

## A 2016-os tanártüntetések szövegeinek feldolgozása és adatvizualizációja interaktív dashboard segítségével

Balogh Kitti<sup>1</sup>, Fülöp Nóra<sup>1</sup>, Szabó Martina Katalin<sup>1,2</sup>

<sup>1</sup>Precognox Informatikai Kft.  
kbalogh@precognox.com; mszabo@precognox.com;  
nora.fulop@gmail.com

<sup>2</sup>Szegedi Tudományegyetem, Orosz Filológiai Tanszék  
szabo.martina@lit.u-szeged.hu

**Kivonat:** A dolgozatban egy automatikus emóció- és szentimentelemzéssel, valamint topik modellezéssel feldolgozott korpusz létrehozásáról, valamint az adatokon végzett elemzésekről és vizualizációs megoldásokról számolunk be. A korpuszt olyan posztokból és kommentekből hoztuk létre, amelyek tematikájukban a 2016. februári és márciusi tanártüntetések eseményeihez kapcsolódnak. A szövegeket automatikusan gyűjtöttük le, melyeken aztán szótáralapú szentiment- és emócióelemzést hajtottunk végre, és topik modellezés módszerével nyertünk ki témákat. Az így feltárt szemantikai adatokat végül egy interaktív dashboard segítségével vizualizáltuk.

### 1 Bevezetés

Jelen dolgozat egy különböző nyelvtechnológiai tartalomelemző módszerekkel feldolgozott korpusz létrehozásáról, valamint a tartalmi adatok vizualizációs megoldásáról, és az azok alapján levonható következtetésekről számol be.

Tárgyaljuk a kutatás során felhasznált eszközöket és módszereket, valamint a feldolgozott korpusz eredményeit egy lehetséges szociológiai megközelítés segítségével. Ismertetjük a saját fejlesztésű adatgyűjtő eszközöket, amellyel lehetőség nyílik Facebookon elérhető adatok gyűjtésére. Emellett bemutatjuk a létrehozott korpuszt, amelynek szövegei tematikájukban a 2016-os tanártüntetések eseményeihez kapcsolódnak, és amelyet szótáralapú automatikus szentiment- és emócióelemzéssel, valamint topik modellezéssel dolgoztuk fel. Az utóbbi módszerek részletes betekintést engednek a korpusz szövegeinek tartalmi sajátágaiba.

A topik modellezés, a szentiment- és az emócióelemzés eredményeinek megértéséhez olyan szociológiai irodalmat használtunk fel, amely a társadalmi mozgalmakban előforduló érzelmekkel foglalkozik. Mindezek mellett a kutatás eredményeit interaktív dashboardon jelenítettük meg, amely vizuális eszközeivel biztosítja az eredmények áttekinthetőségét. A dashboard a <http://labs.precognox.com/fbshiny/> linken érhető el.

## 2 A munka elméleti háttere

A szentimentelemzés segítségével feltárható sajátságokat ma már számos tartalomelemző megoldás hasznosítja [1]. Ennek oka az értékítéletek, vélemények jelentőségének felismerésében keresendő, aminek következtében a szentimentelemzés fontos kutatási és fejlesztési irány az adat- és szövegbányászat területein. Bár a szentimentelemzés mára az egyik legaktívabban művelt NLP-területté nőtte ki magát [2], a magyar nyelvű szövegek értékeléselemzése még mindig jóval kisebb figyelmet kap a nemzetközi gyakorlathoz képest [3], [4], [5], [6], [7].

