Български език 63 (2016), 4, 110-117

ХРОНИКА / CHRONICLE

ВТОРА МЕЖДУНАРОДНА КОНФЕРЕНЦИЯ „КОМПЮТЪРНАТА ЛИНГВИСТИКА В БЪЛГАРИЯ“ (CLIB-2016)

SECOND INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS IN BULGARIA (CLIB-2016)

На 9 септември 2016 г. в Представителството на Европейската комисия (Дом на Европа) в София се проведе второто издание на международната конференция „Компютърната лингвистика в България“ (*Computational Linguistics in Bulgaria – CLIB 2016*). Неин организатор е Секцията по компютърна лингвистика при Института за български език „Проф. Любомир Андрейчин“ (http://ibl.bas.bg) при Българската академия на науките в сътрудничество с Факултета по славянски филологии и Факултета по математика и информатика на Софийския университет „Св. Климент Охридски“. Конференцията се осъществи с подкрепата на Фонд „Научни изследвания“ (договор ДПМНФ 01/9–11.08.2016) и любезното домакинство на Представителството на Европейската комисия в София. Златен спонсор на конференцията беше компанията „Иденрикс“ АД, а бронзов спонсор – „Уеб моушън“ ООД.

Международната конференция „Компютърната лингвистика в България“ е форум, на който български учени, работещи в областта на компютърната лингвистика в България и в чужбина, имат възможност да обменят познания, опит, идеи и постижения с изследователи от други държави, работещи върху проблематика, свързана с компютърната обработка на българския език или други езици, както и с представители на научноизследователския сектор в индустрията. Конференцията има важно значение и за сплотяване на научната общност в областта на компютърната лингвистика, информационните и езиковите технологии. През 2016 г. конференцията премина при голям успех и събра 63-ма участници, включително шестима чуждестранни учени от Сърбия, Хърватия, Румъния и Норвегия, двама известни български учени, работещи в чужбина (Катар и САЩ), и единайсет представители на български компании в сферата на информационните технологии.

Конференцията беше открита от проф. Светла Коева, ръководител на Секцията по компютърна лингвистика и директор на Института за български език „Проф. Любомир Андрейчин“ при Българската академия на науките. В своите приветствени думи тя подчерта, че използването на математически и логически модели върху езикови данни има огромно приложение в съвременните информационни и езикови технологии.

Програмата включваше две сесии от доклади, всяка от тях придружена с пленарен доклад. Представените разработки предизвикаха интерес и доведоха до ползотворни дискусии. Сесията с постери даде възможност на авторите да обсъдят задълбочено работата си с другите участници. В програмата беше включена и технологична демонстрация от фирма „Идентрикс“ АД.

Проблематиката, разглеждана в представените на конференцията доклади, засягаше широк спектър от теоретични и приложни разработки в областта на компютърната лингвистика и езиковите технологии: лингвистична анотация, описание и анализ на различни езикови явления, разработване на езикови ресурси, автоматично извличане на информация, разработване на методи за електронно обучение, автоматично разпознаване на тролове в интернет общуването, различаване на близкородствени езици, извличане на цитати. Представени бяха изследвания със сериозна научна значимост, много от тях с практическо приложение в различни сфери, сред които превод, образование, социология, медийно дело. Тези актуални проблеми имат ключово значение за преодоляване на езиковите бариери и за ефективното управление на информационния поток.

На конференцията присъстваха изтъкнати български учени – преподаватели в Софийския университет като проф. Радка Влахова, проф. Йовка Тишева, проф. Татяна Ангелова и др., изследователи от Института за български език при БАН – доц. Лучия Антонова, доц. Мариана Витанова, доц. Палмира Легурска и др., както и от Института за математика и информатика при БАН, Пловдивския университет и Техническия университет в София. Присъстваха студенти и млади учени от Факултета по славянски филологии и Факултета по математика и информатика на Софийския университет „Св. Климент Охридски“.

В пленарните сесии лекции изнесоха изтъкнати български учени от Катар и САЩ, които имат съществен принос за развитието на компютърната лингвистика в международен план.

Лекцията на д-р Преслав Наков от Катарския институт по компютърни изследвания беше посветена на автоматичното разпознаване на манипулиращи общественото мнение публикации във форуми на интернет издания и блогове. Манипулирането на потребителското мнение за продукти, фирми и политически събития в онлайн форуми и социални мрежи е често срещано явление и може да бъде проследено автоматично. Д-р Преслав Наков и неговият екип предлагат решение, което се основава на анализ на коментарите и мнението на потребителите, въз основа на които се извършва машинно обучение на автоматичен класификатор.

Сутрешната сесия от доклади предложи изследвания, основани на паралелни корпуси и създаване на бази от специфични езикови данни. Проф. Цветана Крстев от Белградския университет изнесе съвместен доклад с проф. Душко Виташ, Любомир Попович и Анжелка Жечевич върху автоматичното различаване на близки езици като сръбски и хърватски. Доц. Иван Держански и д-р Олена Сирук представиха анализ на експлетивното отрицание в изречения с „докато“ и „когато“ в българо-украински паралелни текстове. Фокусът на следващите доклади беше създаването на лингвистични ресурси за автоматична обработка на естествения език. Д-р Вергиника Барбу Митителу от Инс-

титута за изследване на изкуствения интелект към Румънската академия представи извличане на лингвистични данни от корпус със синтактични описания (съвместна разработка с д-р Елена Иримия), а д-р Светлозара Лесева говори за автоматичното идентифициране на конструкции с „леки“ глаголи в български (съвместна разработка с д-р Ивелина Стоянова и д-р Мария Тодорова).

