



Az automatizált tartalomelemzés megvalósíthatósága

NAGY Andor

1. Bevezetés

1.1. Előzmények

2014-ben végeztem el azt a kutatást, amely jelen tanulmányom előzményének is tekinthető. Több webes hírportál rendszeres olvasójaként megfigyeltem, hogy a legtöbb cikkben felfedezhetők olyan egyedi jellemzők, amelyekből kiindulva akkor is azonosíthatók a szerzők, ha névtelenül publikálnak. Ilyen egyedi jellemző lehet például egy tipikus helyesírási hiba, sajátos mondatszerkesztés, fogalmazási stílus stb. Kíváncsi voltam, hogy a sejtésemet alá tudom-e támasztani módszeres vizsgálattal is. Rögtön két problémával is szembesültem: az egyik, hogy a mintaként használt két hírportálon, az *Indexen*¹ és az *Origón*² nagyjából száz–száz cikket publikálnak naponta, ami akkora adathalmazt jelent, amelynek szisztematikus áttekintése túl nagy feladat egy embernek. A másik problémát a saját szubjektivitásom jelentette, ugyanis attól tartottam, hogy akármennyire is próbálnék objektív maradni, az eredmények értékelésében megjelenne a saját véleményem is, én pedig mindenképpen

mérhető, számszerűsíthető adatokkal alátámasztott eredményeket szerettem volna kimutatni. Ezeket figyelembe véve döntöttem úgy, hogy olyan automatizált megoldást dolgozok ki, ami megkönnyíti a dolgot, és közben objektív is maradhatok. Kutatásom eredményeit 2015-ben publikáltam (Nagy, 2015), 2016-ban pedig elkezdtem azt a kutatást, amelynek eredményeiről a jelen tanulmányban számolok be.

1.2. A tartalomelemzésről

A „*content analysis*” (tartalomelemzés) kifejezést először a 20. században használták (Waples, Berelson és Bradshaw, 1940), de a szövegek szisztematikus elemzésével már a 17. században foglalkoztak az egyházak, ugyanis akkoriban kezdtek terjedni a nem egyház által közreadott nyomtatványok, és erre a folyamatra aggódva tekintettek az egyházi vezetők. Először a 17. századi Svédországban végeztek olyan tartalomelemzést, amelynek módszerei hasonlítanak a mai tartalomelemző megoldásokra (Bexter és Babbie, 2003): a *Cion énekei* című 90 tételes, ismeretlen szerzőtől származó pietista zsolnárgyűjteményt elemezték abból a célból, hogy bebizonyítsák, olyan

szektás elemeket tartalmaz, amelyek veszedelmesek a hivatalos papságra. Először csak a szimbólumokat számolták össze, ám ezek számában nem találtak jelentős eltérést a jóváhagyott daloskönyvekhez képest, de amikor a szimbólumokat a kontextusukkal együtt elemezték, akkor már igen.

A ma is használatos technikák csak jóval később, az első világháborút követően alakultak ki, és abban az időben indult ugrásszerű fejlődésnek a számítástechnika is. A tartalomelemzés és a számítástechnika együttes fejlődése nem véletlen. A háborúban szemben álló felek igyekeztek olyan információkat kinyerni egymás kommunikációjából, amelyek feltárják a szöveg elsődleges jelentésén túli látens tartalmat, és ennek egyik legfőbb eszköze a számítógépek segítségével automatizált tartalomelemzés.

1.3. Az automatizált tartalomelemzésről

Fontosnak tartom kiemelni az *automatizált* jelzőt, mivel úgy gondolom, hogy az automatizált tartalomelemzés nem a tartalomelemzési technikák újabb eszköze, hanem egészen más megközelítést igénylő tevékenység. Az évtizedek alatt kidolgozott tartalomelemzési módszertanok még nem számolhattak olyan számítási teljesítménnyel, mint amilyenekre a mai számítógépek képesek. Korábban ugyan léteztek olyan tartalomelemzési technikák, amelyeket kifejezetten az automatizálás céljából hoztak létre, de a legtöbb az emberi érzékszervekre, az emberi gondolkodásra épít, így csak jelentős változtatásokkal ültethetők át modern számítástechnikai környezetbe.

Amikor ma egy nagy teljesítményű számítógépen automatizáltan elemzünk egy szöveget, akkor olyan új megoldásokat alkalmazhatunk, amelyek korábban szóba se jöhettek, mert értelmetlen lett volna kidolgozni olyan algoritmust, amely évek alatt futott volna le. Persze az automatizált tartalomelemzési technikákat veszi alapul, de sok, korábban alapvetésnek tartott megállapítás már nem állja meg a helyét, ha a mai értelemben vett automatizált tartalomelemzésről beszélünk.

2. Az automatizált tartalomelemzés lehetséges alkalmazási területei

A számítástechnika fejlődésének köszönhetően olyan új alkalmazási területek előtt nyílt meg a tartalomelemzés lehetősége, amelyeket korábban azért nem vettek számításba, mert még nem születtek meg a

szükséges technikai vívmányok. Például a nagyfelbontású fényképezőgépeknek hála, a számítógép ma már a festményeken megjelenő ecsetvonásokat is képes kiértékelni és összehasonlítani akár több ezer másik festmény ecsetvonásaival. Ezek alapján egy algoritmus képes lehet összepárosítani azokat a festményeket, amelyek egyazon ecsettel készültek, vagy ugyanazon művész munkái, sőt, az ecset nyomonvonalának mélységéből és irányából akár még az alkotó pillanatnyi lelkiállapotára is lehet következtetni. Mindezt természetesen teljesen automatizált módon, egy számítógépes algoritmus segítségével.

Az értelmezésem alapján tartalomelemzésnek nevezhetünk minden olyan tudatos vagy tudattalan cselekedetet, amely a minket körülvevő világ makro vagy mikro szintű elemzésére irányul. Tartalomelemzést végezhetünk minden olyan emberi cselekedeten, szóbeli vagy írásos megnyilatkozáson, emberi vagy természeti alkotáson, amely mögött valamilyen tartalmat vélünk felfedezni, vagy amelyekben a felszínen nem látható összefüggésekre szeretnénk rávilágítani.

2.1. Automatizált tartalomelemzés könyvtári környezetben

Könyvtárosként elsősorban az foglalkoztat, hogy az automatizált tartalomelemzés módszereit miképpen lehetne hasznosítani könyvtári környezetben. Úgy gondolom, három olyan könyvtárosi tevékenység van, ahol mindenképpen hasznos lenne automatizált tartalomelemzést végezni. Ez a három terület a vizsszakeresés, a tájékoztatás és a kutatás.

1. A visszakereshetővé tétel

Ha a könyvtárak állománya teljesen digitalizált lenne (pl. Norvégiában hamarosan digitális formában is elérhető lesz a nemzeti könyvtár teljes állománya), akkor az automatizált tartalomelemzés módszereivel olyan adatbázist tudnánk az olvasók számára elérhetővé tenni, amelyben a jelenleginél sokkal több szempont szerint szűkíthetnék a találati halmazt. Viszonylag egyszerűen megvalósítható lenne, hogy az olvasó szűrhesse különleges, egyedi szempontok szerint is, mint például, hogy kinek a szemszögéből íródott a regény (pl. lány, fiú, kisgyerek, tinédzser stb.). Ezt a nevekből, a szóhasználatból, a szavak kontextusának vizsgálatából elég nagy pontossággal meg lehetne határozni.

A visszakereshetőséghez tartozik a könyvtári feldolgozó munka, ezen belül is a tartalmi feldolgozás, mint

olyan terület, amely fejleszthető az automatizált tartalomelemzés eszközeivel. Egy dokumentum akkor válik visszakereshetővé, ha megfelelően feldolgozták. Az automatizált tartalomelemzés segítségével a feldolgozó könyvtáros egészen új módokon írhatná le a dokumentumokat. Pontosabb, szofisztikáltabb tárgyszavakkal láthatná el azokat, és akár olyan adatmezők bevezetését is lehetővé tenné, amelybe a szöveg automatizált tartalomelemzése során kinyert adatokat lehetne elhelyezni. A részletek kidolgozása egy önálló kutatást igényelne.