Az emócióelemzés, a szentimentelemzéstől eltérően nem az értékítéletek, hanem az érzelmek automatikus feldolgozását célozza. Ez alapján, bár a két tartalomelemző feladat között van összefüggés, azok nem azonosíthatóak egymással [1], [8]. Az érzelmek szövegalapú elemzésével még nemzetközi szinten is csekély számú nyelvtechnológiai dolgozat foglalkozik [9], [10]. A magyar nyelvű szövegek érzelemelemzésére pedig alig fordítanak figyelmet a kutatók és fejlesztők [1], [8]. Rá kell mutatnunk azonban, hogy az érzelmek bizonyos tudományos diszciplínákban, így a viselkedéstudományban vagy a pszichológiában központi szerepet töltenek be. Kutatási eredményeink [1], [8] alapján úgy véljük, hogy a szövegek emotív tartalmának kinyerése olyan értékes információkat hozhat a felszínre, amelyeket más tartalomelemző módszerek nem tárnak fel. Ezzel összefüggésben, a jelen korpusz esetében is, a szentimentelemzés mellett emócióelemzést is végeztünk a szövegeken. Mivel a szövegek az oktatás átalakításáért küzdő mozgalmak Facebookon is aktív résztvevőitől származnak, nyomon követhetővé válik, hogy a mozgalomban – ahogy a mozgalmak esetében általában – a politikai cselekvés elengedhetetlen velejárói az érzelmek, amelyek segítenek a politikai mozgósításában, a csoporttagok rekrutálásában és az elköteleződés növelésében [12].

A topik modellezés egy nagy népszerűségnek örvendő szövegbányászati módszer, amely nagy mennyiségű szöveges adat rejtett tematikus struktúrájának feltárását célozza [11]. A topik modellek kutatása egy többfelé ágazó, folyamatosan fejlődő terület. Jelen dolgozattal két kutatási irányhoz szeretnénk csatlakozni. Ezek egyike egy a még kevesek által művelt irány, amely a topik modellek eredményeinek interdiszciplináris (szociológiai, jogi, politikai, stb.) felhasználási lehetőségeivel foglalkozik [11]. Kutatásunk során a kinyert témákat, érzelmeket és értékítéleteket szociológiai szempontból értelmezzük. A düh, öröm, meglepettség, félelem, undor és szomorúság mértékének kimutatása mellett fontos eleme az érzelmek vizsgálatának az, hogy mi váltja ki ezeket az érzelmeket: milyen kontextusokban jönnek létre és milyen események vannak rájuk hatással [12]. Ugyanakkor a terület vizsgálata komplex feladat, a megjelenő érzelmek és a különböző témák dinamikus változása módszertanilag nehezen hozzáférhető [12]. A közösségi média oldalairól gyűjthető adatok emóció- és szentimentelemzéssel, valamint topik modellezéssel történő feldolgozása, és az eredmények idősoros követése lehetséges eszközként kínálkozik.

A másik kutatási irány, amelyhez a jelenlegi kutatásunk kapcsolódik, a topik modellek eredményének vizualizációjára és megfelelő user interface-ek létrehozására irányul [11]. E célból hoztunk létre egy interaktív dashboardot, amelyen a vizualizációk a beállításokon keresztül befolyásolhatóak, így a felhasználót érdeklő információk kiemelhetőek. A dashboardon a posztok és a kommentek témáit

idősorososan jelenítettük meg, ami lehetőséget ad az eseményekkel kapcsolatos fontos témák időbeli alakulásának vizsgálatára. A felületen a Facebookról legyűjtött két esemény aktivitás- (posztolás, kommentelés, like- és reakcióadás) és szöveges adatainak vizualizációi is megjelennek. A jelenlegi dolgozatban alapvetően a szöveges adatokra koncentráltunk, az aktivitás adatok feldolgozásáról nem számolunk be.

### **3 A korpusz bemutatása és alapvető adatai**

A korpusz annak a két Facebook-eseménynek a posztjait és kommentjeit tartalmazza, amely oldalakon keresztül a 2016-os tanártüntetésekkel kapcsolatos mozgalom szerveződött.

Eszközeink és a téma metszéspontjánál egy termékeny szociológiai terület bontakozik ki, a társadalmi mozgalmak kutatása és a társadalmi mozgalmakban megjelenő érzelmek vizsgálata. A két kiválasztott Facebook eseményt elemezve – amelyekben jelentős felhasználói aktivitás volt mérhető, és a demonstrációkon is sokan vettek részt – reméljük, hogy képet adhatunk egy mozgalom résztvevőiről, céljaikról és beszédtemáikról.

A korpusz posztjai és kommentjei a 2016. február 2. és március 23. közé eső időszakban keletkeztek, amelyeket automatikusan gyűjtöttük a február 13-i és március 15-i események oldaláról.

