Finding Information Outside the Firewall

Gábor Pécsy

Senior Manager, Data Enrichment

June 25, 2014


MELT
WATER
formula of your health

Oslo, Shack 15 Summer 2001


20,000
clients
900
employees
The global leader
in media intelligence
90
countries
60
offices

Budapest Office

- Opened in 2009
- Research and Development only
  - Content acquisition
  - Data Enrichment
- 15 people, growing

Not-for-profit NGO fully funded and run by Meltwater

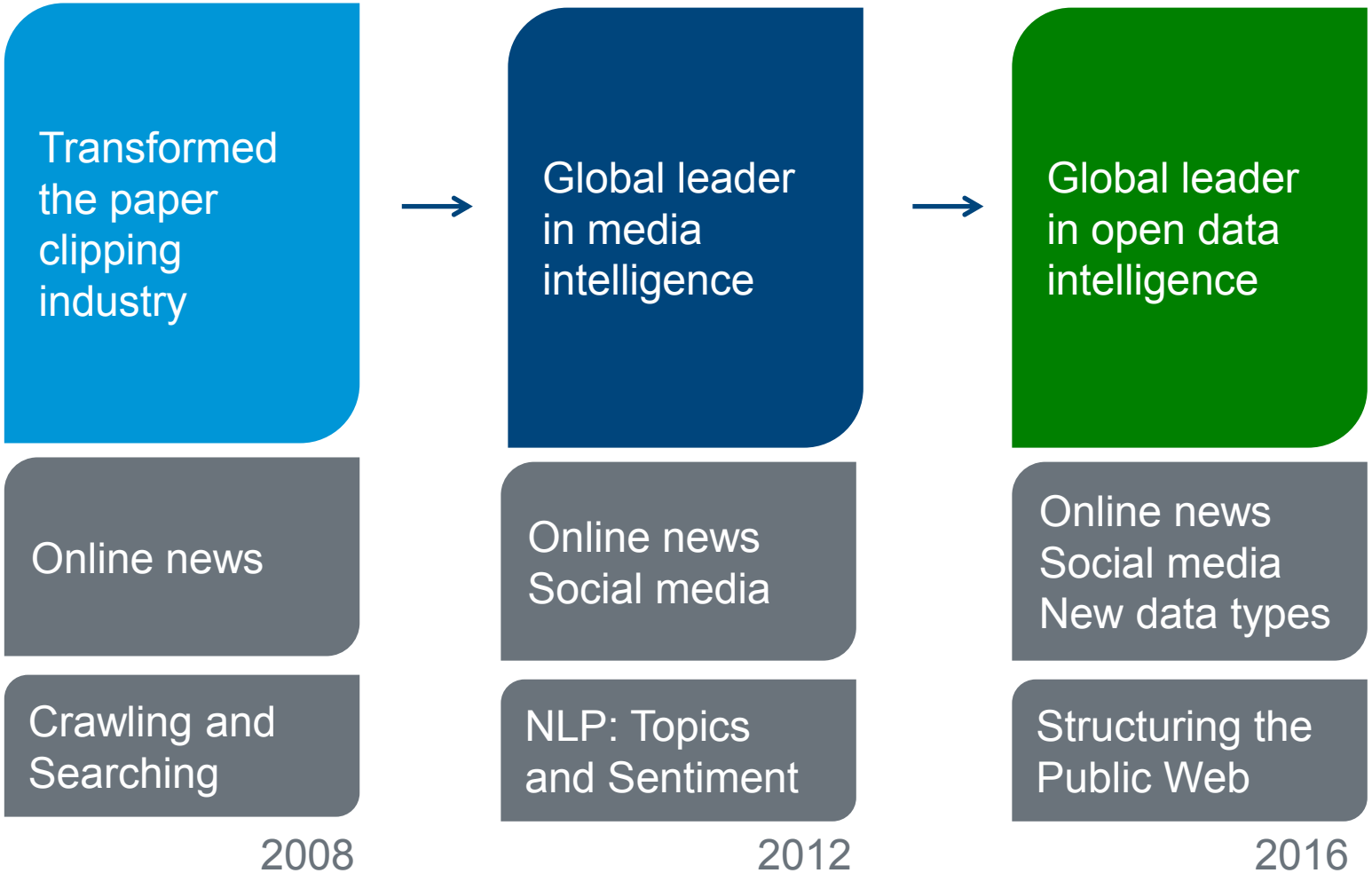
Back to
Finding Information Outside the Firewall

