

**NAGY MENNYISÉGŰ SZAKIRODALOM FELDOLGOZÁSÁNAK TÁMOGATÁSA  
EGY TUDOMÁNYOS FOLYÓIRAT CIKKEINEK TARTALOM SZERINTI  
KATEGORIZÁLÁSÁVAL SZÖVEG-SŰRŰSÉGI MUTATÓK ALAPJÁN A  
SIMULATION & GAMING TUDOMÁNYOS FOLYÓIRAT PÉLDÁJÁN  
SUPPORTING LITERATURE REVIEW WITH ARTICLE CATEGORIZATION OF A  
SCIENTIFIC JOURNAL BASED ON TEXT DENSITY INDICATORS ON THE  
EXAMPLE OF SIMULATION & GAMING SCIENTIFIC JOURNAL**

**Boda Márton Attila**

PhD hallgató

Gazdálkodás- és Szervezéstudományi Doktori Iskola, Szent István Egyetem

E-mail: Boda.Marton.Attila.2@hallgato.szie.hu

### **Összefoglalás**

A tanulmány egy folyóirat példáján mutatja be, hogy számítógéppel támogatott módszerrel hogyan lehet tartalom szerinti elemzéssel csökkenteni az irodalom feldolgozása során szelektálásra fordítandó időt. A kutatás 899 dokumentum, ebből 561 darab kutatási cikk esetében végez szógyakorisága alapján témába történő rendezést, melyet faktor- és klaszteranalízis segít.

### **Abstract**

The paper illustrates a computer-assisted method that can be used to reduce the amount of time spent on screening in literature by content analysis on the example of a journal. The study analyses and sorts to topics 899 documents (of which 561 documents are research articles) based on word frequency supported by factor and cluster analysis.

**Kulcsszavak:** szógyakoriság, irodalmi áttekintés, cikk kategorizálás

**JEL besorolás:** C38

**LCC:** HB131-147

### **Bevezetés**

Az első számítógépre épülő üzleti szimulációk az oktatásban az ötvenes évek Amerikájában jelentek meg (Watson, 1981, Faria & Wellington, 2004), a terület gyors fejlődése már a '60-as évektől folyamatos (Dale & Klasson, 1962, Graham & Gray, 1969, Day, 1968, Roberts & Strauss, 1975, Faria, 1987, Faria & Nulsen, 1996).

Szimulációs, illetve üzleti játékok témakörben, több folyóiratban is publikálnak világszerte a témával foglalkozó kutatók, azonban egyértelműen az 1970 óta megjelenő *Simulation & Gaming* a tématerületen a vezető tudományos nemzetközi folyóirat.

A dolgozat egy konkrét kutatási terület<sup>1</sup> objektív irodalomfeldolgozását segíti, de a dolgozat eredményei, illetve az abban bemutatott módszertan jól általánosítható más területeken, illetve folyóiratok elemzésénél is.

A vizsgálat célja a *Simulation & Gaming* folyóirat tartalmi elemzése és az adott kutatási terület szempontjából releváns kutatási cikkek meghatározása. A dolgozat további célja, hogy nagy

---

<sup>1</sup> döntéstámogató szimulációs eljárások modellezése, mely szorosan kötődik az üzleti szimulációk témaköréhez

mennyiségű szakirodalom feldolgozás előtt álló kutatóknak, szakdolgozat, diploma íróknak módszertani javaslattal szolgáljon, és azt egy gyakorlati példa szemléltesse.

Természetesen a folyóirat tartalmának átolvasása során is megállapítható lett volna, hogy mely cikkek relevánsak. Az adatbázis szintű elemzést indokolja a nagyobb fokú objektivitás biztosítása, valamint, hogy könnyedén kizárásra kerülhessenek az adott kutatási terület szempontjából nem elsődleges fontosságú munkák (pl. a *Simulation & Gaming* folyóiratnak része a különböző orvosi szimulációk is, azonban ez a terület nem tartozik az adott kutatási területhez).

A *Simulation & Gaming* folyóiratot a továbbiakban az S&G formában kerül rövidítésre.

### **Anyag és módszertan**

Az irodalmi áttekintés alapvető követelmény egy kutatás esetén. Tomcsányi Pál (2000) a szakirodalmi áttekintés fontosságát így fogalmazza meg: „*A közzétett eredmények ismerete nélkül újat nem hozhatunk létre, mert (hasonlatként) a falra új téglát csak az tehet, aki a fal tetejére felér. Az újdonság pedig csak a meglévőtől való megkülönböztetéssel bizonyítható be. Kellő ismeretanyag híján ritka kivétel a tudományos eredmény, mert az mindig a szerzett tudásnak és kreativitásnak gyermeke*”.

A Tomcsányi féle intelmeknek megfelelően először megvizsgálásra került, hogy születtek-e és ha igen, milyen típusú munkák készültek, amelyek összegzik a korábbi S&G cikkeket. Természetesen a S&G folyóiratban jelent már meg a korábbi cikkeket összefoglaló, rendszerező munka. Kettő jelentős, több évtizedet átfogó munka kerül összefoglalásra röviden a következőkben.

Bragge, Thavikulwat és Töyli (2010) az 1970 óta megjelent kiadásokat vizsgálták. A 40 éves jubileumig 2096 cikk született, melyből 1046 cikk volt kutatási cikk. A szerzők munkájukban megvizsgálták, hogy mely szerzők voltak a legaktívabbak és ők milyen intézményi kapcsolatokkal rendelkeztek. Ugyancsak vizsgálták a hivatkozott szakirodalmat, a címekben megtalálható jellemző kifejezéseket, valamint a SAGE mint kiadó által publikált cikk leírásokat, illetve az ezekben megtalálható kulcs kifejezéseket.

Ahogy Bragge és társai összefoglalják, „módszertanilag a korábbi munkák rendszerezésére a két véglet formailag a narratív áttekintés és a kvantitatív meta-analízis. (King & He, 2005) A két véglet közötti megoldás lehet például, amikor a szóbeli leírást számszerűsített elemekkel kombinálják. Jó példa erre Faria, Hutchinson és Wellington 2009-es, üzleti szimulációkat rendszerező munkája.”

Faria és társai (2009) munkájukban kimondottan az üzleti szimulációkat vizsgálják, történelmi kontextusba helyezve, bemutatva azok fejlődését, terjedését, stb. A S&G cikkei közül is csak azokat rendszerezték, elemezték, amelyek kifejezetten üzleti szimulációkkal foglalkoznak.

Tartalmi elemzést a törzsszövegben előforduló szavak és kifejezések alapján nem végzett egyik kutatás sem, így módszertanilag mindenféleképpen új elemmel bír ez a dolgozat. Fontos még továbbá, hogy 2009 óta nem született ilyen elemzés a S&G folyóiratban.