A teljes korpusz a szövegműfajok tekintetében 6094 posztból és kommentből áll, amelyekben összesen 15878 mondat, 160589 szó és 1201369 karakter található. Tartalmi szempontból ugyancsak két részre bontható, a bennük tárgyalt események alapján (a februári, valamint a márciusi történésekre). A februári esemény vonatkozásában nagyjából kétszer annyi szöveges megnyilvánulást tettek a felhasználók, mint a márciusi esemény oldalán: a februári eseménynél 4082, a márciusi eseménynél 2012 poszt vagy komment keletkezett a vizsgált időszakban. A két esemény alatti kommentek, illetve posztok közel azonos hosszúságúak az átlagos mondat-, szó- és karakterszám tekintetében. Mindkét eseménynél magasabb volt a kommentek száma a posztokénál. A két eseménynél összesen 4785 kommentet és 1309 posztot írtak a felhasználók. Megfigyelhető továbbá, hogy a februári esemény alatt nagyobb arányban érkeztek kommentek a márciusinál: a februári eseménynél körülbelül 4,5-szer, míg a márciusi eseménynél körülbelül 2,5-szer több kommentet találhatunk, mint posztot.

#### **3.1 A szöveggyűjtés módja**

A korpusz szövegeit saját Facebook-scrapertünkkel gyűjtöttük, amelynek segítségével publikus Facebook-oldalokról nyerhetünk adatokat. Az eszköz nem csupán szöveges adatokat gyűjt, mint például egy adott oldal posztjai és kommentjei, hanem aktivitási adatokat is. A scrapert Python programozási nyelven implementáltuk, ami a publikus Facebook-adatok gyűjtésében egyszerű parancssoros használatot tesz lehetővé. A scraperrel az 1569911826564534 és az 513643295475511 Facebook azonosítóval

rendelkező események oldalairól gyűjtöttük le az adatokat, amelyek letölthetőek voltak az események eltávolításáig. Az utolsó adatokat 2016. március 23-án gyűjtöttük le.<sup>1</sup>

A megfelelő adatvédelmi szempontok figyelembe vétele mellett bármilyen célra szabadon elérhető az eszköz, valamint annak használati módja az alábbi linken: <https://github.com/precognox-admin/FBscraper>.

## 4 A szövegek feldolgozása és a szemantikai információk kinyerése

### 4.1 Nyelvi előfeldolgozás

A leggyűjtést követően, a szöveges adatokat UTF-8 karakterkódolású plain text formátumba konvertáltuk, majd a magyarlanc eszközzel [13] nyelvi feldolgozást végeztünk rajtuk. A szövegeket az eszköz tokenizálta, lemmatizálta, valamint azonosította a lemmák szófaját. Ezután a szövegek lemmáit főnevekre, melléknevekre és ismeretlen szófajú elemekre szűrtük.

A szövegekben az emotikonokat és emojikat is kezeltük, a leggyakrabban használatosakat szóalakokra váltottuk át egy általunk bővített szótár segítségével. Például a “:)” emotikont a “simamosoly” szóalakra alakítottuk át. Ezzel a szófajilag szűrt, lemmatizált és emotikonokat, emojikat kezelő kimenettel fogtunk a szemantikai információk kinyerésébe.

### 4.2 Topik modellezés

Az előfeldolgozott posztokban és kommentekben feltártuk a szövegekben rejlő témákat. Mivel a dashboardot az R statisztikai programnyelv Shiny nevű webapplikációs frameworkjével készítettük el, a topik modell illesztéséhez egy R csomagot, a `topicmodels`-t [14] használtuk.

