

## **Rejtett tartalmak nyomában**

*Mire jó a Textplore dokumentumelemző szoftver?*

*Czibik Ágnes+ - Tóth István János++*

*MTA KTI szeminárium  
2011. május 5. 14:00  
Budapest, Budaörsi út 45. 803.*

---

+: elemző, MKIK GVI, e-mail: [agnes.czibik@gvi.hu](mailto:agnes.czibik@gvi.hu)

++: tudományos főmunkatárs, MTA KTI, e-mail: [tothij@econ.core.hu](mailto:tothij@econ.core.hu)

A Textplore program a TEXTREND (<http://www.textrend.org/>) projekt keretében a Nemzeti Kutatási és Technológiai Hivatal (jelenleg Nemzeti Innovációs Hivatal) Jedlik Ányos programjának támogatásával készül. (Projekt száma: NKFP\_07\_A2-TEXTREND, Szerződés szám: OM-00002/2008). A programot alkotó konzorcium tagjai: Universitas Press Felsőoktatás-kutató Műhely Kft.; Magyar Kereskedelmi és Iparkamara Gazdaság- és Vállalkozáskutató Nonprofit kft.; Glia Számítástechnikai és Tanácsadó Kft.; MTA Számítástechnikai és Automatizálási Kutató Intézet; Eötvös Lóránd Tudományegyetem; Szegedi Tudományegyetem.



Jedlik Ányos program



A Textplore program fejlesztői:

Cserpes Tünde, történész  
Czibik Ágnes, közgazdász  
Farkas Richárd, programtervező matematikus  
Fazekas Mihály, közgazdász-szociológus  
Hajdú Miklós, egyetemi hallgató  
Gyenesse Jenő, programtervező matematikus  
Makó Ágnes, közgazdász-szociológus  
Miskolczi Erna, egyetemi hallgató  
Uhrin Tamás, programtervező matematikus  
Révész Erika, közgazdász  
Várhalmi Zoltán, szociológus  
Zsibrita János, programtervező matematikus  
Tóth István János, közgazdász, kutatásvezető



## Tartalom

<b>Tartalom .....</b>	<b>4</b>
<b>1. A Textplore szoftver rövid leírása .....</b>	<b>5</b>
<b>2. Teszt elemzések – elemzési példák .....</b>	<b>6</b>
<b>3. A matematikai statisztika fogalmainak használata az MTA KTI és az MNB tudományos publikációiban .....</b>	<b>7</b>
3.1. Az elemzés célja .....	7
3.2. Adatok.....	7
3.3. A matematikai statisztika és a statisztikai eszközök használatát jelző kategóriák kialakítása.....	10
3.4. Eredmények.....	11
<b>4. Magyarország képe a New York Times-ban .....</b>	<b>16</b>
4.1. Az elemzés célja .....	16
4.2. Adatok és változók.....	16
4.3. Eredmények.....	19
<b>5. A legnagyobb magyar cégek az on-line sajtóban .....</b>	<b>25</b>
5.1. Az elemzés célja .....	25
5.2. Adatok.....	25
5.3. Eredmények.....	26
<b>Függelék.....</b>	<b>33</b>
A matematikai statisztikai eszközök használata: kategória-szótár .....	33
A legnagyobb magyar vállalkozások keresőszavai .....	35
Korrupció és közbeszerzés kategóriái.....	37
Szókörnyezet statisztikák - Siemens.....	38

## 1. A Textplore szoftver rövid leírása

A Textplore szoftver célja, hogy online módon lehetővé tegye interneten megtalálható dokumentumoknak – például hírportálok cikkei, tudományos publikációk, kutatóműhelyek tanulmányai – és a leendő felhasználók saját dokumentumainak tartalomelemzését.

Tartalomelemzéssel olyan információhoz juthatunk hozzá, amely nem érhető el a dokumentumok tartalmának egyszerű elolvasásával. Áttekinthetjük egy dokumentumhalmaz szerkezetét, hogy elemei milyen főbb témakörökhöz sorolhatók; megfigyelhetjük a különböző forrásból származó, azonos témájú írások szóhasználati – és ezáltal interpretációs – különbségeit, így fényt deríthetünk a dokumentumok rejtett tartalmára.

A Textplore értékei a gyorsaság és áttekinthetőség, mivel segítségével egy rendszeren belül elemezhetőek olyan dokumentumok, amelyek elemzéséhez egyébként először különböző forrásokból kellene összegyűjteni a dokumentumokat, és bonyolult kézi szókeresésekre és további számításokra lenne szükség.

Értéke az egyszerűség, a könnyen kezelhetőség, a kiterjedtség, és a széleskörűség. A rendszeren belül elemezhető dokumentumok száma folyamatosan növekszik, és a felhasználók által is befolyásolható: megadhatnak olyan internetes oldalakat, amelyek dokumentumait elemezni szeretnék, valamint megoszthatják egymással dokumentumaikat, így biztosítva, hogy a rendszeren belül elérhető dokumentumok a lehető legnagyobb mértékben hasznosuljanak.

A felhasználók többféle forrásból származó dokumentumot elemezhetnek.

Elsőként használhatják azt az adatbázist, amelyet automatikusan hoz létre és frissít a Textplore. Ez most több mint 10 magyarországi hírportál mellett a New York Times, a BBC, a Washington Post és a Le Monde oldalairól tölti le az ott megjelenő cikkeket.

Másodsorban minden felhasználó rendelkezik egy saját felhasználói fiókkal, ahova saját dokumentumokat tölthet fel pdf, doc, rtf, txt formák valamelyikében. Ezeket a dokumentumokhoz sem más felhasználó, sem a program adminisztrátorai nem férhetnek hozzá, de a feltöltött dokumentumokat a felhasználói fiók tulajdonosa megoszthatja más felhasználókkal.

Harmadrészt minden felhasználó megadhat tetszőleges számú url-t, amelyekről (amelyről) szeretné, ha a program később letöltéseket végezzon és e letöltések eredményeit a felhasználói fiókjához kapcsolja. Ebben az esetben a felhasználó választhat, hogy milyen kiterjesztésű dokumentum file-ok letöltését szeretné elérni (pdt, rtf, doc, odt, txt, ppt, stb.) Például a <http://www.nber.org/> url cím megadása után a Textplore minden kiválasztott formátumú dokumentumot letölt és automatikusan elvégzi ezeken a szövegelemzéshez szükséges előfeldolgozást (stopszavak elhagyása, szótövezés, szavak összeszámlálása, adatbázis keret létrehozása) is.

A felhasználók a rendelkezésükre álló, illetve a letöltött dokumentumokból elemzésre kiválaszthatnak dokumentum csoportokat, az elemzési eredményeket lementhetik gépükre, illetve az eredményeket a szerveren is tárolhatják.

A rendszer felhasználói között a szövegek mélyebb elemzése iránt érdeklődő elemzők, társadalomkutatók, újságírók jöhetnek számításba, de vállalati vezetők, vállalati elemzők is, mivel a szoftver segítségével elemezhetővé válhat például a vállalatokról és konkurenseikről a médiában megjelenő kép. A másik potenciális felhasználói csoport lehet a kormányzat. Vagy valamilyen kormányzati döntés percepciójának elemzésére, vagy a kormányzat on-line sajtóban való megjelenésének vizsgálatára kézenfekvő megoldás lehet a Textplore.

Ezen kívül a szoftver segítheti a gazdasági, társadalmi kérdésekkel foglalkozó elemzőket abban, hogy egy általuk kevésbé ismert témakör szerteágazó irodalmában gyorsan eligazodjanak a szakirodalom rendszerezése segítségével.

## **2. Teszt elemzések – elemzési példák**

A következőkben három kisebb elemzés révén szeretnénk betekintést nyújtani a Textplore dokumentumelemző szoftver által kínált lehetőségekbe. Mivel célunk főként a szoftver felhasználási lehetőségeinek ismertetése, ezért az elemzésekben a szokásosnál nagyobb szerepet fog kapni egyes technikai részletek bemutatása, míg maguk az elemzések megmaradnak egy kiinduló szinten, kevésbé kidolgozottak, mint ahogyan az a kizárólag tartalmi vetületre koncentráló elemzések esetében elvárható lenne, inkább vázlatosnak mondhatók.

Az első elemzésben a Magyar Nemzeti Bank és az MTA Közgazdaságtudományi Intézetének honlapján megtalálható dokumentumokat elemezzük aszerint, hogy milyen – matematikai statisztikai eszköztár alkalmazására utaló – szavak szerepelnek bennük, és hogy ezek előfordulása összefüggésben van-e publikálásuk dátumával vagy a tanulmányok forrásával.

A második elemzés Magyarország megjelenését vizsgálja a New York Times cikkeiben 2010 novemberétől napjainkig.

A harmadik elemzés témája a legkiterjedtebb: a Magyarországon működő legnagyobb vállalatok sajtó megjelenésére irányul.

### **3. A matematikai statisztika fogalmainak használata az MTA KTI és az MNB tudományos publikációiban**

#### **3.1. Az elemzés célja**

Általános és mindenki által ismert tapasztalat, hogy a matematikai és párhuzamosan a matematikai statisztikai eszközök használata egyre elterjedtebb a közgazdasági kutatásokban. E jelenséget azonban az elméleti megalapozottság, a matematizáltság és az operacionalizálhatóság közötti összefüggések tisztázatlansága és az ehhez kapcsolódó, felerősödő dilemmák kísérik. A matematikai eszközök alkalmazása nem jelent önmagában és szükségképpen elméleti erőt, vagy megalapozottságot (Klein-Romero, 2008), míg a komplex matematikai bizonyítások alkalmazása a operacionalizálhatóság gyengülésével jár együtt (Coelho-McClure, 2005 és Coelho-McCure, 2008).

A statisztikai és matematikai eszközök használatának a közgazdasági kutatásokban való elterjedéséről először Gruber – Boland (1986) gyűjtött és közölt adatokat úgy, hogy egyrészt az American Economic Review-ban 1950 és 1983 között megjelent cikkekben számolta össze, hogy a cikkekben szerepel-e, és ha igen, hány diagram, ábra, tábla és egyenlet, másrészt vezető közgazdasági folyóiratok cikkeit átnézve számolta össze azokat a cikkeket, amelyek matematikai modelleket is alkalmaztak. Coelho-McClure (2005) elemzésükben a matematikai bizonyítás elterjedtségét követték nyomon, azzal, hogy összeszámolták a „lemma” és „multiple equilibrium” szavak hányszor szerepelnek vezető közgazdasági folyóiratok cikkeiben, és vizsgálták a matematizáltság hatását az operacionalizálhatóságra

A matematikai és matematikai statisztikai eszközök használatában meglévő különbségek fényt vethetnek nemcsak a tudományos folyóiratokon belüli különbségekre, de a kutatói közösségek, műhelyek közötti eltérések vizsgálatára is. Mivel a matematikai statisztikai eszközök használatáról nem rendelkezünk magyar empirikus adatokkal, ezért – felhasználva a Textplore programot – megkíséreltünk egy ilyen irányú elemzés első lépését megtenni.

Ehhez két magyar kutatóműhely publikációit vettük alapul. A következőkben azt tekintjük át, hogy az MTA KTI és az MNB honlapjain megtalálható tanulmányokban mennyiben jelennek meg statisztikai, matematikai-statisztikai és ökonometriai eszköztár alkalmazására utaló szavak, és ezek megjelenése milyen jellegzetességeket mutat: időben hogyan változik, és eltér-e a két intézmény esetében.

#### **3.2. Adatok**

A Textplore szoftverrel 2011.03.29-én és 2010.11.18-án végeztünk letöltést az MTA KTI és a Magyar Nemzeti Bank honlapján, ahonnan a feltöltött dokumentumok (pdf, doc) kerültek be adatbázisunkba. A letöltött dokumentumok leszűkítését a tanulmányokra két lépésben valósítottuk meg:

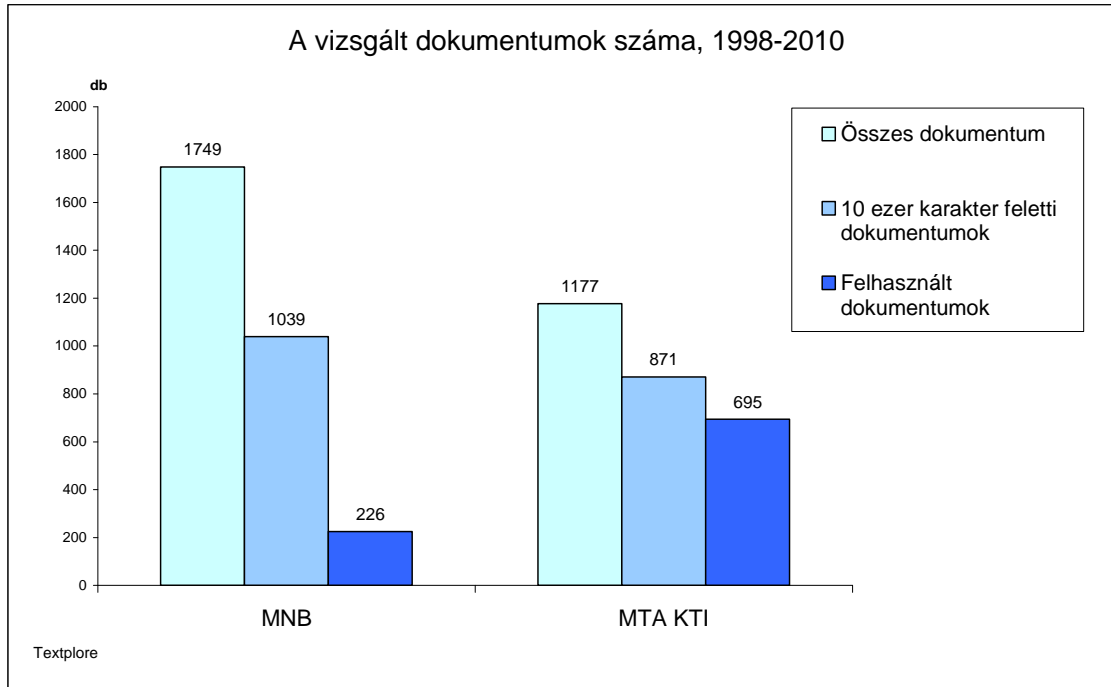
- Kiválasztottuk a Textplore segítségével a tízezer karakternél hosszabb szövegeket, feltételezve, hogy tanulmányok nem rövidebbek ennél.
- A megmaradt szövegeket áttekintettük első 2-3 soruk és file-nevük szerint - és az előbbieket egyértelműségének hiánya esetén a teljes szöveg megtekintésével – és eltávolítottuk a kutatásunk tárgyához nem illeszkedő dokumentumokat.

A dokumentumok kiválasztásának folyamatát és a dokumentumok számát a 3.2.1. ábra mutatja.

A Magyar Nemzeti Bank tízezer karakter feletti dokumentumait erőteljesebben szűkítettük le, mint az MTA KTI dokumentumait, aminek oka az, hogy az MNB honlapján – az intézmény feladataiból adódóan – számos rendelet, lakossági tájékoztató kiadvány, az MNB tevékenységét bemutató rendszeresen megjelenő kiadvány található.

