

ХРОНИКА

Представяне на мрежата META-NET и проекта CESAR в София

Екатерина Търпоманова (София)

Международен форум за представяне на проекта CESAR, част от европейската мрежа META-NET, се състоя на 2 май 2012 г. в хотел Шератон, София. България беше първата спирка от “роудшоутто” CESAR, което ще обиколи всички страни партньори в проекта. Форумът в София беше организиран от Секцията по компютърна лингвистика на Института за български език, БАН, и събра представители на европейските и българските институции, научни центрове и бизнес организации, създатели и потребители на езикови ресурси и системи за тяхното обработване.

META-NET (Multilingual Europe Technology Alliance) е европейска инициатива, която обединява 57 научноизследователски центъра от 33 страни, за да бъде подкрепено изграждането на многоезиково европейско информационно общество. Дейностите на META-NET се финансират от Европейската комисия. CESAR (CEntral and South-East EuropeAn Resources) е един от четирите големи проекта в рамките на META-NET, който има за цел да развие, стандартизира и свърже езикови ресурси и технологии за български, унгарски, полски, хърватски, сръбски и словашки, допринасяйки за създаването на отворена и широко достъпна езикова инфраструктура. От българска страна партньор в проекта е Секцията по компютърна лингвистика (СКЛ) на Института за български език, БАН.

В първия панел на форума беше направен преглед на развитието на езиковите технологии в Европа и бяха представени европейските инициативи за подкрепата и финансирането им. Роберто Ченчони, директор на Отдела за езикови технологии и машинен превод към Дирекцията за информационно общество и медии на ЕК, запозна участниците в събитието с досегашните постижения на Европа в създаването и обмена на езикови технологии и стратегиите за бъдещо развитие в областта – ясна изследователска програма, принос за икономиката и обществото, свързана и мобилизирана партньорска общност, подкрепа на национално ниво. Презентацията на проф. Ханс Ускорайт, координатор на META-NET, изясни ключовата роля на езиковите технологии за преодоляването на комуникационните бариери в многоезиковата Европа и за по-доброто управление и използване на натрупаните знания на човечеството. Проф. Ускорайт изтъкна нуждата от обединяване на усилията на европейската индустрия за езикови технологии и концентрирането върху европейските изисквания и възможности, вместо да се следват модели, определени от американските проучвания. Важна стъпка е създаването на мрежата за върхови постижения META-NET с три насоки за действие: META-VISION – създаване на влиятелна научна общност със споделена визия за научноизследователско развитие; META-SHARE – изграждане на инфраструктура за свободно споделяне и обмен на езикови ресурси и технологии; META-RESEARCH – развитие на интердисциплинарни изследвания. Координаторът на CESAR проф. Тамаш Варади представи ролята и приноса на проекта в европейската мрежа за езикови технологии, както и статистика за езиковите ресурси и програмите за тяхната обработка за всяка от държавите участнички, с акцент върху българското участие.

Докладите от втория панел бяха по конкретни теми, свързани с разработването, приложението и финансирането на езиковите технологии. Тотка Чернаева от Министерството на транспорта, информационните технологии и съобщенията представи програмата “Подкрепа за ИКТ политики”, насочена към ускоряване на развитието на устойчиво конкурентноспособно,

ХРОНИКА

иновативно и приобщаващо информационно общество, възможностите за кандидатстване и финансиране на проекти. Представители на частни фирми запознаха аудиторията с интересни приложения на лингвистичните технологии – в системи за публикации, в динамичното семантично публикуване в спортната уебстраница на Би Би Си, в европейски портал за машинен превод. Връзката на езиковите технологии с науката и образованието беше друга тематична област от втория панел - беше представен метод за модернизация на исторически текстове, чиято цел е интеграцията на културното наследство в съвременните информационни технологии, както и обучението по езикови технологии във Филологическия факултет на Пловдивския университет.

Форумът завърши с панелна дискуссия на тема “Българският език в дигиталната епоха”, на която бяха поставени въпроси за конкуренцията в разработването на езикови технологии, изтичането на мозъци от Европа към САЩ, оцеляването на местни фирми в условията на конкуренция с гиганти като Гугъл. Бяха обсъдени ресурсите, които се създават в европейските научни институти, и проблемите с финансирането и разпространението им. Беше изказано мнението, че силна страна на Европа е разработването на двуезикови ресурси, без да се използва английският като език посредник.

Паралелно с основната програма се проведе изложба и демонстрация на приложения на езиковите технологии, създадени от Института за български език по национални и международни проекти или в партньорство с частния сектор. Беше демонстрирана уеб базираната система за търсене в Българския национален корпус (достъпна на <http://search.dcl.bas.bg/>), създадена от СКЛ, която предлага търсене по наредени и ненаредени комбинации от думи и характеристики на думи и техни булеви комбинации, като разполага и с независим модул за търсене по регулярни изрази. Системата дава възможност на крайните потребители - лингвисти и неспециалисти, да провеждат собствени проучвания и да правят справки за различни цели в Българския национален корпус, който се състои от около 980 милиона думи и е в процес на постоянно разширяване и допълване.

Голям интерес предизвика демонстрацията на системата за проверка на правописа за български език (достъпна на <http://dcl.bas.bg/est/>) в трите ѝ версии - WinEst за операционна система Microsoft Windows XP и Microsoft Windows 7, MacEst за Mac Os и WebEst - уеб услуга за проверка на правописа, която може да бъде интегрирана в различни уеб приложения.

Беше направена и презентация на европейския проект ATLAS (<http://www.atlasproject.eu/>), в който участва СКЛ и който е насочен към създаването на система за публикуване и поддръжка на електронно съдържание на различни езици. Основните услуги, които се предлагат в рамките на проекта, са: автоматична категоризация на документи, автоматично генериране на резюмета на документи, автоматичен превод на резюметата, извличане на важни думи и изрази за даден документ, интелигентно търсене в многоезиково съдържание.