

Knap Árpád: RC2S2, ELTE TáTK – A kutatás az Innovációs és Technológiai Minisztérium ÚNKP-21-3 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült

Tóth Tímea Emese: RC2S2, ELTE TáTK

Barna Ildikó: RC2S2, ELTE TáTK – A kutatást a K-134428 azonosítójú NKFIH-pályázat támogatta.

### **Érzelmek megjelenése a Trianoni békeszerződéssel és a holokauszttal kapcsolatos cikkek szóbeágyazásaiban, az érzelmek automatizált detektálásának lehetséges eszközei**

Előadásunk alapját a Trianoni békeszerződéssel és a holokauszttal kapcsolatos online újságcikkek korpuszán végzett kutatásunk jelenti, amelyben több mint 25 ezer darab, politikai szempontból besorolt írást vizsgáltunk. Ez a két esemény a leginkább megosztó, identitásformáló témák közé tartozik, amely elválasztja egymástól a különböző politikai oldalon lévőket. Az ezekkel kapcsolatos diskurzus nem szorítkozik a történettudományra, hanem aktív részét képezi a politikai nyilvánosságnak és a közbeszédnek is. Elemzésünk célja a fenti megosztottság tematikus és nyelvi megragadása, annak felderítése, hogy a két téma mennyire jelenik meg eltérő hangsúllyal, eltérő keretezésben, és eltérő szóhasználattal a különböző politikai oldalak esetében. A megosztottság karakteresen képződik le az érzelmekben. Ezek pedig jól vizsgálhatók az olyan, különböző politikai oldalak korpuszán létrehozott szóbeágyazásokban megjelenő, erős érzelmi töltettel rendelkező szavak segítségével, mint a fájdalom, trauma, sors vagy szörnyűség.

Előadásunk két részből áll. Az első részben szóbeágyazási modellek, illetve topikmodellek alkalmazásával, valamint a szövegek kvalitatív elemzésével mutatjuk be a politikai oldalak közötti eltéréseket, ami az egyes témákhoz kapcsolódó érzelmek megjelenését, illetve a különböző oldalakra jellemző témapreferenciákat illeti. Prezentációnk második részében pedig az érzelemdetektálás egy célzott, automatizált módszerét mutatjuk be. Ennek alapját egy több mint egy éve megkezdett projekt jelenti, amelyet az RC2S2 kutatócsoport a Digitális Örökség Nemzeti Laboratóriummal együttműködésben végez. A projekt célja egy olyan nyelvi modell megalkotása, amely képes magyar nyelvű szövegekben, aktorokhoz (például helyszínekhez, szervezetekhez, személyekhez, eseményekhez) kapcsolódó szentimentek és konkrét érzelmek klasszifikációjára. A modell tanítóadatbázisát egy tízezer cikkből álló, online újságokból származó, statisztikai mintavétel segítségével összeállított, humán annotált szövegkorpusz jelenti. Prezentációnk ezen részében ismertetjük a projekt kétlépcsős annotálási módszertanát, az érzelemannotálás kihívásait, illetve a kutatás várható eredményeit. Előadásunk végén egy konkrét példán, a Trianoni békeszerződéshez és a holokauszthoz kapcsolódó újságcikkek alá írt kommenteknek a korpuszán mutatjuk be ennek az elemzési eszköznek a társadalomtudományos relevanciáját.