

# Az AVOBMAT (Analysis and visualization of bibliographic metadata and texts) bemutatása esettanulmányokkal

Péter Róbert, PhD  
SZTE BTK Angol-Amerikai Intézet

Networkshop  
Eger, 2018. április 4-6.

SZÉCHENYI 2020

EFOP-3.6.1-16-2016-00008



MAGYARORSZÁG  
KORMÁNYA

Európai Unió  
Európai Szociális  
Alap



BEFEKTETÉS A JÖVŐBE

# Előadás vázlata

1. Digitális fordulat a humántudományi kutatásokban c. EFOP alprojekt

# Előadás vázlata

1. Digitális fordulat a humántudományi kutatásokban c. EFOP alprojekt
2. Az AVOBMAT célja
3. Összehasonlítás és újdonságok

# Előadás vázlata

1. Digitális fordulat a humántudományi kutatásokban c. EFOP alprojekt
2. Az AVOBMAT célja
3. Összehasonlítás és újdonságok
4. Haladó vizualizációk
5. Új tudományos eredmények

**Digitális fordulat a humántudományi  
kutatásokban: új digitális módszerek és  
interdiszciplináris informatikai alkalmazások**

## **Digitális fordulat a humántudományi kutatásokban: új digitális módszerek és interdiszciplináris informatikai alkalmazások**

- Az informatikatudomány, számítógépes nyelvészet, szöveg- és adatbányászat legkorszerűbb eredményeinek humántudományokbeli alkalmazása.
- Új kutatási módszerek és digitális eszközök fejlesztése
- Eddigi humántudományi eredmények és ismeretek újragondolása és új kutatási területek kijelölése

# Kutatók és intézmények

- Dr. Farkas Richárd, Dr. Berend Gábor - MTA-SzTE  
Mesterséges Intelligencia Kutatócsoport
- Dr. Bilicki Vilmos - SzTE Szoftverfejlesztési Tanszék
- Nagy Roland, Simon Gábor informatikus hallgatók
- Sándor Ákos, Dr. Kokas Károly, Nagy Gyula - SzTE  
Klebsberg Könyvtár
- Dr. Labádi Gergely † és Dr. Péter Róbert - SzTE BTK  
Magyar és Angol-Amerikai Intézetek

# Labádi Gergely (1975-2017)



Digitális bölcsészet Szegeden, 2015



# TANIT - Text ANalysis Tools

## TANIT - Text ANalysis Tools

- magyar nyelvű szövegek számítógépes nyelvészeti feldolgozásával dokumentumok összehasonlító elemzéséhez szükséges statisztikákat kigyűjtése

## TANIT - Text ANalysis Tools

- magyar nyelvű szövegek számítógépes nyelvészeti feldolgozásával dokumentumok összehasonlító elemzéséhez szükséges statisztikákat kigyűjtése
- nyelvtechnológiai elemzőlánc kimenetére épülő aggregált statisztikákat számít ki
- témamodelleket épít

## Documents

kef.arcanum\_UTF8

Morph. parse

POS Dist.

Lemma Dist.

Lemma Dist. of associations

Lexical Richness & stats

Topic Model

Download

Back

### Morphological parse

Wordform	Lemma	MSD code	Morphological features
a		NOUN	Case=Nom Number=Sing Number[psor]=Sing Person[psor]=3
köszívű	köszívű	ADJ	Case=Nom Degree=Pos Number=Sing
ember	ember	NOUN	Case=Nom Number=Sing
fiái	fiú	NOUN	Case=Nom Number=Plur Number[psor]=Sing Person[psor]=3
harvan	harvan	NUM	Case=Nom NumType=Card Number=Sing
perc	perc	NOUN	Case=Nom Number=Sing
!	!	PUNCT	-
a	a	DET	Definite=Def PronType=Art
nagytiszteletű	nagytiszteletű	ADJ	Case=Nom Degree=Pos Number=Sing
úr	úr	NOUN	Case=Nom Number=Sing
éppen	éppen	ADV	-
közepén	közép	NOUN	Case=Sup Number=Sing Number[psor]=Sing Person[psor]=3
volt	van	VERB	Definite=Ind Mood=Ind Number=Sing Person=3 Tense=Past VerbForm=Fin
a	a	DET	Definite=Def PronType=Art
tószlak	tószlak	NOUN	Case=Dat Number=Sing
...	...	PUNCT	-

kef.arcanum\_UTF8

Morph. parse

POS Dist.

Lemma

Download

Back

## Parts Of Speech Distribution

### Part of speech Quantity

PRON	2576
AUX	45
PROPN	98
SYM	1
ADJ	10921
NUM	447
SCONJ	66
ADP	257
ADV	4496
PUNCT	36113
PART	4
VERB	20551
CONJ	157
X	234
INTJ	412
NOUN	36299

### Punctuation Quantity

..	1
?!	2
„	594

kef.arcanum\_UTF8

Download

Back

Morph. parse

POS Dist.

Lemma Dist.

L

## Lexical Richness and other statistics

- Sentence quantity: 8640
- Word quantity: 140588
- Average word length: 5.452655987708766
- Lexical richness metrics:
  - Guiraud's R: 45.75268608267229
  - Herdan's C: 0.8225394884569225

## Az AVOBMAT célja

- nagy mennyiségű metaadat és szöveg statisztikai elemzése és adatvizualizációja

# Az AVOBMAT célja

- nagy mennyiségű metaadat és szöveg statisztikai elemzése és adatvizualizációja
- régi hipotézisek tesztelése
- eddig ismeretlen történeti és kutatástörténeti trendek, folyamatok megjelenítése és új kutatási területek kijelölése

# Hasonló programok és újdonságok

- JSTOR Data for Research
- Scopus Analyse Search Results
- Gale Artemis

## **Az AVOBMAT előnyei:**

- felhasználó saját adatbázisa (Zotero csv, rdf, Ep3 xml, db) + ris-Zotero csv konverter
- Jóval több és összetettebb keresési, elemzési és vizualizációs funkciók



# AVOBMAT GUI

- Fazettás keresés

# AVOBMAT GUI

- Fazettás keresés
- Összetett keresés
- Parancssori keresés
- Skálázható fuzzy keresés

## AVOBMAT GUI

- Fazettás keresés
- Összetett keresés
- Parancssori keresés
- Skálázható fuzzy keresés
- Paraméterezhető szó-szomszédsági / szótávolság keresés (proximity search)
- szűkítés után egyszerű és haladó vizualizációk

## AVOBMAT GUI

- Fazettás keresés
- Összetett keresés
- Parancssori keresés
- Skálázható fuzzy keresés
- Paraméterezhető szó-szomszédsági / szótávolság keresés (proximity search)
- szűkítés után egyszerű és haladó vizualizációk
- adatok (csv) és grafikonok exportálása

#### Databases

- 19th\_newspapers\_hungary\_relevance
- bp\_title\_19th\_century\_countries
- worldcat\_freemasonry\_whole
- zspac\_freemasonry
- export\_acta\_xml2
- zotero\_export\_csv\_sample\_for\_test\_purposes

#### Date range yyyy OR yyyy-mm-dd

On  Before  After  Between

From

To

Search

#### Selection criteria

##### Item Type

- book (13)
- film (1)

##### Authors

- Anderson, James: Freemasons. Constitutions. (3)
- Anderson, Eugene L. (1)

#### Advance search

Field	Text	Fuzzy search	Proximity search
Authors	Anderson	1	1

Search

Reset

#### Commandline search

Example: authors:"Vass László"

Search

Sort by Publication date ascending

Total result: 14

- Anderson, James: Freemasons. Constitutions.  
*The Constitutions of the Free-masons : containing the history, charges, regulations, &c. of that most ancient and right worshipful fraternity.* (1723-01-01).  
London, Printed by William Hunter for John Senex . and John Hoake . (1723)
- Anderson, James: Freemasons. Constitutions.  
*The constitutions of the free-masons: containing the history, charges and regulations of that most ancient and right worshipful fraternity. For the use of the lodg*  
(1723-01-01).  
London, New York., Printed by Wm. Hunter for John Senex . R. Macey. (1723)
- Anderson, James: Freemasons. Constitutions.  
*The constitutions of the freemasons, containing the history, charges, regulations, &c., of that most ancient and right worshipful fraternity for the use of the lodg*  
*Printed in London, by William Hunter, for John Senex ... 5723: A.D. 1723.* (1854-01-01).  
Philadelphia, L. Dymeman, (1854)
- Anderson, William T.