Our product vision

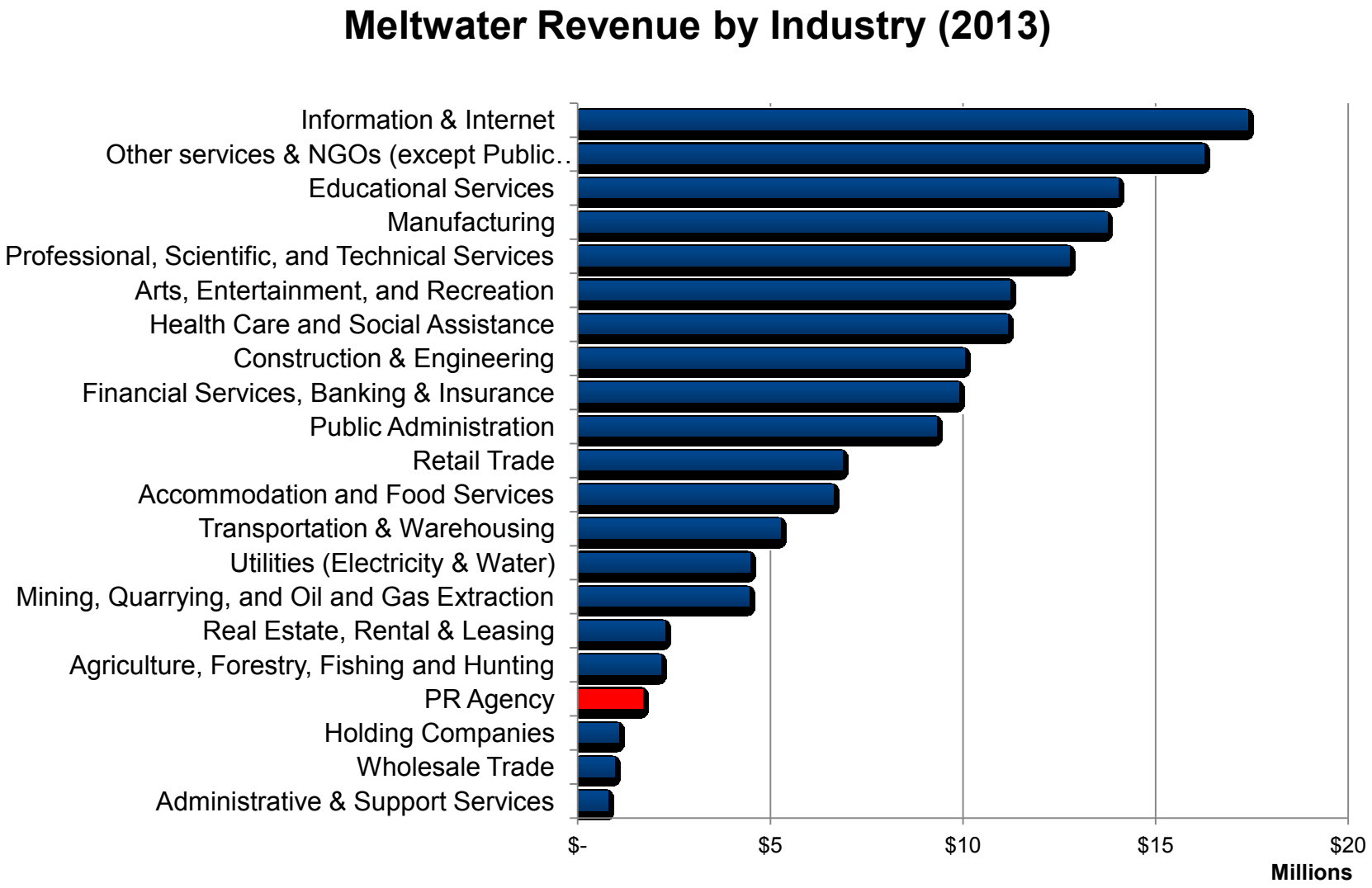
Morning coffee

Informed decisions

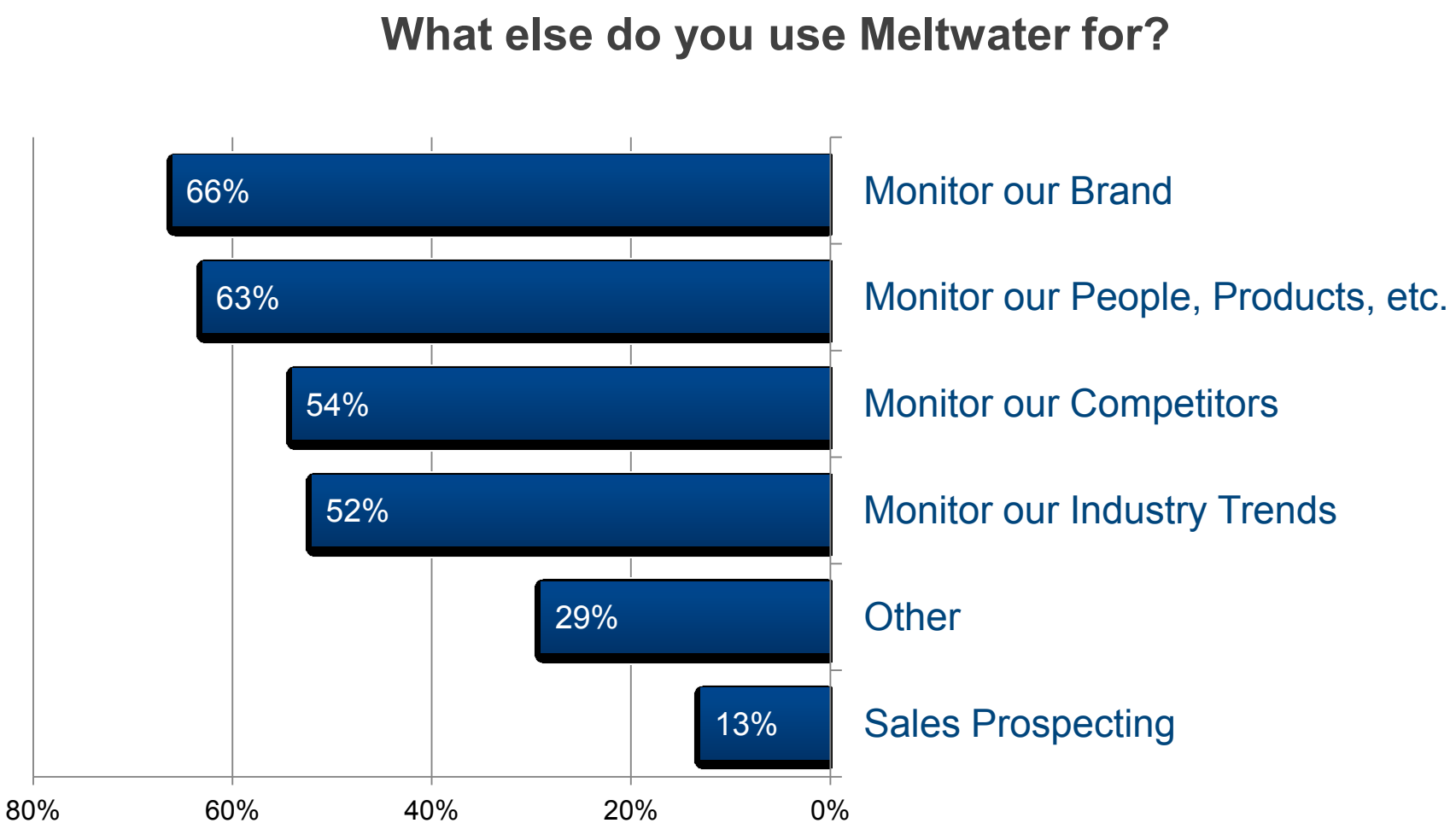
Our Journey


Transformed the paper clipping industry
Online news
Crawling and Searching
2008
Global leader in media intelligence
Online news
Social media
NLP: Topics and Sentiment
2012
Global leader in open data intelligence
Online news
Social media
New data types
Structuring the Public Web
2016

Already much more than PR (by industry)



Already much more than PR (by Use-Case)



Survey of Meltwater clients in Sweden (2011)

Prospecting

Industry: Security Services

Country: Sweden

We use Meltwater to find out about new instances of vandalism and break-ins. Often, the victim is in need of our services."

Product Management

Industry: FMCG

Country: United Kingdom

“Meltwater helps us determine how public perception of certain ingredient chemicals will influence adoption sales.

Performance Measurement

Industry: Television

Country: India

Meltwater is the best way for us to monitor the performance and popularity of our news anchors and programs.”

Competitive Intelligence

Industry: Pharmaceuticals

Country: England

SANOFI

“Surprisingly, we use Meltwater to be alerted of when certain patent will expire in target markets.