A tanulmányok publikálási dátumát a Textplore szoftver alapesetben a pdf vagy doc file létrehozásának időpontja szerint állapítja meg, amely nem minden esetben egyezik a tényleges publikálási dátummal, ezért ezeket a dátumokat felülvizsgáltuk, így már a valós évszámok szerepelnek a dokumentumok meta-adatai között. A vizsgált dokumentumok száma keletkezési évük szerint bontva a következő (lásd 3.2.1. táblázatot).

### 3.2.1. ábra: A dokumentumok száma a kiválasztás egyes lépéseinél



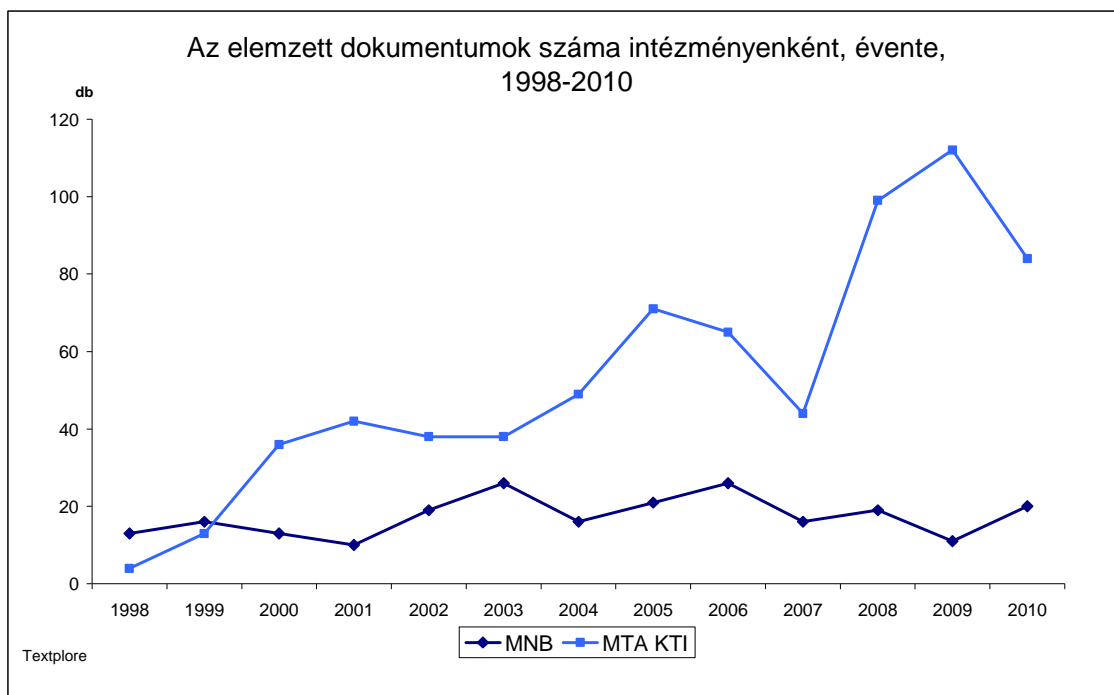


**3.2.1. táblázat: A vizsgált dokumentumok száma a megjelenés ideje szerint**

	MNB	MTA KTI	Összesen
1997	2	0	2
1998	13	4	17
1999	16	13	29
2000	13	36	49
2001	10	42	52
2002	19	38	57
2003	26	38	64
2004	16	49	65
2005	21	71	92
2006	26	65	91
2007	16	44	60
2008	19	99	118
2009	11	112	123
2010	20	84	104
2011	0	17	17
<b>Összesen</b>	<b>228</b>	<b>712</b>	<b>940</b>

Annak érdekében, hogy azonos időszakban keletkezett dokumentumokat tudjunk megvizsgálni és összehasonlítani, az 1997-es és 2011-es tanulmányokat nem vesszük figyelembe az elemzésnél. Ez különösen a 2011-ben megjelent tanulmányok esetében fontos, ahol az MNB-s hiány oka nem az, hogy nem jelent meg tanulmányuk ebben az évben, hanem hogy az MNB honlapját 2010 végén töltöttük le. A végleges dokumentum-gyűjtemény ennek megfelelően 921 tanulmányból áll, amelyek év szerinti megoszlását a következő ábra szemlélteti.

### 3.2.2. ábra: A vizsgált dokumentumok száma a publikálás éve szerint



### 3.3. A matematikai statisztika és a statisztikai eszközök használatát jelző kategóriák kialakítása

A matematikai statisztika használatára utaló, ezt indikáló szavak összegyűjtéséhez 5 százalékos, véletlen mintát vettünk a vizsgált dokumentumokból, az MNB esetében 10, az MTA KTI esetében 30 dokumentumot választottunk ki. Ezeket áttekintve választottuk ki a kategória-alkotásra később felhasznált szavakat. A kiválasztás során törekedtünk rá, hogy magyar nyelven (is) rendelkezésre álló ökonometriai és statisztikai tankönyvek tárgyjegyzékében szereplő szavakat válasszunk ki<sup>1</sup>. A vizsgált dokumentumok között vannak angol és magyar nyelvűek is, így a kiválasztott szavak közül a magyar nyelvűeket lefordítottuk angolra, és fordítva.

A kiválasztott szavakat három kategóriába soroltuk be, a) az ökonometriai, b) a matematikai statisztikai, és c) leíró statisztikai kategóriába. Tudjuk, hogy a kategorizálás önkényes, mivel az ökonometria matematikai-statisztikai eszközöket használ. Ezek a kategóriák valójában átfednek egymással olyan értelemben, hogy a matematikai statisztikai eszköztár részét képezi leíró statisztika alkalmazása is, és az ökonometria is él mindkét másik terület eszközeivel, viszont a kategóriákat diszjunkt módon kell létrehozni. Ezért az ökonometriai kategóriába olyan szavak kerültek, amelyek kifejezetten erre a területre jellemzők, viszont nem kerültek bele azok a matematikai statisztikai fogalmak, amelyeket az ökonometria is rendszeresen használ. Ezek a matematikai statisztika kategóriába kerültek, viszont nem kerültek bele ebbe a kategóriába azok a szavak, amelyek egyszerű leíró elemzések során is

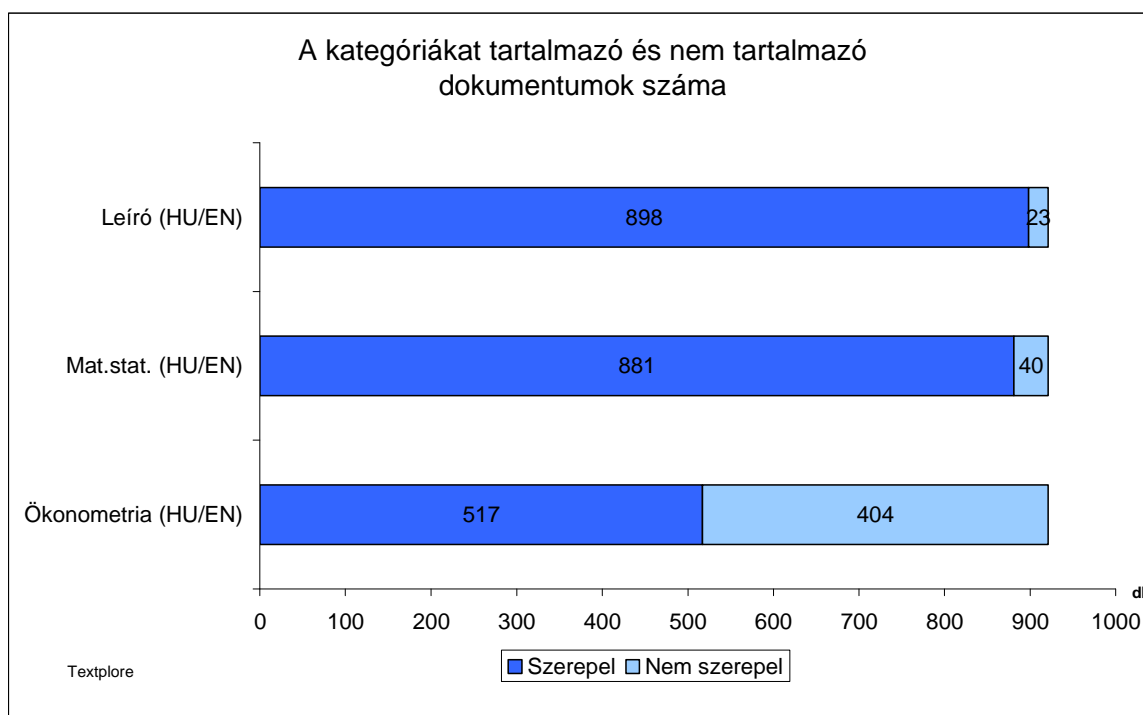
<sup>1</sup> Ehhez Ramanathan (2003) és Hunyadi- Mundruczó-Vita (1997) könyveit vettük alapul.

általánosan használtak. Ezek kerültek a leíró statisztika kategóriába. A kategóriák és a hozzájuk tartozó szavak listája megtekinthető a függelékben. A három kategória együttesére az alábbiakban MSTAT munkanéven hivatkozunk.

### 3.4. Eredmények

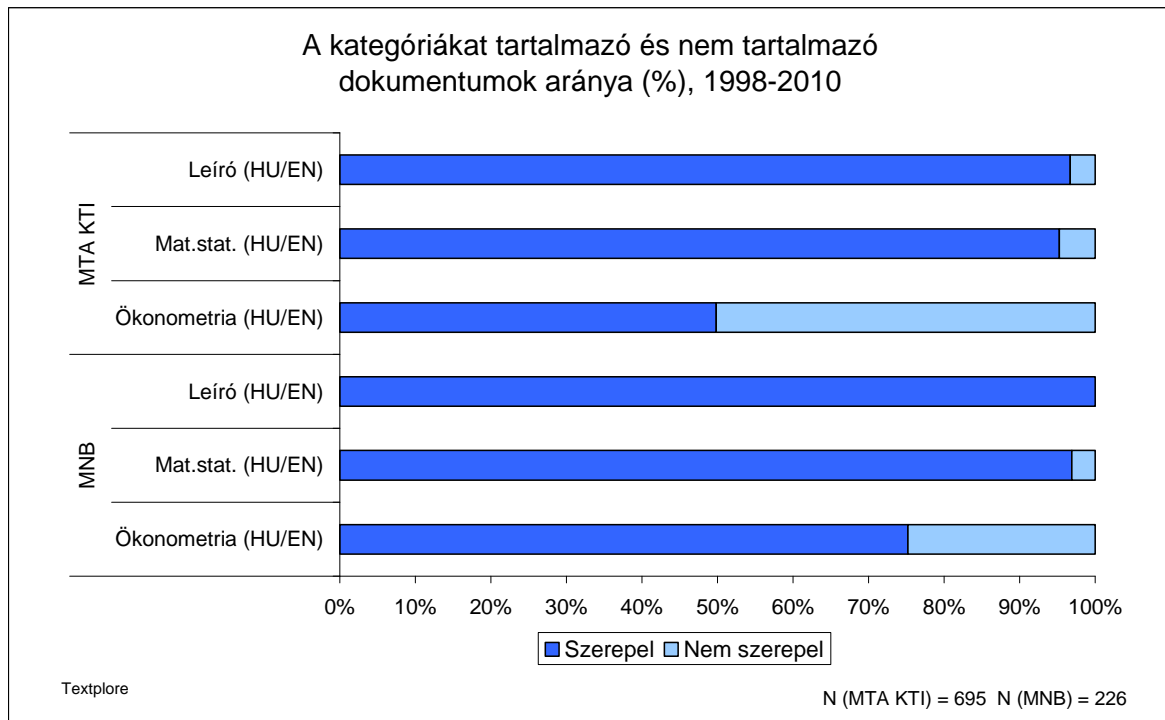
A legalább egy, angol vagy magyar, a MSTAT kategóriába tartozó szót tartalmazó dokumentumok előfordulását dokumentumgyűjteményünkben a 3.4.1. ábra szemlélteti. A leíró és a matematikai statisztika kategória a dokumentumok túlnyomó többségében megjelenik, csak 23 illetve 40 dokumentum nem tartalmaz olyan szavakat, amelyek ezekben kategóriákba tartoznak. Más a helyzet az ökonometria kategória esetében, a dokumentumoknak valamivel több, mint fele, 56 százaléká tartalmaz ilyen szót.

#### 3.4.1. ábra: Az MSTAT kategóriák megjelenése a vizsgált dokumentumokban



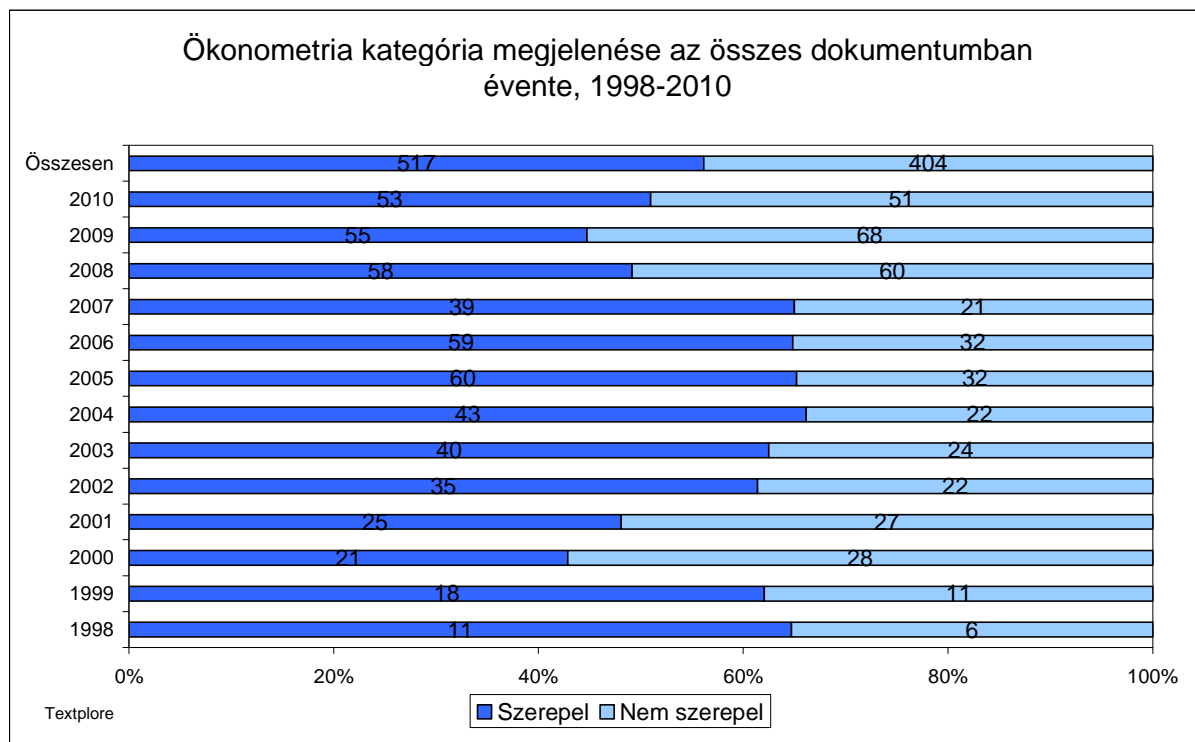
Intézményenként megvizsgálva a kategóriák előfordulását azt figyelhetjük meg, hogy míg a leíró és matematikai statisztika kategória mindkét intézmény esetében a dokumentumoknak több mint 90 százalékában megjelenik, addig az ökonometria kategória az MNB-s dokumentumok 75 százalékában, az MTA KTI-s dokumentumoknak pedig 50 százalékában jelenik meg (lásd a 3.4.2. ábrát). Ez a különbség betudható lehet a két intézmény eltérő kutatási területeinek is.