2. Tájékoztató

Ha a tájékoztató könyvtáros kezében lennének különböző automatizált tartalomelemzési módszereket alkalmazó eszközök, egyszerűbbé válna a munkája, és sokkal több, pontosabb találatot tárhatna az olvasó elé. Ha például az olvasó arra kíváncsi, hogy az elmúlt két évtized magyar miniszterelnökei milyen ügyekkel foglalkoztak a legtöbbit, akkor az automatizált tartalomelemzés módszerével ez könnyen megválaszolható kérdéssé válna. Ehhez az kellene, hogy a könyvtár állományában kellő mennyiségben legyenek reprezentálva a régi folyóiratok digitalizált változatai, és legyen egy olyan automatizált tartalomelemzést végző eszköz, amely képes összekapcsolni a különböző személyekhez köthető gondolatokat.

3. Kutatási feladatok

Ide sorolnám mindazon kutatási feladatokat, amelyeket a könyvtár végez, akár saját célra, akár megrendelésre. Az automatizált tartalomelemzés eszközeivel megválaszolhatók lennének olyan kérdések, ill. feltárhatók lennének olyan összefüggések, amelyek a hagyományos eszközökkel nem.

Természetesen a három felhasználási területen felül még sok könyvtári részterület profitálhatna az automatizált tartalomelemzés által nyújtott újfajta lehetőségekből. Ezek feltérképezése és megtervezése aktív párbeszédet igényel.

3. Az automatizált tartalomelemzés módszerei

Az automatizált tartalomelemzésről, illetve az alkalmazható módszerekről átfogó szakirodalmi feldolgozás nem áll rendelkezésünkre. Sokszor a tartalomelemzést alkalmazó kutatók is csak tapogatóznak, mert nincs olyan módszertani segédlet, amely egyértelműen leírja, hogy milyen feladatra, milyen algoritmust érdemes alkalmazni. Az automatizált tartalomelemzés – ahogyan azt korábban említettem

– az első számítógépek megjelenésével párhuzamosan kezdett el kialakulni, de az igazi fejlődés csak a kétezres évektől kezdődött el, ugyanis a számítógépek teljesítménye akkor érte el azt a szintet, hogy a korábban napokig tartó kiértékelési folyamat néhány órára redukálódott. A kutatásomban az írott szövegek tartalomelemzésének folyamatára fókuszálok, ezen belül is a hírportálok megjelenő tartalmak elemzésére. Teszem ezt azért, mert amatőr programozóként csak az írott szövegek tartalomelemzésének menetét tudom hitelesen bemutatni. A kidolgozott technikák alkalmazhatók más típusú szövegek elemzéséhez is, de az automatizált tartalomelemzés sajátossága, hogy nem létezik olyan univerzálisan hasznosítható megoldás, amely minden műfajú szöveg elemzésekor hatékonyan működik. Az alapelv hasonló egy hír vagy egy irodalmi szöveg elemzésénél, de az alkalmazott technikákat mindig a célnak megfelelően kell módosítani. Léteznek kereskedelmi forgalomban kapható szoftverek, amelyek képesek – elsősorban statisztikai módszerekkel – az automatizált tartalomelemzés egyes formáira, de mivel nem tudjuk, hogy ezek pontosan miként működnek, azaz, nem látjuk a működésük közben kiszámított részeredményeket, ezért nincs lehetőségünk olyan alapos ellenőrzésre, mint amikor egy konkrét adatbázishoz „szabott” programot használunk.

4. Kutatásom célja és hipotéziseim

Kutatásom célja a szövegek esetében alkalmazható legelterjedtebb automatizált tartalomelemző módszerek feltérképezése és működésük bemutatása volt annak érdekében, hogy segítséget adjak azoknak, akik szeretnék bővíteni kutatási módszereik eszköztárát. Természetesen szeretném felhívni a könyvtáros szakemberek figyelmét is az automatizált tartalomelemzésben rejlő lehetőségekre. Lehetetlen lenne valamennyi olyan területre kitérni, ahol az automatizált tartalomelemzés alkalmazható, ezért kifejezetten a *webes hírportálok* tartalomelemzésére fókuszáltam, mivel ez az a terület, ahol nagy mennyiségű szöveges tartalom található meg digitális formában.

Azt a hipotézist fogalmaztam meg, hogy viszonylag egyszerű algoritmusok segítségével is elvégezhetők olyan tartalomelemzések, amelyek az egyszerűségükhöz képest komoly eredménnyel járhatnak. A hipotézis bizonyítása vagy megcáfolása érdekében a tanulmányomban részletezett módszerek egy részét a gyakorlatban is kipróbáltam, és egy igen nagy mintán elvégeztem az automatizált tartalomelemzést.

5. A szövegek tartalomelemzésének a folyamata

A következőkben azt a folyamatot mutatom be részletesen, amelyen mindenképpen végig kell mennie annak a kutatónak, aki valamiféle automatizált tartalomelemzést kíván végezni szöveges tartalmakon. A folyamat leírásához összesen három olyan munkát (Bengtsson, 2016, Graneheim és Lundman, 2004, Varga, 2015) vettem figyelembe, amelyek jó alapot szolgáltatnak a vizsgálatom elvégzéséhez. Igyekeztem a lépéseket gyakorlatiasabb megközelítésből bemutatni úgy, hogy azok kézzelfogható segítséget nyújtsanak egy automatizált tartalomelemző szoftver elkészítéséhez. A folyamatleírást a saját szoftvereim elkészítésével párhuzamosan alakítottam ki, támaszkodva a szakirodalomra és saját gyakorlati tapasztalataimra, észrevételeimre.

5. 1. A forrás kiválasztása

Forrás lehet egy tetszőlegesen kiválasztott digitális szöveg (pl. a hírportálok cikkei vagy akár olyan korpusz, kifejezésgyűjtemény, mint amilyen az MTA Nyelvtudományi Intézetének a Magyar Nemzeti Szövegtára) vagy nyomtatott formában létező szöveg (ezeket optikai karakterfelismerő szoftverrel lehet digitalizálni). Hangos szöveg is lehet az automatizált tartalomelemzés tárgya, ugyanis léteznek olyan megoldások, amelyek elég nagy pontossággal képesek írott szöveggé alakítani a szóban elhangzott tartalmat, de használhatunk feliratsávot is. (Pl. a tévéhíradókhöz a teletexten³ elérhető a hallássérülteknek szánt feliratsáv.)

Kutatásom forrásaiként három hírportál több hónapos cikktermését használtam, ami összesen 40 540 cikket jelentett. A cikkeket automatizált módon mentettem adatbázisba úgy, hogy időzítve megnyitottam a hírportálok RSS-csatornáját, és onnan egy script segítségével adatbázisba mentettem minden tartalmat dátummal, címmel ellátva. Áprilistól júniusig csak az *Index*⁴ és az *Origo*⁵ cikkeit mentettem le, majd júliustól elkezdtem lementeni a *Kuruc.info*⁶ nevű hírportál cikkeit is, mert úgy gondoltam, hogy érdekes ellentétekre derülhet fény a másik két hírportállal szemben. Az adatgyűjtést április 1-jén kezdtem, és október 31-én fejeztem be.

5. 2. A szöveg megtisztítása és a számítógép által értelmezhetővé tétele

Bármilyen írott szöveg tartalomelemzését végezzük, biztos, hogy olyan elemeket is tartalmazni fog, ame-

lyek nem a szöveghez tartoznak. Egy könyvben ilyen az oldalszám, a fejléc és lábléc (élfője, élőláb), a hivatkozások, a hírportálok cikkeinél pedig a reklámok, a szerzők neve és a weboldal más elemei.

Nagyon kell figyelni a szöveg kódolására,⁷ mert azonos nyelven belül is sokféle kódolást használnak az elektronikus szövegek megjelenítésére, és ez problémát jelenthet több különböző forrásból származó szövegek együttes elemzésénél. Ilyen kódolások pl. az *UTF-8* vagy az *ISO-8859*. Különböző kódolásoknál egységesítés után kell felvinni a szövegeket az adatbázisba.