# Tulajdonnév felismerés (NER)

- Farkas R, Berend G. et al „Automatic free-text-tagging of online news archives” *Frontiers in Artificial Intelligence and Applications* (2010), 529-534

# Tulajdonnév felismerés (NER)

- Farkas R, Berend G. et al „Automatic free-text-tagging of online news archives” *Frontiers in Artificial Intelligence and Applications* (2010), 529-534
- új webszolgálatatás

# Tulajdonnév felismerés (NER)

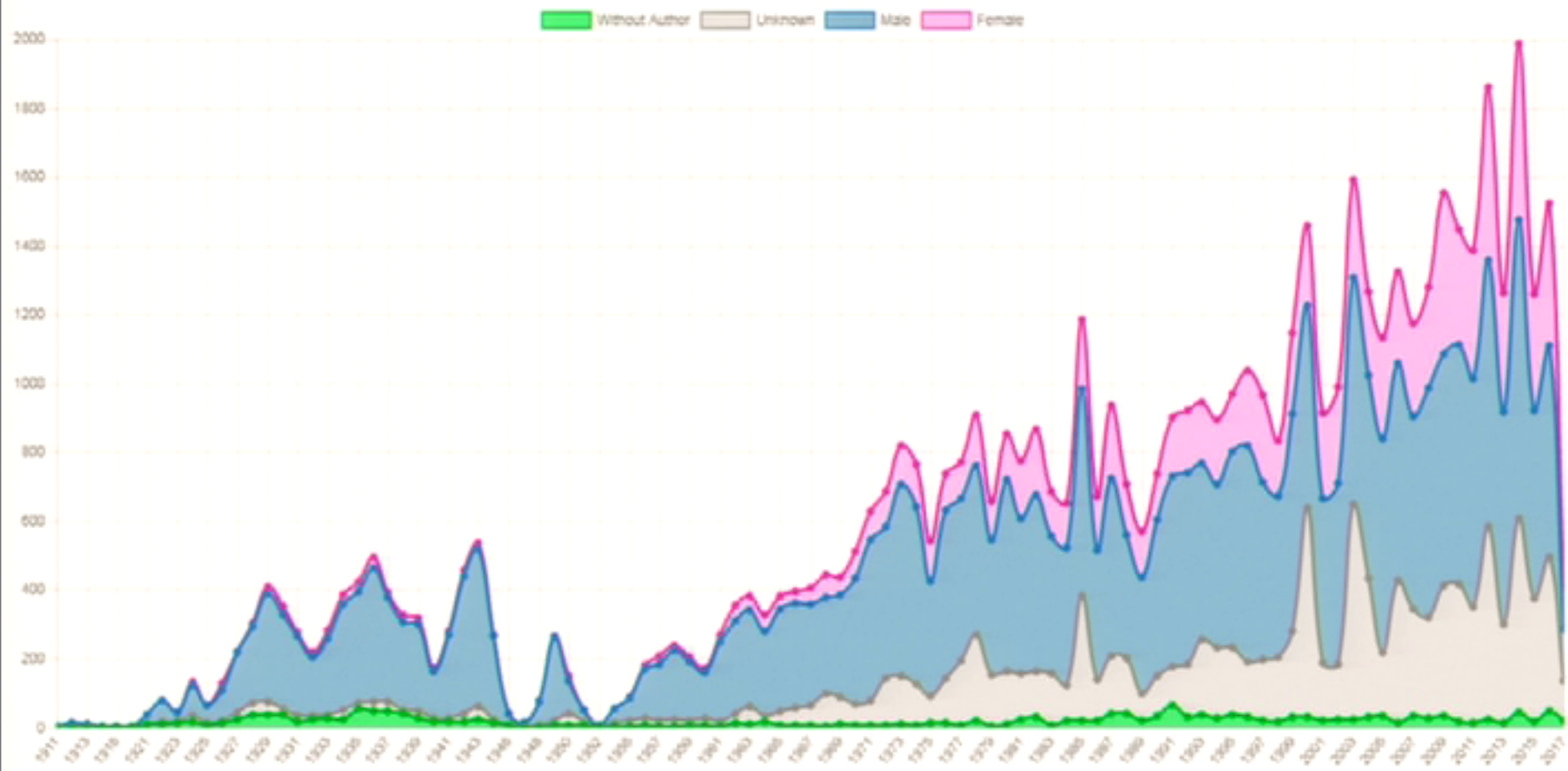
- Farkas R, Berend G. et al „Automatic free-text-tagging of online news archives” *Frontiers in Artificial Intelligence and Applications* (2010), 529-534
- új webszolgálatatás
- id és pdf url / szöveg
- személynevek és helynevek kilistázása és metaadatként való elmentése



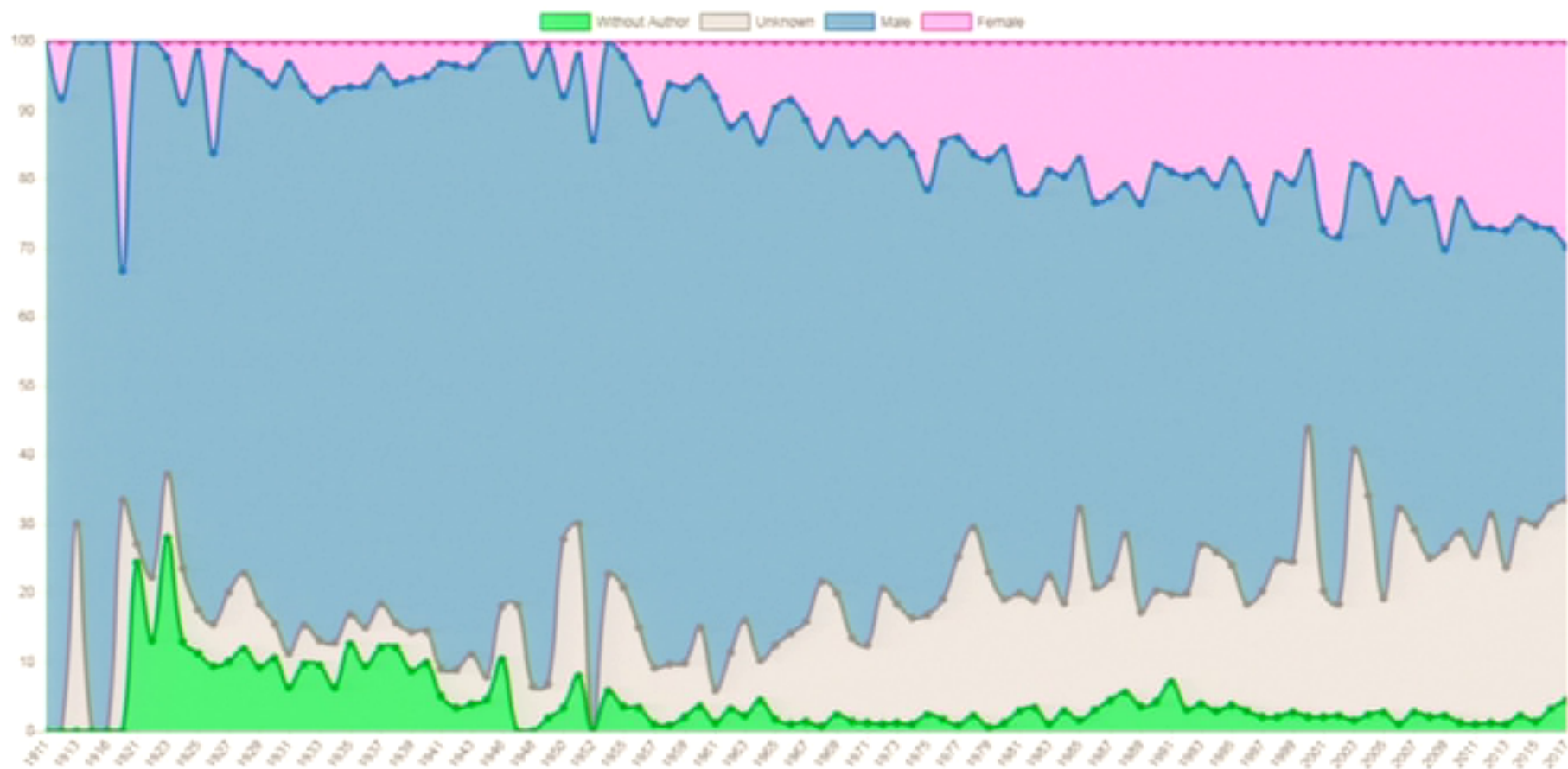
# Tulajdonnév felismerés (NER)

