

ЗАПЛАХА ЗА ЕЗИЦИТЕ И ПРЕДИЗВИКАТЕЛСТВО ПРЕД ЕЗИКОВИТЕ ТЕХНОЛОГИИ

Свидетели сме на дигитална революция, която съществено промени комуникациите и обществото. Съвременното развитие на информационните и комуникационните технологии понякога се сравнява с въвеждането на печатарската преса от Гутенберг. Как тази аналогия ни помага да погледнем в бъдещето на европейското информационно общество и нашите езици?

Дигиталната революция е сравнима с изобретяването на печатарската преса от Гутенберг.

След откритието на Гутенберг истински прелом в комуникацията и обмена на знания са постижения от ранга на превода на библията на говорим език от Мартин Лутер. В следващите столетия се наблюдава сериозно развитие, за да се достигне до компютърна обработка на естествения език и бърз обмен на информация:

- стандартизацията на правописа и граматиката осигурява по-бързо разпространение на идеи и научни открития;
- появата на книжовните езици позволява на гражданите да общуват в рамките на определени (често политически) общности;

- езиковото обучение и преводът улесняват обмена на информация;
- създаването на редакторски и библиографски стандарти осигурява качество и лесен достъп до публикуваните материали;
- появата на различни медии – вестници, радио, телевизия, книги и други – отговаря на нуждата от различни типове комуникация.

В последните 20 години с помощта на информационните технологии много от тези процеси се автоматизират и улесняват:

- Софтуерът за електронно публикуване замества пишешите и печатарските машини.
- Microsoft PowerPoint (и други) заменя прожекционните апарати.
- Документите се изпращат и получават по-бързо по имейл, отколкото по факс.
- Skype (и други) позволява евтини телефонни разговори по интернет и виртуални срещи.
- Форматите за аудио и видео улесняват обмена на мултимедийно съдържание.
- Търсачките осигуряват достъп до уеб страници по ключови думи.
- Онлайн услугите като Google Translate осигуряват бърз приблизителен превод.

- Социалните медии като Facebook, Twitter и Google+ улесняват общуването, сътрудничеството и обмена на информация.

Макар подобни приложения да са полезни, те не могат да предложат достатъчно ефективна подкрепа за устойчиво развитие на многоезиковото европейско общество, в което информация и стоки се придвижват свободно.

2.1 ЕЗИКОВИТЕ ГРАНИЦИ – ПРЕЧКА ПРЕД ЕВРОПЕЙСКОТО ИНФОРМАЦИОННО ОБЩЕСТВО

Не може да се предвиди с точност как ще изглежда бъдещото информационно общество. Много вероятно е обаче поради революцията в комуникационните технологии хора, говорещи различни езици, да се свързват по между си. За да се осигури взаимното разбиране и достъп до съдържание, възниква необходимостта да се учат нови езици и да се създават нови езикови технологии. В глобалното икономическо и информационно пространство обменът на информация на различни езици е по-ефективен с помощта на новите средства за комуникация. Високата популярност на социалните мрежи (Wikipedia, Facebook, Twitter, YouTube, Google+) е само върхът на айсберга.

Глобалната икономика и информационно пространство сблъскват различни езици, хора и информация.

Гигабайтове текст могат да се прехвърлят от една точка на света до друга за няколко секунди, преди

да се установи, че текстът е на език, който не познаваме. Според скорошно проучване, изготвено по поръчка на Европейската комисия, 57% от потребителите на интернет в Европа купуват стоки и услуги на езици, които не са им родни (най-разпространен е английският, следван от френски, немски и испански). 55% от потребителите на интернет четат съдържание на чужди езици, докато само 35% използват чужд език, когато пишат имейли или публикуват съобщения и коментари онлайн [3]. Преди няколко години английският е бил *lingua franca* в интернет. Сега обаче ситуацията е коренно променена след рязкото нарастване на броя на публикациите на други езици (разпространени например в Азия и Средния изток). Дигиталното разединение, произтичащо от езиковите бариери, изненадващо не привлича достатъчно внимание в публичното пространство. И все пак именно то поставя неотложния въпрос: Кои европейски езици ще се развият в съвременното информационно общество, основано на знания, и кои са осъдени да изчезнат?

2.2 РИСКЪТ ЗА НАШИТЕ ЕЗИЦИ

Печатарската преса засилва обмена на информация в Европа, но същевременно води до отмиране на много европейски езици. Текстовете на регионални и малцинствени езици се публикуват рядко, а езици като корнуолски и далматински са ограничени до говоримата си форма. Възможно ли е интернет да има същото влияние?

Близо 80-те езика, които се говорят в Европа, са сред най-богатото и важно културно наследство на Стария континент и съществено компонент от уникалния му социален модел [4]. Популярни езици като английски или испански ще продължат да съществуват в условията на развиващия се дигитален пазар, но

много европейски езици може да се окажат с ограничена употреба в интернет. Това би намалило глобалните позиции на Европа и е в противоречие със стратегическите цели за равно участие на всеки европейски гражданин, независимо от езика, на който говори. Според доклад на ЮНЕСКО, засягащ многоезичието, езиците са важен посредник при осигуряването на фундаментални човешки права, като правото на изразяване на политически убеждения, правото на образование и участие в обществото [5].