В следобедната сесия проф. Драгомир Радев от Университета в Мичиган, САЩ, изнесе лекция на тема „Компютърна обработка на естествения език за целите на анализа на колективния дискурс“. Проф. Радев се включи на живо от САЩ чрез видео връзка. Лекцията представи резултатите от няколко метода за автоматично разпознаване на сходство и извличане на подобни думи и изречения, както и приложението им за решаване на различни задачи за компютърна обработка на езика. Изследването върху т.нар. колективен дискурс (колекции от текстове, произведени от голям брой потребители) обединява резултатите от три проекта. В тях автоматично се анализират коментарите на поредица от карикатури, отстранява се семантичната многозначност при изследване на корпус от кръстословици и автоматично се генерират резюмета на научни статии.

По време на следобедната сесия бяха представени доклади, свързани с компютърно подпомогнатото обучение. Д-р Даша Фаркаш, д-р Матеа Филко и проф. Марко Тадич от Загребския университет представиха HR4EU – обучителен портал за хърватски и приложението на езикови ресурси при компютърно подпомогнатото езиково обучение. Д-р Атанас Атанасов от Софийския университет представи системата за семантична анотация SynTags.

Последните два доклада бяха посветени на автоматичното извличане на информация от интернет. Цветомила Михайлова от Софийския университет представи съвместната си работа с проф. Иван Койчев, д-р Преслав Наков и д-р Ивелина Николова върху автоматичното извличане на информация от онлайн форуми за целите на изграждане на система за автоматично отговаряне на въпроси за български. Във видео представяне д-р Ивелина Стоянова от Института за български език говори за автоматично извличане на цитати от публикации в български медии (съвместна работа с Мартин Ялъмов и проф. Светла Коева).

В постерната сесията гостите и участниците имаха възможност да се запознаят с две изследвания, свързани със съставните лексикални единици – работата на Бистра Поповска и д-р Росица Декова върху прозодията на съставни лексикални единици в английски и български, и проекта на д-р Кешимир Шоят, д-р Даша Фаркаш и д-р Матеа Филко, свързан със съставните лексикални единици с главен конституент глагол в хърватски. Докторантът Тодор Лазаров представи работата си върху разпознаването на глаголни форми за целите на машинния превод от български на английски. В технологичната демонстрация Деян Пейчев от фирма „Идентрикс“ АД представи гъвкава инфраструктура за интегриране на задачите по обработка на данни и автоматичен семантичен анализ.

В заключителната част участниците благодариха на организаторите за отличната работна атмосфера както по време на конференцията, така и на съпътстващите я събития. Дискусиите бяха оценени като много ползотворни, тъй

като осигуриха условия за обмен на опит между лингвисти, компютърни лингвисти и представители на частни компании върху актуалните теоретични и приложни разработки. Бяха оценени усилията за установяване на сътрудничество между научните центрове в подкрепа на развитието на надеждни езиково обосновани технологични приложения. Не на последно място трябва да се отбележи дадената възможност за изява на млади български и чуждестранни учени.

За престижа на конференцията допринесоха интердисциплинарната ѝ тематика, актуалността на проблематиката и присъствието на изтъкнати гост-лектори, участници и представители на специализирани частни компании. Трябва да се отбележи и фактът, че подборът на докладите беше извършен чрез анонимно рецензиране и селекция, за да се осигури високото качество на представените изследователски постижения. Рецензентите бяха 28 учени с международен авторитет, работещи в 12 държави. Сборникът с доклади (публикуван преди началото на конференцията), програмата и снимковият материал са достъпни и електронно на страницата на конференцията (http://dcl.bas.bg/clib/), както и във фейсбук (https://www.facebook. com/CompLingInBulgaria/).

Събитието получи широк отзвук в медиите и беше отразено в календара със събития на Дома на Европа, който е част от страницата на Информационния център на Европейския съюз в България. Новините от конференцията можеха да се прочетат и на уеб страниците на Българската академия на науките, Института за български език, Секцията по компютърна лингвистика, на блога на група „Дейта сайънс съсайъти“ и във форум „Наука“. Събитието беше отразено в новинарските рубрики на Българската телеграфна агенция, както и на уеб страниците на Пловдивския университет и на проекта „Интегриране на електронни форми на обучение в образователния процес по български език“ на Софийския университет „Св. Климент Охридски“.

Третата международна конференция „Компютърната лингвистика в България“ ще се проведе през 2018 г., а домакин ще бъде отново град София. Планира се както широко българско участие на учени, работещи в наши и чужди изследователски центрове, така и чуждестранно участие с цел обмяна на опит и стимулиране на международното сътрудничество.

Мария Тодорова / Maria Todorova

✉ *Гл. ас. д-р Мария Тодорова*
Секция по компютърна лингвистика
Институт за български език „Проф. Л. Андрейчин“ при БАН
бул. „Шипченски проход“ 52, бл. 17, 1113 София, България
maria@dcl.bas.bg

✉ *Assist. Prof. Maria Todorova, PhD*
Department of Computational Linguistics
Institute for Bulgarian Language, Bulgarian Academy of Sciences
52 Shipchenski prohod blvd., bl. 17, 1113 Sofia, Bulgaria
maria@dcl.bas.bg