- Farkas R, Berend G. et al „Automatic free-text-tagging of online news archives” *Frontiers in Artificial Intelligence and Applications* (2010), 529-534
- új webszolgálatatás
- id és pdf url / szöveg
- személynevek és helynevek kilistázása és metaadatként való elmentése
- finomhangolás a szegedi Acta repozitórium 13000 cikkén (16 tudományterület)

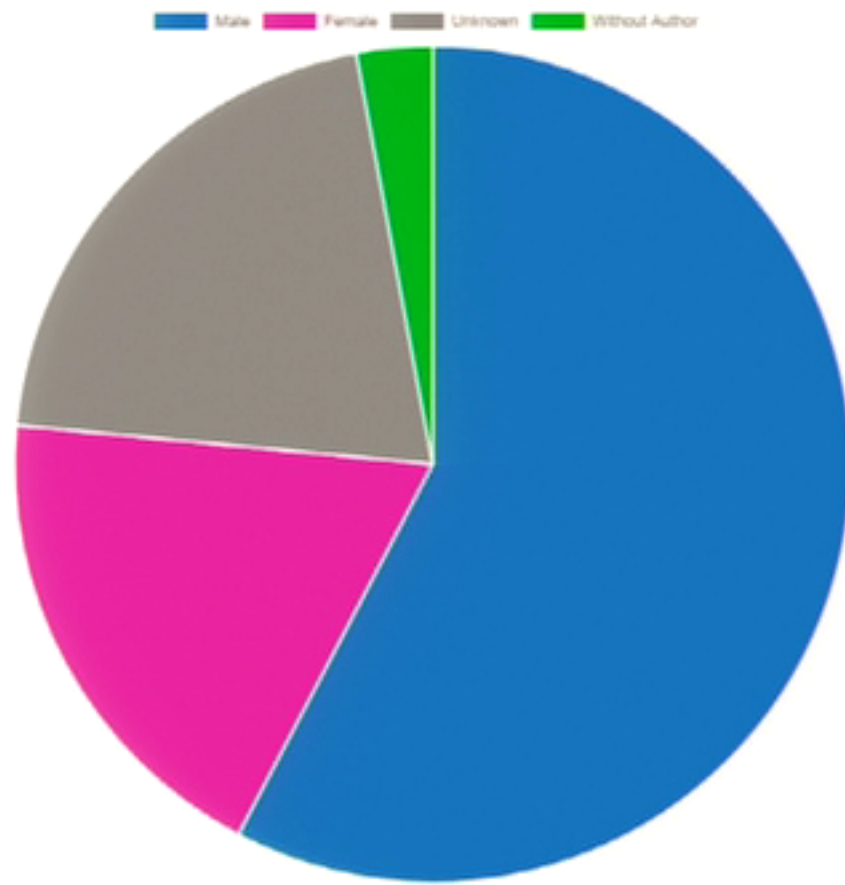
# Női-férfi szerzők azonosítása 1



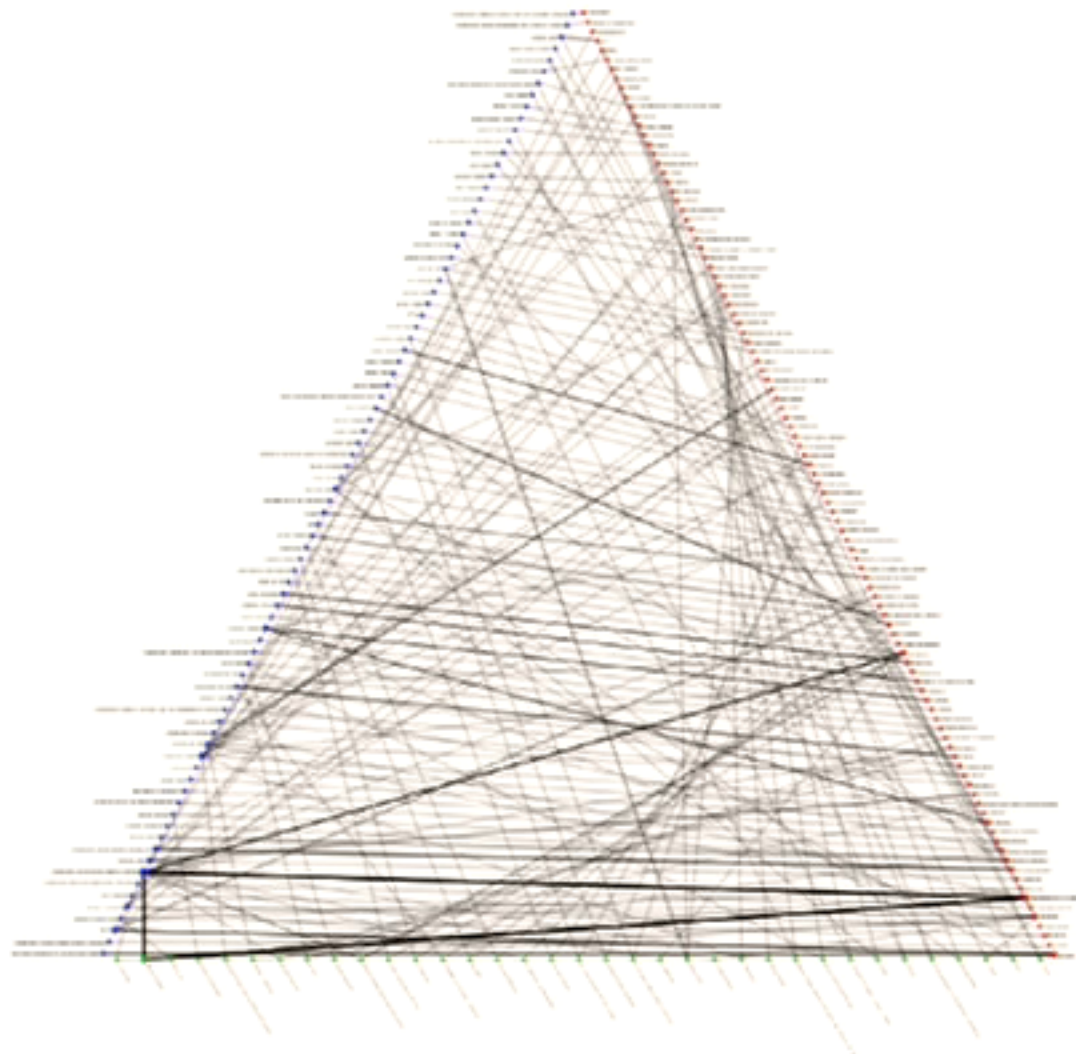
# Női-férfi szerzők azonosítása 2 - normalizált



