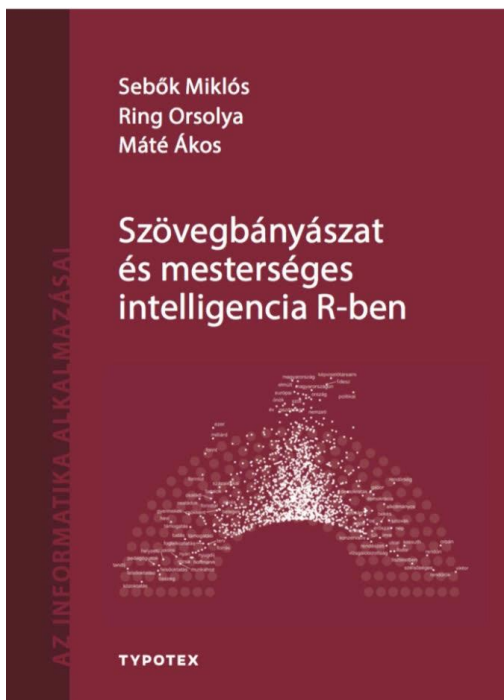


Online konferencia: Szövegbányászat és mesterséges intelligencia R-ben

2021. május 25-én a Társadalomtudományi Kutatóközpont (TK) Politikatudományi Intézete online konferenciát rendezett "Szövegbányászat és mesterséges intelligencia R-ben" címmel. A konferencia apropója a Sebők Miklós, Ring Orsolya és Máté Ákos által szerkesztett [Szövegbányászat és mesterséges intelligencia R-ben](#) című kötet megjelenése (Typotex Kiadó) volt. A kötet a Mesterséges Intelligencia Nemzeti Laboratórium (MILAB) projekt keretében jött létre.

Benczúr András, a MILAB szakmai vezetője, és **Boda Zsolt**, a Társadalomtudományi Kutatóközpont főigazgatója köszöntőikben kiemelték, a kötet megjelenése mutatja, hogy a társadalomtudomány is aktív szerepet vállal a mesterséges intelligencián alapuló kutatási módszerek fejlesztésében. A konferencia résztvevői **Máté Ákos** (tudományos munkatárs, TK PTI poltextLAB) és **Székely Anna** (kutatási asszisztens, TK PTI poltextLAB) könyvismertetőjét követően két különböző tematikájú kerekasztalbeszélgetésen oszthatták meg tapasztalataikat a mesterséges intelligencián alapuló módszerekkel kapcsolatosan.

Máté Ákos és Székely Anna az online elérhető könyv kísérő R csomagját, a [HunMineR](#)-t is bemutatták. A HunMineR egy olyan szövegbányász csomag, amely könnyű, gyors és nyílt hozzáférést biztosít magyar nyelvű szöveges adatbázisokhoz, és amely segítségével az érdeklődők a könyvben szereplő összes elemzést reprodukálhatják.



A konferencia első, **Ságvári Bence** (tudományos főmunkatárs, TK SZI, TK CSS–RECENS) által moderált kerekasztal résztvevői, **Németh Renáta** (tanszékvezető, ELTE TáTK), Berend Gábor (egyetemi adjunktus, SZTE) és **Sebők Miklós** (intézetigazgató, kutatásvezető, TK PTI, poltextLAB,) elsősorban a felsőoktatás és az Natural Language Processing (NLP) viszonyáról, illetve a kutatói utánpótlás helyzetéről beszélgettek. Mindannyian egyetértettek abban, hogy az NLP módszertana még nyugaton sem bevette a társadalomtudományban, így a magyar kutatók nincsenek jelentős lemaradásban európai szinten. Az Egyesült Államoktól eltekintve, a társadalomtudományi konferenciákon sem jellemzően található ilyen tematikájú részleget egyelőre, így óriási a lehetőség a magyar kutatók számára, hogy időben bekapcsolódjanak a módszertani fejlesztésekbe. Ahogy Sebők Miklós kiemelte, egyre több hazai kutató él is a lehetőséggel, a TK például a [COMPTX](#) nemzetközi szövegbányászati konferencia

