Topic trends on the Hungarian Internet - news and academic web presence

György Kampis, László Gulyás,

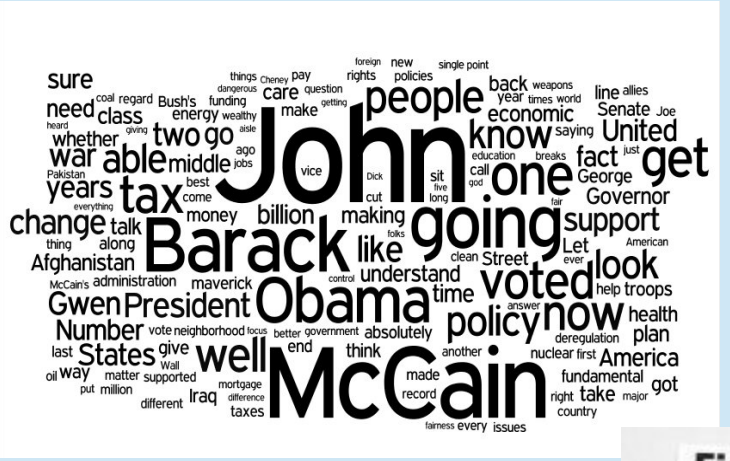
Zsolt Jurányi, **Sándor Soós**

Lorand Eötvös University, Budapest, Hungary

KNOWeSCAPE


© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com
THE NEXT BIG THING
THE NEXT BEST THING
search ID: mban1356

What's On?

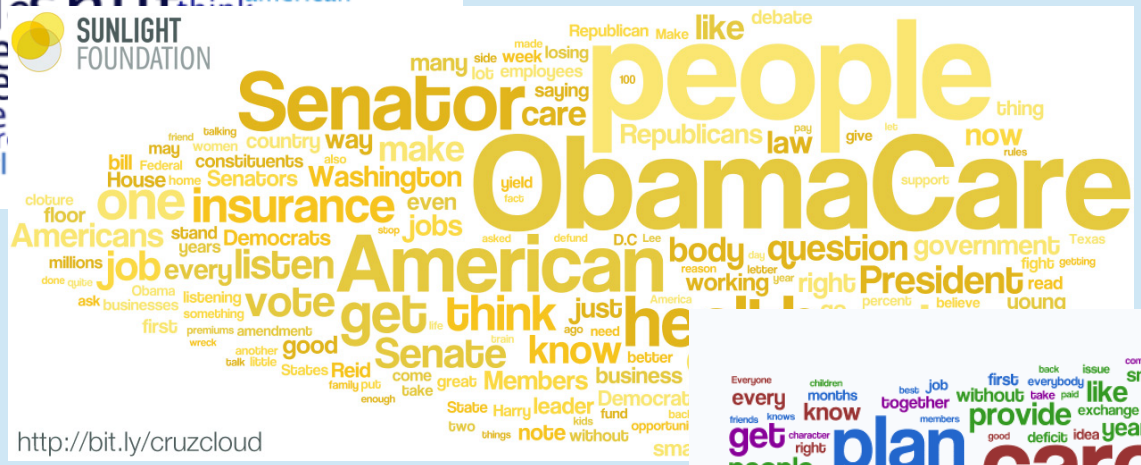

John
Barack
Obama
McCain
people
going
one
know
United
get
tax
change
support
voted
look
President
Gwen
Number
States
well
policy
now
health
plan
America
money
billion
making
like
understand
time
Afghanistan
administration
maverick
years
able
war
two
go
middle
fact
George
Governor
Street
Let
troops
absolutely
deregulation
nuclear
fundamental
got
take
Iraq
taxes
sure
need
class
energy
care
make
economic
Senate
talk
along
thing
think
end
another
made
record
country
issues
every
fairness


social
information
backstage
public
network
people
journalists
sites
private
online
photos
digital
personal
might
reporters
details
place
posting
MySpace
may
news
can
new
using
considered
use
one
consider
million
need
also
way
BBC
News
publish
readers
year
thought
content
students
advertising
friends
Facebook
impact
Journalism
Mediashifts
applying
easy
question
site
made
published
change
generation
Flickr
going
perhaps
intended
masses
editor
known
person
Imagine
UK
time
PDF
popular
York
explosion
concerns
skills
easier
assess
rules
whether
confere
differe
panel
interactions
natives
annual
survey
find
raised
famous
Goffman
McCann
tools
less
Saturday
obscurity
intrusion