Risk Management

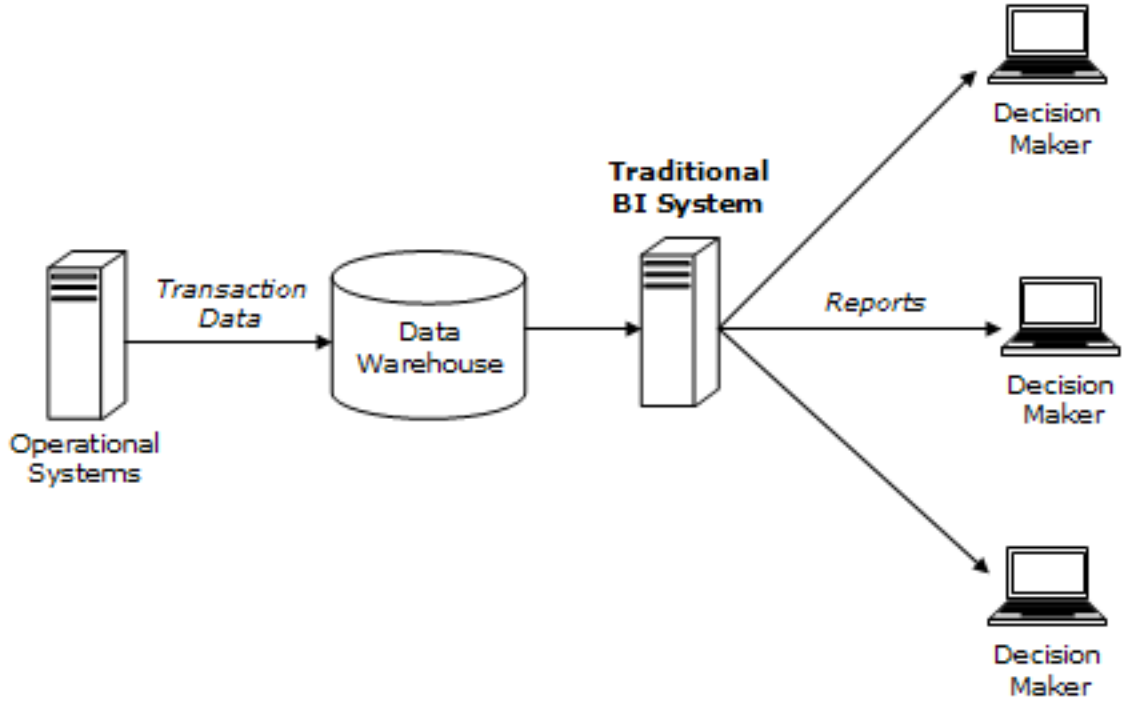
Industry: Precious Metals

Country: Global

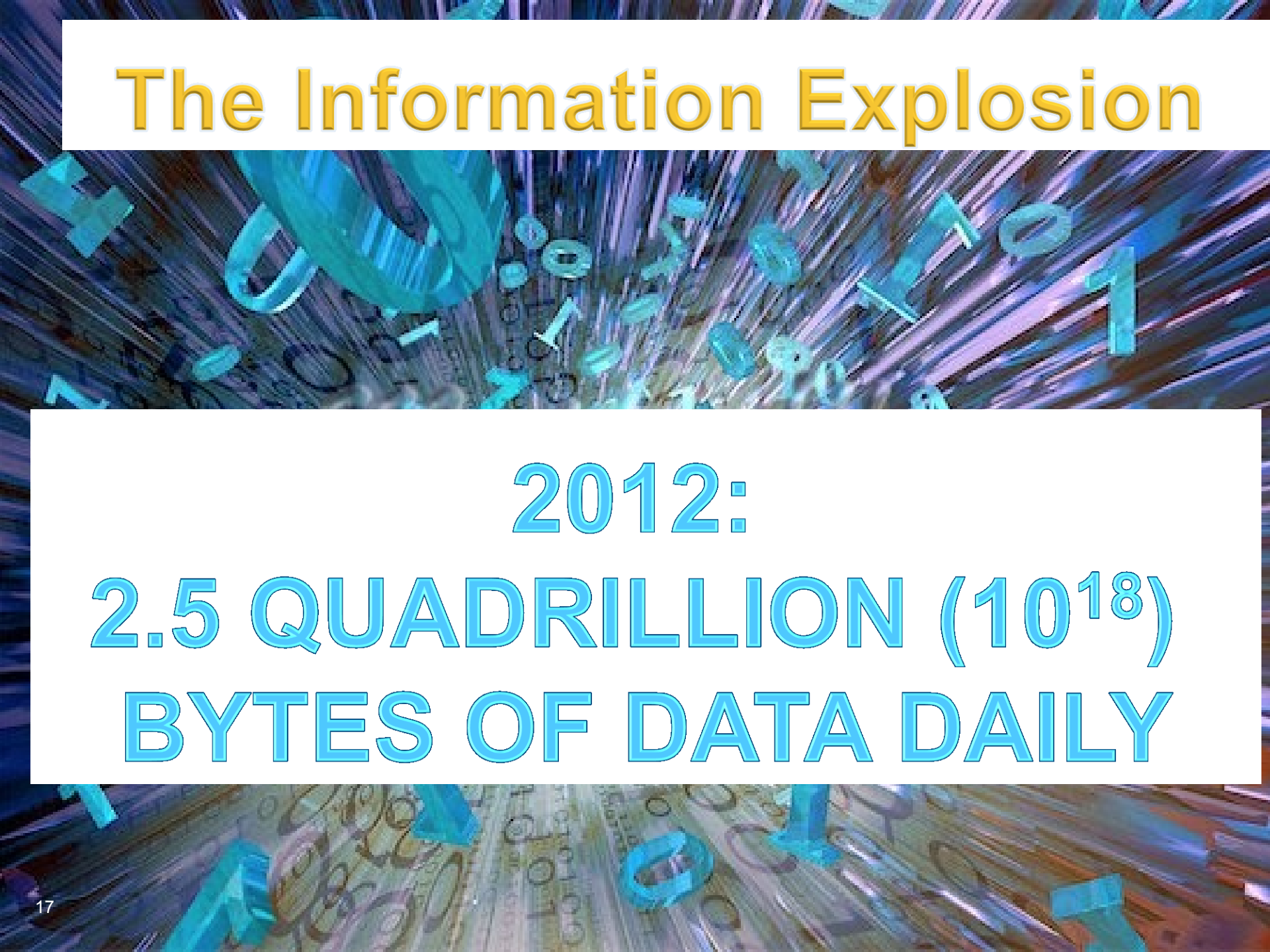
One of the best ways we've found to estimate and prevent infrastructure attacks is deep social listening using Meltwater."

Traditional BI

- Traditionally, BI focused on data owned by the company: **data within the firewall.**
- We have well-established tools for this



The Information Explosion



2012:

2.5 QUADRILLION (10^{18}) BYTES OF DATA DAILY


3.6 Billion people
Talking, tweeting, posting, blogging,
liking, ranting, retweeting, sharing opinions about
YOUR PRODUCTS, COMPANY, MARKET,
COMPETITORS
etc.
An invaluable source
of information

Companies need to learn to ride this information wave

Or it will sweep them away.

A New Industry Category

Classic Business Intelligence

30 years of Business Intelligence and workflow automation based on internal data...

ORACLE

hp

SAP

IBM

salesforce

Open Data Intelligence

…is being disrupted by an explosion of open data.

Real-time
Unfiltered
Benchmarked

Differences to Traditional BI

- Classical BI deals with data
  - Internal, easy to access
  - Small to medium volume
  - Structured data
  - Dense
  - Clean
  - Private
- Data on the Internet:
  - External, not so easy to get
  - Big Data
  - Usually unstructured or semi-structured