Miután már csak a tényleges tartalom jelenik meg az adatbázisban, ki kell szűrni azokat a szavakat, karaktereket, amelyek nem relevánsak a kutatás szempontjából. (Ezek lehetnek pl. a névelők és a központosítás, de mindig a konkrétan elérni kívánt cél határozza meg azt, mit tekinthetünk irrelevánsnak.)

A tartalomelemzés szempontjából releváns részeknél, meg kell adni a szoftver számára, hogy mit tekintsen teljes mondatnak, mondatkezdő nagybetűnek vagy tulajdonnévnek. Ez azért jelent nehézséget, mert a számítógép nem értheti, hogy egy szövegrészletben azért van pont, mert lezár egy mondatot vagy azért, mert rövidítést jelöl. Egy lehetséges megoldás, hogy a pontot akkor tekintjük mondatzáró pontnak, ha utána nagybetű vagy már semmi nem szerepel, de ilyenkor sem biztos a siker, hiszen a pont utáni nagybetű jelölhet mondaton belüli tulajdonnevet is.

5.3. Milyen kérdésekre keressük a választ az automatizált tartalomelemzés segítségével, milyen információkat szeretnénk kinyerni a szövegből?

Olyan témát választottam a szemléltetésképpen elvégzett automatizált tartalomelemzésemhez, amelyről tudtam, hogy sok cikk jelenik meg róla a három hírportálon, és várhatóan az adatgyűjtés hét hónapja alatt is publikálnak még róla. A választott témám a „bevándorlási hullám” volt. Arra voltam kíváncsi, hogyan változik az események tálalása az áprilistól októberig terjedő időszakban, milyen kontextusban írnak a bevándorlásról, és van-e valamilyen különbség abban, ahogy a három hírportál beszámol a fejleményekről.

5.4. A szóba jöhető automatizált tartalomelemzési módszerek kiválasztása

Ehhez kapcsolódóan nem létezik olyan komplex lis-

ta, amely felsorolja az összes lehetséges módszert, mert az automatizált tartalomelemzés ugyan átvette a legtöbb klasszikus tartalomelemző metódust, de a számítógépek alkalmazásával szinte végtelenné vált a lehetséges megoldások tárháza. Kutatásomban a legelterjedtebb, leginkább kézzelfogható módszerekkel foglalkoztam. Ahogy azt korábban kifejtettem, szerettem volna elkerülni, hogy a tanulmányom egy száraz technikai ismertető legyen, ehelyett igyekeztem úgy bemutatni az eredményeimet, hogy azok kézzelfoghatók, a gyakorlatba is könnyen átültethetők legyenek. A következőkben a legelterjedtebb automatizált tartalomelemzési módszereket ismertettem. Ezek közül többet fel is használtam az elvégzett tartalomelemzésekben. (Az eredményekről az egyes módszerek ismertetéséről szóló fejezetekben számolok be.)

5.4.1. Gyakoriság-elemzés

A gyakoriságelemzés a legegyszerűbb módszer, amellyel szám szerint megvizsgáljuk, hogy egy meghatározott adat hányszor fordul elő a szövegben. „*Ez talán az egyik legegyszerűbb, a legkönnyebben gépesíthető módszer.*» Minden olyan esetben, amikor a megfigyelt gyakoriságok arányai 'meglepő'-ek, arra van szükség, hogy explicitte tegyük azt a megoszlást, amellyel ezeket a megfigyeléseket összevetettük, s hogy maga az összevetés valóban igazolható legyen.» (Krippendorff, 1995). „*Vagyis mindig ellenőrizni kell, hogy a gyakoriság mihez képest nagy vagy kicsi. Ennek érvényesnek kell lennie a kulcsszó-meghatározásra is.*” (Varga, 2005. 53. o.)

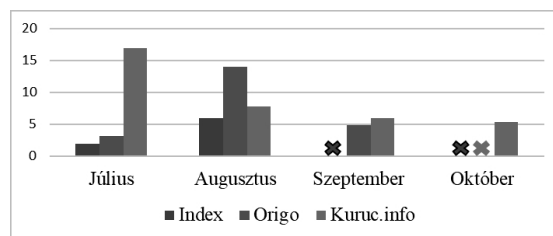
5.4.2. A változók közötti relációk elemzése

A relációk elemzésével bizonyos adatok együttes előfordulását térképezhetjük fel. Így pl. megállapítható, hogy egy adott témáról hogyan vélekedik két különböző ember, mégpedig úgy, hogy a választott témára jellemző szavak környezetét vizsgáljuk külön az egyik és külön a másik személy esetében. Ezt a módszert arra használtam, hogy vettem a „*bevandorló*” kifejezést és annak összes lehetséges szinonimáját, és ezeket azonosként kezelve feltérképeztem azokat a kifejezéseket, amelyekkel rendszeresen együtt szerepelt. Ezeket számszerűsítettem, és két csoportot alkottam belőlük. Az egyikbe az országok és nemzetek nevei kerültek; a másikba pedig a kifejezetten negatív érzelmet kiváltó kifejezések (pl. „terrorista”, „öngyilkos merénylő” stb.).

Aztán hírportálok szerint, hónapokra lebontva meg-

néztem azt a 25 kifejezést, amelyek leggyakrabban szerepeltek együtt a „*migráns*” kifejezéssel és annak szinonimáival. A leggyakrabban előforduló kifejezések ezek voltak: *terrorista, terrorizmus, Németország, Magyarország, Törökország, kormány, befo-gadás, unió, kvóta.*

Ezután kiválasztottam azokat a kifejezéseket, amelyek kifejezetten negatív tartalmat hordoznak, és diagramon ábrázoltam az eredményeket. (Azért júliustól és nem áprilistól indulnak a diagramok, mert csak júliustól kezdve volt adatom mindhárom hírportálról.)