Finding the Keys The
American revolutionaries
Save the Planet
Battleground
TOP of THE
Words Cannot Express
Why a light touch is b
Them All?
The Military Fights Back
INTERNATI
Giving China a bloody nose
Follow the leader?
Last Tuna
Public Financial Management
At last, progress
The Mad Experiment
free in this world
Welcome to My
Hope for the
Transparency
Opportunities in Economic and Financial


The Washing
Nixon
The New York Times
MEN WALK ON MOON
ASTRONAUTS LAND ON PLAIN; COLLECT ROCKS, PLANT FLAG
Chicago Daily Tribune
FINAL
AMELIA MISSING; SEARCH SEA
32 Hurt in Evanston Train Wreck

Topic Trends



Hungary


SZOCIÁLIS
KORMÁNY
POLITIKAI
SZABADSÁG
MAGYAR
DEMOKRATIKUS
ORBÁN
MINISZTERELNÖK
UKRAINE
Miskolc
Nyíregyháza
Kékes
BUDAPEST
Debrecen
Tisza
Szombathely
Székesfehérvár
Balaton
Dunaújváros
Kecskemét
Szeged
Pécs
Danube
SLOVENIA
CROATIA
SERBIA
ROMANIA
magyar
kommunista
párt
ma
előtt
politikai
Imre
Polgártársak
fiatalok
évvel
orosz
http://eo.blog.hu

Pilot Study

- News Portals
  - Public discourse
- Academic Institutions
  - Academic discourse

Started in 2013.

Tools

- Heritix crawler
  - Modified and configured to our purposes
  - Applied by several earlier national archive projects
  - Including at the British Library
- Dell T710 server
  - 2x4 core Xeon E5520
  - 48 GB RAM, 2 TB HDD

Sources

- Two major news venues
  - Hirarxhivum. hu, Index.hu
  - Containing: 155 different online news venues
- Home pages of 48 academic institutions
  - All universities
  - Institutions of the Hungarian Academy of Science
  - Full list is available at http://hungarianscience.org/