### 3.4.2. ábra: A MSTAT kategóriák megjelenése a vizsgált dokumentumokban intézményenként



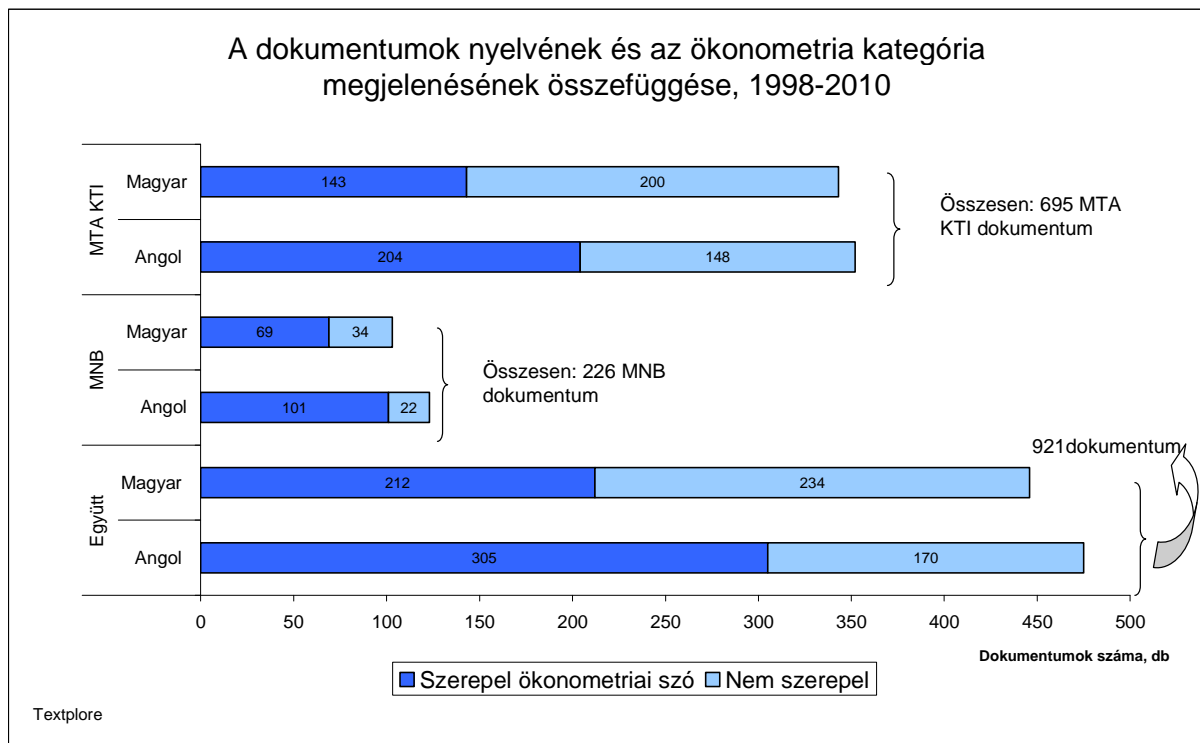
Mivel a leíró és a matematikai statisztika kategória a dokumentumok túlnyomó részében megjelenik, nem kecsegtet újszerű eredményekkel e kategóriák előfordulásainak évenkénti vizsgálata, tendenciáinak feltárása. Az ökonometria kategória időbeli megjelenését az 3.4.3. ábrán követhetjük, azonban egyértelmű tendencia itt sem bontakozik ki.

### 3.4.3. ábra: Az ökonometria kategória megjelenése a vizsgált dokumentumokban évente, 1998-2010



Megfigyelhető, hogy az ökonometria kategória szignifikánsan gyakrabban jelenik meg angol nyelvű dokumentumokban (3.4.4. ábra). Ez az összefüggés mindkét intézmény esetében fennáll.

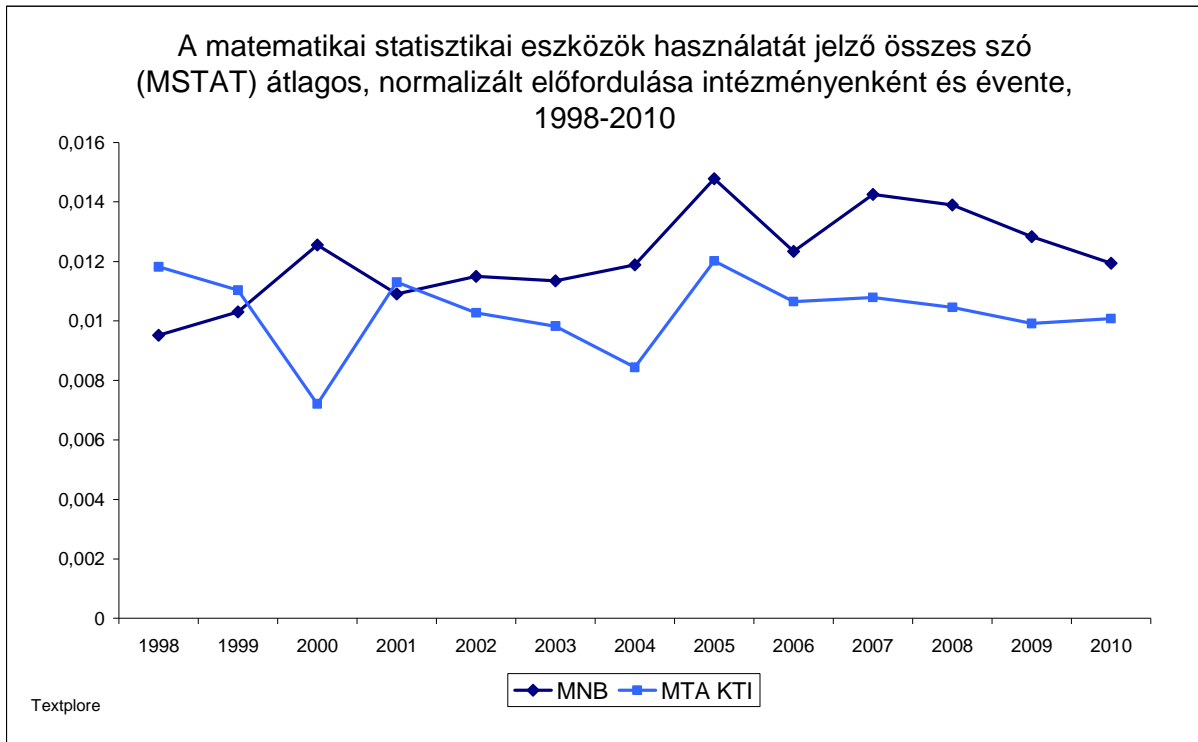
### 3.4.4. ábra: Az ökonometria kategória angol és magyar nyelvű dokumentumokban intézményenként, 1998-2010



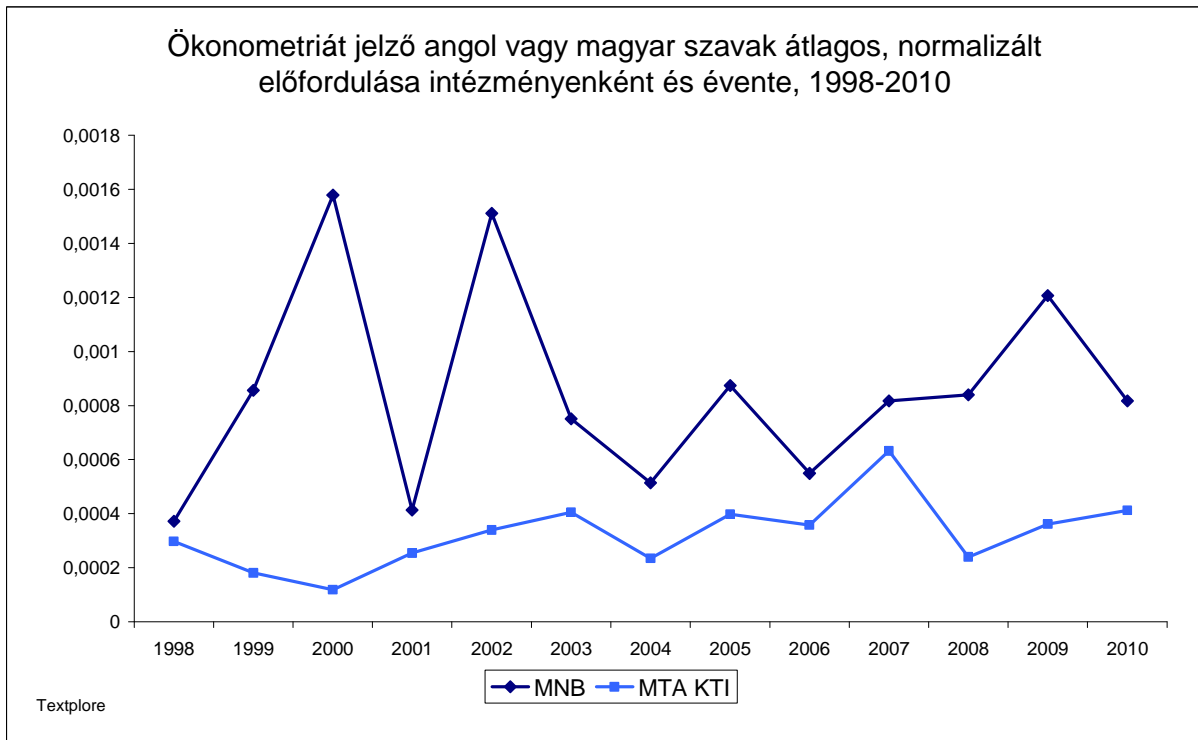
Kísérletet tettünk arra, hogy az információvesztéshez vezető, csak két értéket felvevő változók alkalmazása helyett a szóelőfordulások számának segítségével is megvizsgáljuk azt, hogy a dokumentumokban milyen mértékben támaszkodnak az MSTAT kategóriáiban felsorolt eszközökre. Ehhez minden dokumentum esetében elosztottuk az MSTAT-ban felsorolt szavak előfordulásait a dokumentum szavainak becsült számával. Az így kapott arányok átlagát számoltuk ki évente a 3.4.5. ábrán. Egyértelmű tendencia nem rajzolódik ki: bár enyhe emelkedést mutat a statisztikai és matematikai statisztikai módszerek használatát jelző szavak aránya 1998 és 2010 között. Az enyhén növekvő tendencia inkább az MNB-ben megjelent publikációkhoz köthető, míg az MTA KTI publikációiban a matematikai statisztikai eszközök használatára utaló szavak súlya alig változott az időszak alatt. Az ökonometriai eszközök használata mindvégig az MNB publikációiban volt erősebb az MTA KTI publikációival szemben (lásd 3.4.6. ábra). Ebben a tekintetben mindkét tudományos műhely publikációit az ökonometria enyhén növekvő súlya jellemzi.

Az elemzés következő lépésében meg lehetne vizsgálni az egyes témakörök mentén az MSTAT előfordulásának eltéréseit, illetve a publikációkban közölt referenciák figyelembe vételével azt is, hogy a hivatkozott szerzők milyen csoportjai mellett erősödik fel a matematikai statisztikai eszközök használata.

**3.4.5. ábra: A MSTAT szavainak átlagos, normalizált előfordulása évente, 1998-2010**



**3.4.6. ábra: Az ökonometriai eszközök használatára utaló szavak átlagos, normalizált előfordulása évente, 1998-2010**



## 4. Magyarország képe a New York Times-ban

### 4.1. Az elemzés célja

Az alábbiakban a kvantitatív tartalomelemzés eszközeit felhasználva adunk egy alkalmazási példát a Textplore program használatára.

2010 novemberében építettük be a program jelenlegi verziójába azt a lehetőséget, hogy az általunk kiválasztott on-line hírportálok, újságok oldalairól töltse le az ezeken megjelenő friss cikkeket. Jelenleg több mint húsz portál tartalmát tölti le automatikusan a szoftver (kétpercenként végigböngészve az adott url-eket, és letöltve a frissítéseket), és jelenleg már több mint 650.000 cikk került előfeldolgozásra, és áll így az elemzés rendelkezésére. Ebből a halmazból most a New York Times (NYT) azon cikkeit fogjuk áttekinteni, amelyekben legalább egyszer szerepel a „Hungary” szó. A cikkek témáinak rögzítése nyomán az elemzés azt célozza, hogy az egyes témákban a „Hungary” szó milyen környezetben (pozitív – semleges – negatív tartalmak mellett) jelenik meg, illetve mennyire és hol jelenik meg a kockázat Magyarországgal összefüggésben. Ez alapján vizsgáljuk azt, hogy milyen képet rajzol a NYT Magyarországról.

### 4.2. Adatok és változók

Magyarország többféle szerepben lehet jelen a NYT cikkeiben: úgy, hogy a cikk teljes egésze a magyar helyzet egy vetületét mutatja be, illetve elemzi (a), úgy, hogy a poszt-szocialista országok elemzése közben sorolják fel, illetve érintenek egy magyar vonatkozást (b), illetve úgy, hogy egyéb ország(ok)ról, illetve témáról szól a cikk, amelyben érintőlegesen, felsorolásszerűen van szó Magyarországról (c).

Ennek megfelelően képeztünk egy változót (B), amelynek értéke:

0, ha a cikk Magyarországról szól;

1, ha a poszt-kommunista országokról,

2, ha más országról, témáról.

A másik irány annak megfigyelése, hogy a gazdasági válság közepette mennyire kapcsolódik össze a „kockázat” fogalma a cikkben a Magyarországról szóló képpel, és milyen témák mellett jellemző inkább, hogy a NYT kockázatosnak tartja a magyar helyzetet.

Ennek érdekében a cikkeken három témacsoport előfordulását figyeltük meg: a) gazdasági, b) politikai c) társadalmi témákat. Ezek mindegyikét egymástól



függetlenül tartalmazhat egy-egy cikk. A témák előfordulását három bináris változóval követtük nyomon. Ezek értéke lehet 1, ha a téma szerepel a cikkben, és 0, ha nem:

POL – politikai téma (0,1)

ECON – gazdasági téma (0,1)

SOC – társadalmi téma (0,1).