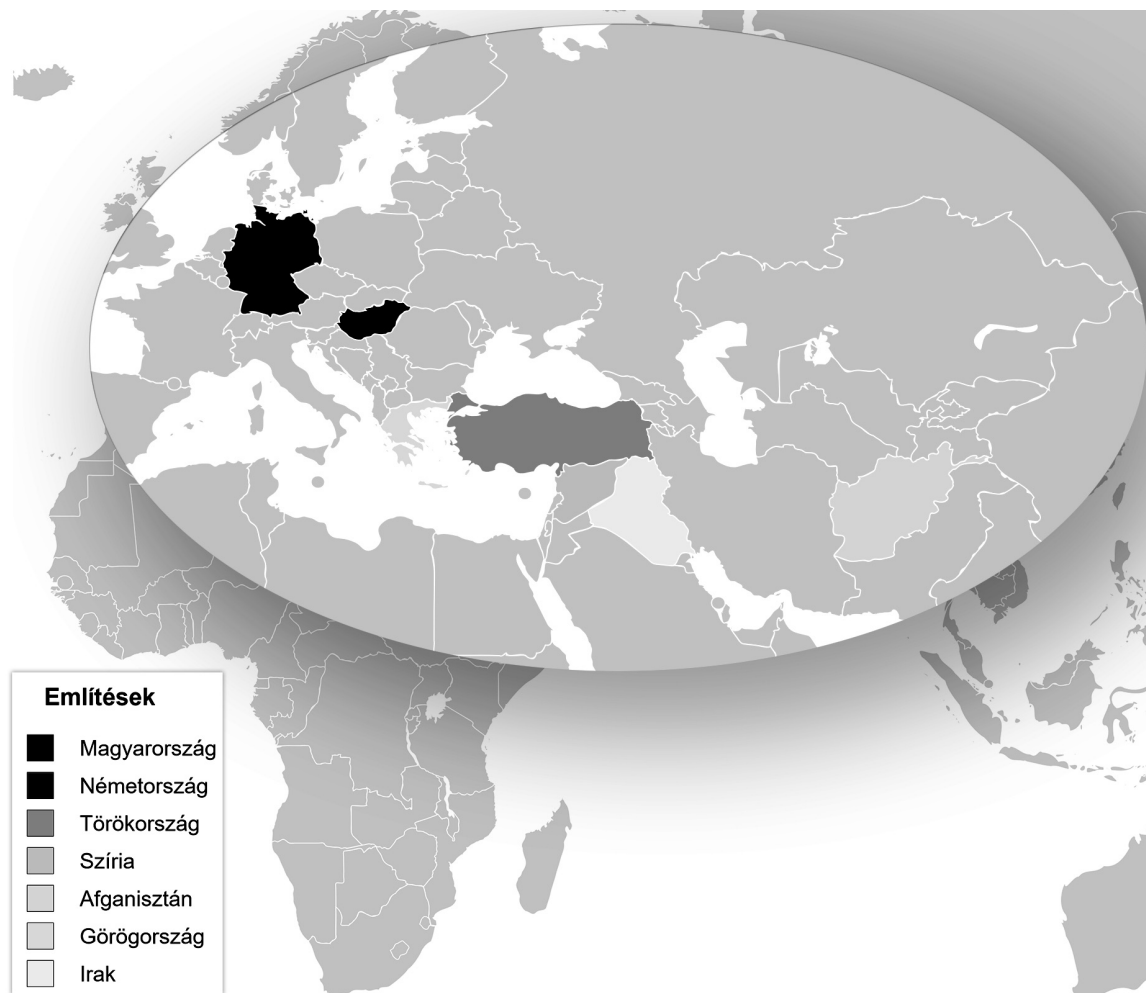


1. ábra

Negatív érzelmet kiváltó kifejezések használata (%)

Az 1. ábrán látható, hogy a hírportálok százalékos értékei nem egymáshoz viszonyulnak, hanem azt mutatják, hogy egyenként, az egyes hírportálokban a „*migráns*” és szinonimáival gyakran együtt használt kifejezések hány százaléka egyértelműen negatív. A diagramon látható, hogy augusztust kivéve a Kuruc.info-n használták a legtöbb negatív érzelmet kiváltó kifejezést a migránsok kapcsán. A sorban a második az Origo, míg az Indexen csak ritkán írtak negatív kontextusban a migránsokról. (Szeptemberben és októberben egyetlen negatív kifejezés sem jelent meg a „*migráns*” és szinonimáinak szöveggörnyezetében). Augusztusban tetőzött a negatív jellegű kifejezések használata, összefüggésben azzal, hogy akkor kezdődött a riói nyári olimpia, és feltehetően a biztonsági intézkedések kapcsán használták gyakrabban a „terrorista”, „terrorizmus”, „szükségállapot” kifejezéseket.

A második csoport az országok és népek neveit tartalmazta. Az egyszerűség kedvéért nem tettem különbséget az országnév és a hozzá tartozó népnév említése között, tehát pl. a „Magyarország” és a „magyar” kifejezéseket azonosként kezeltem. A vizsgált időszakban összesen hét ország (és nép) neve szerepelt együtt a cikkekben a „bevandorló”-val, illetve annak rokon értelmű kifejezéseit tartalmazó szöveggörnyezetben. A gyakoriságvizsgálat eredményét a 2. ábra mutatja.

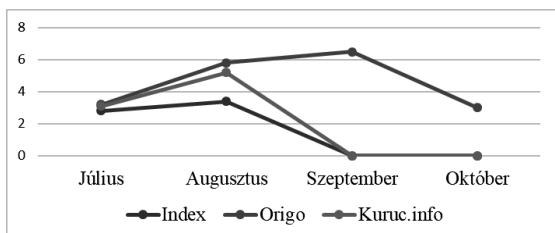


2. ábra

Az egyes országok említésének gyakorisága

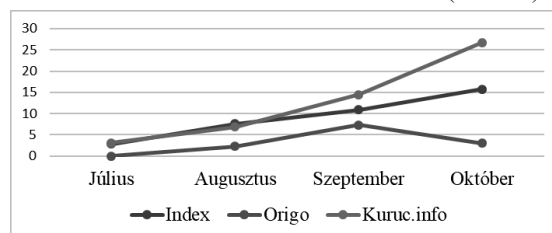
A diagramon látható, hogy a legsötétebb színnel jelölt országot a legtöbbször, a legvilágosabb színnel jelölt országot a legkevésbé hozták szóba a hírportálok.

A Törökországról való diskurzus az Indexen és a Kuruc.infón augusztusban tetőzött, az Origo cikkeiben viszont még szeptemberben is gyakran említették az országot, és ellentétben a másik két hírportállal, októberben is több alkalommal szóba hozták (3. ábra).



3. ábra

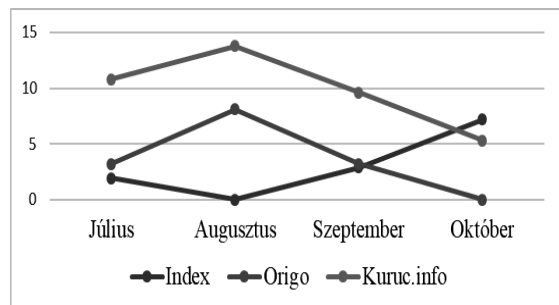
Törökország említésének gyakorisága (súlyozás utáni darabszám)



4. ábra

Magyarország említésének gyakorisága (súlyozás utáni darabszám)

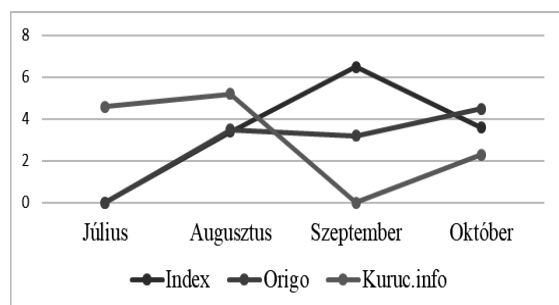
A „Magyarország/magyar” kifejezések előfordulását vizsgálva, ellentétes irány figyelhető meg az előző diagramhoz képest, ami feltehetően az októberi népszavazás⁸ következménye volt, mert a sajtót inkább a helyi események érdekelték (4. ábra).



5. ábra

Németország említésének gyakorisága (súlyozás utáni darabszám)

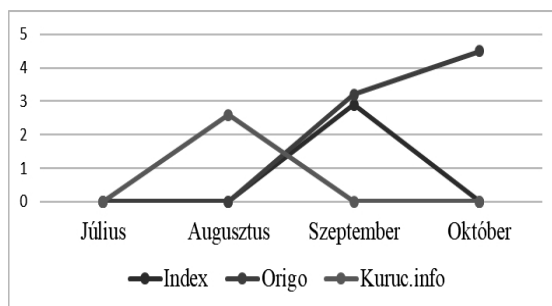
A Németországgal kapcsolatos kifejezéseket az Origo és a Kuruc.info szinte mindig ugyanakkor használta, az Index viszont teljesen szembe ment ezzel a tendenciával. Ennek több oka lehet, pl. elképzelhető, hogy a Kuruc.info és az Origo egymástól vagy egy közös hírforrásból vettek át híreket, esetleg az Index kevésbé tartotta lényegesnek a németországi eseményeket (5. ábra).



6. ábra

Szíria említésének gyakorisága (súlyozás utáni darabszám)

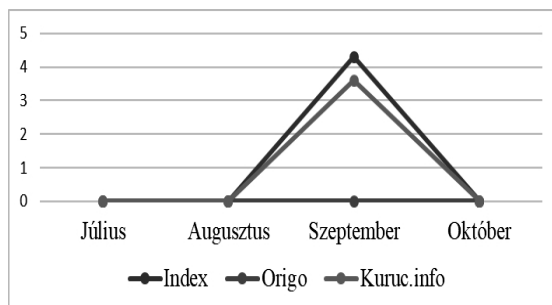
Úgy tűnik, hogy Szíria említésére nagyon különböző időpontokban került sor, de augusztusban és októberben mindhárom hírportál közel ugyanannyi alkalommal említette meg az országot. Utánanéztem, mi lehet ennek az oka; augusztusban többek között arról lehetett olvasni mindhárom hírportálon, hogy a szírek és törökök több települést is visszafoglaltak az Iszlám Államtól,⁹ októberben pedig egy szír bevándorló megkísérelt bombát robbantani egy német repülőtéren,¹⁰ illetve a magyar határnál határőrré támadt egy szír menekült.¹¹ (6. ábra).