HungarianScience.org

THE RESEARCH PERFORMANCE OF HUNGARY: INSTITUTIONAL PATTERNS AND COMPETENCES

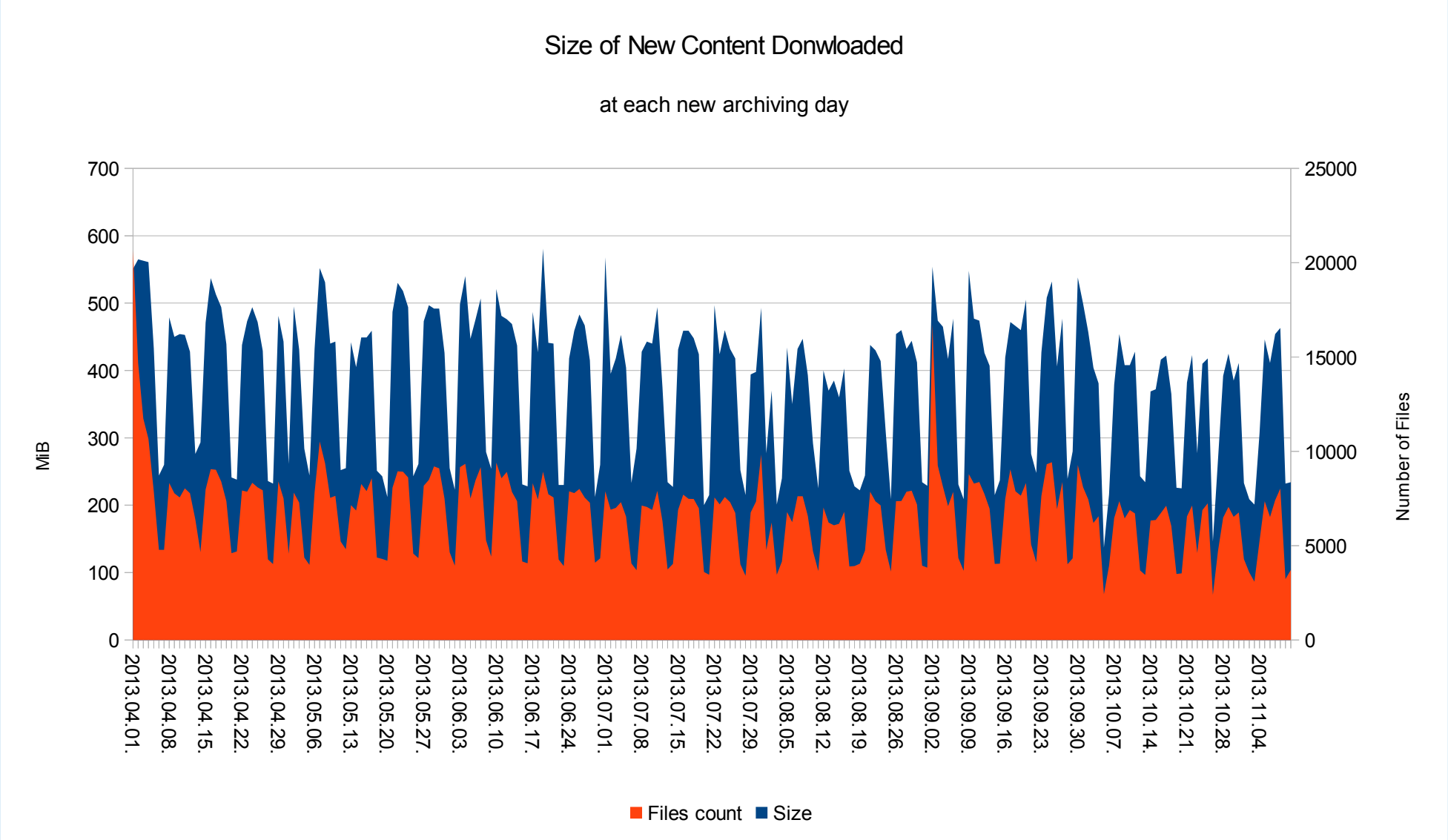
Content Stored

- Sites are revisited
  - ~once in 2 weeks
- Mainly text and video files stored