  - Sparse
  - Noisy, sometimes unreliable
  - Public, semi-public

Challenges – Data Collection

- Lots of open/free data – needs to be found, scraped
  - Requires special experience
  - Copyright issues – is it really free?
- Data aggregators – cost money
  - Can be targeted (more relevant content)
  - Often with enrichments (basic NLP, influence scores, ranking etc.)
  - Don’t cover the full content landscape

Challenges – Data Volume

- Information on Internet is more sparse
- Need to process large quantities of data to get enough information
- Large-scale computing – special competence, expensive hardware and software
  - Service-based models: SaaS, PaaS
- BIG Data technologies

Challenges – Unstructured Data

- Internal data: often structured, e.g. in RDBMS, data warehouse
- Data on internet:
  - Unstructured: information is available as text, images, audio, video
  - Semi-structured: textual data with some known inner structure (e.g. a patent document)
  - Rarely structured, usually costs money
- Extracting information from unstructured data: Natural Language Processing
  - Language specific
  - Social “dialects” – acronyms, emoticons, slang, typos etc.

Basic NLP Services

- Language detection

Basic NLP Services

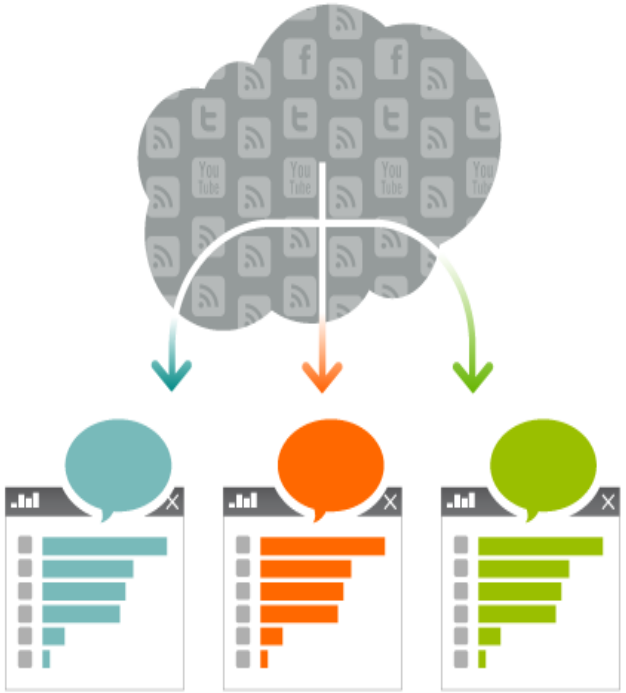
- Language detection
- Sentiment analysis