A korpusz témáinak kinyeréséhez a topik modellek egyik legegyszerűbb tagját, a látens Dirichlet allokációt [16], [17] használtuk, amely az általunk használt csomagban is implementálva van. A modell poszterior eloszlásának közelítéséhez különböző algoritmusokat alkalmaznak, amelyek közül a `topicmodels` csomagban implementált Gibbs mintavételezést használtuk. A korpuszunkban rejlő látens topikok számának meghatározásához egy R-ben implementált függvényt választottunk, a harmonikus átlag módszerét [18], mely már több korpuszunk topik modellezése során bevált [19]. Miután meghatároztuk az optimális témaszámot, unigram alapú modellt illesztettünk a korpuszon. Ezzel megkaptuk minden egyes dokumentum témaeloszlását és a topikok szóeloszlását. A témákat ez alapján, a témákhoz tartozó legjellemzőbb szavak és a témákhoz tartozó legjellemzőbb dokumentumok alapján neveztük el

---

<sup>1</sup> A Facebook-adatok scrapelésére vonatkozó szabályokról bővebben l. [https://www.facebook.com/apps/site\\_scraping\\_tos\\_terms.php](https://www.facebook.com/apps/site_scraping_tos_terms.php)

emberi erőforrás segítségével. A februári esemény posztjaiból 15, kommentjeiből 13 témát nyertünk ki, míg a márciusi esemény posztjaiban 13, kommentjeiben pedig 10 témát tártunk fel.

### 4.3 Szentiment- és emóciótartalom kinyerése

A második lépésben a korpusz szövegeit a saját készítésű szentiment- és emóciósztárainkkal elemeztük [1], [20].

Magyar nyelvű szentimentsztárunkat részben automatikus, részben manuális módszerrel hoztuk létre, magyar nyelvű szövegek automatikus szótáralapú szentimentelemzése céljából [20]. Első lépésben egy angol nyelvű, pozitív és negatív listából álló szentimentsztárat automatikusan magyar nyelvre fordítottunk, majd a fordítás eredményét kézzel ellenőriztük, javítottuk, valamint két szinonimasztár segítségével bővítettük. A szótár készítése során nem csupán melléneveket, hanem főneveket, határozószókat és igéket is felvettük, amennyiben úgy ítéltük, hogy az adott nyelvi elemnek inherens negatív vagy pozitív szentimentértéke van. Az így elkészített szótárunk, amelyet plain text formatumban tárolunk UTF-8 karakterkódolással, összesen 1748 pozitív és 5940 negatív szentimentszót tartalmaz. A szótár kutatási célokra szabadon hozzáférhető (<http://opendata.hu/dataset/hungarian-sentiment-lexicon>).

Az emóciósztáraink létrehozásában ugyancsak egy angol nyelvű lexikonra (*Affective Text*) támaszkodtunk, amelynek automatikusan magyarra fordított anyagát kézzel ellenőriztük, javítottuk és kiegészítettük [1]. Az emóciókifejezések osztályozásában Ekman és Friesen (1969) [21] érzelemtaxonomiáját követtük, tehát azt a hat alapérzelmet vettük alapul, amelyek arckifejezéseit a kutatások alapján kultúrafüggetlenül azonos módon produkáljuk és azonosítjuk. Az alapérzelmek, amelyekre támaszkodva a szótárunk kifejezéseit kategorizáltuk, a következők: öröm, düh, bánat, félelem, undor és meglepődés. Az elkészített szótár a hat kategóriában összesen 1798 emóciókifejezést tartalmaz.

A posztokban és kommentekben megtalálható szentimenteket és emóciókat a `sentiment` R csomag egy a szótárainkkal módosított verziójával azonosítottuk be, amely csomag lehetőséget nyújt a szentimentek és az emóciók szótári illetve gépitánulás-alapú osztályozására is.<sup>2</sup>