  - Excluded: exe, gz, iso, jar, mp3, ogg, ppt, rar, waw, xls, xlsx and zip
- No recreation of archived content (long term goal)

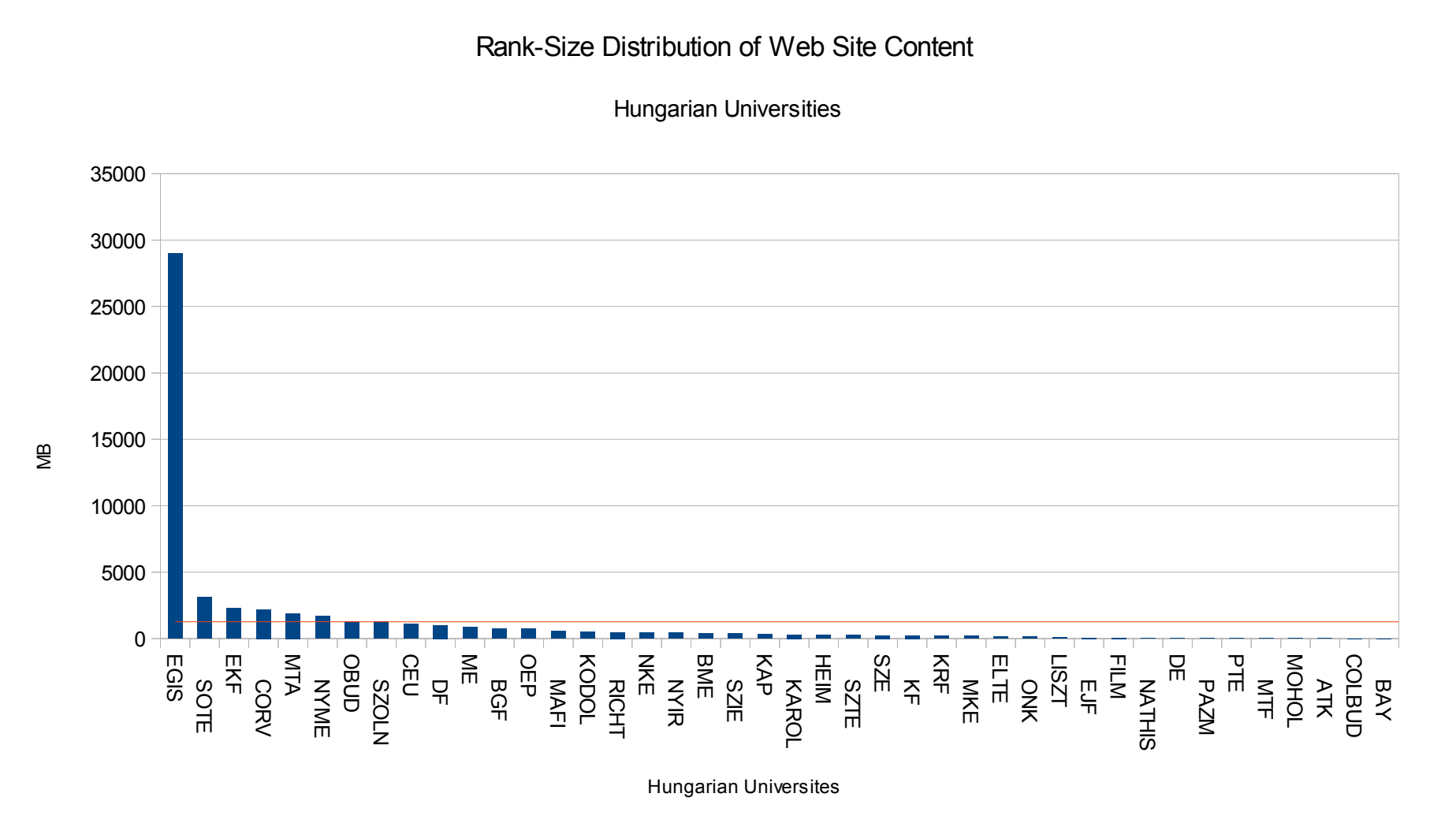
Big Data is Small (for Hungary)