Разнообразието от езици в Европа е една от най-богатите и важни културни ценности.

2.3 ЕЗИКОВИТЕ ТЕХНОЛОГИИ ПРЕДОСТАВЯТ ВЪЗМОЖНОСТИ

В миналото усилията за запазване на езиците са били насочени към езиковото обучение и превода. Според някои изчисления през 2008 г. европейският пазар за писмен и устен превод, локализация на софтуер и превод на уеб сайтове се равнява на 8,4 милиарда евро и се очаква да нараства с 10% на година [6]. Тези данни отразяват малък процент от днешните и бъдещите нужди при комуникацията на различни езици. Пълноценното използване на европейските езици в бъдеще може да се осигури с помощта на подходящи езикови технологии, аналогично на технологиите за осигуряване на енергия, транспорт и равен достъп за хората с увреждания.

Езиковите технологии (насочени към всички видове писмен и говорим текст) помагат на хората да си сътрудничат, да работят заедно, да споделят знание и да участват в социалния и политическия живот независимо от езиковите бариери и компютърните умения.

Тези технологии са вградени в сложни компютърни системи, за да ни помогнат да:

- търсим информация в интернет;
- проверяваме правописа и граматиката в текстообработваща програма;
- разглеждаме предложенията в онлайн магазин;
- слушаме гласови инструкции на навигационна система;
- превеждаме уеб страници онлайн.

Езиковите технологии се използват в някои основни приложения за компютърна обработка на езика, които участват в широка мрежа от продукти. Предназначението на Белите книги на META-NET е да се анализира дали основните езикови приложения са налице за всеки европейски език.

Европа се нуждае от стабилни и финансово достъпни езикови технологии за всички европейски езици.

За да запази предните си позиции в световните иновации, Европа се нуждае от стабилни и финансово достъпни езикови технологии, приложими за всички европейски езици и интегрирани в ключови софтуерни системи. В близко бъдеще работата на потребителя в интерактивна, мултимедийна и многоезикова среда ще е невъзможна без посредничеството на езиковите технологии.

2.4 ПЕРСПЕКТИВИ ПРЕД ЕЗИКОВИТЕ ТЕХНОЛОГИИ

В света на печатното слово технологичната революция започва с бързото възпроизвеждане на текстове с помощта на печатарска преса, задвижвана от механична сила. Трудната работа по проверката, редакцията, превода и обобщаването на съдържание обаче се

е извършвала от хора. Откритието на Едисън позволява да се правят звукови записи – но резултатът са просто аналогови копия.

Днес езиковите технологии улесняват и автоматизират процесите на превод, създаване на съдържание и управление на информация и знания за всеки един от европейските езици. Тези технологии могат да помогнат за развитието на интуитивни езиково базирани приложения за домашна електротехника, съоръжения, автомобили, компютри и роботи. Въпреки че прототипите за някои от тях все още са в начален стадий, подемот на изследванията и разработките през последните години разкрива множество реални възможности. Машинният превод например вече е достигнал сравнителна точност в някои специализирани области, а експерименталните приложения осигуряват възможности за управление на многоезикова информация и знание, както и за създаване на съдържание на много европейски езици.

Езиковите технологии помагат
да се преодолеят „неудобствата“
от езиковото многообразие.

Както при повечето технологии първите езикови приложения – гласов интерфейс и диалогови системи – са предназначени за тясно специализирани области и често са с ограничени възможности. Съществуват значителни пазарни ниши в образователната и развлекателната индустрия за интегриране на езикови технологии в образователно-развлекателни игри и програми, сайтове, посветени на културното наследство, софтуер за библиотеки и издателства, симулационни и тренировъчни системи. Мобилните услуги, приложенията за компютърно подпомогнато езиково или електронно обучение, както и софтуерът за разпознаване на плагиатство и за самооценка, са само някои от областите, в които езиковите тех-

нологии могат да имат съществено приложение. Нарастващата популярност на социалните мрежи като Twitter и Facebook предполага използването на езиковите технологии за мониторинг, обобщение на дискусии, анализ на мнения, разпознаване на емоции, както и за идентифициране на нарушени авторски права.

Езиковите технологии предлагат огромни възможности пред Европейския съюз. Те могат да помогнат при решаването на проблеми, свързани с многоезичието в Европа – например при използване на различни езици във фирми, организации и образователни институции. Жителите на Европа трябва да общуват без езикови граници в единния европейски пазар. Езиковите технологии могат да преодолее тези бариери, утвърждавайки едновременно с това равнопоставеността при използването на различните езици. В по-далечното бъдеще европейските (много)езикови технологии могат да бъдат пример за развитието на многоезиковите общества по света. Езиковите технологии могат да се възприемат като „помощни“ технологии, които помагат за преодоляване на проблемите, произтичащи от лингвистичното многообразие, и за подобряване на комуникацията между общности, говорещи различни езици.