Másrészt a „kockázatosságot” egy kétértékű változóval mértük, amelyben 1 jelentette azt, ha a cikk utal valamilyen kockázatra, és 0, ha nem:

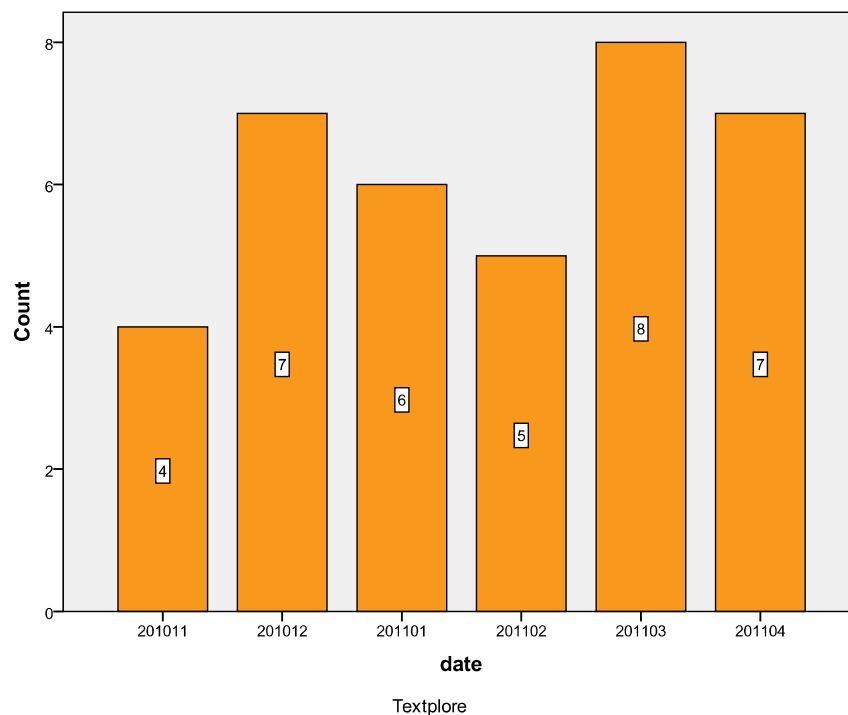
RISK – kockázat (0,1).

A cikk egészének értékelésére egy három kategóriájú változót képeztünk ( E ), amely -1 értéket vett fel, akkor, ha „Magyarország” negatív kontextusban szerepel a cikkben, 0, ha semleges kontextusban és 1, ha pozitív tartalom mellett:

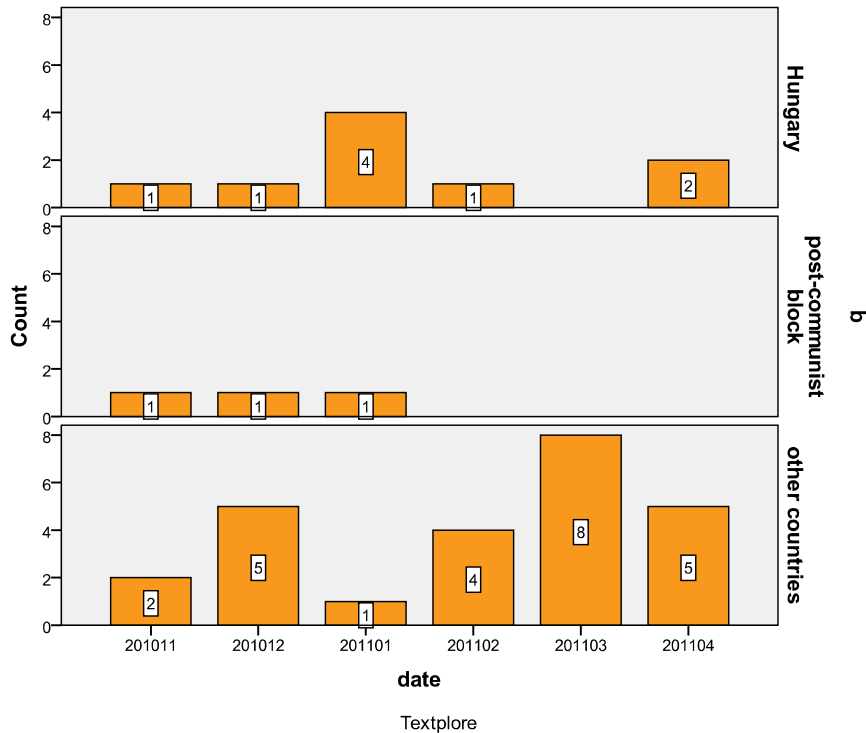
E – értékelés (-1, 0, 1).

Az időszak alatt összesen 37 cikket találtunk a New York Times-ban, amelyben szerepelt a „Hungary” szó (lásd a 4.2.1. ábrát).

#### 4.2.1. ábra: Magyarországot említő cikkek száma a New York Times-ban, 2010. november - 2011. április



4.2.2. ábra: Cikkek téma és a közlés hónapja szerinti megoszlása, amelyekben szerepelt a „Hungary” szó

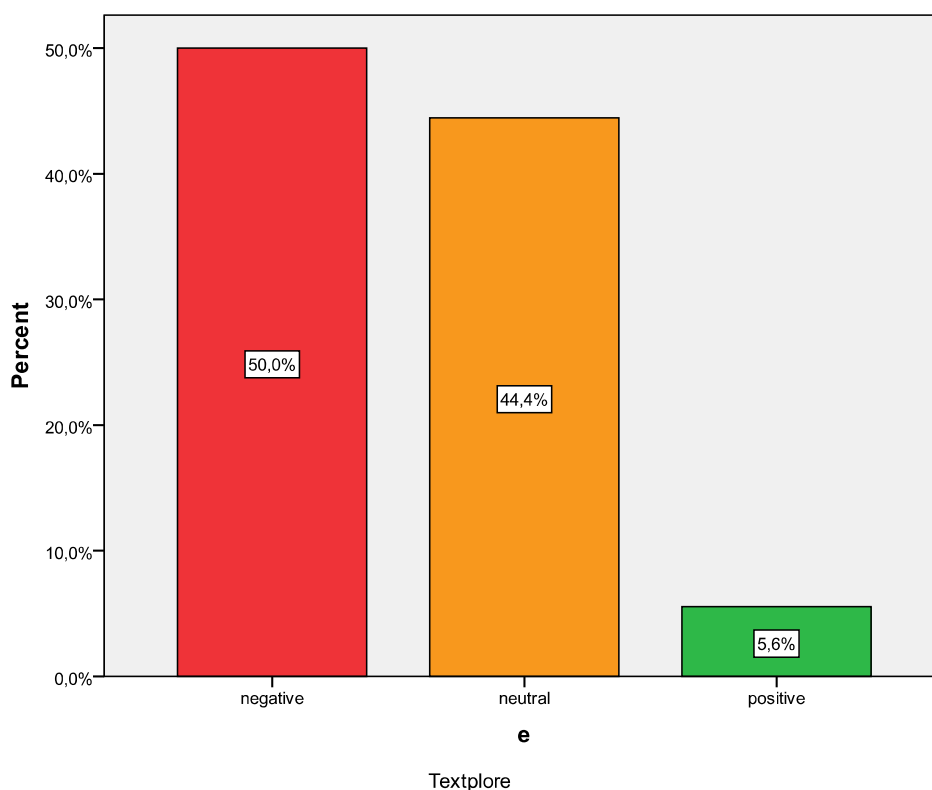


Kilenc cikk kifejezetten Magyarországról szólt, 3 cikk a poszt-kommunista országokról, és a legtöbb cikk, szám szerint 25 egyéb országról, illetve témáról, de amelyekben, akár csak példaként, szó volt Magyarországról is.

### 4.3. Eredmények

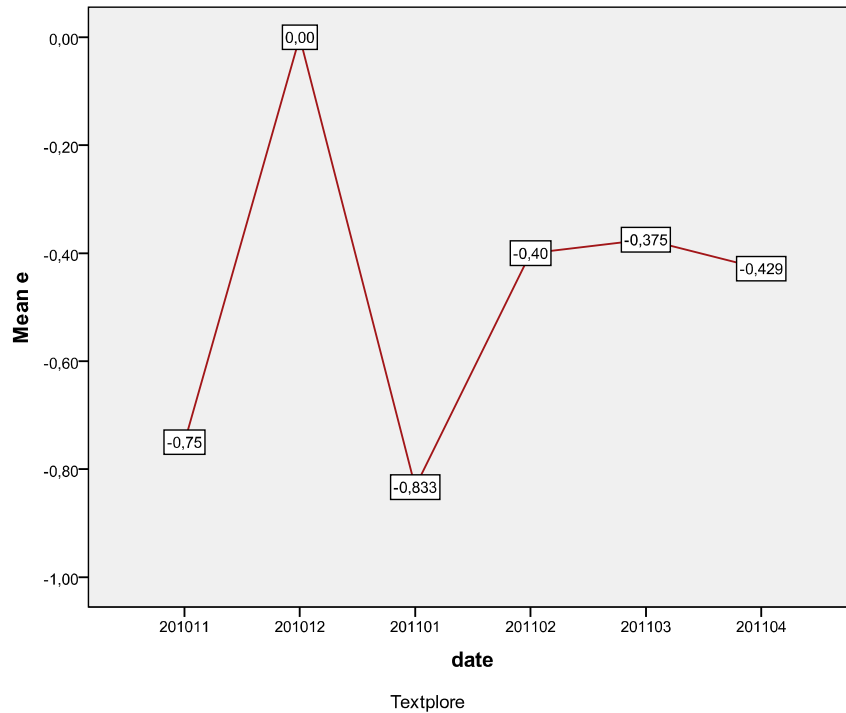
A cikkek többségében, amelyben Magyarország valamilyen formában megjelent, a „Hungary” szó negatív kontextusban szerepelt, miközben elenyésző volt Magyarországot pozitív tartalommal összekapcsoló cikkek aránya (lásd a 4.3.1. ábrát).

**4.3.1. ábra: Magyarország kontextusa a New York Times cikkeiben, 2010. november – 2011. április**

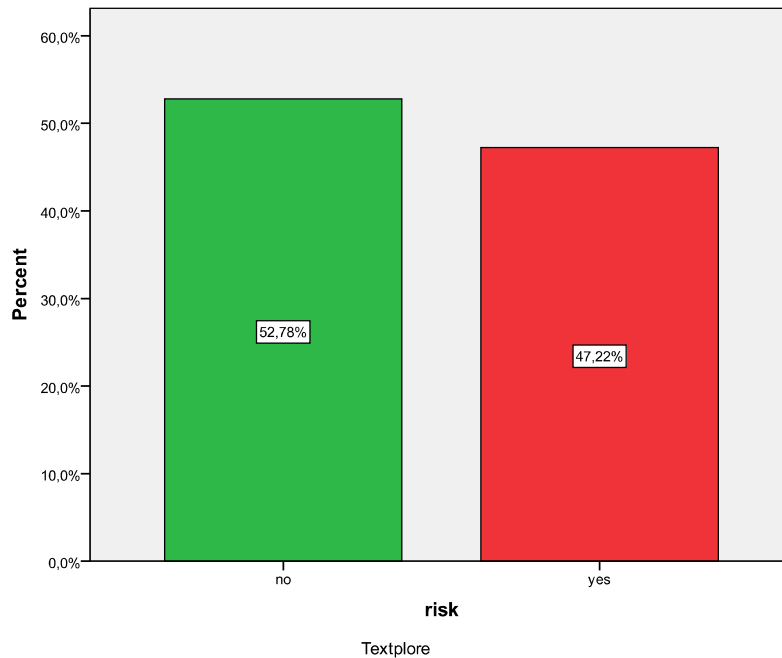


Az egész vizsgált időszakot nézve három minimális értéket lehet megfigyelni, az első az ír válsághoz és az európai gazdasági helyzet bizonytalanságához, a második a médiatörvény elfogadásához és erről az Európa Parlamentben lezajlott vitához, a harmadik pedig az új Alaptörvény elfogadásához kapcsolódik. De az is megfigyelhető, hogy az időszak alatt egyszer sem volt olyan hónap, amikor alapvetően pozitív tartalmú lett volna Magyarország képe a New York Times-ban (lásd a 4.3.3. ábrát).

**4.3.2. ábra: Magyarország kép változása a New York Times cikkeiben, 2010. november – 2011. április**

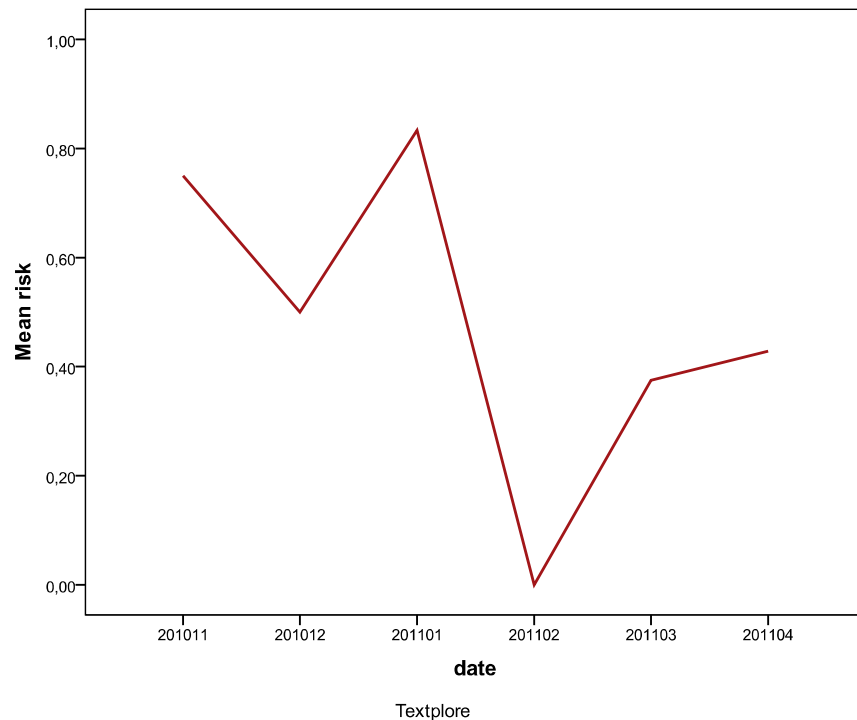


**4.3.3. ábra: Magyarország és a kockázatok együttes előfordulása a New York Times cikkeiben, 2010. november – 2011. április**



A cikkek 47%-ban kapcsolódott Magyarország említése valamilyen (politikai vagy gazdasági) kockázathoz. A kockázatok időbeli alakulásában a legmagasabb átlagos értékeket novemberben, majd januárban a médiatörvény elfogadása után, és áprilisban lehetett megfigyelni.