Bár a S&G folyóirat esetében nem született még ilyen elemzés, maga a vizsgálat módszertana nem új: belépő szintjét jelenti a szövegbányászat (text mining) témakörének. Ahogy Holl (2015) fogalmaz, „a szövegbányászat a nagyméretű, szöveges adatbázisokban való, teljes szövegű kereséssel kezdődik, de további lehetőséget is magában foglal, olyan eljárásokat,

amelyeknek a szöveg, a szövegösszefüggések elemzése is része.”. Holl (2015) megjegyzi még, hogy a tudományban alkalmazott szövegbányászati eljárások jellemzően egy szűkebb szakterülethez vagy egy tudományterülethez sorolható publikációkat vizsgálják. Jó példa erre a hazai gyakorlatból a Magyar Pedagógia neveléstudományi folyóiratán elvégzett elemzés, amely kb. 6500 tanulmány 50 ezer oldalnyi szövegtömegét elemezte (Nagy, 2016a és 2016b). A rengeteg alkalmazási területre néhány példa még Nagy (2016a) tolmácsolásában Cohen és Hersh (2005) valamint Vincze, Szarvas, Farkas, Móra és Csirik (2008) munkái az orvosbiológia területéről, Ghosh, Haider és Sen (2015) üzleti informatikát vizsgáló kutatása, Schneider (2014) számítógépes nyelvészettel foglalkozó elemzése, vagy Nagarkar (2015) könyvtártudománnyal foglalkozó munkája. Jelen dolgozat a S&G folyóirat, mint szűk tématerület publikációit vizsgálja, teljes szövegű kereséssel.

### ***Kiinduló adatmátrix***

A S&G cikkeinek tartalmi elemzésével a cél a cikkek valamilyen kategóriába sorolása volt, hogy könnyebben eldönthető legyen, hogy az adott kutatás során az irodalom feldolgozásban mely munkákat érdemes mindenféleképpen megvizsgálni.

A témakörökbe történő sorolásnál segítséget jelentett az S&G folyóirat saját honlapján található, a folyóiratot bemutató szövege<sup>2</sup>. Ebben 21 tágabb témakört neveznek meg, mint területeket, amelyekkel foglalkozik a folyóirat. A továbbiakban az egyes tématerületekhez tartozó szavak előfordulási aránya kerül megvizsgálásra az egyes cikkekben. Az összehasonlíthatóság miatt az egyes kifejezéseknél 10 ezer szóra jutó darabszám volt a vizsgálat alapja. Ennek következtében nem jelent problémát a különböző hosszúságú cikkek összehasonlítása.

Mivel szülehetnek cikkek a megadott tématerületek átfedésében is, ezért érdemes lesz majd elkészíteni a főkomponens analízist, amellyel csökkenthetjük a komponensek számát összevonásokkal.

A 21 terület az alábbi:

- healthcare,
- sociology,
- decision making,
- psychology,
- language training,
- cognition,
- learning theory,
- management,
- educational technologies,
- negotiation,
- peace and conflict studies,
- economics,
- international studies,
- communication,
- policy and planning,
- organization studies,
- political science,

---

<sup>2</sup> <https://uk.sagepub.com/en-gb/eur/journal/simulation-gaming#aims-and-scope>

- education,
- environmental issues,
- multiculturalism,
- research methodology

### ***Adatgyűjtés adatbázis építéshez***

A Budapesti Corvinus Egyetem könyvtárán keresztül kerültek letöltésre a S&G cikkei az 1999-es évfolyamig visszamenőleg. A könyvtár online adatbázisában eddig voltak elérhetőek a cikkek. Ez összesen 899 darab .pdf kiterjesztésű fájlt jelentett.

### ***Adatok feldolgozása, adatbázis építése***

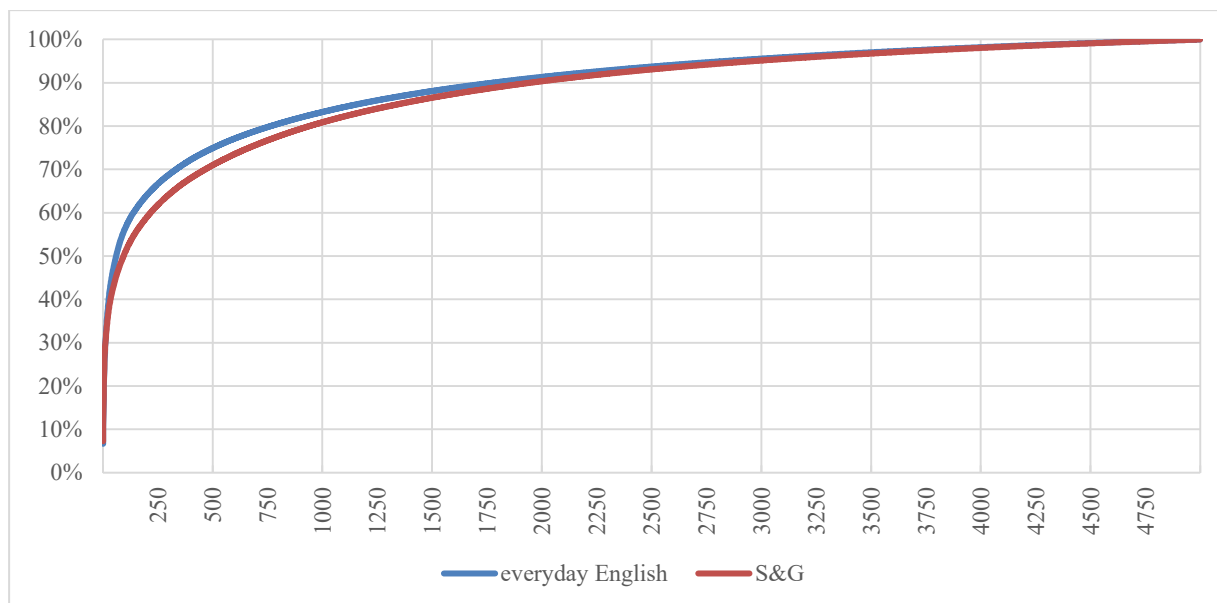
Azt Holl (2015) is megjegyzi, hogy „a szöveg kibontása PDF vagy HTML állományokból nem feltétlenül könnyű”, mivel azok „emberi fogyasztásra készültek”. Holl (2015) javaslatként a gépi formában történő feldolgozásra alkalmas kiterjesztések elérhetővé tételét javasolja a kiadók számára, amely a jövőben egy potenciálisan fontos irány lehet, de jelenleg még ezen folyamat elején tartunk, a kutatás időpontjában a S&G folyóirat sem biztosított lehetőséget, csak .pdf formátumban voltak elérhetőek a cikkek.

