

TUBA MÁRTA

3. EUROPEAN LANGUAGE RESOURCE COORDINATION (ELRC) MŰHELY-KONFERENCIA MAGYARORSZÁGON

2022. február 7-én harmadik alkalommal szervezett a European Language Resource Coordination (ELRC) konzorcium, a Nyelvtudományi Kutatóközpont (ELKH) és a Nyelvtervezési Kutatóközpont (Magyarságtudató Intézet) magyar és angol nyelvű műhely-konferenciát Magyarországon, ezúttal virtuálisan. Fenti szervezetek azért indítottak vitát a magyar nyelvtechnológia helyzetéről és kilátásairól, hogy segítsék a többnyelvű Európa digitális interakcióit. A digitalizáció robbanásszerűen alakítja át az európai állampolgárok kommunikációs, vásárlási és utazási szokásait. A harmadik műhely-konferencia fókuszában a nyelvközpontú mesterségesintelligencia-kutatás, a gépi tanulás és fordítás, a nyelvi adat témakörei álltak. A közel 80 résztvevő több mint fele a kutatás területéről, az egyetemi szektorból érkezett.

A korábbi Nyelvtudományi Intézetben 2015-ben és 2018-ban is rendeztek ELRC-konferenciát a soknyelvű Európa összekapcsolásáról. A mostani eseményt az intézet részéről Váradi Tamás köszöntötte, aki szintén hangsúlyozta, hogy a folyamatos hálózatfejlesztés és digitalizáció, a gépi tanulás és fordítás fejlesztése a nyelvi korlátok lebontását segíti elő.

Az első előadó Philippe Gelin (DG-Connect) volt, aki *Language Technologies Across Europe* címmel tartotta meg beszámolóját. A mesterséges intelligencia segítségével ma már mindenki számára elérhető 4-6 nyelv használata. Nemcsak a bevándorlók, hanem bármely európai állampolgár használhatja a nyelvek digitális kiterjesztésével létrehozott nyelvváltozatokat. Az Európai Unió nyelvi egyenlőséget támogató stratégiai kutatóprogramja a következő prioritásokat tartalmazza: 2030-ra teljes körű digitális nyelvi egyenlőség, ezáltal a veszélyeztetett nyelvek védelmének megvalósulása; a gépi tanulás, az adatfeldolgozás, a nyelvi modellek fejlesztése. A közös szttenderdek lehetővé teszik az ipar, az üzleti és a médiavilág résztvevőinek a nemzetközi piachoz való egyenlő hozzáférést, a személyre szabott szolgáltatások igénybevétele. A kutatóintézetek könnyebben férnek hozzá adatokhoz, a kulturális élet szereplői átláthatóbb információkhoz jutnak. A közigazgatás területén több, egyre javuló minőségű szolgáltatás jelenik meg. Előadása végén Philippe Gelin bemutatta a magyar nyelvet támogató uniós szolgáltatások honlapjait.

Prószycki Gábor (Nyelvtudományi Kutatóközpont) *A nyelvtechnológia és a mesterséges intelligencia lehetőségei – hol tartunk most, és mik a célok?* címmel tartott előadást. A mélytanulásra épülő nyelvtechnológia kutatásnak nagy lehetőségei vannak mind az Európai Unióban, mind Magyarországon. A nyelvtechnológiai fejlesztéssel létrehozott eszközöket széles körben használjuk a marketing, a banki ügyfélszolgálat, a pénzügyi kereskedelem, a mezőgazdaság, az egészségügy, az űrkutatás, a közlekedés terén. A nyelv és intelligencia kutatásának nehézségeiről szólva Prószycki Gábor rámutatott a természetes nyelv és a mesterséges nyelv különbségére.