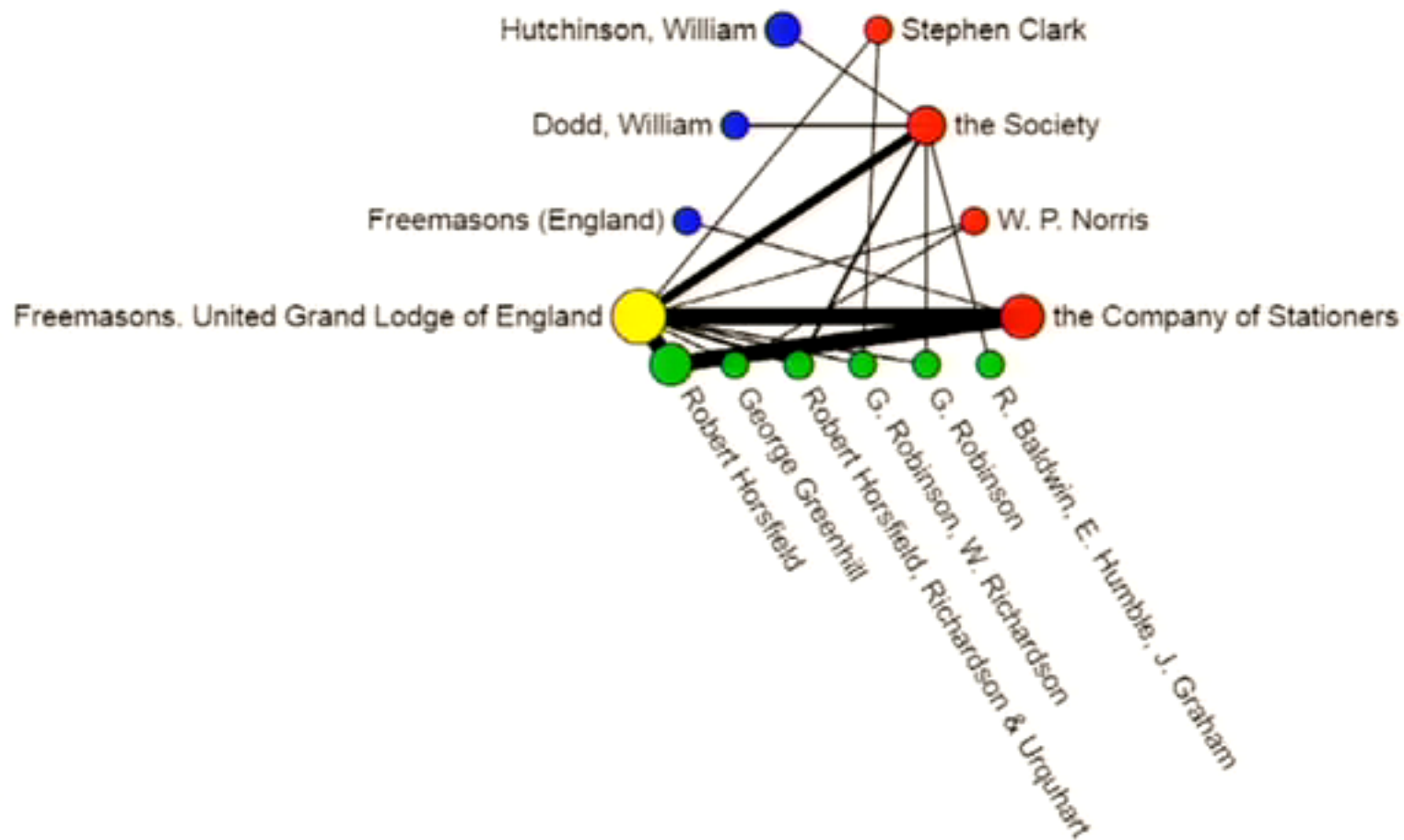
# Női-férfi vizualizáció 3 – paraméterezhető intervallumon



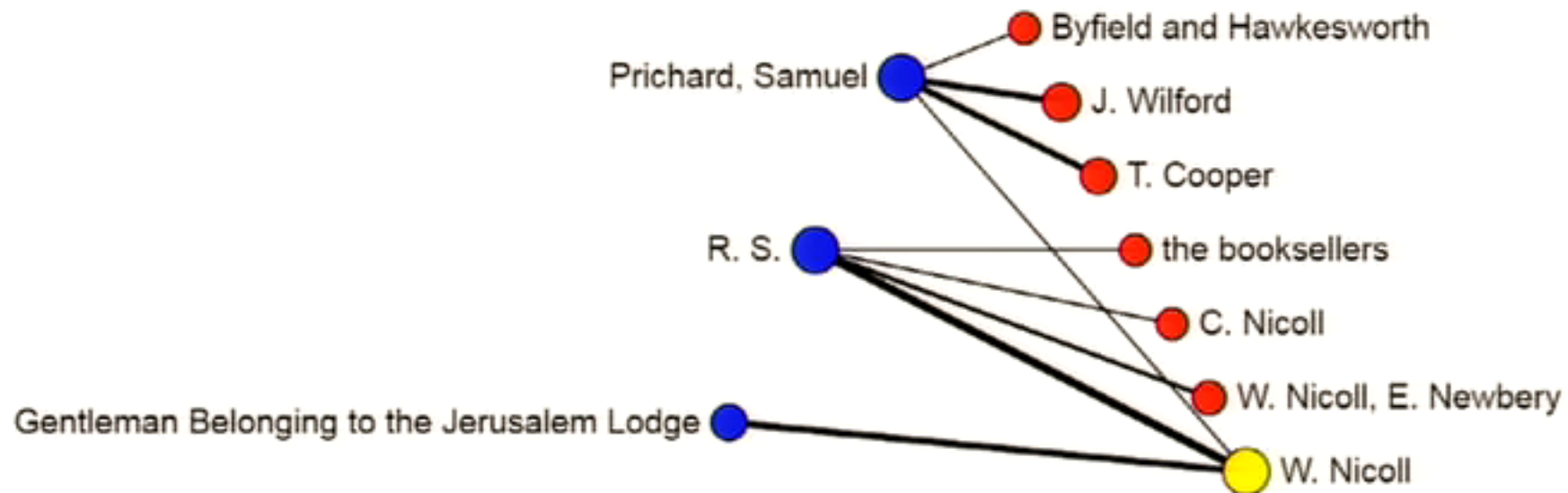
# Szerző-kiadó-eladó hálózat



# A Nagypáholy könyveinek kiadói és eladói



# W. Nicoll (kiadó) és a leleplezett publikált rituálék



# A szabadkőműves újságadatbázis struktúrája

Newspaper	Year	Month	Category	Issue	Locat	Burne	Locatio	Theme	Notes	Extract	FULL TEXT	Word Count
Daily Journal	1730	15.aug	letter to the editor				Rawlinso R1, Knoop	Debate, R	No. 25.—The Daily	Tothe Author of t		1835
Daily Journal	1730	18.aug	letter to the editor		223		Rawlinso CC 29	PMS!!!	letter to tl CC 28	STo the Author c		1612
Daily Courant	1730	02.jan	News	8808	London	2093		MN			On Saturday laft	49
Weekly Journal or Brit	1730	03.jan	News	240	London	3799		MN+			Yesterday 7-Nigl	67
Universal Spectator an	1730	17.jan	News	LXVII	London	1437		M+			On Thurfday lail	36
Daily Post	1730	22.jan	News	3227	London	707		MN+			We hear that laf	140
British Journal (1729)	1730	24.jan	News	108	London	669		MN+			Nirtbampton, Jai	34
Weekly Journal or Brit	1730	24.jan	News	243	London	644		MN+			thrtbajgbtm, Jan	52
Universal Spectator	1730	24.jan	Article	68 (173)	London		BP	MN - Elect	LONDON.		We hear that on	141
Daily Courant	1730	30.jan	News	8832	London	2117		M			Yesterday the Ar	67
London Journal (1720)	1730	31.jan	News	548	London	841		MN, PR - e	Election		On Tburfday laft	139
Weekly Journal or Brit	1730	31.jan	News	244	London	2907		MN, PR - election+			Laft Thurfday th	147
Universal Spectator an	1730	31.jan	News	LXIX	London	2571		MN, PR - election+			On Thurfday laft	138
London Evening Post	1730	31.jan	News	335	London	1875		MN, PR - election			On Thurfday the	138
Monthly Chronicle	1730	01.febr	Arts & Entertainment		London	3551		PU, T+			Register of Bool	48
London Evening Post	1730	05.febr	News	337	London	1829		CH+			Thg Spiciejy of '	24
Daily Journal	1730	09.febr	News	2837	London	478					Command of bis	99
Daily Journal	1730	13.febr	News	2841	London	950		WIII+			Sy Command, o	110
British Journal (1729)	1730	07.márc	News	114	London	2829		MNI+			Thurfday 7 Nigh	74
Weekly Journal or Brit	1730	07.márc	News	259	London	4351		MN+			Thurfday 7-Nigh	55
Universal Spectator	1730	07.márc	Article	74 (173)	London		BP	MN+	LONDON.		O.i Thu*flav lift v	61