Ezt a problémát áthidalva egy Excel makró került megírásra, amely megnyitja az összes .pdf kiterjesztésű cikket, bemásolja a cikkek tartalmát egy fájlba, az abstract előtti részt és az irodalomjegyzék (references) kifejezés utáni résztől megtisztítja a cikket, így a törzsszöveg tartalma marad meg. A makró megtisztította még a szöveget az írásjelektől és a speciális karakterektől. Ezt követően szavanként különböztetésre került a szöveg. Cikkenként összesítésre került, hogy melyik szó hányszor szerepelt az adott cikkben. Az egyes cikkek szógyűjteménye összesítésre került. A korpuszban tisztításra kerültek még az önmagukban értelmezhetetlen rövidítések és a valamilyen ok miatt (pl. elütés) értelmetlen szavak. A fejlécek miatt korrigálásra került a „simulation” és a „gaming” kifejezések Excel makró által megszámlált darabszáma, hiszen ezek minden második oldalon megtalálhatóak voltak. Ezen eljárás során közel 38 ezer egyedi kifejezésből állt a cikkekből felállított korpusz. Elkészítésre került egy összesítő tábla, amely egy táblában mutatja, hogy mely kifejezés mely cikkben hányszor fordult elő, így azt is összesíteni lehetett, hogy összesen hányszor és hány cikkben fordult elő az adott kifejezés.

Miután összeállt az összes kifejezést tartalmazó lista, az elemei gyakoriság szerinti sorba rendezését követően kiszűrésre kerültek a hétköznapi (amerikai) angolban is gyakori szavakat.

Ehhez a wordfrequency.info 5000 szót tartalmazó adatbázisa került felhasználásra. Érdekességként megemlítendő, hogy a „simulation” nem szerepel a listán, míg a „game” a 274. a gyakorisági rangsorban.

Fontos volt, hogy a cikkek szempontjából egyszerre meghatározó jelentőségű és megkülönböztetésre alkalmas szavak kerüljenek kiválasztásra. Ennek két kritériuma volt. Az egyik, hogy ne eseti legyen az adott szó előfordulása, a másik pedig, hogy szakszó legyen. A gyakoriság mérhető, számítógép segítségével automatikusan számolható és feldolgozható. A szakszó jellege az adott szónak úgy lett megállapítva, hogy összevetésre került a hétköznapi nyelvben használt gyakori angol kifejezések és szavak listája a S&G cikkekből nyert kifejezés listával. Feltételezhetően, amennyiben egy S&G-ben található szó a gyakorisági rangsorban jóval előrébb szerepel, mint a hétköznapi angol szavak rangsorában, úgy az speciális vagy szakszónak tekintendő.



**1. ábra: A hétköznapi angol nyelvben és a S&G-ben leggyakrabban használt 5000 szó koncentrációja**

*Forrás: wordfrequency.info, és a Simulation & Gaming tudományos folyóirat cikkeinek saját feldolgozású szógyűjteménye, saját szerkesztés*

A hétköznapi nyelv kifejezéseit tartalmazó lista 5000 szavas, gyakorisági adatokat is tartalmaz. Ezt szemlélteti a következő ábra (1. ábra), kiegészítve a S&G folyóirat ugyanilyen tulajdonságú listájából képzett adatokkal. A két vonaldiagramot összevetve azt látjuk, hogy alapvetően hasonló a két lista koncentrációja, bár a S&G koncentrációja némiképp enyhébb.

A leggyakoribb 10 szó lefedi az 5000 szavas lista 27,6%-át. Az 1. táblázatban a leggyakrabban használt hét szó kerül bemutatásra. Azért a leggyakoribb hét lett kiválasztva, mert ez bár sorrendjében nem, de összességében megegyezik a vizsgált időszakban a S&G cikkeiben használt leggyakoribb hét szóval, és az arány is hasonló az 5000 leggyakoribb kifejezésen belül. Az általános angol nyelvben a hét szó 24,0%, míg a S&G-ben 25,1% aránnyal bír az 5000 leggyakoribb kifejezésen belül. (1. és 2. táblázat)

**1. táblázat: A hétköznapi angol nyelvben leggyakrabban használt hét szó, és ezek aránya a leggyakoribb 5000 szavas listában százalékban, illetve kumulálva**

Sorszám	Szó	Szó aránya [%]	Kumulált szóarány [%]
1	the	6,7%	6,7%
2	be	3,8%	10,5%
3	and	3,3%	13,7%
4	of	3,1%	16,9%
5	a	3,1%	20,0%
6	in	2,1%	22,1%
7	to	1,9%	24,0%

*Forrás: wordfrequency.info, saját szerkesztés*

**2. táblázat: A Simulation & Gaming folyóiratban leggyakrabban használt hét szó, és ezek aránya a leggyakoribb 5000 szavas listában százalékban, illetve kumulálva**

Sorszám	Szó	Szó aránya [%]	Kumulált szóarány [%]
1	the	7,3%	7,3%
2	of	4,0%	11,3%
3	and	3,4%	14,8%
4	to	3,0%	17,7%
5	be	2,6%	20,4%
6	in	2,4%	22,8%
7	a	2,3%	25,1%

*Forrás: a Simulation & Gaming tudományos folyóirat cikkeinek saját feldolgozású szógyűjteménye, saját szerkesztés*

A hétköznapi angol nyelvben is népszerű szavak kiszűrését követően a 3000 darab, leggyakrabban előforduló szavak közül kiválogatásra kerültek azok, amelyek többé-kevésbé egyértelműen besorolhatók valamely témakörbe a 21 megadott közül. Ennek megfelelően végül 134 kifejezés gyakoriságvizsgálatával kerültek elemzésre a cikkek az SPSS programcsomag segítségével.

***Az adatok elemzésére alkalmazott módszer kiválasztása***

A 10 ezer szóra jutó kifejezések száma 20 témacsoportba lett rendezve miután az education és education technologies témakörök összevonásra kerültek. Ezt követően a faktoranalízis módszere került alkalmazásra, mivel feltételezhető volt, hogy a témacsoportok valószínűsíthető átfedései miatt további adattömörítésre lesz lehetőség, illetve fontos, hogy ez a módszer az adatstruktúra feltárásában is segít azzal, hogy a kiinduló változókat kevesebb számú faktorváltozóba vonja össze, amelyeket közvetlenül nem lehetett volna előállítani más módszerrel. Ezeket a változókat használva a további elemzésekben a klaszteranalízis segítségével kerültek csoportosításra a megfigyelési egységek, azaz a S&G folyóirat 1999 és 2017 között született kutatási cikkei. Az elemzés elején kitűzött cél az volt, hogy objektív mérőrendszer szerint, (viszonylag) homogén csoportokba kerüljenek besorolásra a cikkek, hogy megállapítható legyen, hogy melyek az adott kutatási tématerület szempontjából relevánsnak tekintendő cikkek. A klaszteranalízis úgy rendezi csoportokba a megfigyelési egységeket, hogy a megfigyelési egységek minél jobban hasonlítsanak az azonos csoportba kerülő megfigyelési egységekhez, de minél jobban eltérjenek a más csoportokba kerülő megfigyelési egységektől (Sajtos & Mitev, 2007). Az elemzések az SPSS programcsomag segítségével lettek végrehajtva.