7. ábra

Görögország említésének gyakorisága (súlyozás utáni darabszám)

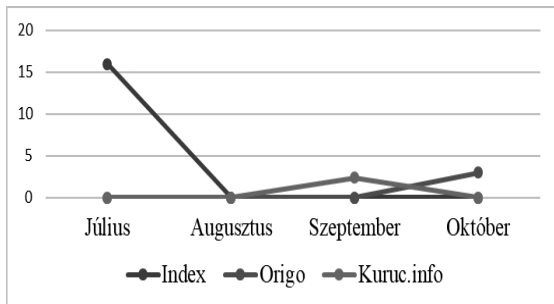
Görögországot viszonylag kevés alkalommal említették az elmúlt négy hónapban a bevándorlók kapcsán, de szeptemberben az Index és Origo szinte ugyanannyiszor írt az országról. A gyakori említésnek az egyik görög menekülttáborban történt gyűjtogatás lehetett az oka.¹² (7. ábra).



8. ábra

Irak említésének gyakorisága (súlyozás utáni darabszám)

Irak említését az Origo szinte teljesen mellőzte az elmúlt négy hónapban (legalábbis a „bevándorló” és annak szinonimái vonatkozásában). Az Indexnél és Kuruc.infónál is csak szeptemberben írtak az országról. Rövid kutatómunka után kiderült, hogy az amerikai és iraki katonai csapatokról publikált¹³ olyan híreket az Index és a Kuruc.info, amelyet az Origo kevésbé tartott fontosnak (8. ábra).



9. ábra

*Afganisztán említésének gyakorisága
(súlyozás utáni darabszám)*

Afganisztánról nem sokat lehet olvasni a magyar sajtóban a bevándorlókkal összefüggésben, de a júliusi kiugró értékének utánanézttem. Úgy tűnik, hogy a kiugró értéket az afgán menekültek helyzetéről szóló hosszú, helyzetelemző cikkel¹⁴ érte el az Index. Nem konkrét időponthoz kötődő eseményről számolt be, tehát a másik két hírportálon ezért nem említették Afganisztánt a bevándorlók említése kapcsán (9. ábra).

5.4.3. Kontingencia-elemzés

Ezzel a módszerrel az együttes előforduló kifejezések közötti kapcsolatot vizsgálhatjuk meg. Megnézhetjük például, hogy egy-egy párban álló kifejezés milyen gyakran jelenik meg egy megadott szövegrészen belül vagy, hogy az együttes megjelenésük kizárja-e más szövegpárok felbukkanását.

Olyan algoritmust készítettem, amellyel listáztam az összes együttes előfordulást az Index és az Origo teljes áprilisi és októberi cikktermésében, illetve a Kuruc.info júliusi és októberi cikkeiben. Reméltem, így kiderül, hogy az egyes hírportálok mely kifejezéseket szeretik más kifejezésekkel együtt említeni (pl. milyen kontextusban használják a különböző közszerplők neveit).

Az algoritmus elkészítése, lefuttatása és az eredmények kiértékelése után azt tapasztaltam, hogy az eredményekből sajnos nem olvasható ki semmilyen összefüggés vagy mélyebb tartalom. Lehet, hogy ez az algoritmusom hibája volt, de az is lehet, hogy ebben a formában a módszer nem volt célravezető. Bár sajnálom, hogy kontingencia-elemzéssel nem tudtam feltárni semmilyen hasznos információt, de legalább egy tévutat is be tudok így mutatni, szemléltetve, milyen buktatók lehetnek az automatizált tartalomelemzés folyamatában. Az egyik buktató tehát az, hogy csak az algoritmus megírása után derül ki, érdemes

volt-e elkészíteni a programot, vagyis kimutatható-e az elemezni kívánt szövegből bármilyen összefüggés, látens tartalom. Nincs rá garancia, hogy a program megírásába fektetett idő és pénz végül megtérül, de ezzel számolni kell.

5.4.4. Klaszterálás

Amikor már túl sok együttes előfordulást találtunk, lesz egy értelmezhetetlen adathalmazunk. Ilyenkor segít a klaszterálás technikája. A folyamat során az összetartozó, hasonló jelentéssel bíró csoportokat összevonjuk, és a későbbiekben együttesen kezeljük őket. Nem vizsgáljuk meg mindig külön-külön, hogy milyen más szó párokkal állnak kapcsolatban, hanem a nagyon hasonló szó párokat egyszerűen azonosnak tekintjük. Ez kicsit hasonlít a könyvtári osztályozásra.

5.4.5. Kontextuális osztályozás

Ezzel a módszerrel a kifejezések szövegek környezetét vizsgáljuk, és annak alapján próbálunk rokon értelmű szövegrészeket, kifejezéseket találni, hogy mennyi közös vonás van az egyes kifejezések nyelvi környezetében. Nyilván minél több a közös vonás két különböző szó nyelvi környezetében, annál inkább értékelhetjük szinonimaként azokat. Ha pl. a „kutyá” és az „eb” kifejezés is rendszeresen a „kutyaház” kifejezéssel szerepel együtt, akkor beállíthatjuk a tartalomelemző szoftverben azt, hogy amennyiben ez sokszor megismétlődik, akkor kutyát és az ebet kezdje el azonos kifejezésként kezelni.

5.4.6. Szótár alapú tartalomelemzés

A szótár alapú tartalomelemzés viszonylag egyszerű módszer, de a számítógép szempontjából eléggé erőforrás-igényes. A lényege, hogy veszünk valamilyen tematikus szótárat, amely egy meghatározott témakört fed le, és az ott szereplő összes szót összehasonlítjuk a tartalomelemzés tárgyát képező szövegekkel.

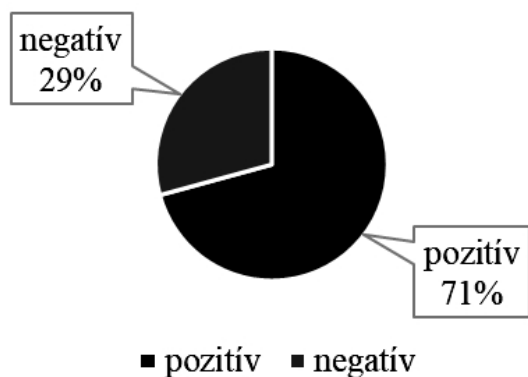
A módszert annak megállapítására használtam, hogy a három hírportálon milyen megoszlásban jelennek meg pozitív és negatív tartalmú cikkek. Pozitív hír például, ha egy zsiráf bébi születéséről írnak, negatív pedig, amikor bankrablásról.

Az első lépéshez az ingyenesen használható *Szósablya*¹⁵ nevű weboldalt hívtam segítségül. A weboldalt a Budapesti Műszaki és Gazdaságtudományi Egyetemen fejlesztik. A weboldal felületén be lehet írni egy szót, amire az oldal válaszként néhány másik adat

kísérletében megadja a kifejezés szótövét és szófaját. Egy script¹⁶ segítségével összesen 9 millió 831 ezer szót küldtem be százaz tömbökben¹⁷, és így olyan adatbázist építettem fel, amelyben minden egyes szó szótöve és szófaja is szerepel.