# Tematikus fejezetek

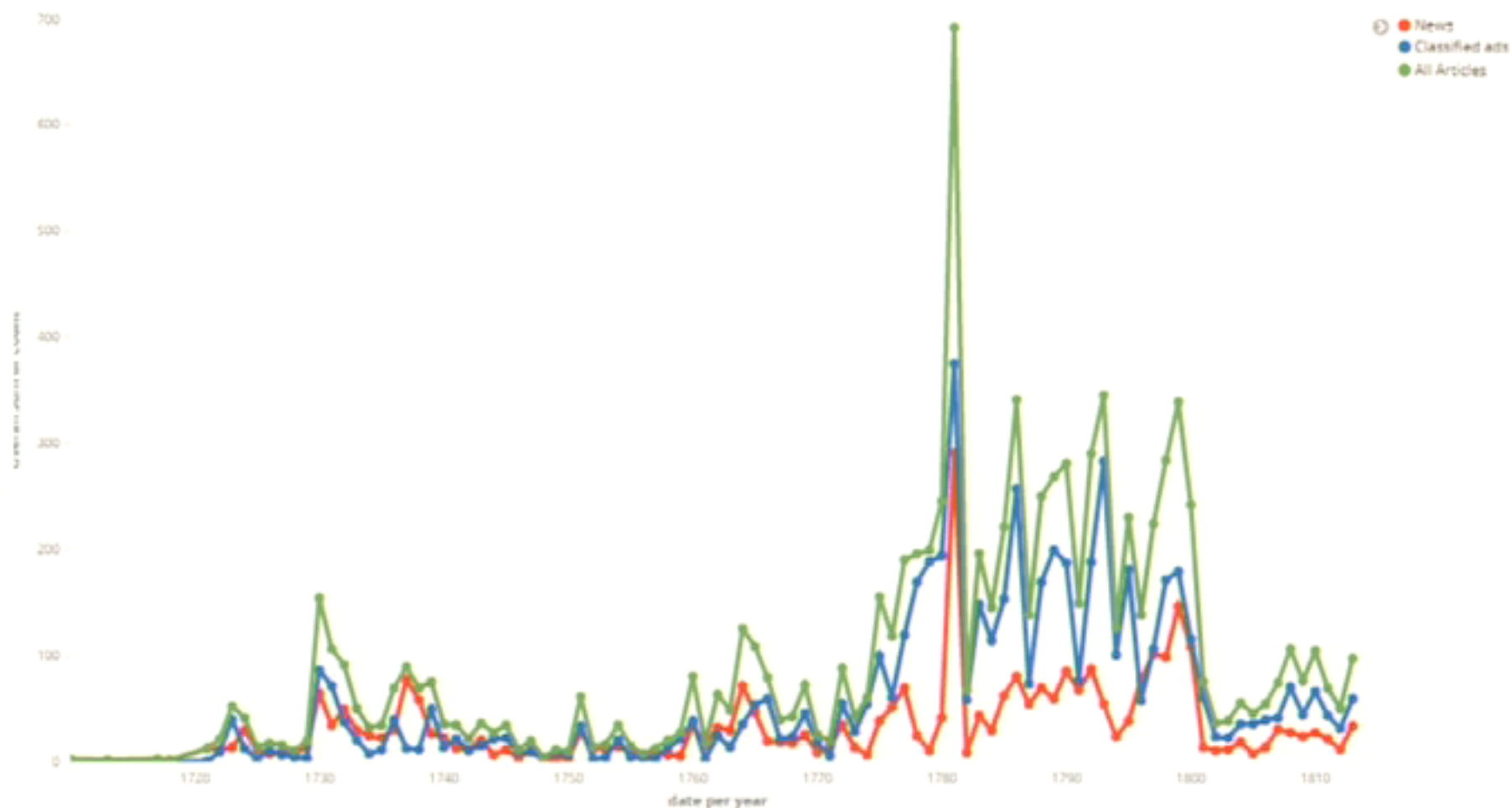
- Admissions and lodge meetings
- Processions
- Theatre
- Debates and conflicts
- Women
- British fraternal societies and the response to Grand Lodge Freemasonry

***British Freemasonry, 1717-1813***  
**(London / New York: Routledge, 2016)**

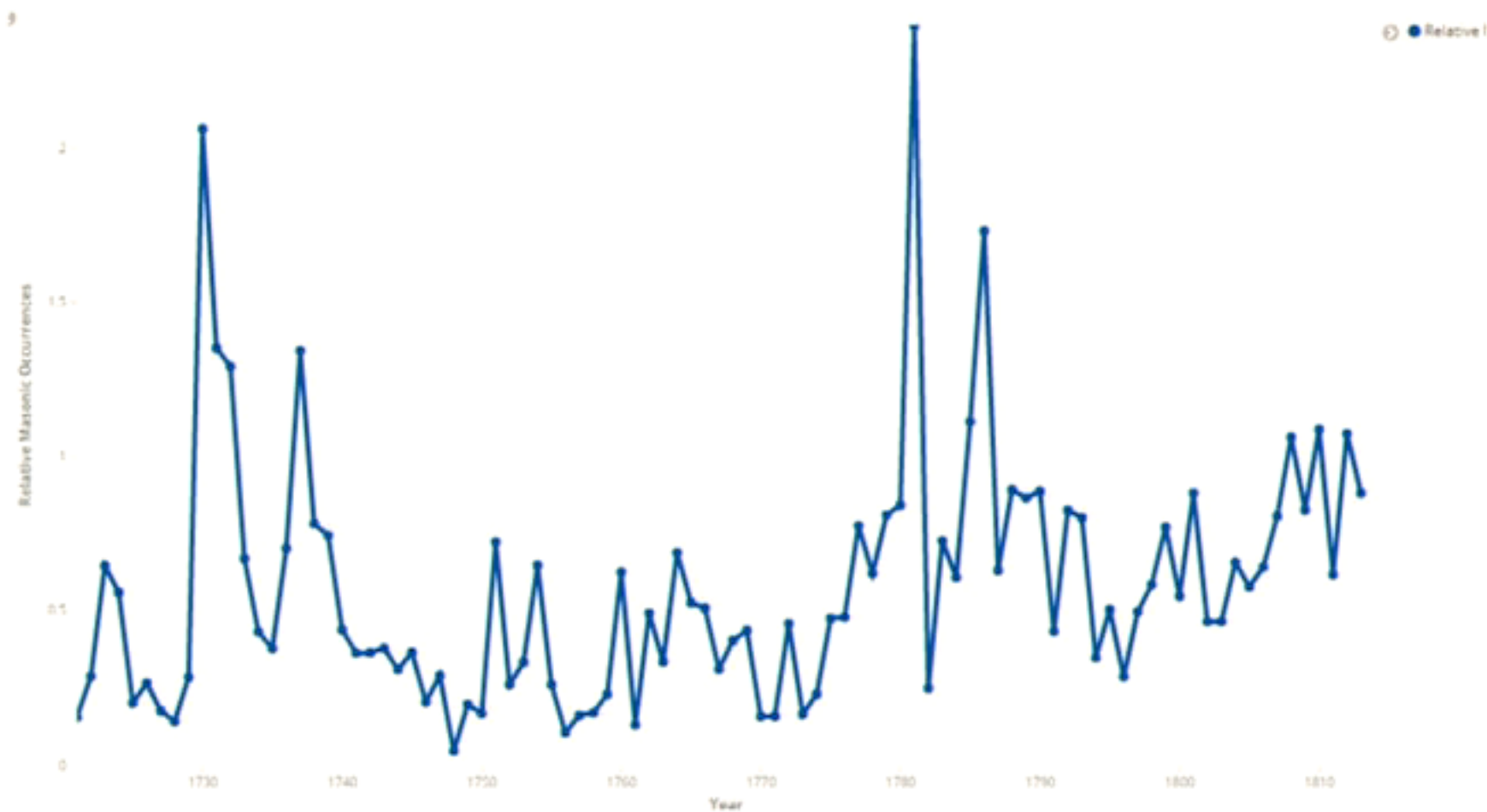
2606 o., 5967 szerkesztői jegyzet, 71 o. két oszlopos  
mutató

- Vol. 1. Institutions (Cécile Révauger)
- Vols. 2-3. Rituals (Jan A. M. Snoek)
- Vol. 4. Debates
- **Vol. 5. Representations** (480 o., 486 sajtó és folyóiratcikk, 1680 szerkesztői jegyzet)