Накрая, област на активни изследвания е използването на езиковите технологии при спасителни операции в райони, засегнати от природни бедствия, където точността на информацията може да е въпрос за живот и смърт. Бъдещите интелигентни роботи с възможности за комуникация на много езици ще имат потенциал да спасяват човешки живот.

2.5 ПРЕДИЗВИКАТЕЛСТВА ПРЕД ЕЗИКОВИТЕ ТЕХНОЛОГИИ

Макар че езиковите технологии отбелязват съществен напредък през последните години, технологичният прогрес и създаването на иновационни продукти значително изостават. Широко използваните технологии като програми за проверка на правопис и граматика, интегрирани в съвременните текстообработващи системи, обикновено са едноезикови и достъпни само за ограничен брой езици.

Технологичният прогрес трябва да се ускори.

Онлайн услугите за машинен превод, макар че са полезни за бързо генериране на приблизителен превод, са напълно неподходящи за цялостна и точна интерпретация. Поради комплексността на човешките езици моделирането им в компютърни програми и тестването на моделите в реалния свят е продължителна и скъпа дейност, която изисква трайни финансови ангажименти. Европа трябва да поддържа водещата си роля при посрещане на технологичните предизвикателства на многоезиковото общество посредством нови ефективни методи за ускорено развитие. Това включва както напредък в информационните технологии, така и използване на съвременни средства при програмирането.

2.6 КАК ХОРАТА И МАШИНИТЕ УЧАТ ЕЗИК?

За да разберем как компютрите обработват езиковата информация и защо се затрудняват с тази задача, нека хвърлим поглед към начина, по който хората научават родния си език и чужди езици. След това нак-

ратко ще бъде представен начинът, по който работят системите, базирани на езикови технологии.

Човек усвоява езика по два различни механизма. Детето научава родния си език, като слуша разговори между носители на езика – родители, братя, сестри и други членове на семейството. От около двегодишна възраст децата започват да формулират първите си думи и кратки фрази. Това е възможно, тъй като хората имат генетичната способност да възпроизведат, както и да осмислят и комбинират това, което са възприели.

Усвояването на втори език на по-късна възраст изисква повече усилия, тъй като детето обикновено не живее сред носители на чуждия език. В училище чуждите езици най-често се изучават с помощта на абстрактни правила, парадигми и примерни текстове, илюстриращи граматическата структура, речника и правописа на даден език. С възрастта усвояването на чужд език става все по-трудно.

Хората придобиват езикови умения по два различни начина: чрез научаване на примери и чрез научаване на езикови правила.

Двата основни типа компютърни системи, базирани на езикови технологии, „усвояват“ езика по подобен начин. Статистическите („базирани на данни“) подходи извличат лингвистични знания от големи корпуси от текстове. Докато за трениране на програма за проверка на правописа са достатъчни текстове само на един език, то за трениране на система за машинен превод са нужни паралелни текстове на два (или повече) езика. Алгоритмите за извличане на информация всъщност „научават“ моделите, по които думи, кратки фрази или цели изречения се превеждат от един на друг език.

Статистическите методи понякога обработват милиони изречения, като качеството на резултатите се увеличава заедно с големината на анализирания

текст. Това е една от причините, поради които компаниите, разработващи системи за търсене на информация, се стремят да съберат колкото може повече езикови данни. Програмите за проверка на правописа и услуги като Google Search и Google Translate разчитат на статистически подходи. Основното предимство на статистическите подходи е, че езиковите модели се извличат много бързо чрез последователни тренировъчни цикли, макар че качеството на тези методи може да варира в значителна степен.

Вторият подход при езиковите технологии и особено към машинния превод се основава на лингвистични правила. Експерти – лингвисти, компютърни лингвисти и програмисти, формулират граматически правила (правила за превод) и съставят речникови списъци (лексикони). Изграждането на системи, основани на правила, изисква много време и човешки труд. Някои от водещите системи за машинен превод, базирани на правила, се изграждат в продължение на повече от 20 години. Основното предимство на тези системи е, че експертите могат в по-голяма степен да контролират обработката на езиковата информация. Това осигурява възможност за бърза корекция на грешки в софтуера и съобра-

зяване с изискванията на потребителите, особено в случаите, когато системите, базирани на лингвистични правила, се използват за езиково обучение. Поради финансови ограничения подобни приложения се разработват само за по-разпространените езици. Тъй като силните и слабите страни на статистическите системи и системите, базирани на лингвистични правила, обикновено се допълват взаимно, в момента изследванията са ориентирани към хибридни подходи, които комбинират двата метода. Засега хибридните подходи се прилагат по-слабо в практическите разработки в сравнение с научните изследвания.

Както личи от тази глава, много от компютърните приложения, които се използват в момента в информационното общество, разчитат в значителна степен на езиковите технологии, особено в европейското икономическо и информационно пространство. Макар че през последните години езиковите технологии отбелязват значителен напредък, все още има сериозен потенциал за повишаване на качеството на техните приложения. Следващата глава ще хвърли светлина върху мястото на българския език в европейското информационно общество и върху състоянието на езиковите технологии за български.

The Bulgarian Language in the Digital Age

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 81 p. 24 illus. in color., Softcover

ISBN: 978-3-642-30167-4