Erre a melléknév-szótáram elkészítéséhez volt szükség. Kézi módszerrel kiválasztottam ötszáz olyan melléknévet, amelyről viszonylag egyértelműen megállapítható, hogy pozitív vagy negatív tartalmat hordoz. Ilyen pl. a „díjnyertes”, „hűség”, „megbízható”, a másik oldalon pedig a „radioaktív”, „igazságtalan”, „fasiszta”. Természetesen a szövegtörzstől függően egy-egy pozitívnek vagy negatívnak vélt melléknév jelenthet egészen mást is, mint amire a szótár összeállításánál gondoltam, de nem is azt vártam, hogy egy ilyen melléknév-szótárral teljes biztonsággal tudom majd kategorizálni a cikkeket pozitív és negatív tartalom szerint, hanem kísérletként tekintettem rá. Kíváncsi voltam, hogy mennyire működőképes egy ilyen megoldás.

Miután elkészítettem a pozitív/negatív melléknév-szótárat, összehasonlítottam a 40 ezer cikk 9 millió szavával, és így megkaptam, melyik cikkben hány pozitív és hány negatív tartalmú melléknév jelenik meg. Azt a cikket ítéltam negatív vagy pozitív tartalmúnak, amelyben legalább öt szó felbukkant a melléknév-szótáramból, és legalább 60%-kal több pozitív melléknévet tartalmazott, mint negatívát (vagy fordítva).



10. ábra

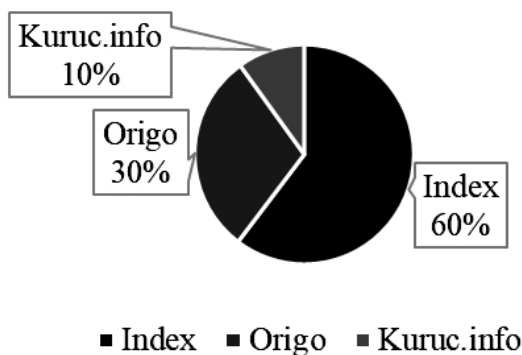
A pozitív és negatív töltetű cikkek megoszlása a három hírportálon, összesítve

Meglepő módon a módszer nagyon sikeresnek bizonyult: az algoritmus nagyon sok cikkről pontosan meg tudta állapítani, hogy pozitív vagy negatív (a hiábaarány mindössze 5%-os volt). Az viszont kiderült, hogy az ötszáz szavas melléknévszótár nagyon kicsi

ahhoz, hogy ilyen sok cikkre alkalmazható legyen. Így csak néhány száz olyan cikket talált a programom, amelyben szerepeltek kifejezések a melléknév-szótáramból (10. ábra).

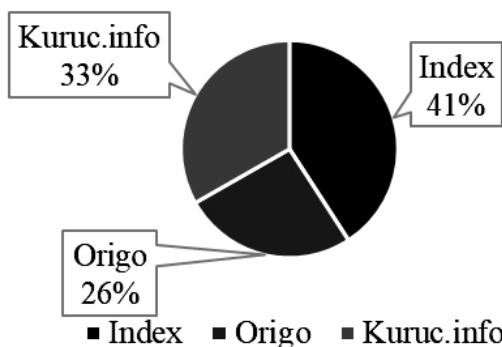
Ez nem meglepő eredmény: a statisztikában a tudományos, kulturális és sportrovatok cikkei is szerepelnek; ha csak a politikai és gazdasági híreket néznénk, akkor valószínűleg máshogy nézne ki a diagram.

Mivel az Index és Origo publikált tartalmait áprilistól gyűjtöttem, a Kuruc.infoét pedig csak júliustól – ráadásul nem is azonos számú cikk jelent meg az egyes hírportálokon –, ezért valahogy meg kellett oldanom, hogy összehasonlíthatók legyenek az adatok. Ehhez kiszámoltam, hogy az egyes hírportálokon havonta átlagosan hány cikk jelenik meg, majd úgy súlyoztam az értékeket, hogy a pozitív és negatív melléknévszámok összehasonlíthatóak legyenek (11. ábra)



11. ábra

A pozitív töltetű cikkek megoszlása a három hírportál között (súlyozás után)



12. ábra

A negatív töltetű cikkek megoszlása a három hírportál között (súlyozás után)

Érdekes, hogy az Index mind a pozitív, mint a negatív melléknévekben vezet, ez valószínűleg azért van,

mert vannak rovatai, amelyekben szinte csak pozitív hangvételi cikkek jelennek meg, és vannak olyanok, amelyekben nagyon sok a borús képet festő cikk, ezek valószínűleg a Belföld és a Külföld rovatban foglalnak helyet (12. ábra).

Az Origón fele annyi pozitív tartalmú cikk jelent meg, mint az Indexen, viszont negatív sincs sok, ebből arra következtetek, hogy sok olyan cikket publikált, amelyek semleges hangvételűek vagy a melléknév-szótáram viszonylag kis mérete miatt nem kerültek be a statisztikába.

A Kuruc.infón mindössze 10% a másik két hírportálhoz viszonyított pozitív cikkek aránya, az összes többi vagy negatív hangvételi vagy semleges.

5.4.7. Több módszer összefűzése

Korábban már leírtam, melyek azok a klasszikus tartalomelemző módszerek, amelyeket részben az automatizált tartalomelemzés is átvett. Amikor automatizált módszerekkel szeretnénk különféle összefüggésekre fényt deríteni, és feltárni egy-egy szöveg látnens tartalmát, szinte soha nem elég csak egyféle módszert alkalmazni. Erre jó példa az, hogy az ismertett tartalomelemzések során minden esetben éltem a *gyakoriságelemzés* módszerével is, még akkor is, ha nem konkrétan a gyakoriságelemzés módszertanát szerettem volna szemléltetni.

Készítettem egy olyan programot is, amely egyszerre támaszkodik a *gyakoriságelemzés* módszerére, a *szótár alapú tartalomelemzésre* és a *klaszterálásra*, pedig a program „mindössze” azt vizsgálja meg, hogy 80 nap leforgása alatt hány olyan esemény volt, amelyről nagyon hasonló módon írt cikket az Index és az Origo.

Mivel a számítógép, amelyen a tartalomelemzést végeztem, korlátozott erőforrásokkal rendelkezik, ezért csak a cikkek egynegyedével, 10 000 cikkel dolgoztam. (Ez a mennyiség a 2016. március 21. és május 31. közötti időszakot öleli fel).

A hírportálok rendszeres olvasójaként az volt a hipotézisem, hogy naponta több olyan cikk is megjelenik a két weboldalon, amelyek közel azonosak nemcsak a téma szerint, hanem a szóhasználat tekintetében is. Ennek több oka is lehet, pl. az, hogy közös hírforrást használnak (pl. a Magyar Távirati Irodát)¹⁸, de egymástól is vehetnek át híreket.

A programom működési mechanizmusa a következőképpen alakult.

A korábban már említett Szószablya nevű weboldal segítségével megállapítottam a tízezer cikk összes

szavának a szótóvét, és kiszűrtem mindazokat, amelyeket az elemzés szempontjából nem tartottam relevánsnak (névelők és speciális karakterek).

Eltávolítottam a cikkeken belül ismétlődő szóelőfordulásokat. Erre azért volt szükség, mert megfigyeltem, hogy a kezdeti próbálkozásaim sikertelensége arra vezethető vissza, hogy a két hírportál közül az egyikben sok esetben terjedősebben számoltak be az egyes eseményekről, és így hiába egyezett meg szinte teljes egészében két cikk, a szóismétlések miatt a tartalomelemző algoritmusom nem tekintette egyezőnek cikkeket. Többféleképpen korrigálhattam volna az algoritmuson, de végül néhány próbaelemzés után arra jutottam, hogy az ismétlődő szóelőfordulások eltávolításával tudom elérni a legpontosabb eredményt.

A fenti lépések után gyakorlati szempontból mind a tízezer cikk egy-egy szótárrá vált. Ezután már csak össze kellett hasonlítanom az összes „szótár” szókészletét a többi „szótárával”.