Basic NLP Services

- Language detection
- Sentiment analysis
- Key phrase extraction

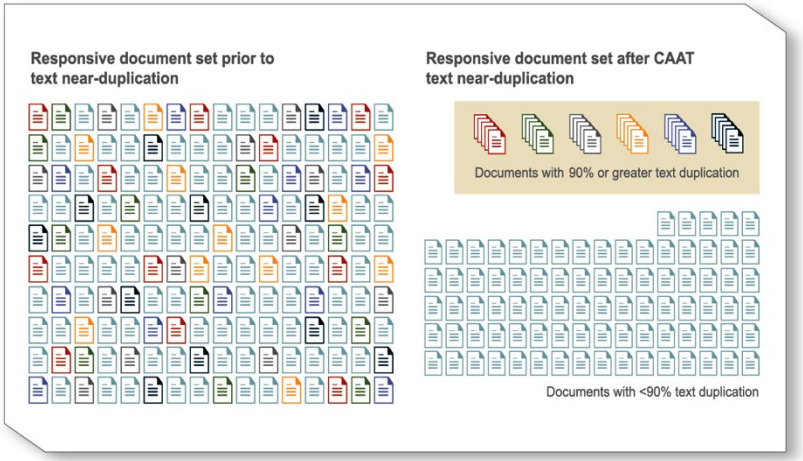
Basic NLP Services

- Language detection
- Sentiment analysis
- Key phrase extraction
- Content Categorization



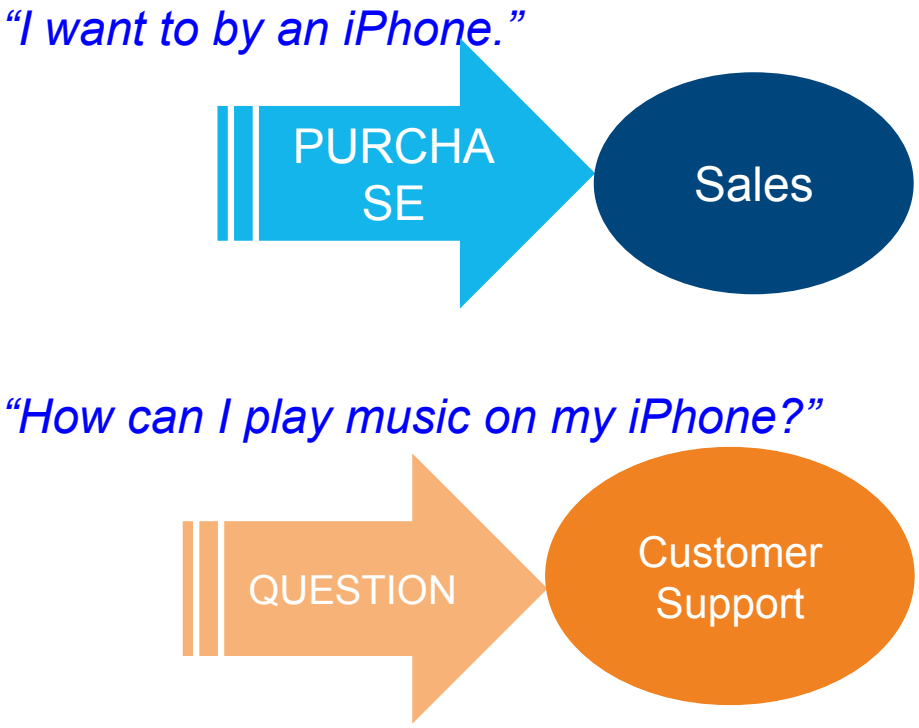
Basic NLP Services

- Language detection
- Sentiment analysis
- Key phrase extraction
- Content Categorization
- Near duplicate detection



NLP Services

- Language detection
- Sentiment analysis
- Key phrase extraction
- Content Categorization
- Near duplicate detection
- Intent detection

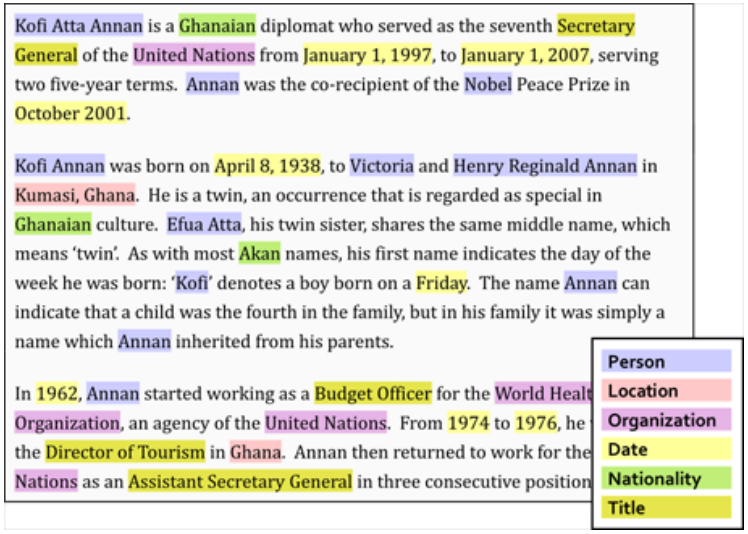


Using Intent Detection

- Direct customer interactions
  - Purchase Intent: sales offer
  - Question Intent: customer support
- Analytics
  - Correlation of purchase intents and sales figures in a region.
  - Comparisons to competitor products