***A megfigyelési egységek további szűrése az első elemzések eredményei alapján***

Az első elemzések lefuttatását követően, a klaszteranalízis eredményeit elemezve a nagyon kis elemszámmal bíró klasztereket megvizsgálva megállapításra került, hogy ezek többségében nem kutatásokat tartalmazó fájlok voltak, hanem valamilyen, „editorial”, „tribute”, „news&notes”, „call for papers”, azaz egyéb kategóriába sorolható cikkek, összefoglalóan tehát nem kutatási cikkek. Mennyiségét tekintve ez 338 darabot jelentett, melyek a további vizsgálatoknak nem képezték részét, mivel csak a kutatási cikkek kerültek vizsgálatra. Kutatási cikkből 561 darab maradt, a további elemzésekben csak ezek vesznek részt.

## Eredmények

### A faktoranalízis eredményei

A KMO mutató és a Bartlett próba táblázat alapján került vizsgátra a változók alkalmassága a többváltozós adatelemzési módszerekkel történő elemzéshez. A KMO mutató 0 és 1 között lehet, 0,5-től mondható elfogadhatónak a rendszer a főkomponens vizsgálathoz. A 3. táblázatból kiolvasható, hogy a KMO értéke 0,556, a Bartlett próba szignifikancia szintje pedig alacsonyabb, mint 1%, így a rendszer megfelel a főkomponens analízis elvégzéséhez szükséges kritériumoknak.

### 3. táblázat: KMO és Bartlett teszt eredménye alapján a változók többváltozós adatelemzési módszerekkel történő elemzésre való alkalmasságának vizsgálata

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,556
Approx. Chi-Square		765,868
Bartlett's Test of Sphericity	df	190
	Sig.	,000

Forrás: SPSS, saját számítás

A kommunalitásokat mutató táblázatban azt tudhatjuk meg, hogy a megfigyelt változók információtartalmának hány százalékát tudják a keletkezett faktorok megmagyarázni. A legkisebb mértékben a „MULTICULTURALISM” előfordulásának szóródását magyarázzák a faktorok. A legnagyobb mértékben a „MEDICAL” változó alá rendelt kifejezések vannak magyarázva, információtartalmának 79,6%-át adja vissza a látens változórendszer.

### 4. táblázat: Kommunalítások táblázata: a keletkezett faktorok a megfigyelt változók információtartalmának hány százalékát magyarázzák

Communalities		
	Initial	Extraction
SOCIOLOGY	1,000	,550
RESEARCH & METHODOLOGY	1,000	,396
PSYCHOLOGY	1,000	,658
POLITICAL SCIENCE	1,000	,577
POLICY & PLANNING	1,000	,487
PEACE & CONFLICT	1,000	,611
ORGANIZATION STUDIES	1,000	,711
NEGOTIATION	1,000	,692
MULTICULTURALISM	1,000	,385
MEDICAL	1,000	,796
MANAGEMENT	1,000	,623
LEARNING THEORY	1,000	,578
LANGUAGE	1,000	,621
INTERNATIONAL STUDIES	1,000	,636
ENVIRONMENTAL ISSUES	1,000	,675
EDUCATION	1,000	,622
ECONOMICS	1,000	,673
DECISION-MAKING	1,000	,616
COMMUNICATION	1,000	,672
COGNITION	1,000	,470

Extraction Method: Principal Component Analysis.

Forrás: SPSS, saját számítás

Az 5. táblázat adataiból leolvasható az egyes faktorok magyarázó ereje. 9 faktor tekinthető jelentősnek. A megfigyelt változók információtartalmának 8,176%-át tömöríti az első faktor, 7,257%-át a második, 6,927%-át a harmadik, 6,926%-át a negyedik, 6,708%-át az ötödik,

6,575%-át a hatodik, 6,361%-át a hetedik, 5,817%-át a nyolcadik, és 5,501%-át a kilencedik. Az összes információtartalom 60,247%-a van beletömörítve az első kilenc faktorba. Ezek rotálás utáni adatok, kis mértékben változott az összetétel a kiinduló állapothoz képest.

**5. táblázat: Az egyes faktorok magyarázó ereje**

Component	Total Variance Explained								
	Initial Eigenvalues			Loadings			Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1,976	9,880	9,880	1,976	9,880	9,880	1,635	8,176	8,176
2	1,626	8,131	18,012	1,626	8,131	18,012	1,451	7,257	15,433
3	1,602	8,009	26,021	1,602	8,009	26,021	1,385	6,927	22,360
4	1,321	6,604	32,626	1,321	6,604	32,626	1,385	6,926	29,286
5	1,196	5,981	38,607	1,196	5,981	38,607	1,342	6,708	35,994
6	1,187	5,934	44,541	1,187	5,934	44,541	1,315	6,575	42,569
7	1,085	5,427	49,968	1,085	5,427	49,968	1,272	6,361	48,930
8	1,040	5,198	55,166	1,040	5,198	55,166	1,163	5,817	54,746
9	1,016	5,081	60,247	1,016	5,081	60,247	1,100	5,501	60,247
10	,943	4,716	64,962						
11	,902	4,512	69,475						
12	,864	4,320	73,795						
13	,814	4,069	77,863						
14	,756	3,782	81,645						
15	,740	3,701	85,346						
16	,657	3,283	88,629						
17	,643	3,216	91,845						
18	,594	2,971	94,815						
19	,527	2,633	97,448						
20	,510	2,552	100,000						

Extraction Method: Principal Component Analysis.

*Forrás: SPSS, saját számítás*

A 6. táblázat a megfigyelt és a látens változók kapcsolatrendszerét mutatja. Amennyiben nem különíthetők el határozottan a változók, úgy van lehetőség a mutatórendszer rotálására, amely több iterációt követően már jobban biztosítja, hogy a változók jól elkülönüljenek egymástól. A következő táblázat már a rotálás utáni állapotot mutatja.

