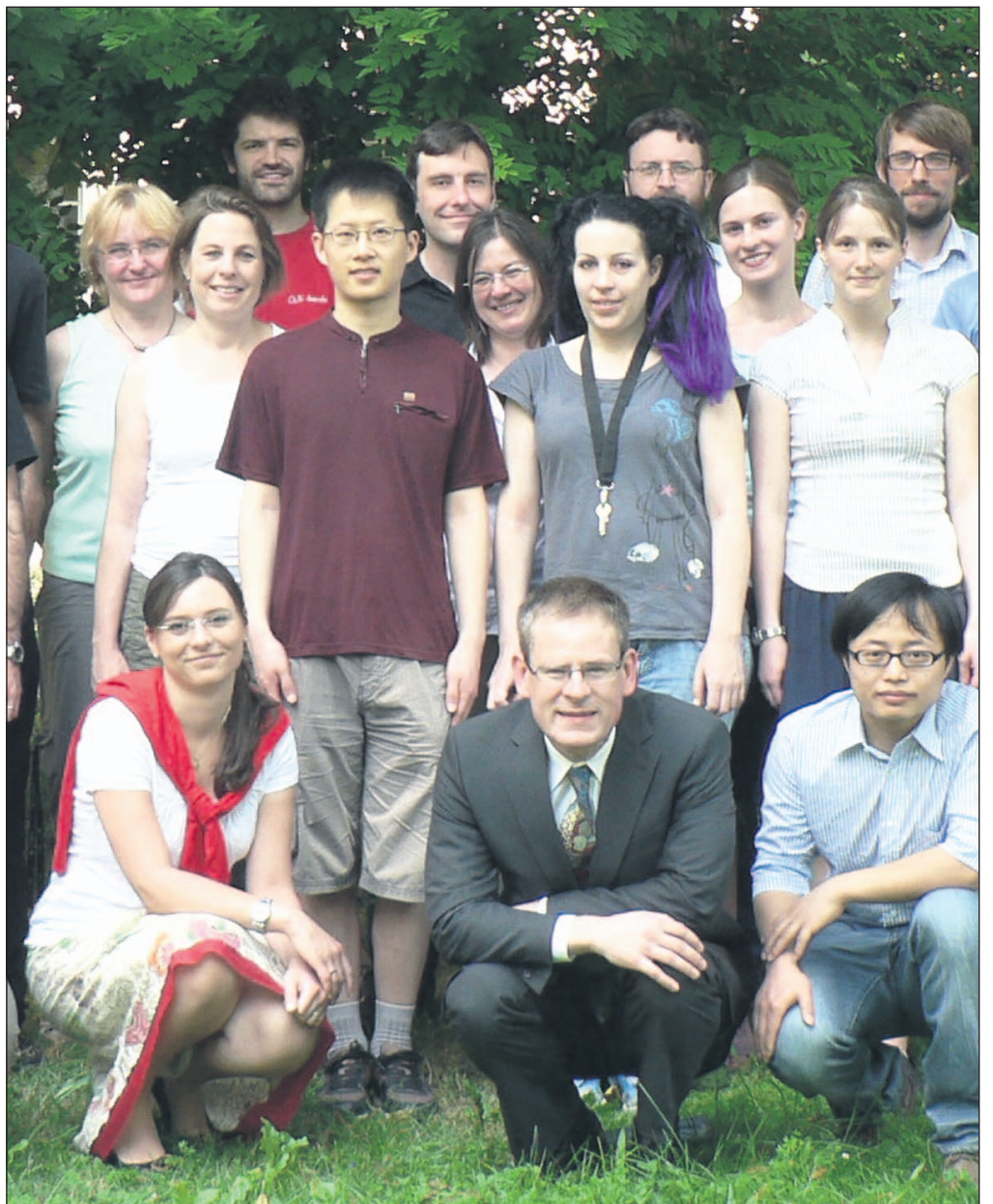
компютърната лингвистика. Това ще улесни работата ни за всички езици и ще спести много усилия и пари.

– Споменахте национално и европейско финансиране на изследванията. Има ли и подкрепа от страна на бизнеса?

– Обикновено бизнесът инвестира средства, когато те ще доведат до конкретни резултати в обозримото бъдеще. В това няма нищо лошо. Но има дългосрочни изследвания, чийто хоризонт е пет и повече години. В случая е необходимо национално или европейско финансиране, което да подсирира научното развитие. А когато се стигне до откритие, което може да намери комерсиално приложение, бизнесът подава ръка.

– Възможно ли е в бъдеще програмите за машинен превод да станат толкова добри, че да нямаме нужда от преводачи? Ще преодолеем ли езиковите бариери между хората?

– Не мисля. Бариерата не е само езикова, но и културна. Ако не познаваш културата, не можеш да общуваш с хората, които имат различен опит и познания. Това различие не може да се компенсира от машинния превод, така че винаги ще имаме нужда от преводачи. Разбира се, ще продължаваме да усъвършенстваме автоматичния превод. Убеден съм, че след време програмите ще превеждат отлично текстове на справочници или закони. Но винаги, когато са намесени културни взаимоотношения, машинният превод ще си остане ограничен. Книгите и стихотворенията винаги ще се превеждат от хора.



Професорът (в средата на първия ред) сред свои колеги от Университета в Мюнхен

Нивото на компютърната лингвистика в региона е много високо, особено в Българската академия на науките.