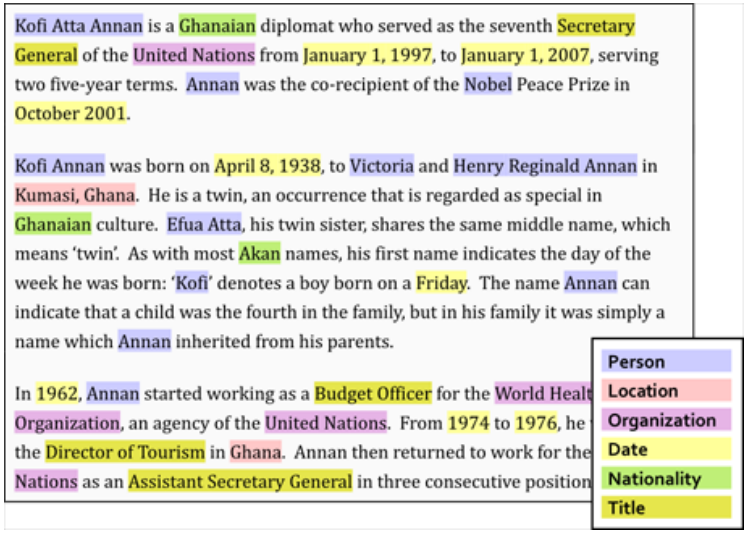
NLP Services

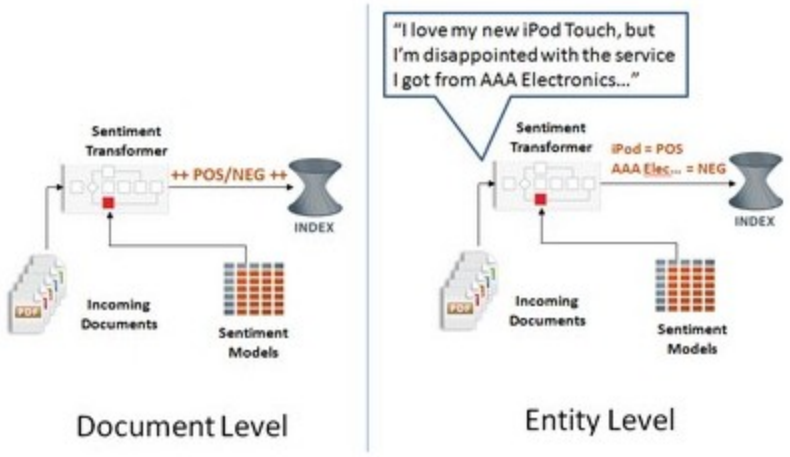
- Language detection
- Sentiment analysis
- Key phrase extraction
- Content Categorization
- Near duplicate detection
- Intent detection
- Named Entity Recognition



NLP Services

- Language detection
- Sentiment analysis
- Key phrase extraction
- Content Categorization
- Near duplicate detection
- Intent detection
- Named Entity Recognition
- Entity-level Sentiment



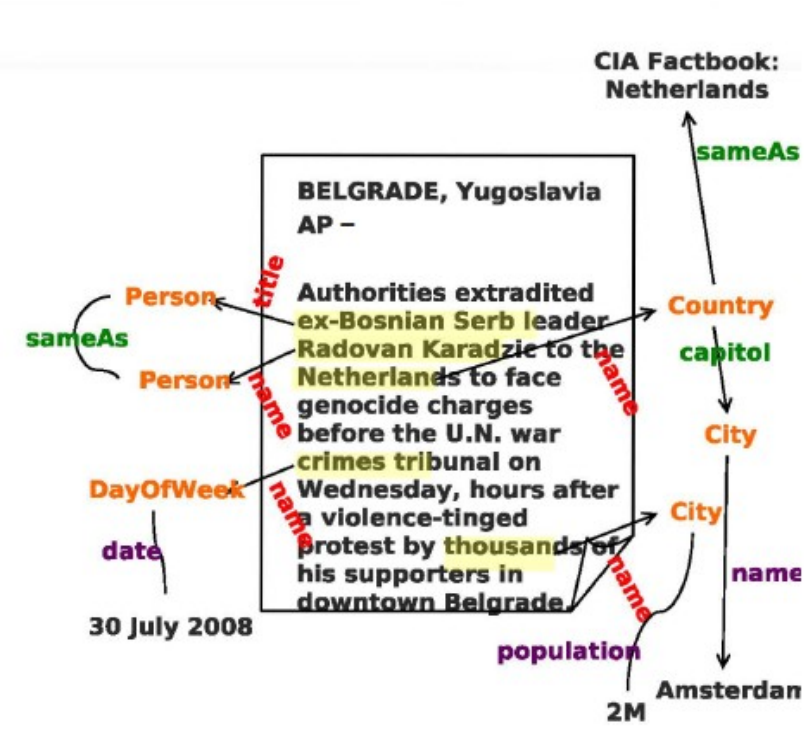


NLP Services

- Language detection
- Sentiment analysis
- Key phrase extraction
- Content Categorization
- Near duplicate detection
- Intent detection
- Named Entity Recognition
- Entity-level Sentiment
- Named Entity Disambiguation