Ezt követően meg kellett határoznom, mikor tekinthető egyezőnek két „szótár” (cikk). Ezt egyszerű kísérletezéssel próbáltam megállapítani. Első körben akkor tekintettem egyezőnek két cikket, ha azok szókészlete legalább 30%-ban megegyezett. Így sok olyan cikket is egyezőnek tekintett az algoritmus, amelyek valójában teljesen másról szólnak, ezért addig növeltem a határt egy-egy százalékkal, mire közel teljesen pontos eredményeket kaptam. Ez 49%-nál következett be, tehát elmondható, ha pusztán a cikkek szókészlete összehasonlításának útján szeretnénk összepárosítani a közel azonos tartalmú cikkeket, akkor legalább a kifejezéseik 49%-ának kell megegyezniük ahhoz, hogy összetartozónak tekinthessük azokat.

Ezekre az eredményekre jutottam:

Átnézve a 2016. március 21. és május 31-e közötti időszakot, összesen 10 000 cikkel dolgoztam. A cikkek az Index és az Origo hírportálokról származnak, de nem teljesen egyenlő az eloszlásuk, a vizsgált időszakban 5425 cikk jelent meg az Indexen, 4575 az Origón.

A 10 000 cikk között 1383 volt olyan, amelyet a tartalomelemző programom egyezőnek tekintett. Ez azt jelenti, hogy az Index 5425 cikkének 25,5%-a hasonló formában megtalálható az Origón is. Az Origo 4575 cikke között 30,2% volt azon cikkek aránya, amelyek nagyon hasonlítottak az Index cikkeire.

Természetesen ezek az eredmények nem tekinthetők teljesen pontosnak. Az is elképzelhető, hogy több, de

az is, hogy valamivel kevesebb a szinte teljesen egyező cikkek aránya a két hírportálon. Az első kétszáz egyező cikk alapos szemrevételezése után 14 esetben találtam olyan párt, amelyet nem tartok egyezőnek. Ez 7%-os hibaarány, és bár ez nem akkora minta, hogy általánosíthatassunk, arra alkalmas, hogy lássuk, már egy viszonylag egyszerű algoritmus is aránylag nagy pontossággal képes összepárosítani az összetartozó szövegeket.

Pontosabb eredményeket csak egy jóval kifinomultabb automatizált tartalomelemzési algoritmus segítségével lehetne elérni. A tanulmányomnak nem az volt a célja, hogy tökéletesen pontos adatokkal szolgáljak a magyar hírportálokról; hiszen a programokat azért készítettem, hogy hitelesen és kézzelfogható, közérthető módon tudjak írni az automatizált tartalomelemzés módszereiről, lépéseiről, és mindezek hasznosságát konkrét példákon keresztül szemléltessem.

5.5. Az elkészített algoritmusok futtatása az összes szövegen és az eredmények tárolása az adatbázisban

Figyelni kell arra, hogy az eredményeket úgy tároljuk el, hogy azok később számszerűsítve is felhasználhatók legyenek, pl. a felismert főnevek legyenek 1-essel jelölve az adatbázisban, a melléknevek 2-essel és így tovább. Számokkal mindig egyszerűbb dolgozni: pl. ha átvisszük a kapott adatokat Excelbe, akkor ott is sokkal könnyebb diagramokat, kimutatásokat készíteni, ha minden tulajdonságot egy meghatározott számmal látunk el.

5.6. Ellenőrzés

Ez az egyik legfontosabb lépés az automatizált tartalomelemzés során, és nem szabad abban bízni, hogy ha egy algoritmus működött egy konkrét szövegcsoporton, akkor egy másikon is ugyanúgy működni fog. Az automatikusan kiszámolt eredmények közül néhányat érdemes manuálisan is átszámolni, illetve mindig meg kell nézni, mi az oka a kiugró értékeknek.

6. Következtetések

Tanulmányom írása során az jelentette a legnagyobb nehézséget, hogy úgy fogalmazzak meg megállapításokat az automatizált tartalomelemzés módszereiről és folyamatáról, hogy közben tisztában voltam azzal,

bármit is írok le, minden kijelentésemet kezdhetem úgy, hogy „például”, vagy úgy, hogy „jelen pillanatban”. Annyira dinamikusan változó tudományterületről van szó, hogy mire valaki alaposan kidolgoz egy automatizált tartalomelemzési módszert, elkészíti a dokumentációját, megírja a szükséges algoritmust, és publikálja eredményeit, lehet, hogy addigra már el is veszítette az aktualitását. A programozási nyelvek folytonos átalakulása, a számítógépek kapacitásának exponenciális fejlődése és az új technikai vívmányok megjelenése következtében újra és újra át kell értékelnünk azokat a megállapításokat az automatizált tartalomelemzés módszereit illetően, amelyeket korábban tényként kezeltünk. Persze mindig lesznek olyan alapvető módszerek, amelyeket valamilyen formában a jövőben is használni fognak a szövegek automatizált elemzése során, de ezek technikai megvalósítása az informatikával együtt fog fejlődni, átalakulni.

A hipotézisem az volt, hogy viszonylag egyszerű algoritmusok segítségével is olyan – hasznos – eredményekhez juthatunk, amelyeket nem lehetett volna elérni a szövegek „manuális” elemzése során. Úgy gondolom, hogy a hipotézist igazoltam, hiszen sikerült olyan összefüggéseket feltárnom az Index, az Origo és a Kuruc.info nevű hírportálok tartalmaiban, amelyek számszerűek, pontosak és objektívek. A bemutatott tartalomelemzési módszerekkel és technikai megoldásokkal természetesen nemcsak webes hírportálok automatizált tartalomelemzése valósítható meg, hanem átültethetők bármilyen más területre is, ahol szövegekkel dolgoznak.

Nagyon fontosnak tartom hangsúlyozni, hogy a tanulmányomban ismertetett módszerek csupán egy kis részét fedik le mindannak, amit automatizált tartalomelemzésnek tekinthetünk. Kutatásomban arra kerestem választ, hogyan segíthetnénk a tudományos munkát egyszerű algoritmusok segítségével. A kézzelfogható példák bemutatásával szeretném felkelteni az automatizált tartalomelemzés iránti érdeklődést, elsősorban könyvtári vagy a könyvtárhoz közel álló területeken. Úgy gondolom, hogy az automatizált tartalomelemzés alkalmazása a könyvtári szolgáltatások minőségi javulását eredményezné.

Jegyzetek

1. Index <http://index.hu> (2016. okt. 1.)
2. Origo <http://origo.hu> (2016. okt. 1.)
3. PCMag <http://www.pcmag.com/encyclopedia/term/52714/teletext> (2016. dec. 27.)
4. Index <http://index.hu> (2016. okt. 01.)
5. Origo <http://origo.hu> (2016. okt. 01.)
6. Kuruc.info. <http://kuruc.info> (2016. okt. 1.)
7. TechTerms <http://techterms.com/definition/characterencoding> (2016. okt. 20.)
8. Magyarország Kormánya <http://nepszavazas2016.kormany.hu> (2016. nov. 20.)
9. Index http://index.hu/kulfold/2016/08/24/meg_egy_varost_felzabadiottak_a_sziriai_felkelok (2016. nov. 5.)
10. Origo <http://www.origo.hu/nagyvilag/20161008-nagyszabasu-rendori-akcio-a-nemetorszagi-chemnitzben-robbantasra-keszulhettek.html> (2016. nov. 5.)
11. Index. http://index.hu/belfold/2016/10/16/karoval_tamadt_egy_hatarvadasra_a_szir_menekult (2016. nov. 5.)
12. Index http://index.hu/kulfold/2016/09/19/tuz_miatt_kiuritetek_egy_gorog_menekulttabort (2016. nov. 5.)
13. Kuruc.info <https://kuruc.info/r/4/163910> (2016. nov. 10.)
14. Index http://index.hu/kulfold/2016/07/27/europa_teljesen_felkeszuletlen_pedig_ujabb_menekulthullam_johet (2016. nov. 10.)
15. Szószablya <http://szotar.mokk.bme.hu/szoszablya/searchw.php> (2016. dec. 5.)
16. TechNetwork <https://pcforum.hu/szotar/?term=script&tm=miaz> (2016. nov. 5.)
17. Carnegie Mellon University web <https://www.cs.cmu.edu/~adamchik/15-121/lectures/Arrays/arrays.html> (2016. okt. 29.)
18. Magyar Távirati Iroda <http://www.mti.hu/mti/Default.aspx> (2016. nov. 10.)