- Overall Quantity is Small
  - Update frequency is also modest
- News (one copy)
  - Size: ~ 600 MiB
  - Num of Files: ~8000 files
- Academia (one copy)
  - Size: 53 Gb
- Yet, distributions are long-tailed

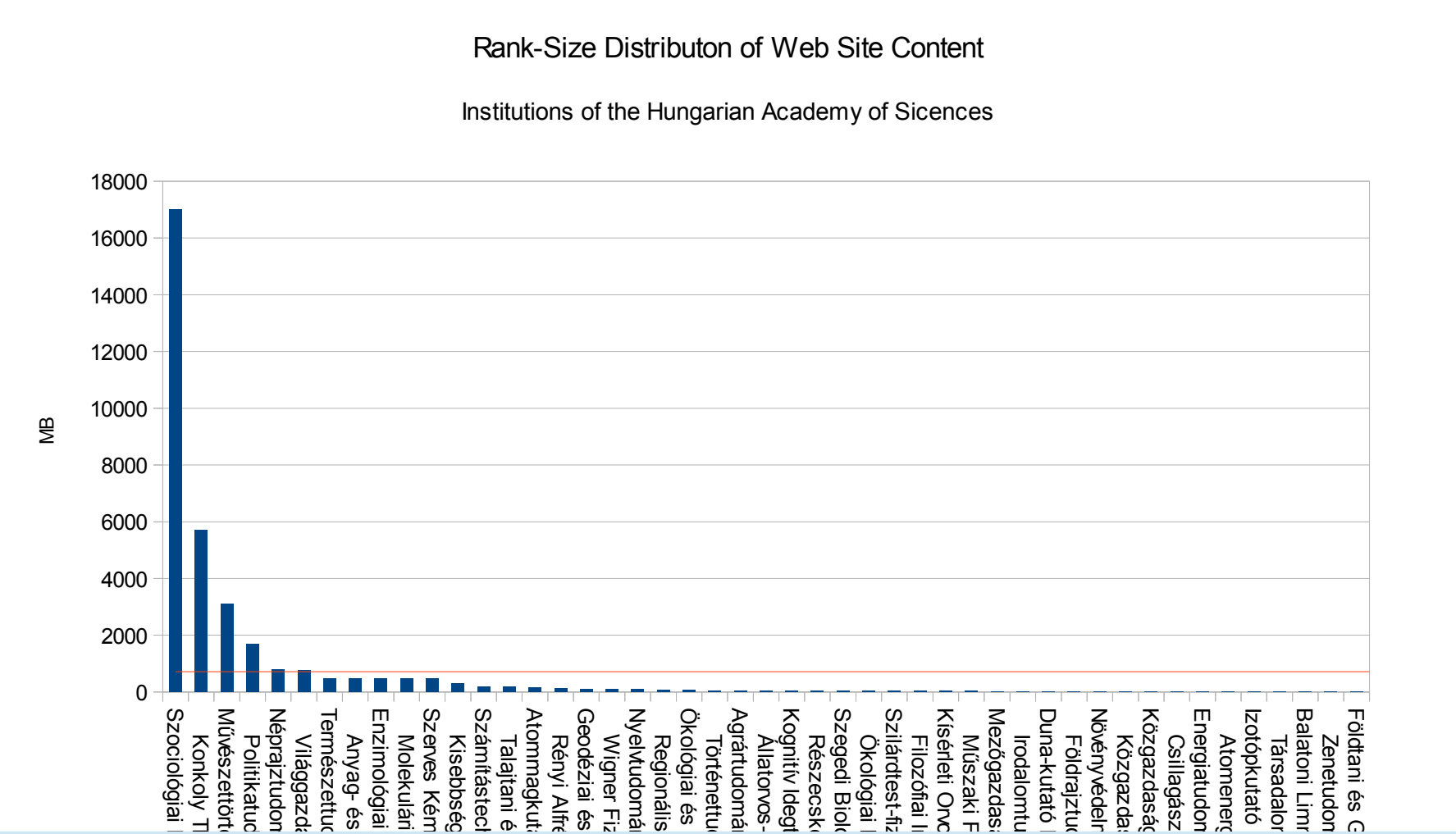
Size of New Content on News Sites



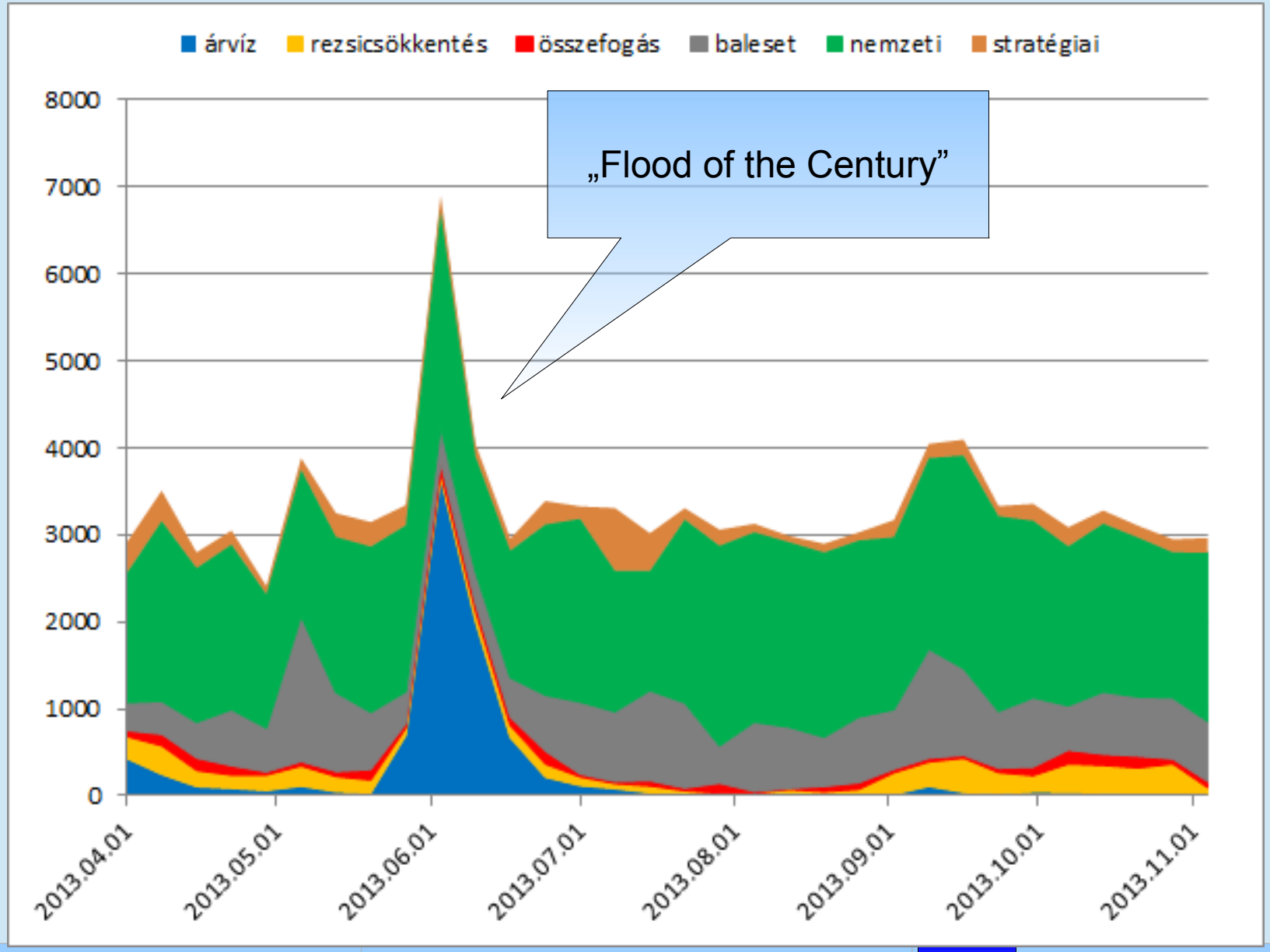
Rank-Size Distribution of University Web Sites