NLP Services

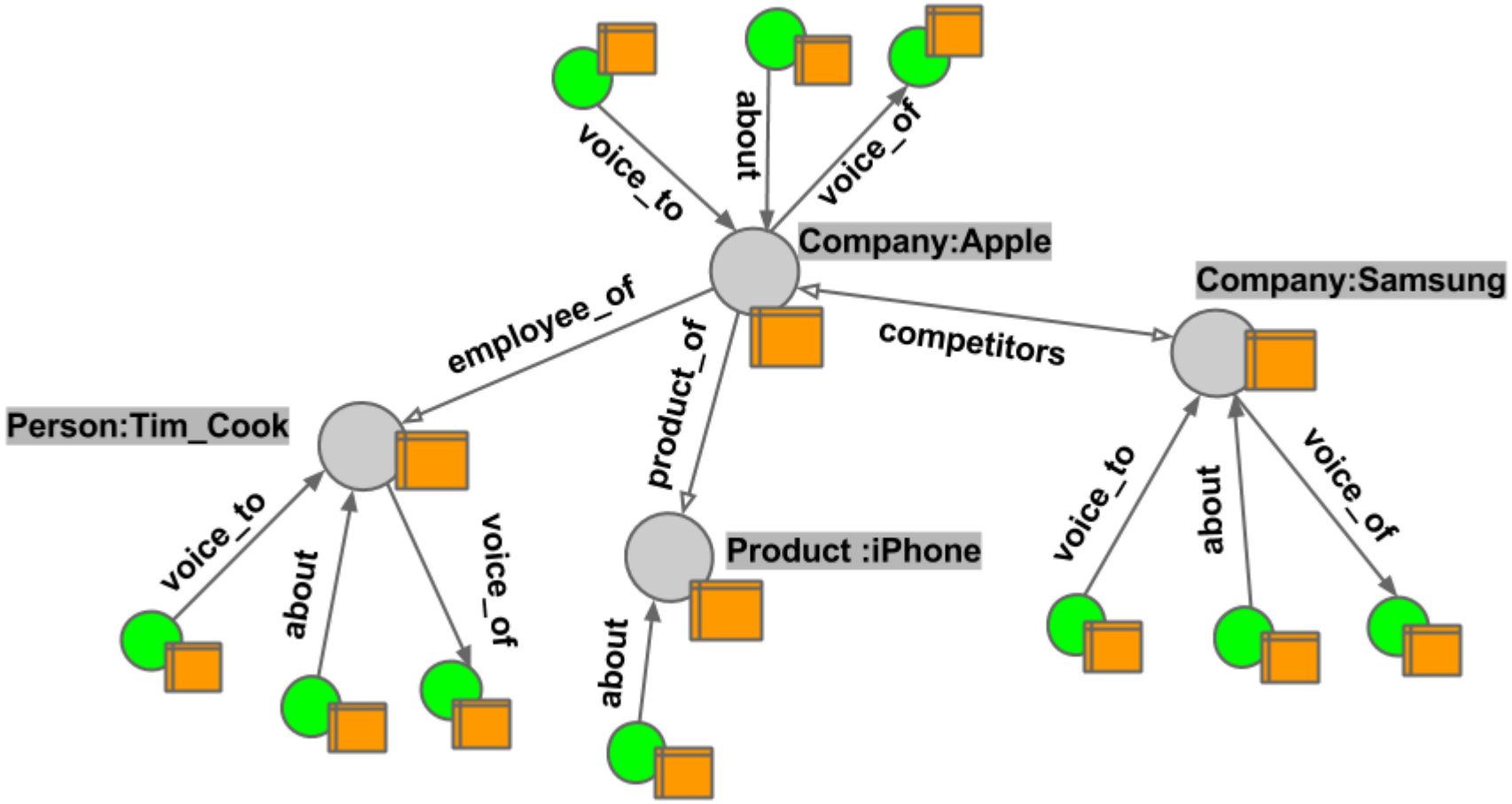
- Language detection
- Sentiment analysis
- Key phrase extraction
- Content Categorization
- Near duplicate detection
- Intent detection
- Named Entity Recognition
- Entity-level Sentiment
- Named Entity Disambiguation
- Entity Relationship Extraction



Graph search

facebook

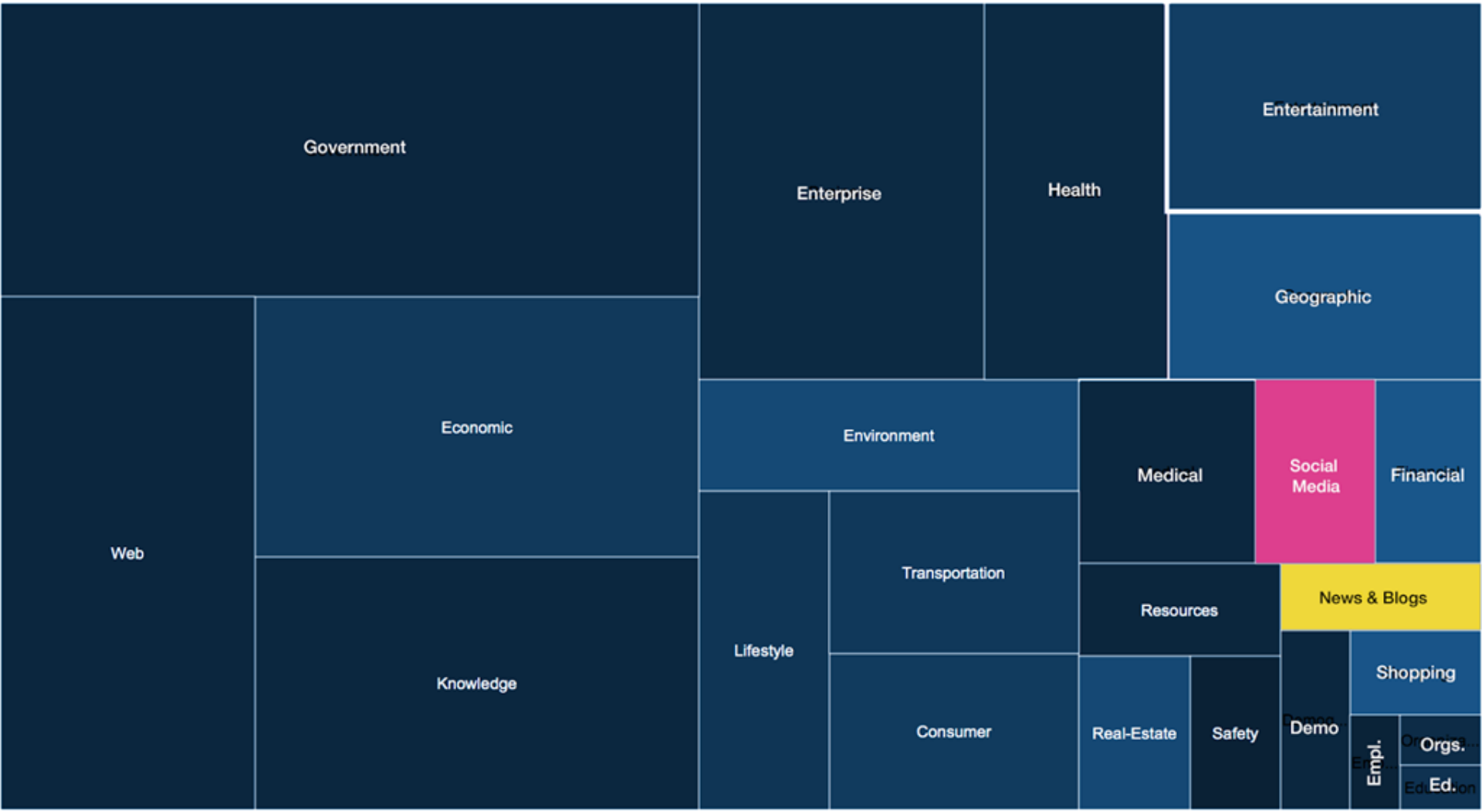
Object Graph is underlying structure for simplifying and processing complex queries



