

Németh Renáta: RC2S2, ELTE TáTK – A szerzőt a K-134428 azonosítójú NKFIH pályázat támogatta kutatási munkájában.

Nyelvi polarizáció kutatása NLP-vel: módszertani kihívások egy scoping review alapján (általánosíthatóság, szakterületi tudás szerepe, kevert módszerek, oksági következtetés)

A "szöveg mint adat" mozgalom részeként a természetes nyelvfeldolgozási (natural language processing, NLP) technikák új lehetőségeket kínálnak a politikai polarizáció nyelvi megnyilvánulásainak feltárására. Az e témában indult, NKFIH-támogatta projektünket módszertanilag megalapozó scoping review-nkban áttekintést végeztünk a 2010 óta megjelent tanulmányokról. Meghatároztuk az adatforrásokat és az elemzési technikákat, áttekintettük a polarizáció különböző alkalmazott konceptualizációit és operacionalizációit, és szintetizáltuk a megközelítések, lehetséges hibák és megoldások széles skáláját. Reményeink szerint eredményeink tanulságosak lehetnek az NLP más területen végzett társadalomkutatási alkalmazásai számára is.

154 releváns tanulmányt azonosítottunk. A tanulmányok több mint fele amerikai adatokat használt. Minden harmadik tanulmány gépi tanulást alkalmazott a politikai nézetek szövegekből történő előrejelzésére. A legtöbb tanulmány nem alkalmazott close reading-et, és csak néhány tanulmány tárgyalta a szöveges adatokból történő oksági következtetések során felmerülő lehetséges problémákat. Bár az elmúlt évtizedben nőtt a tanulmányok száma, csak kisebb részük használt szakterületi tudást (domain knowledge) a kutatásban (45%-ban nem volt szakterületi szerző). Amikor azonban, éppen fordítva, a szerzők között csak társadalomtudósok szerepeltek, gyakran előfordult, hogy az általuk alkalmazott egyszerű módszertan nem használta ki teljes mértékben az NLP-ben rejlő lehetőségeket. Ezek az eredmények, valamint a nem interdiszciplináris tanulmányok magas aránya vélhetően az NLP társadalomtudományokon kívüli intézményesülésének következménye.

Az áttekintés alapján kijelenthetjük, hogy az NLP-ben rejlő lehetőségek a politikai polarizáció kutatásában nyilvánvalóak. Az is kitűnt azonban, hogy szakterületi ismeretek nélkül az NLP olyan technika maradhat, ami kevés tudományos értéket ad hozzá a társadalomtudományokhoz. Reméljük, hogy áttekintésünk érveket szolgáltat a kutatás multidiszciplinaritás irányába történő elmozdítása mellett, mivel ez lehet a kulcsa társadalmunk mélyreható változásainak jobb megértésének.