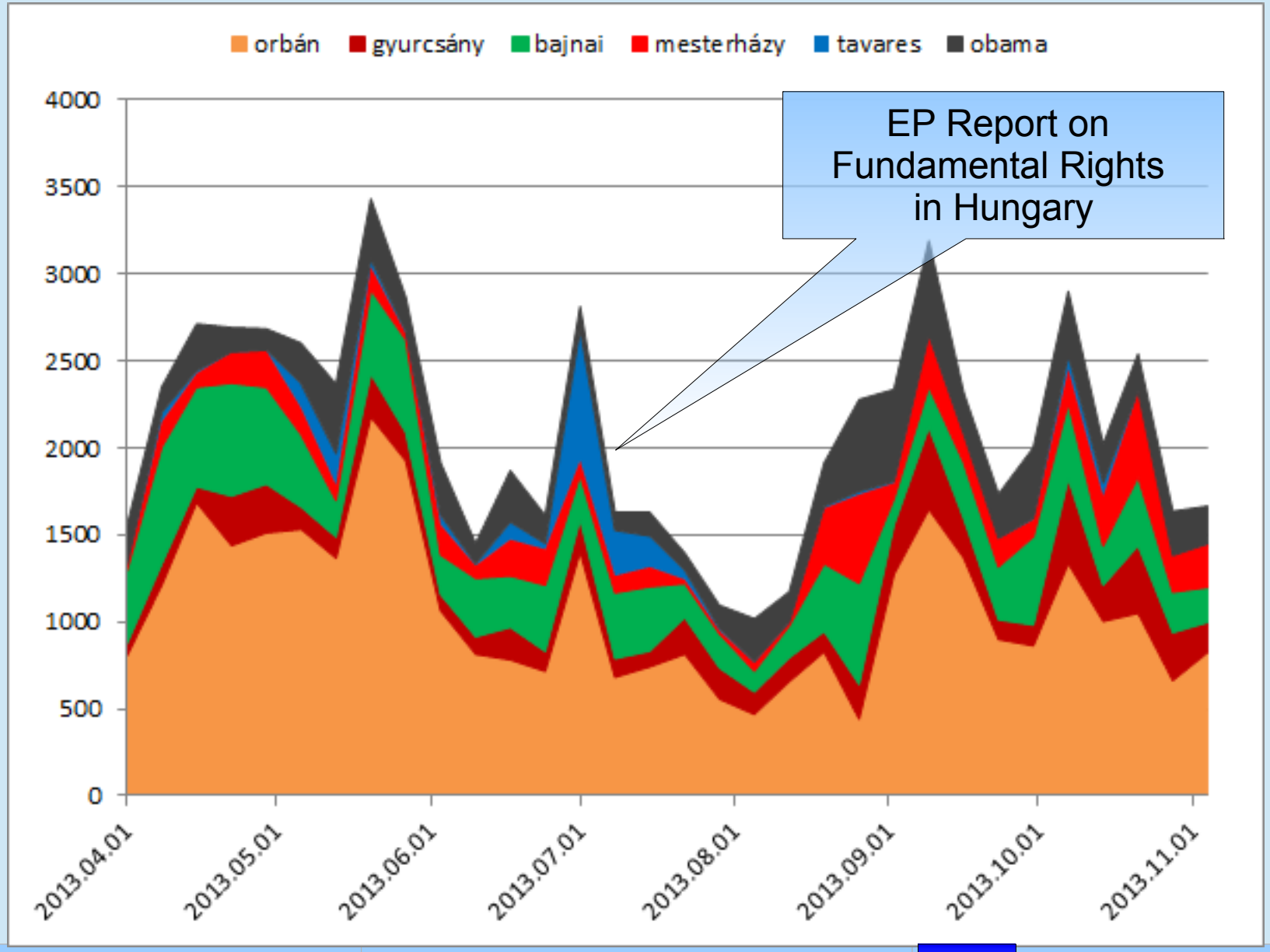
Rank-Size Distribution of Hungarian Academy of Sciences Web Sites