Az emberi nyelv hatékony, rugalmas és összetett: többjelentésű szavakkal dolgozik, parafrázisokat alkalmaz, a jelentés kontextustól függ, állandó a jelentésátvitel. A kutatás egyre több kontextust (adatot) igényel. Mára a kutatásfejlesztés sok megbízható nyelvtechnológiai szolgáltatást eredményezett: automatikus beszédfelismerés, beszédgenerálás, szentimentelemzés, névelem-felismerés, relációkinyerés, összefoglaló-generálás. A hétköznapi felhasználók jól ismerik a dialógusrendszereket (chatbotok) és a levélszemétszűrést. Ami tíz évvel ezelőtt még csak ígéret volt, ma valóság, például a véleménybányászat. Az előadás azzal a bejelentéssel zárult, hogy a Nyelvtudományi Kutatóközpont, a Pécsi Tudományegyetem és a Microsoft Hungary fejlesztésében elkészült a magyar nyelv digitális modellje, amit eddig csak tíz nyelvre dolgoztak ki a világon.

A konferencia első kerekasztal-beszélgetése a *Nyelvtechnológia Magyarországon* témában hangzott el Váradi Tamás vezetésével. Résztvevői: Bessenyei Gábor (MorphoLogic Lokalizáció Kft.), Horváth-Varga János (T-Systems Magyarország), Körmendi György (Clementine), Szekeres Péter (Neticle Zrt.), Vadász Pál (Montana Zrt.), Varga Gábor (Microsoft Magyarország). A résztvevők kiemelték, hogy a számítógépek a kép- és beszédfelismerés, valamint a szövegértés terén elérték azt a teljesítményt, amit az ember tud. Magyarország készen áll a mesterséges intelligencia fogadására, nem a technológia hiánya miatt nem törnek át komplex megoldások a magyar piacra. A kutatásra felhasználható pénzügyi források tekintetében célszerű regionális szinten gondolkodni. A fő probléma az adathoz való hozzáférés, de már a minőségi adat számít, nem a mennyiség. Ma már adatmegismerés nélkül is lehetséges az adatfeldolgozás úgy, hogy biztosítva van az adatvédelem.

Farkas Ágnes (DG-Translation) *A CEF gépi fordítási platform* címmel tartott előadást. Elmondta, hogy az eTranslation olyan neurális gépi fordítóeszköz, amely online közszolgáltatásokban elérhető nemcsak a közigazgatás szereplői, hanem minden európai polgár számára. Az eTranslation dokumentumok gépi nyersfordítását végzi el, amely műveletnek a célja a szövegek tartalmi vázának megismerése. Ezután a fordításokat lektoráltatni szükséges. A gépi fordítás előnye, hogy integrálható akár munkafolyamatokba (pl. miniszterek ülése), akár weboldalakba (fordítóablakok használata). További előnye, hogy ingyenes és biztonságos. A szövegeket az Európai Bizottság tűzfala védi. 24 óra után törlik az adatokat, de lehetséges azonnali adattörlés is. Az eTranslation több domainban alkalmazható, tíz motorral kereshető, sok támogatott formátumban elérhető. Harminc nyelvre tud fordítani, és egyszerre több dokumentum fordítható több nyelvre is. Az eszköz az uniós szövegeken a leghatékonyabb, mert saját, magas minőségű adatbázison tanították be. Farkas Ágnes a résztvevők figyelmébe ajánlotta a <https://language-tools.ec.europa.eu> és <https://webgate.ec.europa.eu/etranslation> honlapokat, ahol regisztrálhatnak a gépi fordítás igénybeviteléhez. Az eszköz az egyetemi oktatáshoz is használható.

A második kerekasztal-beszélgetést Bódi Zoltán (Nyelvtervezési Kutatóközpont) vezette a *Nyelvtechnológia az állami szektorban* kérdésköréről. Bódi Zoltán áttekintette, mely közigazgatási területeken hasznosul jelenleg a mesterséges intelligen-