## 5 Az adatok vizualizációja – a dashboard létrehozása és az eredmények értékelése

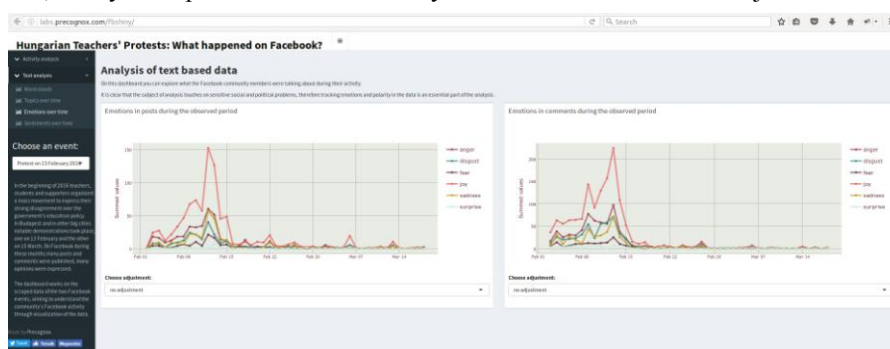
A szövegek többszintű feldolgozását követően a Shiny nevű webapplikációs framework (<http://labs.precognox.com/fbshiny/>) segítségével készítettünk egy interaktív dashboardot az adatok vizualizálásának és elemzésének céljából.

<sup>2</sup> A csomag a következő címen érhető el: <https://github.com/timjurka/sentiment>. Az eszköz használatáról többek között l. <https://www.r-bloggers.com/sentiment-analysis-with-machine-learning-in-r/>.

A <http://labs.precogno.com/fbshiny/> linken elérhető dashboard két részből áll. Egyrészt prezentálja és elemzi a két esemény oldaláról nyert aktivitásadatokat, másrészt a szöveges adatokat vizualizálja és elemzi.

A szöveges adatokból kinyert szemantikai információkat négy részre osztva jelenítettük meg a dashboardon. Az első részen a posztok és a kommentek leggyakoribb szavai jelennek meg szófelhő formájában. A februári esemény alatt a többek között a “pedagógus”, a “tüntetés”, a “gyermek”, a “kormány”, az “oktatás” és az “ember” szavak jelennek meg hangsúlyosan. Márciusban hasonló képet láthatunk, azonban az 1848. március 15-i események felelevenítése miatt a “március” szó is nagy szerephez jut. A jellemző szavak és a megjelenő topikok mind jól mutatják a mozgalom retorikai készletét, amely döntő fontosságú a célok és a szerepek definiálásában, ezen keresztül pedig a mozgósítás alapjává válik [12].

A társadalmi mozgalmak vizsgálatának fontos szempontja a politikai lehetőségek érzékelése [23], hiszen ez nagy hatással van a tagok elköteleződésére. Az idősorokon keresztül tetten érhetőek azok a témák, amelyek láthatóan kiugrásokat okoztak az érzelmek és szentimentek adataiban. A szöveges adatokban megjelenő szentimentek idősorainál megfigyelhető, hogy mindkét eseménynél a tüntetés időpontjában, illetve pár nappal előtte van a legtöbb negatív és pozitív szentimentet tartalmazó poszt vagy komment, igazolva a politikai cselekvések érzelmileg felfokozott állapotát. A posztok számával arányosított idősorokon más jellegű eseményekre is felfigyelhetünk. Például a februári idősor esetében március 6-án egy nagy negatív csúcs több olyan posztot jelez, amelyben a posztot írók a kormánnyal szembeni ellenérzésüket fejezik ki.



1. ábra. A dashboardon található emóció-idősorok

Az emóciós idősorok alapján a februári eseménynél a posztokban és a kommentekben végig az öröm a legerőteljesebb érzelm, míg a második legjellemzőbb az esemény időpontja körül a szomorúság, majd átveszi a szerepet az undor. A márciusi eseménynél szintén az öröm volt a legerőteljesebb emóció, azonban a második legjellemzőbbnek a düh bizonyult. A szakirodalom szerint az öröm jellemző a társadalmi mozgalmakra, hiszen a tagok így fejezik ki az összefogással kapcsolatos érzelmeiket, amely fontos összetartó erővel bír a kollektíva egészére nézve [12]. Ha azonban az elvárt lehetőségek nem következnek be, könnyen az elkeseredettség vetheti fel fejét, valamint a váratlan csalódásra válaszként a düh [12]. Az eredmények színes képet festenek a tanártüntetések Facebook eseményeinek

érzelmi ökonómiájáról, azonban tekintettel kell lennünk arra is, hogy az eredményeket befolyásolhatják az egyes emóciókhoz tartozó szótárak hosszúságai.