**6. táblázat: A megfigyelt és a látens változók kapcsolatrendszere**

	Rotated Component Matrix <sup>a</sup>								
	Component								
	1	2	3	4	5	6	7	8	9
LANGUAGE	,760	-,042	-,096	-,099	-,029	,036	,012	-,058	-,129
COMMUNICATION	,739	-,065	,223	,085	-,140	-,085	-,043	,127	,141
EDUCATION	,542	,047	-,243	-,089	,398	-,010	,177	-,246	,091
DECISION-MAKING	-,015	,747	,128	,046	-,026	,164	-,014	,079	,071
ECONOMICS	-,084	,731	-,212	,052	-,069	-,139	-,011	-,062	-,234
PSYCHOLOGY	,014	,079	,758	-,107	,163	,042	-,014	-,151	,119
SOCIOLOGY	-,018	-,122	,682	,033	-,039	-,019	,065	,126	-,216
ORGANIZATION STUDIES	,016	-,120	,003	,823	,055	,073	-,059	,048	-,069
MANAGEMENT	-,077	,247	-,068	,733	,014	-,073	,049	-,075	,027
LEARNING THEORY	,293	-,188	-,186	,062	,581	-,152	-,126	,112	-,174
RESEARCH & METHODOLOGY	-,036	,054	,220	,187	,540	-,042	,042	-,059	,098
COGNITION	-,188	-,230	,205	-,176	,522	-,068	-,138	-,064	-,091
MULTICULTURALISM	,217	-,233	,144	,014	-,401	-,147	,046	-,073	-,270
POLITICAL SCIENCE	,033	,071	-,016	-,097	-,020	,735	,117	,073	-,038
PEACE & CONFLICT	-,148	-,199	,081	,121	-,223	,579	-,018	-,378	-,026
POLICY & PLANNING	-,043	,221	-,007	,171	,032	,469	-,004	,431	,026
ENVIRONMENTAL ISSUES	-,027	-,035	-,008	-,052	-,049	,002	,053	,815	-,035
NEGOTIATION	-,039	,019	,065	,013	-,097	-,098	,810	,042	,093
INTERNATIONAL STUDIES	,082	-,049	-,023	-,027	,010	,295	,718	,015	-,150
MEDICAL	,021	-,107	-,064	-,035	,007	-,067	-,019	-,036	,879

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 12 iterations.

*Forrás: SPSS, saját szerkesztés*



A faktorok az egyes változókkal mutatott kapcsolat erőssége alapján a következőképpen lettek elnevezve:

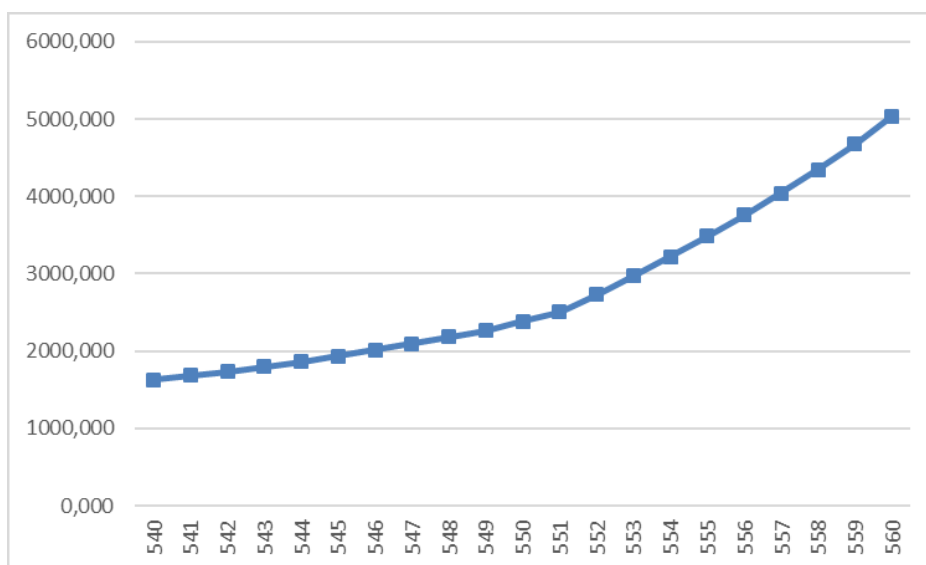
1. faktor: Nyelvi oktatás: A „Nyelvi oktatás” faktorba a „LANGUAGE”, a „COMMUNICATION” és az „EDUCATION” változók kerültek. Ezek között egyirányú a kapcsolat.
2. faktor: Üzleti döntéshozatal: Az „Üzleti döntéshozatal” faktorba a „DECISION-MAKING” és az „ECONOMICS” változók kerültek nagy magyarázó erővel. A kettő közötti kapcsolat egyirányú.
3. faktor: Pszichológia és szociológia: A harmadik faktorba a „PSYCHOLOGY” és a „SOCIOLOGY” változók kerültek, szintén a korábbiakhoz hasonlóan egyirányú kapcsolattal a kettő között.
4. faktor: Szervezet és menedzsment tanulmányok: A negyedik faktorba az „ORGANIZATION STUDIES” és a „MANAGEMENT” változók kerültek. Itt is egyirányú a kapcsolat a két változó között.
5. faktor: Tanulás elmélet és kutatás módszertan: Az ötödik faktorba a „LEARNING THEORY”, a „RESEARCH & METHODOLOGY”, a „COGNITION” és a „MULTICULTURALISM” változók kerültek. Ezek közül az első három közötti kapcsolat egyirányú, míg a „MULTICULTURALISM” ezekkel ellentétes irányú. Ez alapján azt lehet mondani, hogy a multikulturális témákkal foglalkozó cikkek további témái nem a tanulási elmélet és nem kutatási módszertan vagy a megismerés, mint módszertan felől közelítik a témát. Megjegyzendő azonban, hogy a „MULTICULTURALISM” változó mindegyik faktorial közepeknél gyengébb kapcsolatot mutatott. Azért az ötödik faktorba került besorolásra szakmailag, mert ezzel mutatta a legszorosabb kapcsolatot.
6. faktor: Politikatudományok és konfliktuskezelés: A hatodik faktorba a „POLITICAL SCIENCE”, a „PEACE & CONFLICT” és „POLICY & PLANNING” változók kerültek, ezek között egyirányú kapcsolat figyelhető meg.
7. faktor: Környezetpolitika: A hetedik faktorba a „NEGOTIATION” és az „INTERNATIONAL STUDIES” változók kerültek, a kapcsolat iránya egyirányú a két változó között.
8. faktor: Nemzetközi tárgyalások: A nyolcadik faktorba szintén bekerült a „POLICY & PLANNING” változó, csak kicsivel gyengébb a nyolcadik faktorban a súlya, mint a hatodikban. Ebben a faktorban az „ENVIRONMENTAL ISSUES” változó szerepel még, jóval nagyobb magyarázó erővel. A két változó között a kapcsolat egyirányú.
9. faktor: Egészségügyi tanulmányok: A kilencedik faktorba egyedül a „MEDICAL” változó került, azaz valószínűsíthető, hogy az egészségügyi témájú cikkek a S&G folyóirat vizsgált időszakában jól elkülönülnek tartalmilag a többi kutatási cikktől, ami a számos egészségügyi szakkifejezésnek köszönhetően nem is meglepő.