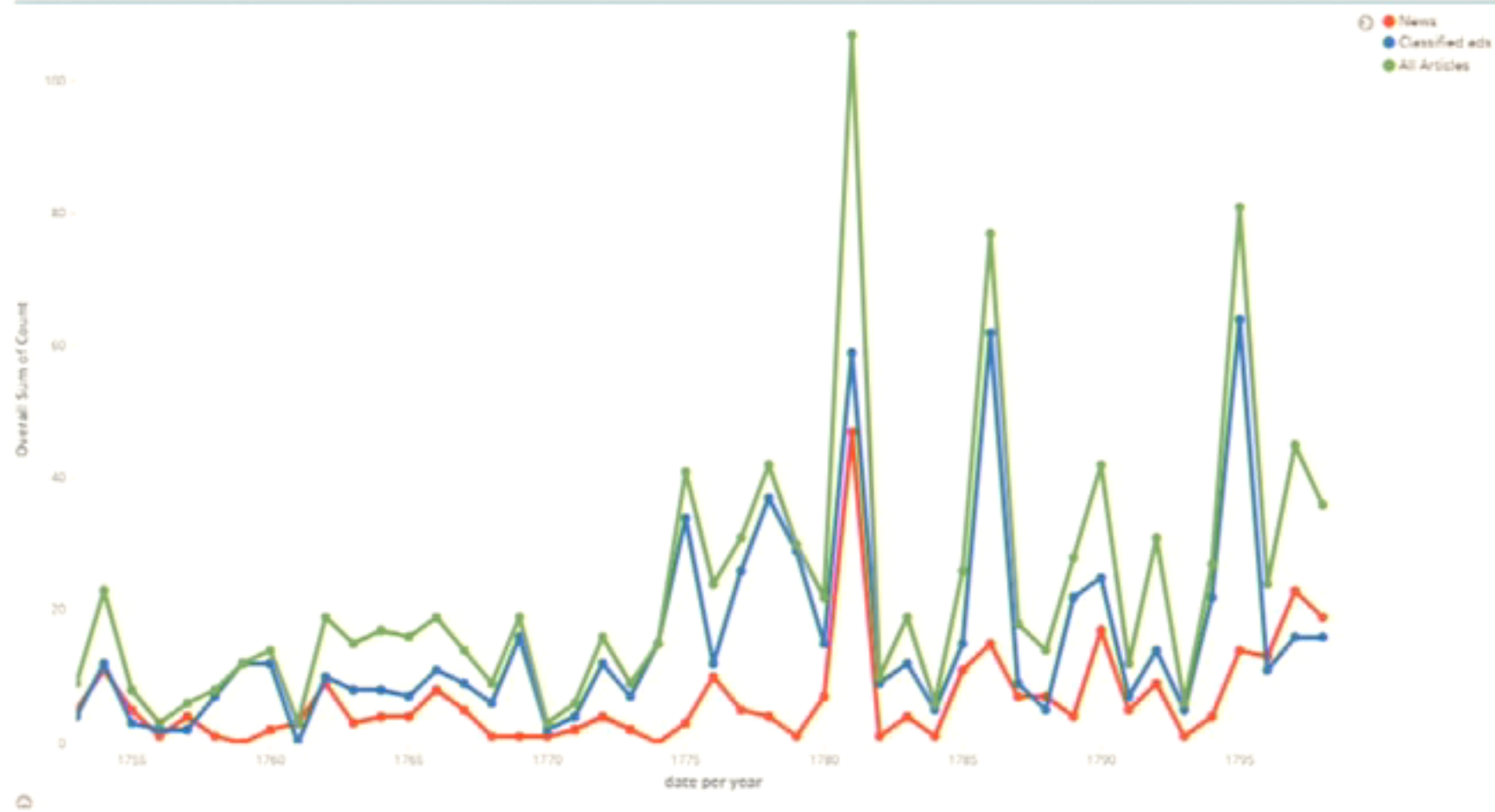
# Szabadkőművességre utaló cikkek száma, 1720-1800



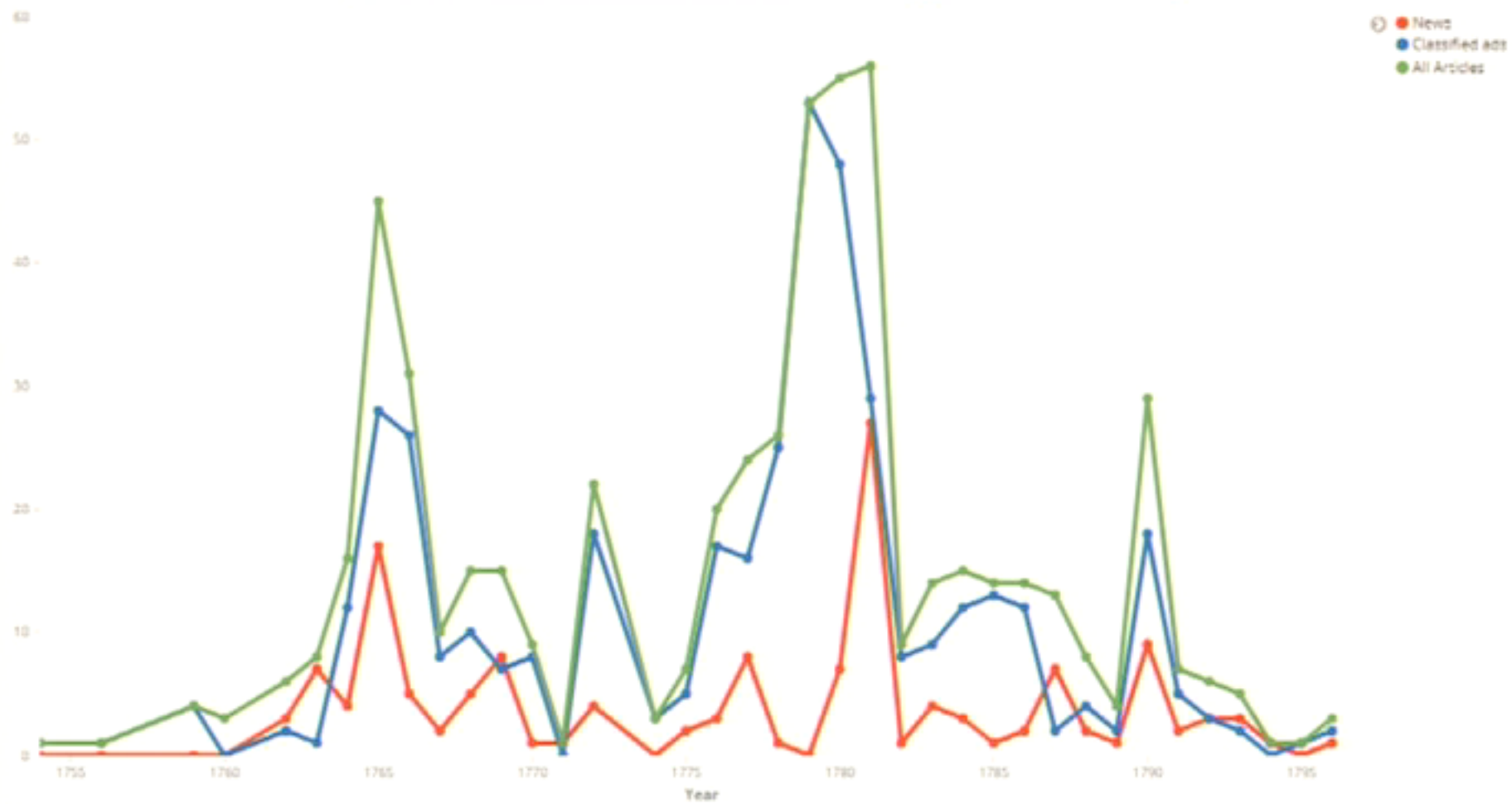
# Szabadkőművességre utaló cikkek – relatív statisztika