**4.3.4. ábra: Magyarország és a kockázatok együttes előfordulása a New York Times cikkeiben, 2010. november – 2011. április**



A Magyarországgal kapcsolatos tartalom értékelése ( E ) ott negatív a legnagyobb arányban, ahol a cikk kifejezetten Magyarországról szól (lásd a 4.3.1. táblázatot). De hozzá kell tenni ehhez, hogy azokban a cikkekben is, amelyek elsősorban más országról, illetve más témáról szólnak, nagyobb azon cikkek száma, amelyekben Magyarország említése negatív tartalommal jár együtt, mint amelyekben pozitív tartalommal.

**4.3.1. táblázat: A cikk témája és Magyarország megítélése a New York Times-ban**

**b \* e Crosstabulation**

		e			Total	
		-1,00 negative	,00 neutral	1,00 positive		
b	,00 Hungary	Count	8	1	0	9
		% within b	88,9%	11,1%	,0%	100,0%
		Adjusted Residual	2,7	-2,3	-,8	
1,00 post-communist block		Count	2	1	0	3
		% within b	66,7%	33,3%	,0%	100,0%
		Adjusted Residual	,6	-,4	-,4	
2,00 other countries		Count	8	14	2	24
		% within b	33,3%	58,3%	8,3%	100,0%
		Adjusted Residual	-2,8	2,4	1,0	
Total		Count	18	16	2	36
		% within b	50,0%	44,4%	5,6%	100,0%

Az egyes témacsoportokon belül a magyar helyzet értékelésében nincs különbség azon cikkek között, amelyekben szerepel, és amelyekben nem szerepel gazdasági téma. Ugyanez a helyzet a társadalmi kérdéseket tárgyaló cikkek esetében is. A leggyakrabban negatívak azok a cikkek, amelyeknek témája a magyar politikai helyzet (11 cikkből 9 tartozik ide). Azaz, ha a NYT egy cikkében negatív tartalom mellett jelenik meg Magyarország, akkor ez leggyakrabban akkor következik be, ha a cikk a magyar politikai helyzetről (is) tudósít.

#### 4.3.2. táblázat: Témacsoportok és Magyarország megítélése a New York Times-ban

		Crosstab				
		e			Total	
		-1,00 negative	,00 neutral	1,00 positive		
econ	,00	Count	10	6	1	17
		% within econ	58,8%	35,3%	5,9%	100,0%
		Adjusted Residual	1,0	-1,0	,1	
	1,00	Count	8	10	1	19
		% within econ	42,1%	52,6%	5,3%	100,0%
		Adjusted Residual	-1,0	1,0	-,1	
Total		Count	18	16	2	36
		% within econ	50,0%	44,4%	5,6%	100,0%

		Crosstab				
		e			Total	
		-1,00 negative	,00 neutral	1,00 positive		
soc	,00	Count	9	8	1	18
		% within soc	50,0%	44,4%	5,6%	100,0%
		Adjusted Residual	,0	,0	,0	
	1,00	Count	9	8	1	18
		% within soc	50,0%	44,4%	5,6%	100,0%
		Adjusted Residual	,0	,0	,0	
Total		Count	18	16	2	36
		% within soc	50,0%	44,4%	5,6%	100,0%

		Crosstab				
		e			Total	
		-1,00 negative	,00 neutral	1,00 positive		
pol	,00	Count	9	15	1	25
		% within pol	36,0%	60,0%	4,0%	100,0%
		Adjusted Residual	-2,5	2,8	-,6	
	1,00	Count	9	1	1	11
		% within pol	81,8%	9,1%	9,1%	100,0%
		Adjusted Residual	2,5	-2,8	,6	
Total		Count	18	16	2	36
		% within pol	50,0%	44,4%	5,6%	100,0%

A Magyarországgal kapcsolatos kockázatok említése azokban a cikkekben a leggyakrabban, amelyek kifejezetten Magyarországgal foglalkoznak (lásd 4.3.2. táblázatot).

#### 4.3.3. táblázat: A cikk témája és Magyarországgal kapcsolatos kockázatok megjelenése a New York Times-ban

		b * risk Crosstabulation			
		risk		Total	
		,00 no	1,00 yes		
b	,00 Hungary	Count	2	7	9
		Adjusted Residual	-2,1	2,1	
	1,00 post-communist block	Count	1	2	3
		Adjusted Residual	-,7	,7	
	2,00 other countries	Count	16	8	24
		Adjusted Residual	2,4	-2,4	
Total		Count	19	17	36

Azok a cikkek kapcsolják össze leggyakrabban Magyarországot a kockázatokkal (11 cikkből 8), amelyek a jelenlegi magyar politikai helyzetet elemzik, miközben a társadalmi kérdésekkel foglalkozó cikkek az átlagosnál kevésbé utalnak kockázatokra (4.3.4. táblázat).

A Magyarországgal kapcsolatban megjelenő kockázatok tehát inkább a politikai helyzetre, mint a magyar gazdaság helyzetére vonatkozó témák mellett jelennek meg a New York Times-ban.

Ebben az elemzési példában a cikk és a cikkben szereplő téma volt a megfigyelés egysége, de ez magától értetődően csak elnagyolt eredményekre ad módot. Ha a kutató célja ennél behatóbb vizsgálat, akkor szavak, majd kategóriák és ezt követően kijelentések<sup>2</sup> szintjén kell az elemzést tovább folytatni. Ezek után a kvantitatív eszközökről a kvalitatív elemzés irányába elmozdulva, fel lehetne tárni a NYT cikkeiben a Magyarországgal kapcsolatos állítások típusait és egymással való összefüggéseit.

<sup>2</sup> A magyar szakirodalomban például Csigó (1998) a kijelentések elemzésén keresztül vizsgálta a Borkos csomaggal kapcsolatos megszólalások összefüggéseit, míg Kovács – Tóth (1992) a szavak és a kategóriák elemzésére helyezte a hangsúlyt az 1990-es választás pártprogramjainak elemzésénél.

#### 4.3.4. táblázat: Témacsoportok és Magyarországgal kapcsolatos kockázatok megjelenése a New York Times-ban

		Crosstab			
		risk		Total	
		,00	1,00		
econ	,00	Count	9	8	17
		% within econ	52,9%	47,1%	100,0%
		Adjusted Residual	,0	,0	
	1,00	Count	10	9	19
		% within econ	52,6%	47,4%	100,0%
		Adjusted Residual	,0	,0	
Total		Count	19	17	36
		% within econ	52,8%	47,2%	100,0%

		Crosstab			
		risk		Total	
		,00	1,00		
pol	,00	Count	16	9	25
		% within pol	64,0%	36,0%	100,0%
		Adjusted Residual	2,0	-2,0	
	1,00	Count	3	8	11
		% within pol	27,3%	72,7%	100,0%
		Adjusted Residual	-2,0	2,0	
Total		Count	19	17	36
		% within pol	52,8%	47,2%	100,0%

		Crosstab			
		risk		Total	
		,00	1,00		
soc	,00	Count	6	12	18
		% within soc	33,3%	66,7%	100,0%
		Adjusted Residual	-2,3	2,3	
	1,00	Count	13	5	18
		% within soc	72,2%	27,8%	100,0%
		Adjusted Residual	2,3	-2,3	
Total		Count	19	17	36
		% within soc	52,8%	47,2%	100,0%



## 5. A legnagyobb magyar cégek az on-line sajtóban

### 5.1. Az elemzés célja

Az elemzés annak vizsgálatára irányul, hogy milyen a legnagyobb magyar cégek megjelenése a magyar online médiában. Ha egy kutatás azt tűzné ki feladatul, hogy bemutassa az egyes cégek nevei milyen témák mellett, milyen szöveggörnyezetben jelennek meg, azaz a cég által képviselt márkához milyen tartalmak társulnak, milyen kép rajzolódik ki a sajtóban róla, akkor az első lépés annak felderítése, hogy a cégnevek milyen gyakran és hogyan jelennek meg a sajtóban. Ezután kerülhet sor annak behatóbb vizsgálatára, hogy a cég neve mellett megjelent tartalmak, témák hogyan oszthatóak, milyen tematikus térben helyeződik el a cég maga és ez a tematikus tér hogyan változik az időben. Mindez fontos háttér-információként szolgálhat egy média kampány indításához, illetve a cég márkájának a fogyasztók számára vonzóbb kontextusba helyezéséhez.

### 5.2. Adatok

Az elemzéshez az Index 2000. január 3. és 2011. április 21. közötti, valamint az Origo 1998. december 1. és 2011. április 21. közötti cikkeket, több mint 550.000 cikket használtuk fel. Ezekben a cikkekben a Figyelő TOP100 kiadványából az 2009-es árbevétel alapján vezetőnek minősülő 20 vállalatot kiegészítettük az OTP-vel, a Siemens-szel és a MÁV-val – és ezen cégek megjelenéseit kerestük a cikkekben.

A kiválasztott cégek az alábbiak voltak:

Audi Hungária Motor Kft.  
Bosch Csoport  
Budapesti Elektromos Művek  
E.on Földgáz Trade Földgázkereskedő Zrt.  
E.on Hungária Energetikai Zrt.  
Flextronics International Kft.  
GE Hungary Ipari és Kereskedelmi Kft.  
Magyar Suzuki Zrt.  
Magyar Telekom Távközlési Nyrt  
MÁV  
MOL  
MVM  
Nokia Komárom Kft.  
OTP  
Panrusgáz Gázkereskedelmi Zrt.  
Philips Magyarország  
Richter Gedeon Vegyészeti Gyár Nyrt.  
Samsung Electronics Magyar Zrt.  
Sanofi-aventis/Chinoin  
Siemens  
Spar Magyarország Kereskedelmi Kft.  
Tesco-Global Áruházak Zrt.

## TIGÁZ Tiszántúli Gázszolgáltató Zrt.

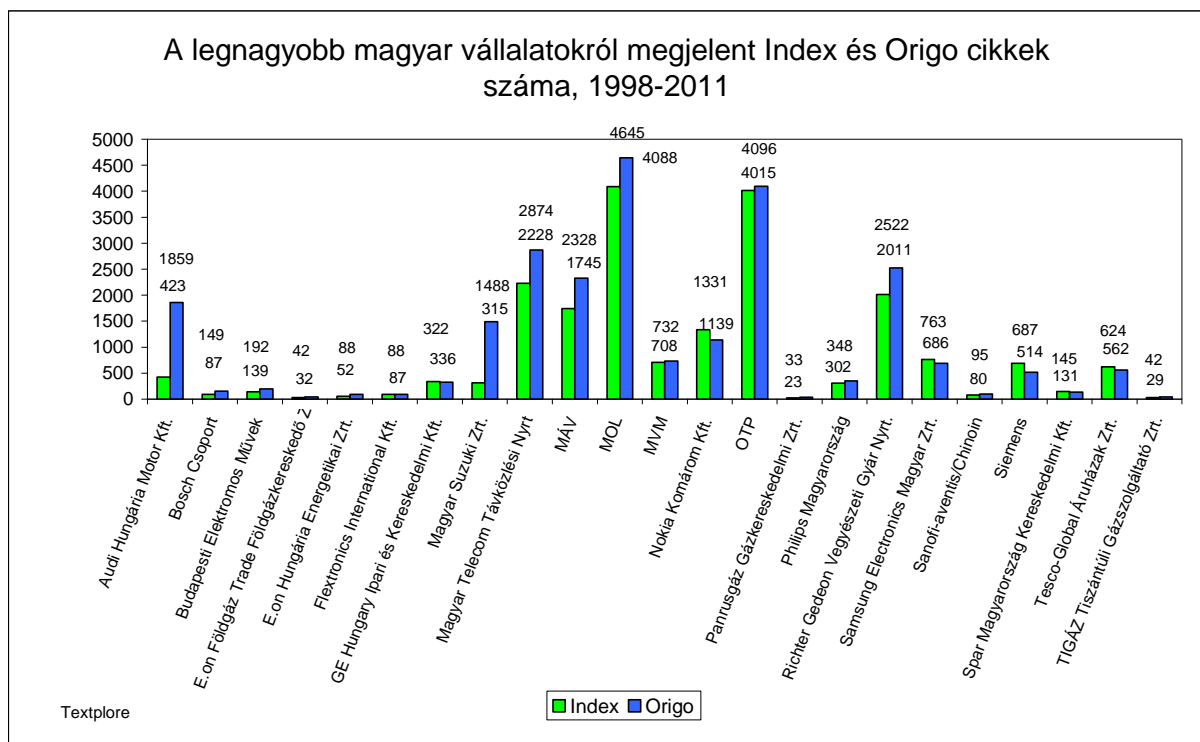
A vállalatokról szóló cikkek azonosításához egy-egy vállalat esetében több keresőszót is használtunk. Ezeket a függelékben közöljük.

A Textplore program segítségével pillanatok alatt ki lehet válogatni azokat a cikkeket, amelyekben a keresőszavak egyike legalább egyszer szerepel: 45.225 ilyen cikket találtunk, amelyből 20.250 az Indexen, míg 24.975 az Origón jelent meg.

### 5.3. Eredmények

A vizsgált cégek közül a legtöbbször a MOL-t és az OTP-t említette cikkeiben az Index és az Origo. Több mint 8000 cikkben jelent meg a két vállalat külön-külön az időszak alatt. Legkevésbé az energetikai cégek, az E.on, a Panrusgáz, valamint a TIGÁZ szerepelt az online médiában: az egyes cégek legfeljebb 50-130 cikkben szerepeltek külön-külön.

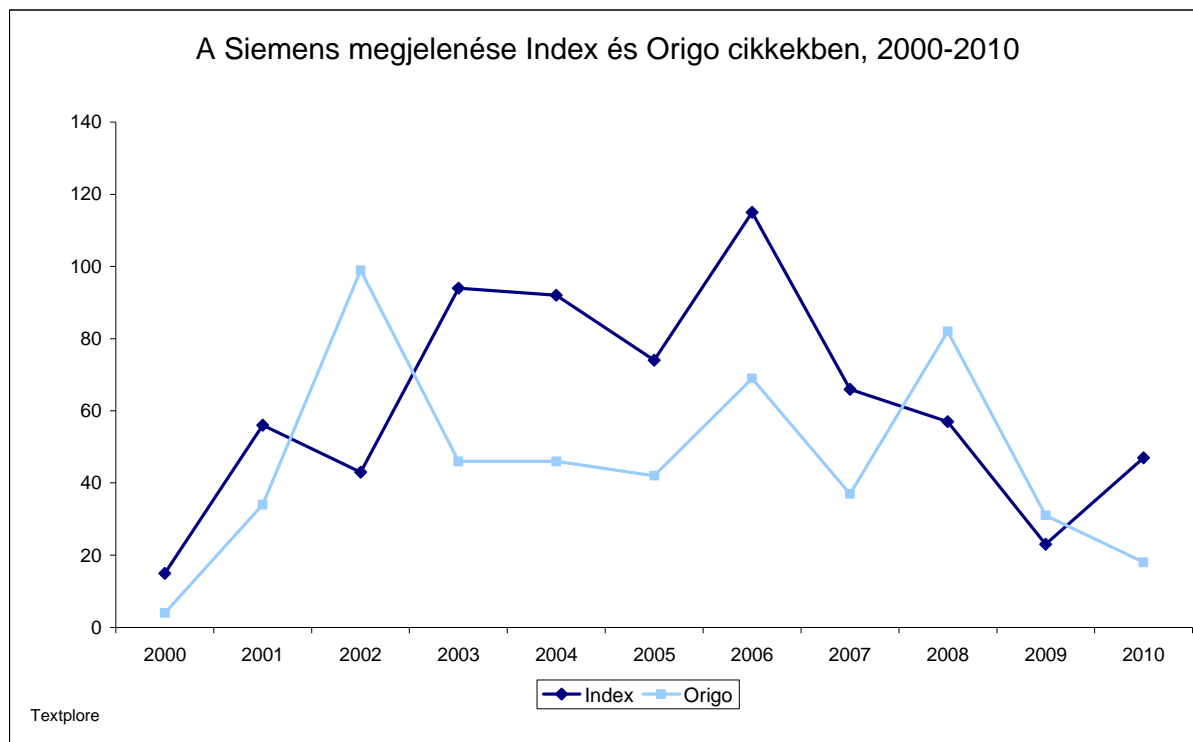
#### 5.3.1. ábra: A legnagyobb magyar vállalatok megjelenése az Indexen és az Origón, 1998-2011



A cégek közül kiválasztottunk egy vállalatot, a Siemens-t, amelynek adatait kissé részletesebben is elemezni fogunk az alábbiakban. A cikkek körét leszűkítettük annak érdekében, hogy a két különböző forrásból származó cikkek ugyanarra az időszakra vonatkozzanak, és ne szerepeljen köztük töredék év. Így a következőkben csak a 2000 és 2010 közötti cikkeket fogunk vizsgálni, és eltekintünk 11 darab 2011-

es és a 2000 előtti cikkektől. A Siemens a következőképpen jelenik meg a két hírportálon a vizsgált időszakban (lásd az 5.3.2. ábrát).