### ***A klaszteranalízis eredményei***

A következőkben a keletkezett faktorok segítségével elvégzett klaszteranalízis kerül bemutatásra. A klaszterelemzés az első kilenc faktorial készült, a magyarázó ereje ennek a kilenc faktorialnak (60,247%) jónak mondható.

A hierarchikus klaszteranalízis lefuttatása után megállapításra került, hogy hány klasztert érdemes képezni. A klaszterek számának meghatározásához az Agglomeration Schedule tábláját egy Excel táblázatba átmásolva az utolsó 20 összevonást vonaldiagram szemlélteti (2. ábra).

A klaszterek fokozatos összevonásával azok belső heterogenitása egy ponton érezhetőbben megnő. Ez a pont akkor van, amikor 10 klaszter után áttérünk 9 klaszterre (az 552. lépésben). Ez azt jelenti, hogy a 10 klaszter jól elkülönül egymástól, így az elemzést 10 klaszterrel lett lefuttatva.



2. ábra: Klaszterek számának meghatározása

Forrás: SPSS adatok alapján saját szerkesztés

A 10 klaszter soknak tűnhet a későbbi elemzések szempontjából, azonban érdemes megjegyezni, hogy 561 darab vizsgált elem van, és a cikkek lehető legjobb csoportokba rendezése a cél, ezért kifejezetten előnyös lehet, ha relatíve kisebb elemszámú klaszterek jönnek létre.

A klaszterek számának meghatározása után, újra lefuttatásra kerültek az elemzések, beállítva, hogy 10 klasztert képezzen az SPSS. A keletkezett klaszterek szakmai leírásához az egyes klaszterek átlagai kerültek vizsgálatra az első kilenc faktor szerint. (7. táblázat)

7. táblázat: A klaszterek átlagai az első kilenc faktor szerint

Statistic	Descriptives									
	Ward Method									
Mean	1	2	3	4	5	6	7	8	9	10
Nyelvi oktatás	,08992	-,63430	-,24344	3,19906	-,24068	-,26644	-,12749	-,17984	,08331	,36485
Üzleti döntéshozatal	-,33994	-,86625	-,41219	-,11817	1,91169	-,06347	-,12162	-,01636	-,44314	,65434
Pszichológia és szociológia	-,42385	,06540	-,07334	,19862	-,35935	-,00832	-,01885	,38906	-,66796	5,62268
Szervezet és menedzsment tanulmányok	-,31280	,38127	-,23615	-,16859	-,05647	-,12612	2,93186	,00970	-,25508	-,64523
Tanulás elmélet és kutatás módszertan	1,45389	-,106328	-,43645	-,22288	-,34517	-,06890	,17561	-,01336	-,06225	,72320
Politikatudományok és konfliktuskezelés	-,25198	4,05197	-,18581	-,14379	-,30481	,40246	-,24877	,43602	-,32707	-,17178
Környezetpolitika	-,06015	-,31444	-,21017	-,36753	-,14212	4,50604	-,07713	-,02873	-,05676	,25677
Nemzetközi tárgyalások	-,30322	-,1,90723	-,23350	,02904	-,17231	-,08399	-,17176	1,29856	-,34837	-,75161
Egészségügyi tanulmányok	-,23908	-,17083	,08103	,03500	-,28966	,06621	-,14249	-,15756	6,01222	,18900

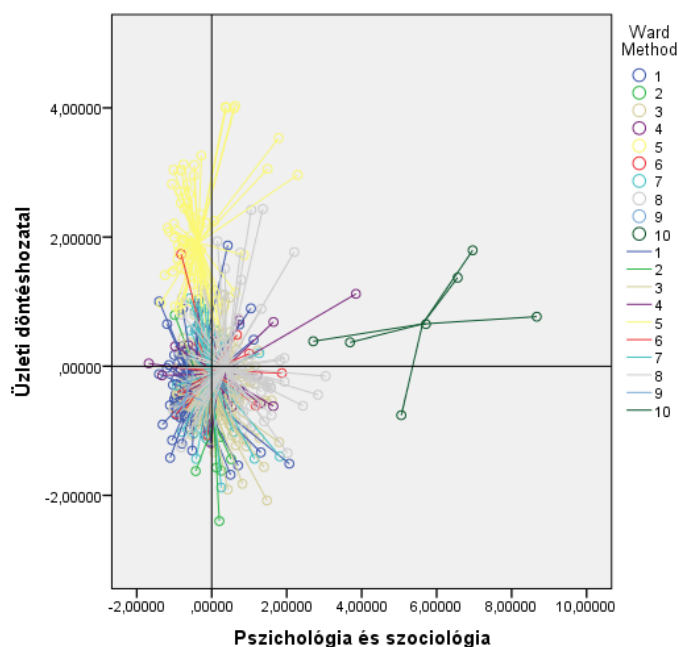
Forrás: SPSS, saját szerkesztés

Az egyes faktorok standardizált változók, vagyis az átlaguk 0, a szórásuk pedig 1. A pozitív klaszterátlagok a teljes adatbázis átlagától való magasabb, a negatívak pedig alacsonyabb átlagokat jelentenek. Minél jobban eltérnek a klaszterátlagok egymástól, annál jobban érzékeltetik az egyes klaszterek közötti különbséget.

A 10 klaszter, amelybe besorolásra kerültek a S&G 1999 és 2017 közötti kutatási cikkei, az alábbiak szerint alakul:

1. klaszter: 87 db cikk, témakörét tekintve leginkább a „Tanulás elmélet és kutatás módszertani” témákat fedik le.
2. klaszter: 12 db cikk, témakör szerint a „Politikatudományok és konfliktuskezelés” témaköreit dolgozzák fel, és kifejezetten nem jellemző ebben a klaszterben az átfedés a „Nemzetközi tárgyalások” témakörrel.
3. klaszter: 208 db cikk, a 9 főkomponens egyike sem mutat szorosabb kapcsolatot ezen cikkekkel, azaz nem sorolható be ezen kategóriák szerint.
4. klaszter: 28 db cikk, elsősorban „Nyelvi oktatás” témakörhöz kapcsolható cikkek.
5. klaszter: 70 db cikk, Gazdaság, üzlet és döntéshozatal témakörökhöz kapcsolódó cikkek.
6. klaszter: 17 db, „Környezetpolitika” témakör szerinti cikk.
7. klaszter: 30 db, „Szervezet és menedzsment tanulmányok” tárgykörhöz kapcsolható cikk.
8. klaszter: 95 db cikk, amelyek elsősorban Nemzetközi tárgyalások témakörhöz, de gyengébb erősséggel kapcsolódnak a Politikatudományok és konfliktuskezelés valamint a Pszichológia és szociológia témakörökhöz is.
9. klaszter: 7 db cikk, amelyek egészségügyi tanulmányok.
10. klaszter: 7 db, amelyek elsősorban „Pszichológia és szociológia” témájúak, de megjelennek a „Tanulás elmélet és kutatás módszertan” valamint az „Üzleti döntéshozatal” témakörök is.