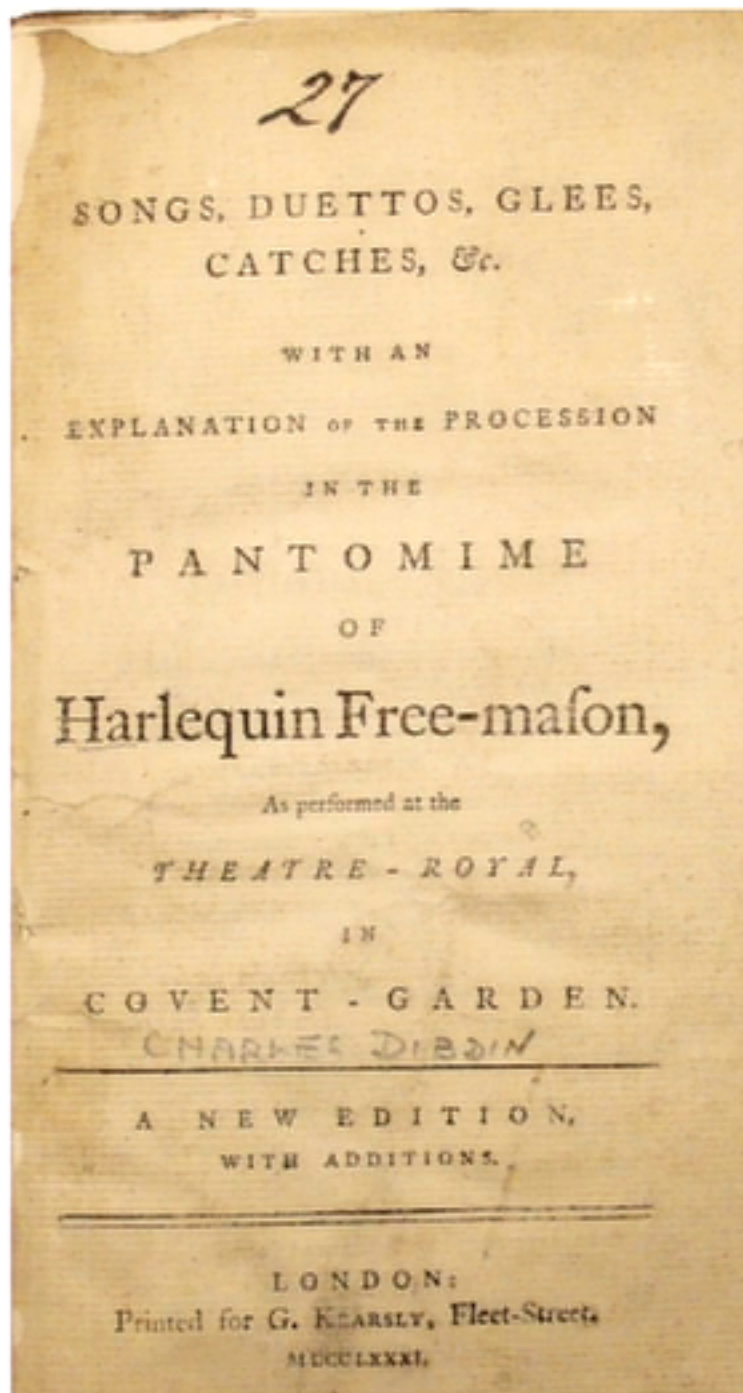
# Public Advertiser



## Gazetteer and London Daily Advertiser/ Gazetteer and New Daily Advertiser







## Harlequin Free-mason

- Sikeres pantomim 63 előadással 1780-1781
- Charles Dibdin (1745–1814)



## A szabadkőművességre legtöbbször és leggyakrabban utaló londoni újságok

<b>Newspaper Title</b>	<b>Number of References</b>	<b>Period</b>	<b>Average reference per year</b>
<b>Public Advertiser</b>	<b>846</b>	<b>1753 - 1793</b>	<b>20,6</b>
<b>Morning Chronicle</b>	<b>666</b>	<b>1791-1813</b>	<b>28,9</b>
<b>Gazetteer and New Daily Advertiser</b>	<b>550</b>	<b>1764-1796</b>	<b>16,7</b>
<b>Morning Post and Daily Advertiser</b>	<b>528</b>	<b>1772- 1792</b>	<b>25,1</b>
<b>World (1787)</b>	<b>459</b>	<b>1787-1794</b>	<b>57,3</b>



## James Perry (1745-1831)

- *Morning Chronicle* szerkesztője 1791-1821
- kormányellenes újság
- 'a professzionális újságírás megteremtője'
- Az Ősiek Nagypáholyának helyettes nagymestere (1787-1790)
- a szabadkőműves történetírás teljesen ignorálta
- A tradicionális kutatási módszerek nem alkalmasak ilyen eredmények megjelenítésére

Cím	Év	Hónap & nap	Kategória	Pédányszám	Hely	Szavak száma
His Mailesties Declaration Touching His Proceedings in the Late Assemblie and i	1622	01.jan	News		London England	3522
23 of May Weekly Newes from Italy	1622	23.máj	Classified ads		London England	63
23 of May Weekly Newes from Italy	1622	23.máj	News		London England	2107
28 of June Weekly Newes from Italy	1622	18.jún	Classified ads		London England	60
28 of June Weekly Newes from Italy	1622	18.jún	News		London England	1496
2 of September Two Great Battailes Very Lately Fought	1622	02.szept	News		London England	5678
Ninth of September 1622 Count Mansfields Proceedings since the Last Battaile	1622	09.szept	News		London England	3899
3 of Octob 1622 A True Relation of the Affaires of Europe	1622	04.okt	News		London England	4378
Newes from the Palatinate	1622	18.okt	News		London England	8460
October 30 1622 No 4 A Continuation of the Weekly Newes from Bohemia	1622	30.okt	News		London England	3719
November 5 1622 Numb 5 A Continuation of the Newes of This Present Weeke	1622	05.nov	News		London England	3622
February the 11 Numb 18 Weekly Newes	1623	11.febr	News	18	London England	2854
March 31 1623 Numb 24 Weekly Newes from Forraine Parts	1623	20.febr	News		London England	1767
February 28 Numb 20 The Newes of Forraine Partes	1623	28.febr	News	20	London England	1923
Aprill 8 1623 Numb 25 A Relation of the Last Newes	1623	08.ápr	News	25	London England	175
Aprill 8 1623 Numb 25 A Relation of the Last Newes	1623	08.ápr	News	25	London England	3944
Aprill 17 Numb 26 The Continuation of Our Former Newes	1623	17.ápr	News	26	London England	2644
April 21 Numb 27 The Continuation of Our Weekly Newes	1623	21.ápr	News	27	London England	243
April 21 Numb 27 The Continuation of Our Weekly Newes	1623	21.ápr	News	27	London England	6005
May 2 Numb 29 The Continuation of Our Weekly Newes	1623	02.máj	News	29	London England	5697
May 7 1623 Numb 30 A Relation of the Duke of Brunswicks March	1623	07.máj	News	30	London England	5143
May 12 Numb 31 The Newes of This Present Weeke	1623	12.máj	News	31	London England	5353
May 17 Numb 32 The Last Newes Relating These Particulars	1623	17.máj	News	32	London England	1951
May 26 1623 Numb 33 A Relation of Count Mansfeilds Last Proceedings	1623	26.máj	News	33	London England	5305

**Magyar vonatkozású cikkek az angol  
sajtóban, 1622-1800**

## Magyar vonatkozású cikkek az angol sajtóban (36816 cikk)

<b>Newspaper Title</b>	<b>Number of References Period</b>		<b>Average reference per year</b>
<b>Public Ledger or the Daily Register of Commerce and Intelligence</b>	<b>386</b>	<b>1760-61</b>	<b>193,00</b>
<b>Daily Gazetteer (Country Edition)</b>	<b>171</b>	<b>1741</b>	<b>171,00</b>
<b>Daily Gazetteer (London Edition)</b>	<b>1629</b>	<b>1735-45</b>	<b>148,09</b>
<b>General Advertiser (1744)</b>	<b>1019</b>	<b>1744-52</b>	<b>113,22</b>
<b>Weekly Journal or Saturday's Post</b>	<b>794</b>	<b>1717-25</b>	<b>88,22</b>
<b>Penny London Morning Advertiser</b>	<b>85</b>	<b>1744</b>	<b>85,00</b>
<b>Daily Post</b>	<b>2101</b>	<b>1719-1746</b>	<b>75,04</b>

# Magyar hamuzsír eladása Londonban, 1760-1774

For SALE by the CANDLE,  
At GARRAWAY'S Coffee-house in Exchange  
Alley, Cornhill,

On Wednesday the 13th of February, 1760,  
Immediately after Mess. Bliffon and Co. and Mark  
Hudson's Sale of Indigo,

THE following GOODS, viz.