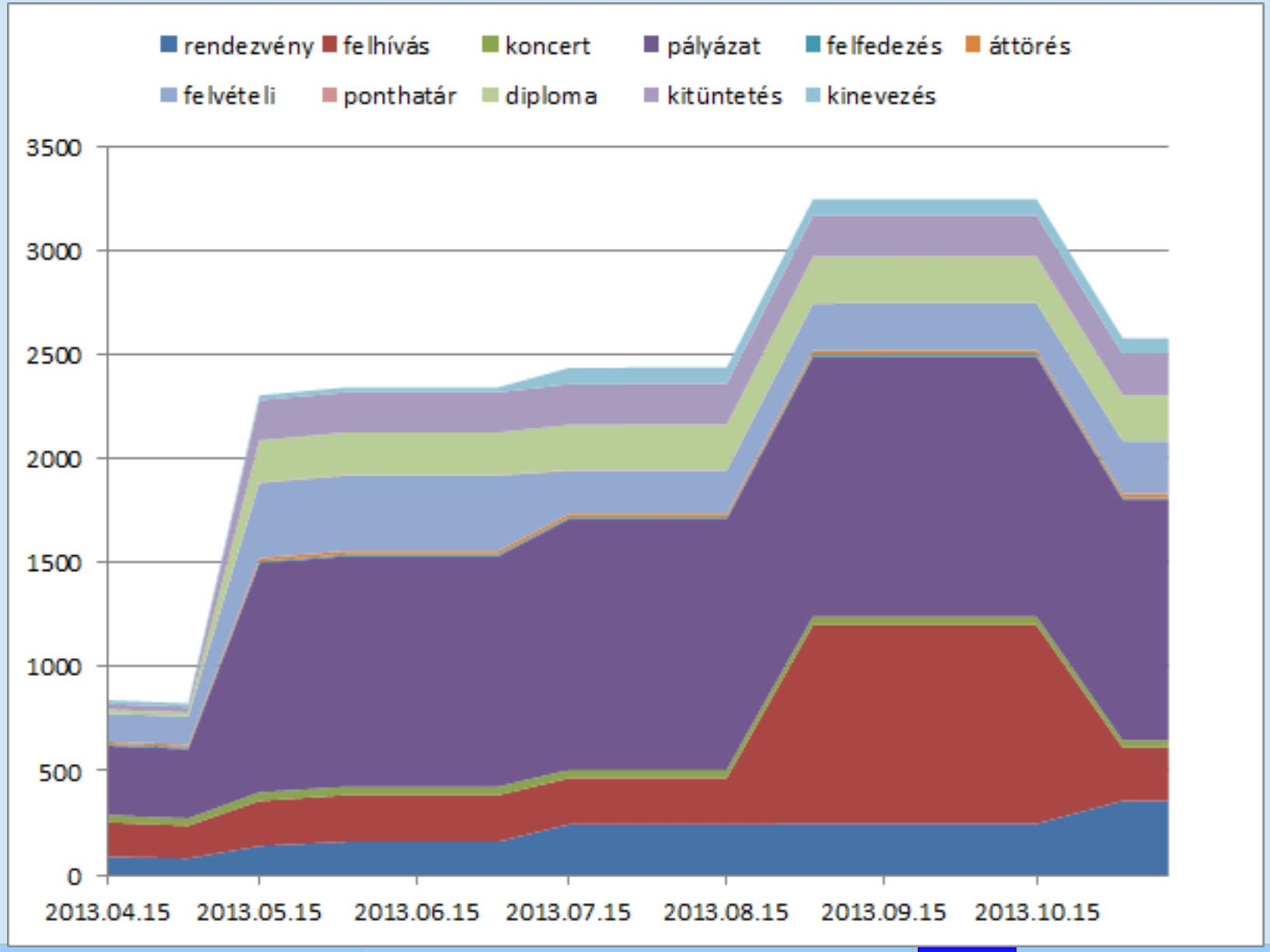
Topics Trends in News



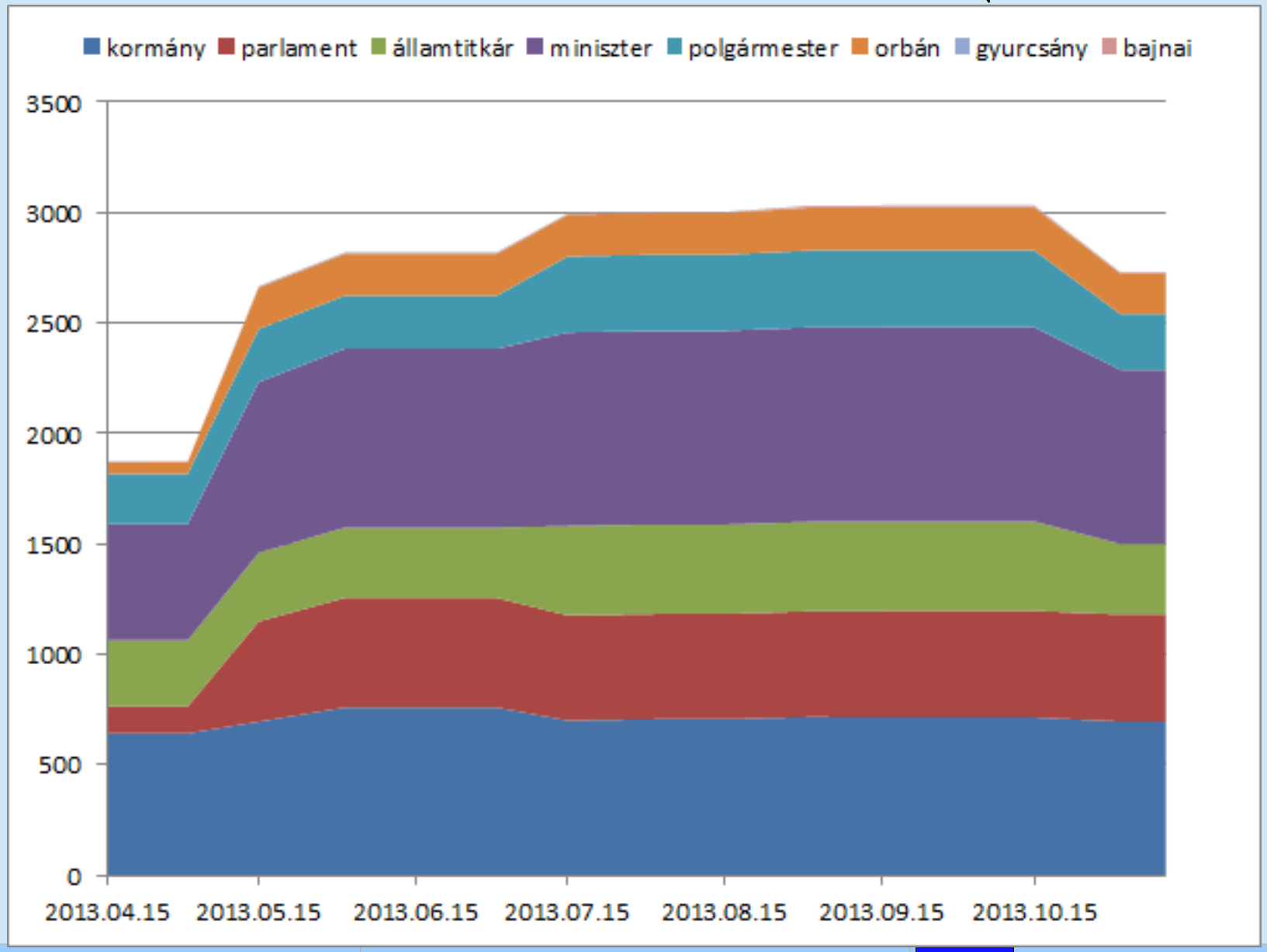
Topics Trends in News



Topics Trends in Academia



Topics Trends in Academia (political)



Summary and Conclusions

- A pilot study in Hungary targeting public internet content of
  - Public news portals
  - Academic research institutions
- Tools and methods
  - In continuous development
- "Big Data" is unexpectedly small
  - But has long tails

Future Works

- More sophisticated longitudinal analysis
  - More data
  - Seasonal corrections
  - Etc.
- Change rate analysis
  - Speed
  - Correlation between (academic) web site activity and scientific output / success

Acknowledgements

- This work was partially supported by the European Union and the European Social Fund through project (TAMOP-4.2.2.C-11/1/KONV- 2012-0013).

- Building on technology from

- With contributions from
  - Zsolt Jurányi, Balázs Bálint, Attila Pálmai

HungarianScience.org

THE RESEARCH PERFORMANCE OF HUNGARY: INSTITUTIONAL PATTERNS AND COMPETENCES