ciára épülő nyelvtechnológia. A napi adatfeldolgozás műveleteivel – automatikus üzenetválasz, a megkeresések irányított továbbítása, a telefonhívások összefoglalása – számos hivatalban találkozunk. A törvényalkotást befolyásolja a közvélemény követése elektronikus konzultációk alapján. Az üzleti intelligenciához hozzátartozik a digitális piackutatás és szentimentelemzés. A közigazgatás digitális átalakítása adatvezérelt, bizonyítékokon alapuló döntéshozatal, ezáltal nagyobb hatékonyságot eredményez. A résztvevők – Boa László (MI Koalíció), Jobbágy László (Digitális Jólét Nonprofit Kft.), Szviridov István (Idomsoft), Tarcsi Ádám (Nemzeti Tudásközpont) – a nyelvtechnológia integrálásának mértékéről beszélgettek. A kormány támogatja az elektronikus ügyintézés, közgyűjtemények digitalizálását, mesterségesintelligencia-alapú közigazgatási modellek felállítását. Létrejött a Digitális Kazinczy Műhely, amelynek célja, hogy hosszú távon elérhetőek legyenek jó minőségű digitális szövegek. Azért stratégiai terület a hazai nyelvtechnológia, mert a nagy techcégek rendelkeznek annyi digitális tartalommal a magyar nyelvről, hogy nélkülnk is fejlesszenek. A piaci és állami vegyes fejlesztés eredményeként az állampolgároknak folyamatosan kényelmi szolgáltatásokat kell felajánlani. Az adatvédelmet jogszabályi háttérrel szükséges megtámogatni. A kutatás és a közszféra szereplőiben is megvan a szándék, hogy a magyar nyelvet alkalmassá tegyék a számítógépes szövegfelismerésre és döntéshozatalra.

A konferencia délutáni részét Feldmann Ádám (Pécsi Tudományegyetem) nyitotta a *The Value of Data for the Development of Top Quality LT* című előadásával. Azért növekszik a havonta megjelenő újabb és újabb nyelvmodellekhez szükséges adatok mennyisége, mert nagy korpuszhoz nem szükséges tanítóprogram. Továbbá a tanítóadat mérete és a modell teljesítménye összefügg. Nagy adatmennyiség esetén ellenőrizhetetlen, hogy a modell mit tanult meg. A nyelvmodellekben megjelenik a transfertanulás jelensége: amit megtanult angolul, azt át tudjuk vinni a magyar nyelvre anélkül, hogy megtanítanánk neki. A nagy nyelvmodellek az összegyűjtött adatok multimodális megjelenítésére (szöveg-kép, narratíva-vidéo párok) használhatók.

A harmadik kerekasztal-beszélgetés a *Nyelvi adatok létrehozása, kezelése és megosztása: bevált gyakorlatok és kihívások Magyarországon* témakörben folyt. Csósz Gergely (Szerzői Jogi Szakértő Testület), Farkas Richárd (MI Nemzeti Laboratórium), Feldmann Ádám (Pécsi Tudományegyetem), Schin Lotár Csaba (OTP Bank), Sebők Miklós (Társadalomtudományi Kutatóközpont) részvételével. A beszélgetést Váradi Tamás és Bódi Zoltán közösen vezette. A szakemberek megállapították, hogy egyes mondatok, címek, annotált szövegek, dalszövegek, a gépi tanítóprogramnak, valamint a gépi beszédkeltés szavainak jogi védelme ugyan sok további problémát vet fel, de Magyarország betartotta a szerzői jogi törvény módosítására vonatkozó 2021. június 7-i határidőt: a szöveg- és adatbányászat szabályozása bekerült a törvénybe.

Ezt követően Jelencsik-Mátyus Kinga ismertette a European Language Equality (ELE) projekt magyar nyelvet érintő erőforrástérképét. Az ismertetés után rövid demonstrációk következtek. Körmendi György (Clementine) bemutatta a Hello,

Hanga! chatbot működését. Vadász Pál (Montana Zrt.) áttekintette a chatbotok alkalmazását a jogi szövegek átvizsgálásában és a jogi tanácsadásban. Bessenyei György (MorphoLogic) a gépi fordítómotorok szűk spektrumú (domain adapted) fejlesztésének szükségességéről beszélt. Szekeres Péter (Neticle Zrt.) a szentimentelemzésre hozott példát.

A konferenciát Váradi Tamás zárta, aki köszönetet mondott a szervezőknek és minden résztvevőnek, ugyanis a közönség kérdőívek kitöltésével járult hozzá a vélemények megismeréséhez. Az előadásokról készült felvétel elérhető a www.lr-coordination.eu/ honlapon.

Tuba Márta

középiskolai tanár, PhD

Gárdonyi Géza Általános és Középiskola, Érd

E-mail: dr.tuba.marta@gmail.com

<https://orcid.org/0000-0002-8264-2108>