### 5.3.2. ábra: A Siemens megjelenése az Indexen és az Origón, 2000-2010



Az 5.3.2. ábra egyrészt azt mutatja, hogy meglehetősen változékony, hogy a cég egy év alatt mennyiszer jelenik meg az online média általunk vizsgált szegmensében. Másrészt azt, hogy bár a két hírportál a Siemens-szel kapcsolatos híreinek száma gyakran együtt mozog, olyan éveket is találunk, amikor az előző évhez képest ellentétes tendencia jelenik meg a két hírportálon a cikkek számát tekintve. Ez a helyzet 2002-ben, valamint 2009-2010-ben is. Míg 2009-ben viszonylag közel áll a cikkek száma egymáshoz (Index: 23, Origo: 31), addig 2010-ben az Indexen nőtt a Siemens-et említő cikkek száma (47 cikk), míg az Origón ez csökkent csökkent (18 cikk).

Ennek okait kvalitatív elemzéssel, a megadott szavak szókönyezetét bemutató funkcióval lehet megkísérelni. A Textplore lehetőséget ad arra, hogy mi magunk állítsuk be, hogy 10 szavas, 20 szavas, vagy 50 szavas szókönyezetet szeretnénk-e kapni. Az elemzéshez a Siemens szó 20 szavas környezetét vizsgáljuk meg az Index és Origo 2009-2010-es cikkeiben. A két évre vonatkozó szókönyezet statisztikákat a függelékben közöljük.

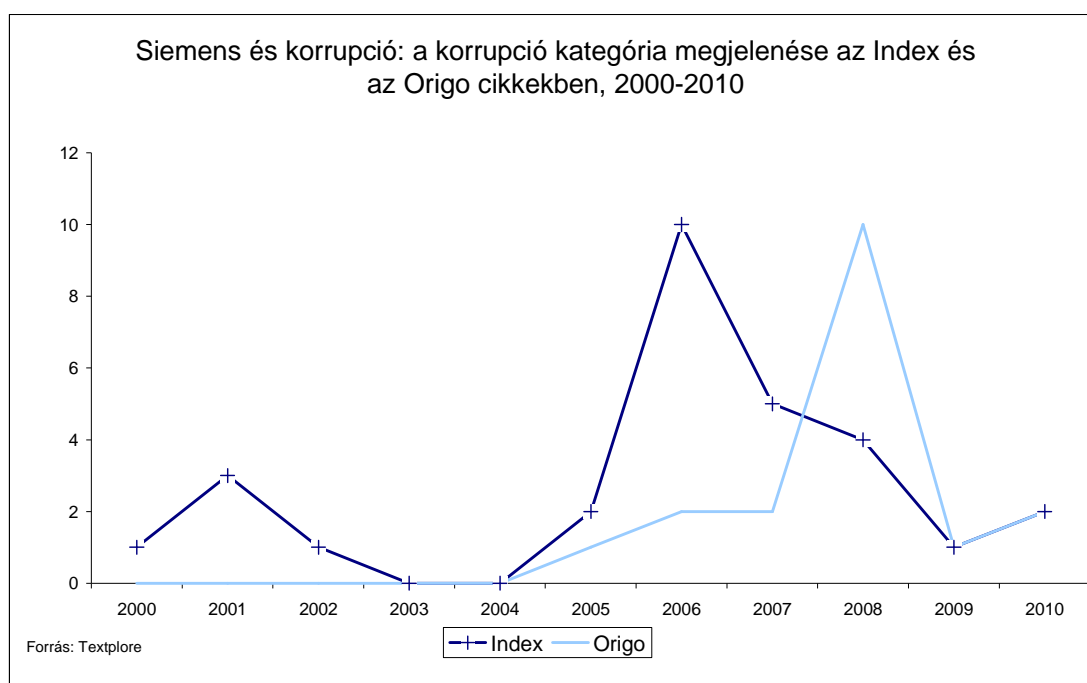
A szókönyezetek alapján átfogó képet kaphatunk arról, hogy milyen témák jelennek meg a cikkekben, jelen esetben szembevetően a technológiai és műszaki cikkekkel kapcsolatos témák megjelenése, emellett pedig találunk utalást a vállalat üzleti helyzetére, megrendelőikkel kapcsolatos konfliktusokra és korrupciós ügyekre is:

(„...egyelőre ismeretlen donortól - írja a BBC. Az ügy a 2004-es athéni olimpiára való felkészüléssel hozható összefüggésbe. A Siemens azért kente meg a politikust és más hivatalnokokat, mert azt remélték, hogy egy zsíros szerződést köthetnek...).

A cikkek teljes tartalmi kategorizálását ebben az elemzésben nem végezzük el, hanem a korrupcióval, és a közbeszerzéssel kapcsolatos tartalmakra koncentrálnunk.

A 3. ábrán megfigyelhető, hogy a két hírportálon nem ugyanabban az évben tetőzött a Siemens-szel kapcsolatos, korrupcióval (is) foglalkozó cikkek száma. Az Index 2006-ban, a Siemens Németországban kirobbant korrupciós botrányának évében, az Origo pedig 2008-ban, a Siemens-perек kezdetekor tartotta kiemelkedően fontosnak, közlésre érdemesnek ezt a témát<sup>3</sup>.

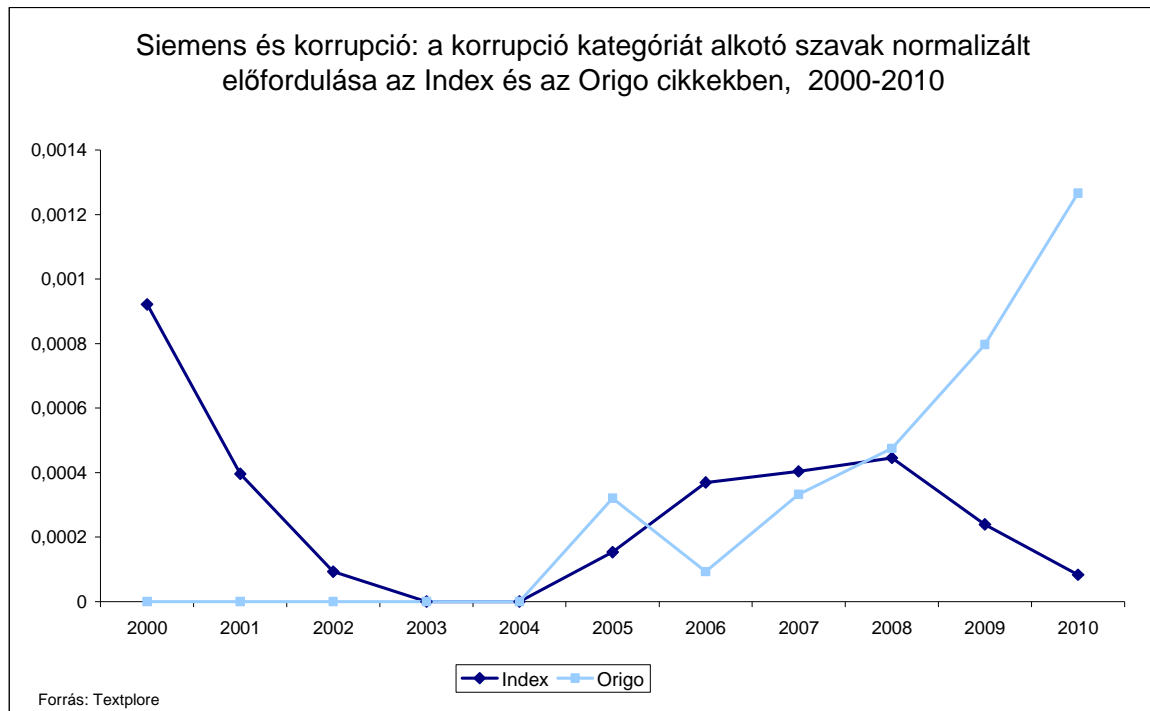
### 5.3.3. ábra: Siemens és a korrupció: a 'korrupció' kategória megjelenése az Index és az Origo cikkeiben, 2000-2010



Az 5.3.4. ábrán a 'korrupció' kategóriát alkotó szavak számának előfordulásait mutatjuk be a cikkek hosszával normalizálva, forrásonként, éves átlagban. Megfigyelhető, hogy míg 2010-ben ugyanannyi (két) cikk jelent meg a két portálon a Siemens-szel kapcsolatban, korrupció témakörben, addig az Origo esetében jóval magasabb a korrupcióra utaló szavak átlagos aránya a cikkeiben. Ennek oka, hogy az Indexen az Origóhoz képest sokkal több olyan Siemens-ról szóló cikk jelent meg, amelyek nem érintették a korrupció témakörét, így a korrupcióra utaló szavak aránya viszonylag alacsony volt.

<sup>3</sup> Ezekben az időszakokban a nemzetközi sajtóban is jelentős szerepet kapott a Siemens botránya. Például: Spiegel: <http://www.spiegel.de/international/spiegel/0,1518,455234,00.html>  
Reuters: <http://www.reuters.com/article/2008/12/15/us-siemens-settlement-idUSTRE4BE4AH20081215>  
New York Times: <http://www.nytimes.com/2008/12/21/business/worldbusiness/21siemens.html>  
BBC News: <http://news.bbc.co.uk/2/hi/business/6444797.stm>

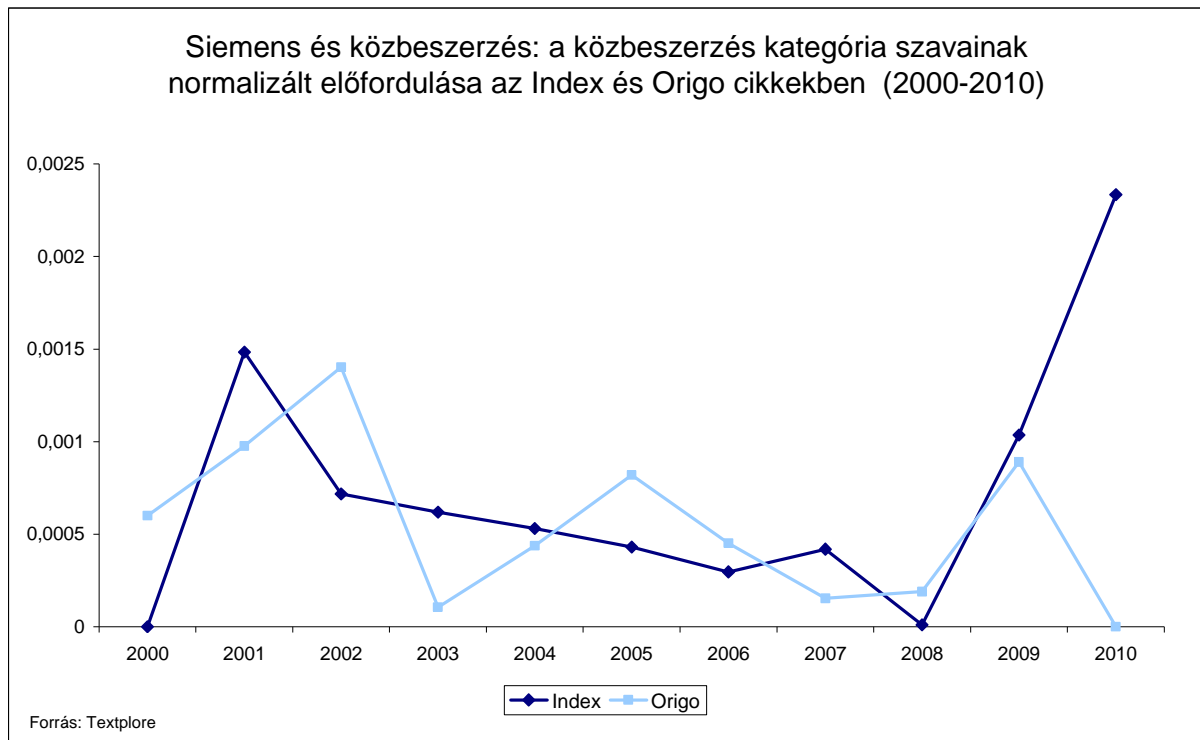
### 5.3.4. ábra: Siemens és a korrupció: a 'korrupció' kategóriát alkotó szavak normalizált előfordulása az Index és az Origo cikkeiben, 2000-2010



A Siemens neve eltérő időpontokban kapcsolódott intenzívebben össze közbeszerzési eljárásokkal a két hírportál cikkeiben. Az Origo 2002-ben, egy 9,4 milliárd forintos posta-megrendelés – az új országos levél- és csomagfeldolgozó központ felépítése – kapcsán írt az elmúlt évtizedben a legtöbbet Siemens-ről a közbeszerzéssel összefüggésben, míg az Index a már korábban említett német Siemens-botrányt kapcsolta össze a közbeszerzés témakörével.

A 5.3.5. ábrán a közbeszerzés kategóriába tartozó szavak éves átlagos előfordulásait tekinthetjük meg cikkek hosszával normalizálva, külön az Index és az Origo esetében. Figyelemre méltó, hogy míg 2006-ban az Index több cikket jelentetett meg közbeszerzésre vonatkozó tartalommal a Siemenssel összefüggésben, mint az Origo. De az előbbi szóhasználata nem volt ilyen intenzív: átlagosan kevesebbszer írt le a közbeszerzésre utaló szavakat, mint az Origo.