A klaszterek ábrázolása is jól szemlélteti, hogy az üzleti döntéshozatalhoz kapcsolható cikkek első sorban az 5. klaszterben találhatóak, továbbá a többi klaszter átlagait vizsgálva leginkább a 10. klaszter elemei állnak közel az üzleti döntéshozatal témaköréhez.

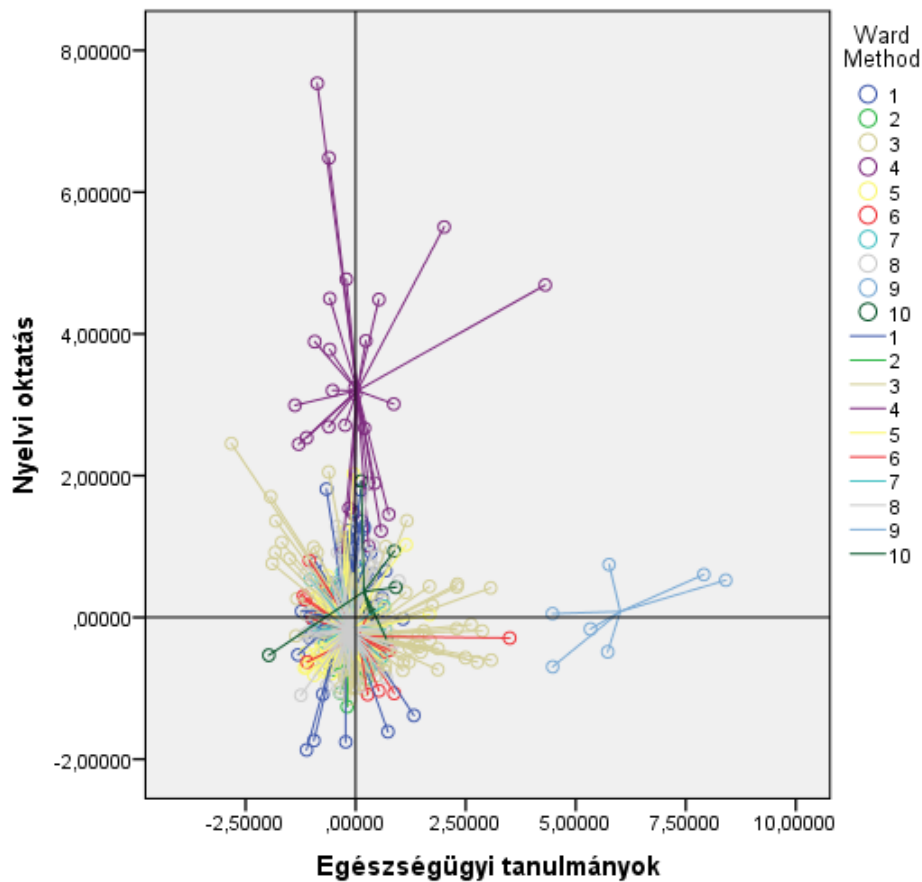


**3. ábra: A klaszterek ábrázolása az „Üzleti döntéshozatal” és a „Pszichológia és szociológia” dimenziók mentén**

*Forrás: SPSS, saját szerkesztés*

Egyes cikkek tekintetében az összes klaszter elemei mutatnak átfedést az üzleti döntéshozatal témakörében. Az is jól látható az ábrán, hogy az 5. klaszterbe sorolt cikkek nincsenek közelebb

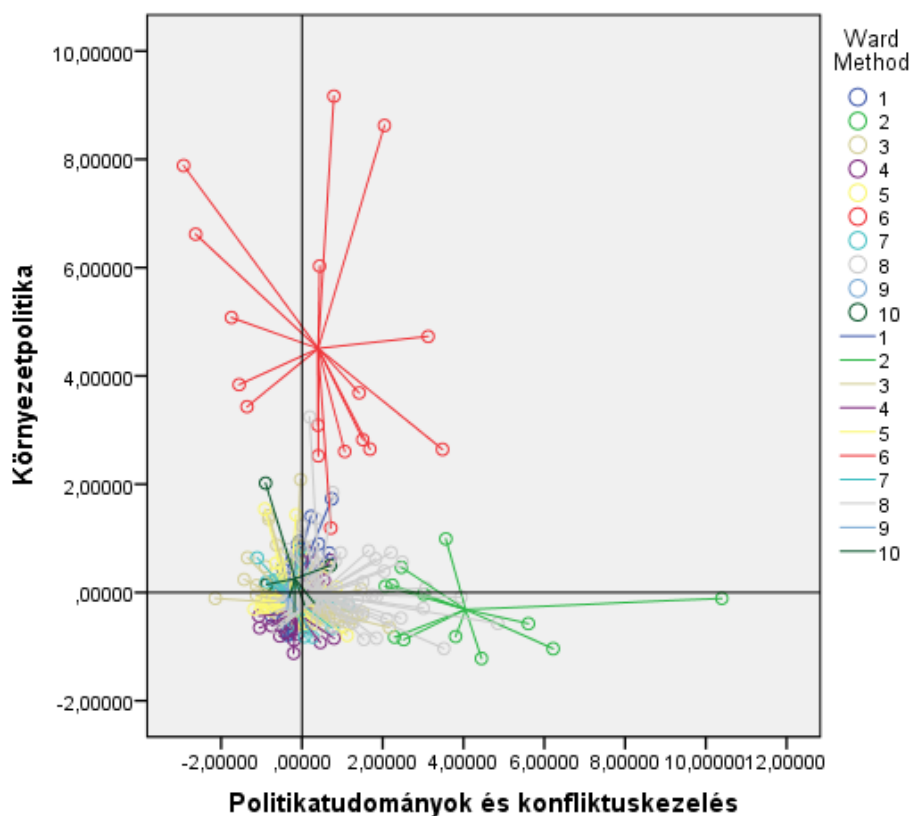
a pszichológia és a szociológia területéhez, mint a többi klaszter (a 10. klasztert kifejezetten ez a téma jellemzi).