alapítói és főszerzői között is szerepel. Az utóbbi években a magyar kutatók Európa élcsapatába kerültek, a kérdés azonban az Sebők szerint, hogy milyen szélessé tudja kinőni magát a mögöttük lévő mozgalom, ami kétségtelenül összefüggésben van az egyetemi képzések naprakésziségeivel. Ami a felsőoktatást illeti, jelzésértékű, hogy ahol a diákoknak lehetőségük van önszerveződő módon kurzusokat indítani, ott évről évre új szövegbányászati képzéseket rendeznek, melyekből aztán olyan projektek születnek, mint a kötetben is bemutatott kutatások.

A belépési korlátok azonban továbbra is magasabban vannak Magyarországon, mint a nyugati országokban, így a következő évek fókuszában, Sebők szerint, az alapstruktúra kialakítása és a magyar korpuszok építése kell álljon, hogy a hazai fiatal kutatók egyre könnyebben tudják alkalmazni az NLP módszertanát. A felzárkózási folyamatban fontos szerepe van az olyan interdiszciplináris együttműködésnek teret adó fórumoknak is, mint a MILAB.

A második kerekasztal résztvevőivel a véleményelemzés mesterséges intelligenciára alapuló megoldásairól beszélgetett **Szabó Martina Katalin** (tudományos munkatárs, TK CSS–RECENS). A beszélgetőtársak, **Ring Orsolya** (tudományos munkatárs, TK PTI poltextLAB, TK CSS–RECENS), **Farkas Richárd** (egyetemi docens, SZTE, MILAB-NLP alaprojekt vezetője), **Gyires-Tóth Bálint** (egyetemi docens, BME) és **Feldmann Ádám** (egyetemi adjunktus, PTE) különböző kutatási területekről jutottak el a szentimentelemzés módszertani lehetőségeinek felismeréséhez. A kerekasztal több résztvevője az akadémia mellett ipari projekteken is dolgozik. Nem meglepő a piaci érdeklődés, hiszen a véleménykinyerés mesterséges intelligenciára épülő megoldásai kiválóan alkalmazhatók olyan területeken is, mint a piacelemzés, különösen annak meghatározásában, hogyan viszonyulnak a fogyasztók a termékek különböző aspektusaihoz.

A beszélgetőtársak egyetértettek abban, hogy ugyan a szentimentelemzés klasszikus formája a hosszabb szövegek pozitív/negatív/semleges klasszifikációja, a komplexebb megközelítésekben, például a mondat- és targetszintű elemzésekben jóval több lehetőség rejlik. Sőt, bizonyos kutatási témák esetében a pozitív/negatív/semleges klasszifikáció szinte lehetetlen vagy kevésbé szerencsés, mint más megoldások. Ring Orsolya és Szabó Martina újságcikkeket elemző kutatása, amely szentimentelemzésnek indult például, azt a meglepő eredményt hozta, hogy a pozitív/negatív/semleges klasszifikációnál sokkal nagyobb egyetértést tud elérni az annotátorok között egy többemeű kategóriarendszer, amely a szövegeket inkább 12 különböző érzelmi kategóriába sorolja.

Ami a mesterséges intelligenciára épülő szentimentelemzés jövőjét illeti, a kerekasztal résztvevői a kontextus alapú, beágyazott modellek terén várják a legnagyobb áttörést. A kutatók tapasztalatai alapján a közösségi médiában megjelenő posztok, kommentek és tweetek elemzésénél gyakran lehetetlen kontextusból kiragadva pontos megállapításokat tenni. A résztvevők megfigyelései alapján sok esetben jelent nehézséget a domainfüggetlen szótárak alkalmazása is. A domain-specifikusságból eredő problémák áthidalásában azonban pusztán a szentimentelemzés módszertanát használó kutatások számának gyarapodása is áttörést hozhat: minél több domaint tárnak fel a kutatók, annál hatékonyabban lehet majd adaptálni az eddigi eredményeket és megtalálni az adott domainre legjobban alkalmazható megoldásokat.