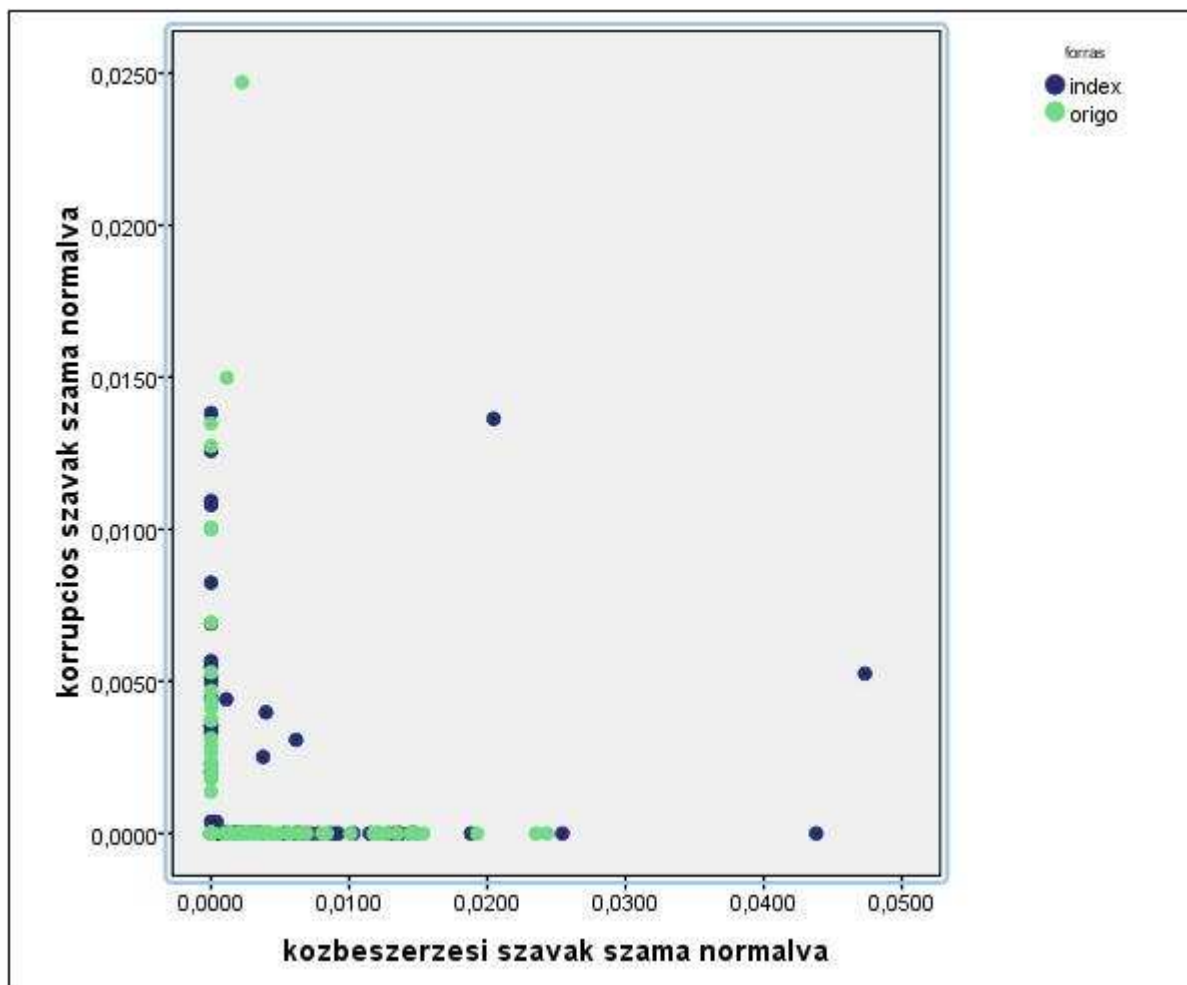
### 5.3.5. ábra: A 'közbeszerzés' kategóriába tartozó szavak átlagos, a cikkek hosszával normalizált előfordulása évente és forrásonként



A két témakör kapcsolatának illusztrálására összevetettük, hogy milyen gyakran jelennek meg együtt egy cikkben belül a korrupció és a közbeszerzés kategóriát alkotó szavak. A vizsgálathoz a cikkek hosszával normált előfordulási arányokat használtuk. Az 5.3.6. ábrán azt figyelhetjük meg, hogy a két kategória együtt-előfordulása nem gyakori, azonban amikor mégis együtt- említést figyelhetünk meg, ez az Index cikkeihez köthető inkább.

Az elemzést lehetne természetesen tovább folytatni akár a Siemens további vizsgálatával: milyen témák kapcsán esik szó a cégről (szakszervezetekkel folytatott vita, határidő csúszás, minőségi hibák, új technológia bevezetése), illetve milyen kijelentésekkel, jelzőkkel írják le a Siemens-t, mint céget, és a cég tevékenységét.

**5.3.6. ábra: A korrupció és közbeszerzés kategóriát alkotó szavak normalizált együttes előfordulása**



## Irodalom

- Antal László (1976): *A tartalomelemzés alapjai*, Magvető Kiadó, Budapest.
- Coelho, Philip R. P. – James E. McClure (2008): The Market for Lemmas: Evidence that Complex Models Rarely Operate in Our World, *Econ Journal Watch*, Vol. 5, Number 1, January 2008, pp. 78-90.
- Coelho, Philip R.P. – James E. McClure (2005): Theory versus application: does complexity crowd out evidence?, *Southern Economic Journal*, 71(3), pp. 556-565.
- Csigó, Péter (1998): A gazdasági stabilizációs diskurzus – a konszolidáció diskurzusa, *Szociológiai Szemle*, 3., 121-151. oldal, <http://www.szociologia.hu/dynamic/9803csigo.htm>
- Grubel, Herbert G. – Lawrence A. Boland (1986): On the Efficient Use of Mathematics in Economics: Some Theory, Facts and Results of an Opinion Survey, *Kyklos*, Vol. 39, Issue 3, pp. 419–442, August
- Hunyadi László - Mundruczó György - Vita László (1997): *Statisztika*, Aula Kiadó, Budapest.
- Kovács Éva – Tóth István János (1992): Ki mit mondott 1990-ben? Választási pártprogramok tartalomelemzése, *Politikatudományi Szemle*, 1. szám, 99-123 oldal
- Ramanathan, Ramu (2003): *Bevezetés az ökonometriába alkalmazásokkal*, PANEM Kft., Budapest.
- Szántó Zoltán - Tóth István János (2008) (szerk.): *Korrupciós kockázatok az üzleti szektorban*, Kutatási Háttér tanulmányok, Transparency International, BCE Korrupciókutató Központ, Budapest, [http://www.crc.uni-corvinus.hu/download/korruptio\\_2008\\_tanulmanykotet\\_080619.pdf](http://www.crc.uni-corvinus.hu/download/korruptio_2008_tanulmanykotet_080619.pdf)
- Weber, Robert Philip (1990): *Basic Content Analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 49), SAGE Publications, Inc., Newbury Park, CA



## Függelék

### A matematikai statisztikai eszközök használata: kategória-szótár

<b>Ökonometria</b>	<b>Econometrics</b>
adf	adf
aic	aic
arima	arima
arma	arma
autokorreláció	autocorrelation
dekompozíció	decomposition
egységgyök	single unit root
endogén	endogenous
exogén	exogenous
garch	garch
granger okság	granger causality
heteroszkedaszticitás	heteroscedasticity
homoszkedaszticitás	homoscedasticity
idősor	time series
kovarianciamátrix	covariance matrix
kpss	kpss
multikollinearitás	multicollinearity
ols	ols
pp	pp
stationer	stationery
trendszűrés	detrending
<b>Matematikai statisztika</b>	<b>Mathematical statistics</b>
becslés	estimation
csúcsosság	kurtosis
egyenlet	equation
együttható	coefficient
ferdeség	skewness
független változó	independent variable
függő változó	dependent variable
főkomponens	principal component
indikátor	indicator
khi-négyzet	chi-square
khi <sup>2</sup>	chi <sup>2</sup>
koefficiens	coefficient
konstans	constant
korreláció	correlation
kovariáns	covariate
képlet	formula
logaritmikus	logarithmic
logit	logit
magyarázó változó	explanatory variable
marginális	marginal

modell	model
multinomiális	multinomial
nem-paraméteres	nonparametric
nemlineáris	nonlinear
normális eloszlás	normal distribution
négyzetösszeg	sum of squares
paraméter	parameter
probit	probit
referenciakategória	reference category
regresszió	regression
szabadságfok	degree of freedom
szignifikancia	significance
szignifikáns	significant
sztochasztikus	stochastic
szórás	standard deviation

### **Leíró statisztika**

%  
arány  
esetszám  
hányad  
kvartilis  
medián  
százalék  
táblázat  
ábra  
átlag

### **Descriptive**

%  
ratio  
case number  
ratio  
quartile  
median  
percentage  
table  
figure  
average

## A legnagyobb magyar vállalkozások keresőszavai

Audi Hungária Motor Kft.

- Audi

Bosch Csoport

- Bosch

Budapesti Elektromos Művek

- Budapesti Elektromos Művek
- ELMŰ

E.on Földgáz Trade Földgázkereskedő Zrt.

- E.on Földgáz Trade
- EON Földgáz Trade

E.on Hungária Energetikai Zrt.

- E.on Hungária
- Eon Hungária

Flextronics International Kft.

- Flextronics

GE Hungary Ipari és Kereskedelmi Kft.

- GE
- General Electric

Magyar Suzuki Zrt.

- Suzuki

Magyar Telekom Távközlési Nyrt

- Magyar Telecom
- Magyar Telekom
- T-Home
- T-Systems
- T-Mobile

MÁV

- MÁV
- Magyar Államvasutak

MOL

- MOL
- Magyar Olaj- és Gázipari

MVM

- MVM
- Magyar Villamos Művek
- MVM-Adwest
- HUPX
- Paksi Atomerőmű
- Mátrai Villamos Művek
- Bakonyi Villamos Művek
- KárpátEnergo
- MIFŰ
- Hungarowind

- Mavir

Nokia Komárom Kft.

- Nokia

OTP

- OTP

- Országos Takarékpénztár

Panrusgáz Gázkereskedelmi Zrt.

- Panrusgáz

Philips Magyarország

- Philips

Richter Gedeon Vegyészeti Gyár Nyrt.

- Richter Gedeon

- Richter

Samsung Electronics Magyar Zrt.

- Samsung

Sanofi-aventis/Chinoin

- sanofi-aventis

- sanofi

- aventis

- Chinoin

Siemens

- Siemens

Spar Magyarország Kereskedelmi Kft.

- Spar

Tesco-Global Áruházak Zrt.

- Tesco

- Tesco-Global

TIGÁZ Tiszántúli Gázszolgáltató Zrt.

- TIGÁZ

- Tiszántúli Gázszolgáltató

## **Korrupció és közbeszerzés kategóriái**

A korrupció kategória kialakításához felhasználtuk a Budapesti Corvinus Egyetem Korrupciókutató Központjának a korrupció médiában való megjelenésével kapcsolatos tanulmányát (Szántó Zoltán - Tóth István János, 2008).

Korrupció =

- korrupt
- korrupció
- csúszópénz
- kenőpénz
- veszteget (+ különféle szófajú alakjai)
- pénzmosás (+ különféle szófajú alakjai)

Közbeszerzés =

- közbeszerzés (+ különféle szófajú alakjai)
- közpénz
- kbt
- nyertes

## Szókörnyezet statisztikák - Siemens

### Részlet a Siemens említések szókörnyezet statisztikájából, Index - 2009

#	ID	szókörnyezet
1	263241	) cikksorozata a pécsi kultúrfőváros-projekt visszasságairól. A 2009-es Soma-díjat megosztva ítélték oda: Brückner Gergelynek (Figyelő) a <b>Siemens</b> korrupciós ügyeiről szóló cikksorozatáért, és Pethő Andrásnak (Origo) a Welt 2000 Kft. pályázatánál elnyert
2	263241	hogyan a szerző két cikke - Számlagyáros hajléktalanok, Mozgalmas purgatórium - feltárta azt a mechanizmust, amelynek révén a <b>Siemens</b> hajléktalanokkal kötött álszerződésekkel és teljesítmény-igazolásokkal fedezte korrupciós kiadásait, illetve adatokat szerzett a hajléktalanokat tucatszámra beszerző, a
3	264402	épül, és emiatt sikeres lehet. A pályázat győztesét a hálózati eszközök gyártói is erősen támogatják. A Nokia <b>Siemens</b> Networks csütörtökön bejelentette, hogy nemcsak műszaki megoldásokkal, de pénzügyi partner bevonásával is az új szolgáltató mögé
4	272811	azóta megszűnt londoni leányvállalatánál, a Hungarian International Banknál volt ösztöndíjas. A Magyar Telekom igazgatósági tanácsi tagsága mellett a <b>Siemens</b> magyarországi leányvállalatának felügyelő bizottságában is szerepet vállalt. Patai rövid ideig a Kádár Jánost balról támadó Biszku Béla
5	274323	hogyan tartozásuk fejében a szovjetek építsék meg a dél-budai metró. A rendszerváltás idején már több külföldi érdekeltségű cég ? <b>Siemens</b> , AWG, Skanska, Intermetro, Matra Transport ? is jelentkezett, hogy részt vennének a metróépítésében
6	275195	Miniszteri székét, ha minden igaz, Vahl Tamásnak adja át, egy informatikai szakembernek, aki többek között a <b>Siemens</b> , az IBM, a SAP Hungary vezető munkatársa volt, illetve jelenleg a német-magyar kereskedelmi kamara elnöke
7	275289	tanulmányait. Első munkahelye a Metrimex volt, később a Konsumex, illetve a Medicor üzletkötőjeként dolgozott. Ezután a <b>Siemens</b> Kft.-vé átalakuló Sicontacnál, a Siemens által gyártott hardverek és szoftverek értékesítését végezte. 1991-től a Berliner Bank számítástechnikai
8	275289	volt, később a Konsumex, illetve a Medicor üzletkötőjeként dolgozott. Ezután a Siemens Kft.-vé átalakuló Sicontacnál, a <b>Siemens</b> által gyártott hardverek és szoftverek értékesítését végezte. 1991-től a Berliner Bank számítástechnikai leányvállalatának magyarországi képviselőjét irányította, majd
9	275304	hogyan Bajnai Gordon korábbi gazdasági miniszteri székét egy informatikai szakembernek, Vahl Tamásnak adja át, aki többek között a <b>Siemens</b> , az IBM, a SAP Hungary vezető munkatársa volt, illetve jelenleg a német-magyar kereskedelmi kamara elnöke
10	276571	Viszik a drótot Az utóbbi hónapokban nagyszabású megállapodások születtek Nagy-Britanniában. Az Orange körülbelül ötszáz dolgozót adott át a Nokia <b>Siemens</b> Networksnek, és a következő öt évben az NSN szakértői menedzselik, bővítik és optimalizálják az Orange brit
11	278539	a Föld természetes mágneses tere. A hatalmas mágneses mezőt egy ötvenhét tonnás szupravezető mágnes segítségével állítják elő. A <b>Siemens</b> Medical Solutionsszal szoros együttműködésben kifejlesztett MRT egy kombinált eszköz része, amelyhez egy pozitron emissziós tomográf (PET