- 24 Casks French Prize Indigo garbled.
- 2 Ditto ditto Duff.
- 3 Ditto Carolina Indigo.
- 100 Tons chipt Honduras Logwood.
- 30 Tons unchipt ditto.
- 2 Tons Jamaica ditto.
- 2 Casks Tortoisshell.
- 31 Casks Hungary Pearl Ashes.
- 10 Tons Brazilwood.

The aforesaid Goods to be seen from Monday  
Morning to the Time of Sale at the following Places,  
viz. the Indigo and Tortoisshell at No. 2. Dyer's-  
Hall; the Brazil and Pearl Ashes, Lot. 15 and 17,  
at the Steel-yard, and Lot 18 and 24, at Seething-  
Lane Warehouses under the Care of the Bull Porters;  
the 100 Tons chipt Honduras Logwood at Mr.  
Willis's Wharf, Westminster Bridge; and Lot. 11  
and 12, twenty Tons unchipt Honduras Logwood, at  
Mess. Smith and Speck's Wharf, and the remainder  
at Sir Peter Thomson and Co's Wharf St. Saviour's  
Dock; where Catalogues may be had, at the Place of  
Sale, and of

MARK HUDSON, Broker,  
Laurence Poultney Lane.

1760 február 9

For SALE by the CANDLE,  
At GARRAWAY'S Coffee-house in Exchange-  
Alley,

This Day, the 5th Instant, immediately after  
Mess. Bliffon and Co. Sale of the Cargo of the  
Anna Rufina,

THE following GOODS, viz.

- 110 Tons Honduras Logwood.
- 45 Tons Jamaica ditto.
- 55 Casks Hungary Pearl Ashes.
- 12 Ditto Dantzick Ditto.
- 5 Ditto amage! Argol.
- 187 Sicks Vêrjigresic.
- 27 Bags St. Domingo Cotton Wool.

The aforesaid Goods to be seen at the following  
Places, viz. The Honduras Logwood at Capt.  
Stevens's Wharf; the Jamaica Logwood at Smith and  
Speck's Wharf, opposite the Tower; the Hungary  
Ashes, at a Warehouse No. 4. in Heydon-Yard in  
the Minories, under the Care of Messrs. Thomas  
Mudge, and Co. The Dantzick Ashes, at Seething-  
Lane Warehouses, and the Remainder of the Goods  
at the Steel-Yard, under the Care of the Bull-  
Porters, where Catalogues may be had, at the Place  
of Sale, and of

MARK HUDSON, Broker.

1760 március 5

1837  
ECTIONS  
FOR MAKING *Emerson*  
CALCINED  
OR *Ministry*  
PEARL-ASHES,  
AS PRACTISED IN  
HUNGARY, &c.  
WITH  
A COPPER-PLATE Drawing of a Calcining  
FURNACE.

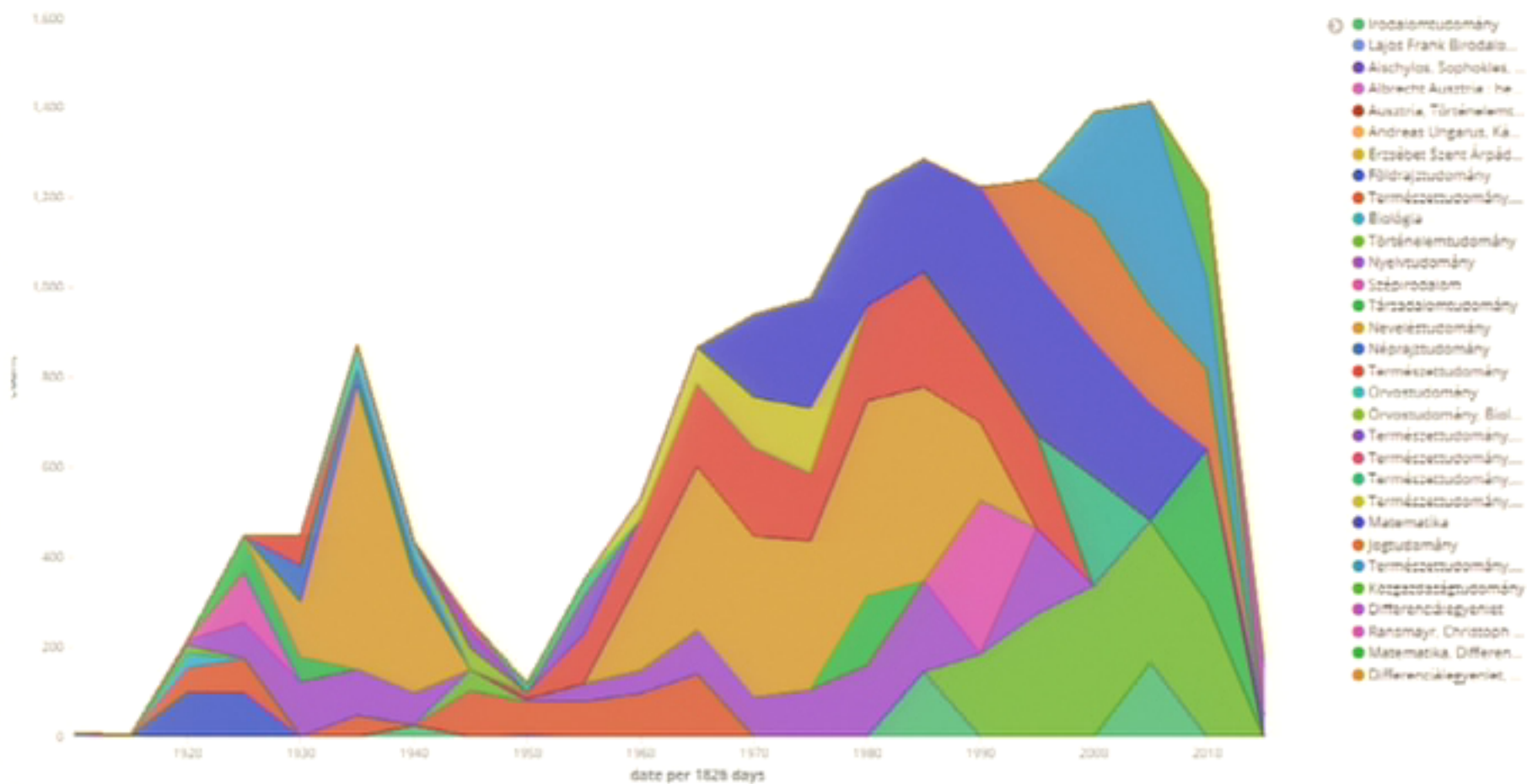
---

---

BOSTON:  
Printed for and sold by JOHN MERRIN, at the Lon-  
don Book-store, second Door above the British Caf-  
fee-house, North-side of King-street.

MDCCLXVI.

# Metaadatok javítása, pontosítása



## További fejlesztési tervek

- Női-férfi szerző azonosítás 40 nyelven



## További fejlesztési tervek

- Női-férfi szerző azonosítás 40 nyelven
- Új repozitóriumok és adatbázisok feltöltése (MEK, DIA ?)
- Feltölthető fájlformátumok bővítése
- Király Péter - metaadat ellenőrző webszolgáltatás beépítése a feltöltésnél?
- Jogosultságkezelés
- A könyvészeti anyag digitalizációja hamarosan befejeződik hazánkban – módszertani fordulat szükséges az adatbázisok építésében

# Köszönöm a figyelmet!

EFOP-3.6.1-16-2016-00008

SZÉCHENYI  2020



Európai Unió  
Európai Szociális  
Alap



BEFEKTETÉS A JÖVŐBE

# Módszer és esettanulmányok

- **“Researching (British Digital) Press Archives with New Quantitative Methods”** *HJEAS* 17.2 (2011): 283-300.
- “Representations of anti-masonry in eighteenth-century London newspapers” CMRC (2010)
- “Hidden Numbers: the Concept of Esoteric(ism) in the Nineteenth-century British Press” ESSWE (2011)
- “Eighteenth-century Hungary through English eyes - a terra incognita?” BSECS (2013)
- “New sources and novel digital methods in the study of English-Hungarian relations” HUSSE (2013)
- “The Digital and Methodological Turn in the Study of the Press: the distant reading of Hungary-related 18th-century English press material” *Aetas* 30 (2015), pp. 5-30.