A topik modell illesztésének eredményei átfogó képet adnak a két esemény természetéről. A februári esemény posztjait, kiváltképp közvetlenül a demonstráció előtt és után meghatározó téma a közszolgáltatási szférában dolgozók helyzete. Az oktatás az egészségügy helyzetével összefonódva jelenik meg. Az oktatáspolitikai követelések mellett szó van béremelési követelésekről mind a két ágazatban dolgozók számára, valamint megszólítanak más, a szociális rendszertől függő rétegeket is (nyugdíjasok). Ahogy a posztokban, úgy a kommentekben is láthatóak kormányellenes megnyilvánulások. A februári kommentekre jellemző, hogy a magyar politikai életet tematizálják, és ellenzik a politikusok beleszólását az oktatásügybe. Emellett, talán meghatározóbban az érintettekről szólnak a kommentek: a gyerekekről, a tanárokról, a szülőkről. A februári esemény oldalán jellemzően reálpolitikai követeléseket tematizálnak, egyetlen szimbolikus elem jelenik meg, a szabadság, amely azonban a szabad iskoláztatáshoz kötődik. Mindezzel ellentétben a márciusi esemény oldalán kevés reális követelést olvashatunk, a tüntetést egy történelmi-politikai narratívába helyezik, az 1848-as forradalomhoz hasonlítják, a nép és hatalom konfliktusa váltja fel a szakpolitikai követeléseket. Igaz, a kommentek között megjelennek a gyerekek érdekei (szembeállítva a kormány érdekeivel), de az eredeti konfliktus háttérbe szorul. Mind a két esemény kapcsán megfigyelhető a magyar kultúra szimbolikus használata: sok irodalmi művet és történelmi eseményt idéznek vagy említnek a felhasználók.

## 6 Összegzés

A dolgozatban egy korpusz létrehozásáról számoltunk be, amelynek anyagát automatikus emóció- és szentimentelemzéssel, valamint topik modellezéssel dolgoztuk fel. Az eredményeket egy interaktív dashboard segítségével vizualizáltuk és elemeztük a megfelelő társadalomtudományi irodalom figyelembe vétele mellett.

A korpuszt olyan posztokból és kommentekből hoztuk létre, amelyek tematikájukban a tanártüntetések eseményeihez kapcsolódnak, és a 2016. február 2. és március 23. közé eső időszakban keletkeztek. Az adatokat egy saját fejlesztésű scraperrel automatikus módszerrel gyűjtöttük, majd a magyarul eszközzel dolgoztuk fel. Ezt követően a korpuszon szótáralapú automatikus szentiment- és emócióelemzést hajtottunk végre, és topik modellezés módszerével kinyertünk azok témáit is. A korpusz így feltárt szemantikai adatait végül egy interaktív dashboard segítségével vizualizáltuk.

A létrehozott interaktív dashboard szabadon hozzáférhető az alábbi linken: <http://labs.precognox.com/fbshiny/>

### Köszönetnyilvánítás

A jelen kutatás Az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programjának támogatásával valósult meg.