**4. ábra: A klaszterek ábrázolása a „Nyelvi oktatás” és a „Egészségügyi tanulmányok” dimenziók mentén**

*Forrás: SPSS, saját szerkesztés*

Érdekességgént érdemes még megnézni a nyelvi oktatás és az egészségügyi tanulmányok koordináta rendszerben történő vizuális ábrázolását a klaszterek besorolásának. Jól látható, hogy a 4. klaszterbe sorolódnak egyértelműen a nyelvi oktatáshoz tartozó cikkek, míg a 9. klaszterbe az egészségügyi tartalommal rendelkező tanulmányok. A többi klaszter pedig viszonylag jól elkülönül ezen klaszterektől.



**5. ábra: A klaszterek ábrázolása az „Környezetpolitika” és a „Politikatudományok és konfliktuskezelés” dimenziók mentén**

*Forrás: SPSS, saját szerkesztés*

Ugyanez a párhuzam jól megfigyelhető a Környezetpolitika és Politikatudományok és konfliktuskezelés témakörök vonatkozásában, ahol a 6. és a 2. klaszter különül el jól láthatóan ezen témák mentén a többi klasztertől.

### Következtetések

A vizsgálat megmutatta, hogy egyértelműen elkülönültek egymástól a vizsgált témakörök alapján a cikkek. Ezt a megállapítást az a tény sem gyengíti, hogy a legtöbb, szám szerint 208 elem a 3. klaszterbe került besorolásra, ahol igazán markáns témakör szerinti besorolást nem tudott adni az elemzés. Ez ugyanakkor csak azt jelenti, hogy a *S&G* folyóirat által meghatározott témakörök egyikébe sem lehetett ezeket a cikkeket egyértelműen besorolni.

Megállapítható, hogy az elemzés előtt 899 fájlt tartalmazó lista jelentősen szűkíthető témacsoport szerint. A kutatási cikkek kritériumával ez a szám már csak 561 cikket jelent. Kutatási terület szűkítésével pedig objektíven megállapítható, hogy elsősorban az 5. és a 10. klaszterbe tartozó 70, illetve 7 darab cikket érdemes értő olvasással feldolgozni a bevezetésben megadott kutatási témára történő szűkítésnek megfelelően, ami azt jelenti, hogy a kutatási cikkek 13,7%-a, és a teljes fájlmenyiség 8,6%-a igényli a mélyebb szakirodalmi áttekintést. Bár az elemzések sok időt vettek igénybe, ezek nagy része gépi munka volt. A nyers adatok előállításával járó Excel makrók futtatása volt időigényes folyamat, ehhez azonban nem volt szükség emberi felügyeletre a futtatás során, csak a makrók megírása volt emberi energiát igénybe vevő folyamat. Összességében tehát megállapítható, hogy idő takarítható meg a szakirodalom tartalmi besorolásának számítógéppel támogatott feldolgozása során.

**Irodalomjegyzék**

1. Bragge, J., Thavikulwat, P., & Töyli, J. (2010). Profiling 40 years of research in Simulation & Gaming. *Simulation & Gaming*, Vol. 41(6), 869-897. doi:10.1177/1046878110387539
2. Cohen, A. M., Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, 6(1) 57-71. online elérhetőség: <https://dmice.ohsu.edu/hersh/briefings-05-cohen.pdf>
3. Faria, A. J., Hutchinson, D., & Wellington, W. J. (2009). Developments in business gaming: A review of the past 40 years. *Simulation & Gaming: An Interdisciplinary Journal*, Vol. 40, 464-487. doi: 10.1177/1046878108327585
4. Day, R., (1968). Beyond the Marketing Game-New Educational Uses for Simulation, *Proceedings of the American Marketing Association*, 581-588
5. Dale, A. G., & Klasson, C. R., (1962). *Business Gaming: A Survey of American Collegiate Schools of Business*, Austin, TX: Bureau of Business Research, University of Texas
6. Faria, A. J., (1987). A Survey of the Use of Business Games in Academia and Business, *Simulation & Games*, Vol. 18(2): 207-224. doi:10.1177/104687818701800204
7. Faria, A. J., & Nulsen, R., (1996). Business Simulation Games: Current Usage Levels, a Ten Year Update. *Developments In Business Simulation & Experiential Exercises*. Vol. 23
8. Faria, A. J., & Wellington, W. J., (2004). A survey of simulation game users, former-users, and never-users. *Simulation & Gaming*. Vol. 35(2): 178-207 doi:10.1177/1046878104263543
9. Ghosh, R., Haider, S., Sen, S. (2015). An integrated approach to deploy data warehouse in business intelligence environment. *Computer, Communication, Control and Information Technology* 1-4. online elérhetőség: <https://ieeexplore.ieee.org/document/7060115>
10. Graham, R. G., & Gray, C. F., (1969). *Business Games Handbook*, New York: American Management Association.
11. Holl, András, (2015). Szövegbányászat, adatbányászat, ismeretfeltárás Új lehetőségek a tudományos kommunikációban. Kézirat. Megjelent: *Magyar Tudomány* 2015(6): 680-685
12. King, W. R., & He, J. (2005). Understanding the role and methods of meta-analysis in IS research. *Communications of the AIS*, Vol. 16:665-686.
13. Nagarkar, S. P. (2015). Text mining: an analysis of research published under the subject category 'Information Science Library Science' in Web of Science Database during 1999-2013. *Library Review*, 64(3):248-262, doi:10.1108/LR-08-2014-0091
14. Nagy, Gyula (2016a). Tudományometriai és tartalmi elemzések szövegbányászati módszerekkel, Networkshop 2016 – Debrecen, 2016. április 1., online elérhetőség: [http://publicatio.bibl.u-szeged.hu/13216/1/Nagy\\_Gyula\\_2016\\_networkshop.pdf](http://publicatio.bibl.u-szeged.hu/13216/1/Nagy_Gyula_2016_networkshop.pdf)
15. Nagy Gyula (2016b). Tudománymetria és neveléstudomány. *Iskolakultúra*, 26(2): 50–62
16. Roberts, R. M., & Strauss, L., (1975). Management Games in Higher Education 1962 to 1974 -An Increasing Acceptance, *Proceedings of the North American Simulation and Gaming Association*, 381-385.
17. Sajtos, L., & Mitev, A., (2007). *SPSS Kutatási és adatelemzési kézikönyv*. Alinea Kiadó, Budapest

18. Schneider, G. (2014). *Applying Computational Linguistics and Language Models: From Descriptive Linguistics to Text Mining and Psycholinguistics*. Department of English Institute of Computational Linguistics University of Zurich.
19. Tomcsányi, P., (2000). *Általános kutatósmódszertan*. Gödöllő–Budapest, Szent István Egyetem, Országos Mezőgazdasági Minősítő Intézet
20. Vincze, V., Szarvas, G., Farkas, R., Móra, G., Csirik, J. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(Suppl 11): S9. doi:10.1186/1471-2105-9-S11-S9
21. Watson, H. J., (1981). *Computer Simulation in Business*, New York: John Wiley & Sons Publishing Company.