Irodalom

- ANTAL László: A tartalomelemzés alapjai. Budapest, Magvető Könyvkiadó, 1976. 151 p.
- BAXTER, Leslie A. – BABBIE, Earl R.: The basics of communication research. Belmont, Wadsworth Publishing, 2003. 240. p.

- BENGTSSON, M.: How to plan and perform a qualitative study using content analysis = NursingPlus Open, 2. vol. 2016. 8–14. p. <http://www.sciencedirect.com/science/article/pii/S2352900816000029> (2018. jan. 15.)
- Ehmann Bea – Balázs László – László János – Gushin, V. (szerk.): Izolált kiscsoportok pszichodinamikája: a Mars-500 űranalóg szimuláció legénységi kommunikációjának tartalomelemzéses vizsgálata. In: Vargha András (szerk.): A tudomány emberi arca. A Magyar Pszichológiai Társaság XXI. Országos Tudományos Nagygyűlése. Kivonatkiötet. Budapest, Magyar Pszichológiai Társaság, 2012, 239. p.
- GÉRING Zsuzsanna: Tartalomelemzés: A virtuális és a valós világ határán: Egy vállalati honlap-elemzés bemutatása = Kultúra és Közösség, 5. évf. 2014, 1. sz. http://publikaciotar.repositorium.bgf.hu/611/1/Gering_tartalomelemzes_2014.pdf (2016. dec. 29.)
- GRANEHEIM, Ulla H. – LUNDMAN, B.: Qualitative content analysis in nursing research: concepts, procedures and measure to achieve trustworthiness = Nurse Education Today, 24. vol. 2004, 2. no. 105–112. p. [http://www.nurseeducationtoday.com/article/S0260-6917\(03\)00151-5/abstract](http://www.nurseeducationtoday.com/article/S0260-6917(03)00151-5/abstract) (2017. szept. 29.)
- HSIEH, H. F. – SHANNON, S.E.: Three approaches to qualitative content analysis = Qualitative Health Research, 15. vol. 2005. 9. no. 1277–1288. p.
- KRIPPENDORFF, Klaus: A tartalomelemzés módszertanának alapjai. Budapest, Balassi Kiadó, 1995. 119 p.
- KRIPPENDORFF, Klaus: Content analysis. An introduction to its methodology. London, SAGE Publishing, 2013.
- KRIPPENDORFF, Klaus: An examination of content analysis. A proposal for a framework and an information calculus for message analytic situations. [Doctoral thesis]. Urbana, University of Illinois, 1967. 412 p. <https://pdfs.semanticscholar.org/67b7/3bc81b28bbafc7319505ce9f081cb8a2bb04.pdf> (2018. jan. 15.)
- László János – Ehmann Bea (szerk.): LAS Verticum. Egy szó feletti tartalomelemző szoftver In: Magyar Számítógépes Nyelvészeti Konferencia. Szeged, MSZNY, 2003. 288–289. p.
- LOWE, Will: Understanding wordscores = Political Analysis, 16. vol. 2008. 4. no. 356–371. p. <http://faculty.washington.edu/jwilker/559/Lowe.pdf> (2018. jan. 29.)
- NAGY Andor: Az automatizált tartalomelemzés lehetőségei = Tudásmenedzsment, 16. évf. 2015. 1. sz. 132–139. p. http://epa.oszk.hu/02700/02750/00037/pdf/EPA02750_tudasmenedzsment_2015_01_132-139.pdf (2018. jan. 15.)

NEUENDORF, Kimebly A.: The content analysis guidebook. London, SAGE Publishing, 2016. 120 p.
https://books.google.hu/books?id=huPVtmu4sigC&printsec=frontcover&redir_esc=y#v=onepage&q&f=false (2018. jan. 29.)

TESCH, Renata.: Qualitative research. Analysis types & software tools. Bristol, Falmer Press, 1990. 330 p.

VARGA Katalin: Szöveg és tartalom az információs társadalomban. Pécs, Pécsi Tudományegyetem Felnőttképzési és Emberi Erőforrás Fejlesztési Kar, 2005. 164.p.

WAPLES, Douglas – BERELSON, Bernard – BRADSHAW, Franklyn. R.: What reading does to people. A summary of evidence on the social effects of reading and a

statement of problems for research. Chicago, University of Chicago Press, 1940. 9. p. <https://archive.org/details/whatreadingdoest00wapl>

WEBER, Robert P.: Basic content analysis. Newbury Park, Sage Publications, 1990. 96 p. https://wrt303sp08.files.wordpress.com/2008/01/weber_content-analysis.pdf (2018. jan. 29.)

A tanulmány első változata megjelent a Diáktudósok. A XXXIII. OTDK 13. Szekciójának első helyezett pályamunkái (Szegedi Egyetemi Kiadó ; Juhász Gyula Főiskolai Kiadó, Szeged, 2017.) c. kötetben.

Beérkezett: 2017. szeptember 29.



Ösztöndíj hátrányos helyzetű diákoknak Celldömölkön

A támogatási forma létrehozásának tervét Németh Tibor könyvtárvezető jelentette be még 2017-ben a Magyar Kultúra Napja városi ünnepe során. A civil szervezetekre vonatkozó törvényi szabályozás változásával egyidejűleg a Kresznerics Ferenc Könyvtári Kultúráért Alapítvány e feladatot beemelte tevékenységi körébe. A felhívás 2017 novemberében jelent meg: a kemeszaljai általános iskolák olyan 3–7. osztályos tanulói pályázhattak, akik hátrányos szociális helyzetük dacára felvillantották tehetségüket.

A kiírás szellemiségének alapját a celldömölki városi könyvtár névadójának, a nyelvész Kresznerics Ferenc 1817-ben kelt végrendeletének 7. pontja képezi, amely rövidítve így fogalmaz egy ösztöndíjról: „Stipendium adassék egy tanuló Ifjúnak [...] És a' ki életemnek nagyobb részét a' tanításban töltöttem el, még holttom után is ezen Stipendium által tanítani fogok.” A jövőépítés jegyében alkotó módon őrizve a tudós pap szellemi örökségét a pályázati eredmény kihirdetésére minden évben a Magyar Kultúra Napján kerül sor a könyvtárban. A támogatásról döntő kuratóriumi ülésre Kresznerics Ferenc egykori és Koltai Jenő esperesplébános jelenlegi székhelyén, az alsósági plébánia épületében került sor a tudós akadémikus sírjának megkoszorúzását követően.

Az ösztöndíj pénzügyi fedezetét hatvanezer forinttal támogatta Celldömölk Város Önkormányzata, valamint több magánszemély. A 2018. évi 240 000 forintos keretösszeg biztosításához ötven-ötvenezer forinttal járultak hozzá a Kresznerics Ferenc Könyvtári Kultúráért Alapítvány kuratóriumának tagjai: Esztergályos Jenő, dr. Koltai Jenő és Németh Tibor.

Négy tanuló nyújtotta be pályázatát és a kuratórium valamennyi jelölt támogatását indokoltnak tartotta. A diákok hatvan-hatvanezer forint ösztöndíjban részesülnek, amelynek folyósítására két részletben kerül sor. A Támogatási Alap döntéséről szóló oklevelet a tanulók Fehér László polgármestertől, Németh Tibortól, a kuratórium elnökétől és Esztergályos Jenőtől, az Alapítvány alapítójától vehették át 2018. január 22-én, a Magyar Kultúra Napja városi ünnepe során.

A szervezők bíznak abban, hogy kezdeményezésükkel a könyvtár demokratikus szolgáltatási rendszerén túl a helyi társadalom hatékony tehetségmentő folyamatainak katalizátorai lehetnek. (A mentés szó jogosságához pedig nem férhet kétség akkor, amikor támogatottak között van diák, akinek a családjában az egy főre eső jövedelem huszonhatezer forint...)

(Németh Tibor tájékoztatása alapján)