## Hivatkozások

1. Szabó M.K., Morvay G.: Emócióelemzés magyar nyelvű szövegeken. In Gecső T., Sárdi Cs. (szerk.). *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához* 177. Budapest, Tinta (2015) 286-292.
2. Liu, B.: *Sentiment Analysis and Opinion Mining*. Draft (2012) Elérhető: <http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>.
3. Berend, G., Farkas, R.: *Opinion Mining in Hungarian based on textual and graphical clues*. Proceedings of the 8th conference on Simulation, modelling and optimization. Stevens Point, Wisconsin, USA, World Scientific and Engineering Academy and Society (WSEAS) (2008) 408-412.
4. Miháltz, M.: *OpinHu: online szövegek többnyelvű véleményelemzése*. In Tanács, A., Vincze, V. (szerk.). VII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2010). Szegedi Tudományegyetem, Szeged (2010) 14-23.
5. Hangya V., Farkas R., Berend G.: *Entitásorientált véleménydetekció webes híryanagyokból*. In Tanács A., Varga V., Vincze V. (szerk.). XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015). Szeged, Szegedi Tudományegyetem (2015) 227-234.
6. Szabó M. K., Vincze V.: *Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai*. In Tanács, A., Varga, V., Vincze, V. (szerk.) XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015). Szegedi Tudományegyetem, Szeged (2015) 219-226.
7. Szabó M.K., Vincze V., Simkó K., Varga V., Hangya V.: *A Hungarian Sentiment Corpus Manually Annotated at Aspect Level*. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). Portoroz, Szlovénia Portoroz: European Language Resources Association (ELRA) (2016) 2873-2878.
8. Szabó M.K., Vincze V., Morvay G.: *Magyar nyelvű szövegek emócióelemzésének elméleti nyelvészeti és nyelvtechnológiai problémái*. In Reményi A. Á., Sárdi, Cs., Tóth, Zs. szerk. *Távlatok a mai magyar alkalmazott nyelvészetben*. Budapest: Tinta (2016)
9. Strapparava, C., Mihalcea, R.: *Learning to identify emotions in text*. SAC 2008. <http://web.eecs.umich.edu/~mihalcea/papers/strapparava.acm08.pdf>.
10. Mulcrone, K.: *Detecting Emotion in Text*. Elhangzott: UMM CSci Senior Seminar Conference. Amerikai Egyesült Államok, University of Minnesota: Morris. 2012. április 28. <https://wiki.umn.edu/pub/UmmCSciSeniorSeminar/Spring2012Talks/KaitlynM%20ulcrone.pdf>
11. Blei, D.: *Probabilistic topic models*. Communications of the ACM. 55(4) (2012) 77-84.
12. Goodwin, Jeff; Jasper, James M. *Emotions and Social Movements*. In: J. E. Stets, J. H. Turner szerk. *Handbook of the Sociology of Emotions* (2007) 611-631.
13. Zsibrita, J., Vincze, V., Farkas, R.: *magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian*. Proceedings of RANLP 2013 (2013) 763-771.
14. Grün, B., Hornik, K.: *topicmodels: An R Package for Fitting Topic Models*. Journal of Statistical Software. 40(13) (2011) 1-30. <http://www.jstatsoft.org/v40/i13/>.
15. McCallum, A. K. *MALLET: A Machine Learning for Language Toolkit* (2002) <http://mallet.cs.umass.edu>.
16. Griffiths, T. L., Steyvers, M.: *Finding scientific topics*. Proceedings of the National Academy of Sciences (2004) 5228-5235.
17. Blei, D., Ng, A. and Jordan, M. *Latent Dirichlet allocation*. Journal of Machine Learning Research, 3 (2003) 993-1022.
18. Ponweiser, M.: *Latent Dirichlet Allocation in R*. Diploma Thesis. Institute for Statistics and Mathematics, Wirtschaftsuniversität Wien (2012)
19. Balogh, K.: *A látens Dirichlet allokáció társadalomtudományi alkalmazása. A kuruc.info romaellenes megnyilvánulásainak tematikus elemzése*. Szakdolgozat. Survey Statisztika mesterképzés, Eötvös Loránd Tudományegyetem (2015) Elérhető:



- [http://labs.precognox.com/kurucinfo\\_adatviz/A\\_latens\\_Dirichlet\\_allokacio\\_tarsadalomtudo\\_manyi\\_alkalmazasa\\_Balogh\\_Kitti.pdf](http://labs.precognox.com/kurucinfo_adatviz/A_latens_Dirichlet_allokacio_tarsadalomtudo_manyi_alkalmazasa_Balogh_Kitti.pdf).
20. Szabó M.K.: Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai és dilemmái. In Gecső T., Sárdi Cs. szerk. *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához* 177 (2015) 278-285.
  21. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1 (1969) 49-98.
  22. Cho, M., Schweickart, T., Haase, A.: Public engagement with nonprofit organizations on Facebook. *Public Relations Review*, 40(3) (2014) 565-567.
  23. Goodwin, J., Jasper, J. M., Polletta, F.: Return of the Repressed the Fall and Rise of Emotions in Social Movement Theory. *Mobilization: An International Journal*. 5(1) (2000) 65-83.