## Részlet a Siemens említések szókörnyezet statisztikájából, Index - 2010

#	ID	szókörnyezet
1	241109	szó, ám gyorsan rájöttek, hogy ennél sokkal speciálisabb kártevő a Stuxnet, mely kifejezetten egy rendszerre, a <b>Siemens</b> ipari területen használt Simatic WinCC-re készült. A vírus több jellemzője, így a frekvenciaválasztási tartomány ? 807
2	317630	kutatói pedig az adatok útvonalát szabályozták. Fontos, hogy más cégek is csatlakozzanak a kezdeményezéshez, hiszen a Nokia <b>Siemens</b> Networks és az Ericsson, az iparág vezető gyártócégei is sok fejlesztésükkel csökkentik a hálózatok fogyasztását. Sok
3	318688	hálózat. Az online cégek életében drámai változást hoz majd a 4. generációs, úgynevezett LTE mobiltechnológia. A Nokia <b>Siemens</b> Networks tavalyi bemutatóján előben tapasztalhattam, hogy az új hálózaton nagyon rövid a válaszidő, a webes rendszerek
4	318698	a távolság azonban továbbra is relative apró maradt. Igaz, egy másik, ?nagyobb távolságot? (a <b>Siemens</b> pontos adatot nem közölt) felőlelő kísérletben öt led együttes alkalmazásával 100 Mbit/s-os sebességet is mutattak
5	318698	nem közölt) felőlelő kísérletben öt led együttes alkalmazásával 100 Mbit/s-os sebességet is mutattak a műszerek. A <b>Siemens</b> a VLC (Visible Light Communication) néven jegyzett technológia előnyei között külön kiemelte a bosszantó wifi-interferencia elkerülésének
6	318812	lehetett, a legelőnyösebb ajánlat lett volna a nyertes. Az első forduló értékelésekor az ajánlatkérő kizárta a tenderből a <b>Siemens</b> Zrt.-t valamint az NJL Konzorciumot (az NJL Kft. és Szenzor Kft. közösen indult). A
7	318812	a Siemens Zrt.-t valamint az NJL Konzorciumot (az NJL Kft. és Szenzor Kft. közösen indult). A <b>Siemens</b> esetében az volt az indok, hogy az alvállalkozójának a nettó árbevétele az előző két évben nem érte
8	318816	, míg a következő 3 évben üzemeltetnie kell a rendszert a nyertesnek. Az eredeti 2006 októberi határidőig a Nokia <b>Siemens</b> Networks Kft. és a Kapsch&Nortel Konzorcium jelentkezett a közbeszerzésre. Mindkettő tovább jutott a második fordulóba, amelyben
9	318816	feltételeket. A KDB helyt adott a panasznak, de a kiírást nem változtatták meg, így csak a Nokia <b>Siemens</b> pályázott. Ezt az ajánlatot azonban érvénytelennek nyilvánította a MÁV. Ekkor a Nokia Siemens fordult a KDB-hez
10	318816	, így csak a Nokia Siemens pályázott. Ezt az ajánlatot azonban érvénytelennek nyilvánította a MÁV. Ekkor a Nokia <b>Siemens</b> fordult a KDB-hez, amely most igazat adott a cégnek, és jogszerűtlennek ítélte a MÁV mindkét érvénytelenségi
11	319269	használt eszközök mind laptopok, az egyik egy viszonylag erős HP Pavilion dv5, a közepes konfiguráció egy hároméves Fujitsu <b>Siemens</b> Amilo Pro, a legkisebb pedig egy Atom processzoros netbook, ami éppen csak teljesíti a Windows 7
12	319280	üzleti évben 20 ezer munkatársat bocsátott el a világszerte közel 400 ezer embert foglalkoztató társaság. Peter Löscher elnök-vezérigazgató a <b>Siemens</b> keddi közgyűlésén hangsúlyozta: a válság piaci szerkezetváltást hozott magával, és az alkalmazkodás további létszámleépítést tesz szükségessé
13	319647	reagáltak a elnökség bejelentésére. Emlékeztettek arra: a konszern éves beszámolójában épp a napokban utalt arra, hogy a <b>Siemens</b> a vártnál nagyobb nyereségre tett szert, és "fényes jövő" előtt áll.
14	319725	elektronikus készletnyilvántartó eszközt nevez iPadnak. Más országokban is levédtek a nevet, de teljesen más iparágakban: Németországban a <b>Siemens</b> használja az autóiparban, Kanadában a Coconut Glove pushup melltartója iPad. Ezek aligha jelentenek gondot, de
15	320000	hogy a Fidesz kampányindulója legyen. . A szerzeményt Orbán Viktor rendelte meg a zeneiparban is tevékenykedő Szentkúti Lászlótól, a <b>Siemens</b> vezetőjétől, és a mű - egyes információk szerint Demjén Ferenc előadásában - várhatóan a Fidesz hivatalos kampánynyitóján

## Részlet a Siemens említések szókörnyezet statisztikájából, Origo - 2009

#	ID	szókörnyezet
1	477298	időpontját követő 60 hónapos élettartamra biztosít szervizelést Magyarország területén a mozgó alkatrészt nem tartalmazó hardverelemek meghibásodása esetén. A Fujitsu <b>Siemens</b> Computers arra az általános hazai gyakorlatra kívánt ezzel a szolgáltatásával reagálni, amely szerint a szerverszállítók három év
2	477298	garancia nemcsak biztosítékot kínál a rendkívül költséges javításokkal szemben, hanem jelzi azt a bizalmat is, amit a Fujitsu <b>Siemens</b> Computers fektet Primergy szervereibe, azaz azt a meggyőződést, hogy meghibásodások ritkán fognak bekövetkezni. A Fujitsu
3	477298	Fujitsu Siemens Computers fektet Primergy szervereibe, azaz azt a meggyőződést, hogy meghibásodások ritkán fognak bekövetkezni. A Fujitsu <b>Siemens</b> Computers Primergy szerverei kimagasló megbízhatósággal rendelkeznek, ami a precíz német mérnöki munkának és a japánok minőség iránti
4	477298	fordulatszámú redundáns ventilátorok. Az alacsonyabb hőmérséklet pedig egyet jelent a hosszabb élettartammal. A hűtés hatékonysága mellett a Fujitsu <b>Siemens</b> Computers nagy hangsúlyt fektet az alacsony fogyasztásra is, a megfelelő alkatrészek alkalmazása révén. Az alacsonyabb fogyasztású
5	477298	ugyanis a hatékonyság leesik, ráadásul a nagyobb teljesítményű darabok terheletlen fogyasztása is magasabbról indul. Kimagasló hatékonyság A Fujitsu <b>Siemens</b> Computers energiatudatosságának bizonyítékául a vezető gyártók közt a Primergy elsőként törte át az iparágilag elfogadott tesztek kidolgozó Standard
6	477298	31 decibeles zajszintje irodai környezetben is tökéletesen használhatóvá teszi, ennél a normál emberi beszéd is hangosabb. A Fujitsu <b>Siemens</b> Computers életciklus garanciájával kapcsolatos további információkért látogasson el a cég weboldalára.
7	477686	. A díjat idén megosztotta a zsűri, a másik nyertes Brückner Gergely, a Figyelő újságírója, aki a <b>Siemens</b> gazdasági ügyeiről írt cikkeket. (Pethő - szintén az NFÜ-cikkért - ebben a hónapban megkapta a Minőségi
8	478048	Commerzbank is zuhant 10,66 százalékot. Nyertese nem volt a napnak, jelentős eséssel zárt a MAN járműipari és a <b>Siemens</b> technológiai részvény is (-8.96 és -12,04 százalék). Párizsban is bank lett, a Soci�t� G�n�rale
9	478681	hordozható készülék. Ez utóbbiból ugyancsak nőtt még vagy 23 százalékkal az eladás. Vesztesek és nyertesek Leginkább a Fujitsu <b>Siemens</b> itta meg a piac szűkülésének a levét. A cég jelentősen veszített részesedéséből, ez azonban azzal is
10	479271	évvel korábbi 18,4 milliárd euróról. A cég megerősítette a folyó pénzügyi évre vonatkozó 8-8,5 milliárd eurós nyereségvárakozását. A <b>Siemens</b> alaptervekenységből származó működési nyeresége 20 százalékkal 2,01 milliárd euróra emelkedett. A kedden közölt adatok meghaladták a szakértői
11	479534	. A teljesség igénye nélkül készült listán az elbocsátók között olyan cégeket találunk, mint a Hugo Boss, a <b>Siemens</b> , a Volkswagen vagy az SAP. Ezek alapján kijelenthető, hogy nem csak néhány ágazat van krízisben
12	480023	Sakk Strip Póker Póker Texas Hold'em Dáma Torpedó Passziánsz Sudoku Wild West IQ kártya Pexeso Puzzle Hoki Keresés A német <b>Siemens</b> óriásvállalat kedden, Moszkvában felkérést kapott Vlagyimir Putyin orosz miniszterelnöktől arra, hogy szorosabb atomenergetikai együttműködésről tárgyaljon az
13	480023	jelentésében a Reuters hírügynökség. Putyin az orosz nukleáris energiaágazat fejlesztésében, illetve nemzetközi piaci tevékenységében jelentős szerepet szán a <b>Siemens</b> cégnek. Közölte, teljes partnerkapcsolatot szeretnének a két vállalat között. Az együttműködést nemcsak az orosz és
14	480023	. Az együttműködést nemcsak az orosz és a német piacon képzei el Putyin, hanem további országokban is. A <b>Siemens</b> és a Rosatom kisebb projektekben már 1990 óta együttműködik. Az új kooperáció konkrét formáiról április vége előtt
15	480023	1990 óta együttműködik. Az új kooperáció konkrét formáiról április vége előtt döntünk - jelentette ki Peter Loescher, a <b>Siemens</b> vezérigazgatója.



## Részlet a Siemens említések szókörnyezet statisztikájából, Origo - 2010

#	ID	szókörnyezet
1	97959	atomprogramhoz kötődnek. Átállítja az urándúsító centrifugákat A világszerte százezer számítógépet megfertőző vírusról már felfedezése után kiderült, hogy a <b>Siemens</b> SCADA nevű ipari szoftverének sebezhetőségeit használja ki. A kutatók eredetileg csak azt feltételezték, hogy a Stuxnet
2	509546	üzemeltetési szempontból is jelentős előrelépést várnak a tervezők az új lokomotívtól, melyet nem titkoltan a Bombardier Traxx és a <b>Siemens</b> EuroSprinter mozdonycsaládok méltó ellenfelének szánnak a franciák. A tesztidőszak lezárása 2010-ben várható, ezt követően 2011-ben indulhat
3	509650	AMD az első, aki a gépre kívülről csatlakoztatható izmos grafikai hardvereken gondolkodik. Rovatunk is tesztelte korábban a Fujitsu <b>Siemens</b> SA3650-es laptopját, amihez szintén járt a gyári turbó. A megoldás meglepően jól működött, csupán azt
4	511188	európai elektronikai cég csütörtökön bejelentette, hogy németországi üzemeiben 2000 munkahely megszüntetését irányozta elő. Az ezzel kapcsolatos döntést a <b>Siemens</b> azzal indokolta, hogy az elmúlt időszakban jelentősen visszaesett a kereslet a konzern termékei iránt. A szakszervezetek
5	511188	jelentősen visszaesett a kereslet a konzern termékei iránt. A szakszervezetek, illetve az alkalmazotti üzemi tanácsok felháborodottan reagáltak a <b>Siemens</b> elnökségének bejelentésére. Emlékeztettek arra: a konzern éves beszámolójában épp a napokban utalt arra, hogy a
6	511188	a Siemens elnökségének bejelentésére. Emlékeztettek arra: a konzern éves beszámolójában épp a napokban utalt arra, hogy a <b>Siemens</b> a vártnál nagyobb nyereségre tett szert, és "fényes jövő" előtt áll. A konzern vezetői
7	511188	legélesebben bírálta a leépítéssel kapcsolatos terveket. A konzern németországi alkalmazottainak száma megközelíti a 130 ezret, míg világméreteken a <b>Siemens</b> több mint 400 ezer embert foglalkoztat.
8	511556	A nemzeti csapat jövő pénteken Romániával, szombaton a házigazda litvánokkal, vasárnap pedig a fehérorosz ligaválogatottal találkozik a vilniusi <b>Siemens</b> Arénában. A magyar jégkorong-válogatott 34 fős kerete: Kapusok: Budai Krisztián (FTC), Hetényi
9	511571	EGPR) nevet viselő, rendszerváltás előtti géppark mását felépítsék. Nem az volt a cél, hogy egy ősrégi <b>Siemens</b> 4004-en olvassuk le a szalagokat - magyarázta a Nyílt Társadalomért Alapítvány (OSA) konferenciáján Barczy Imre informatikai
10	514268	rendelt metrókocsikat, a BKV pedig már nem hagyhatja sokáig futni a régi orosz kocsikat. Ráadásul a vezérlőrendszert építő <b>Siemens</b> kötbért követelhet a BKV-n, ha a kocsik (amelyeknek már hónapok óta futni kellene) nem állnak
11	516845	eljárás indíthatnak minden olyan külföldi vállalat ellen, amelyet jegyeznek a New York-i tőzsdén. A Daimler előtt egyébként a <b>Siemens</b> is hasonló sorsra jutott, és több száz millió dollárt fizetett. Ezt teszi a stuttgarti gyár is
12	519295	hogy a tesztelés időszaka, vagyis 2008 óta nem jelent meg a Saab gépjárműveiben. Hasonló elven működő rendszert a <b>Siemens</b> már 2006-ban bemutatott: a Driver Attention Systemet elsődlegesen a teherszállításban tervezték bevezetni, de ez sem történt
13	522592	egyelőre ismeretlen donortól - írja a BBC. Az ügy a 2004-es athéni olimpiára való felkészüléssel hozható összefüggésbe. A <b>Siemens</b> azért kente meg a politikust és más hivatalnokokat, mert azt remélték, hogy egy zsíros szerződést köthetnek
14	527418	komolyan súlytott. Ha az amerikai szabályozó testületek jóváhagyják, a várhatóan az év végéig lezáruló akvizíció révén a Nokia <b>Siemens</b> a mobil hálózati eszközök második legnagyobb szállítójává válik majd, és 50 új mobilszolgáltatót ügyfelet kaparint meg.
15	527418	a Motorola tulajdonában maradnak vezeték nélküli szabadalmai, illetve az egyes amerikai szolgáltatóknál használt iDEN nevű technológiája. A Nokia <b>Siemens</b> viszont a GSM, WCDMA, Wimax és LTE technológiájú eszközök gyártásával foglalkozó üzemek tulajdonosává válik.