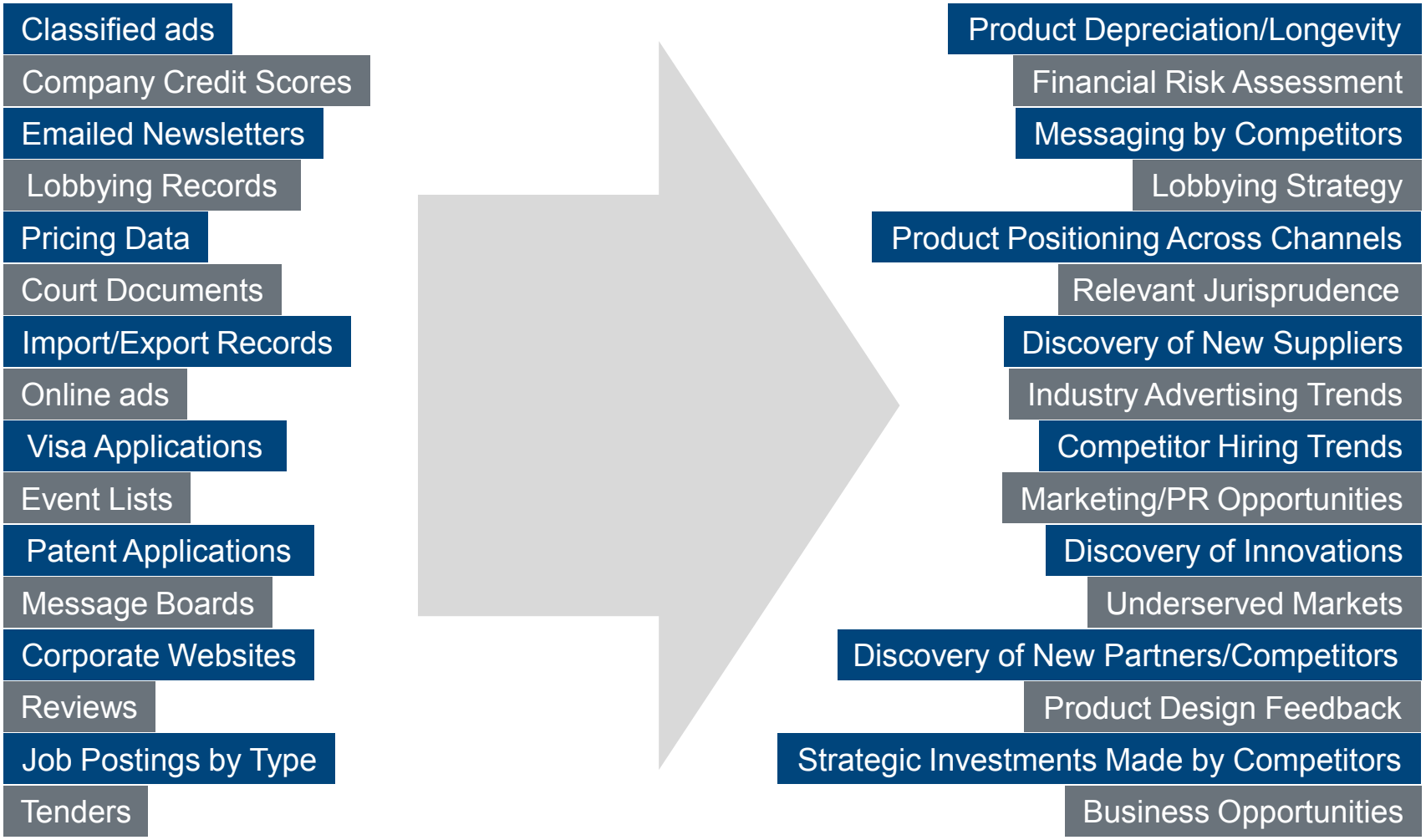
Graph Query Examples

- Give me a document feed from all the influencers who compare ios and android
- Graph and compare the sentiment of support complaints between Apple and LG
- Give me a word cloud of the topics most used by both Apple and Samsung in the past 30 days
- Give me a word cloud of keywords the Apple CEO is using that other smartphone leaders are not
- Show me the top 3 most rapidly changing topics associated with tablets being talked about by influencers in the mobile market

The Explosion is Driven by New Types of Data



Because it carries valuable insights

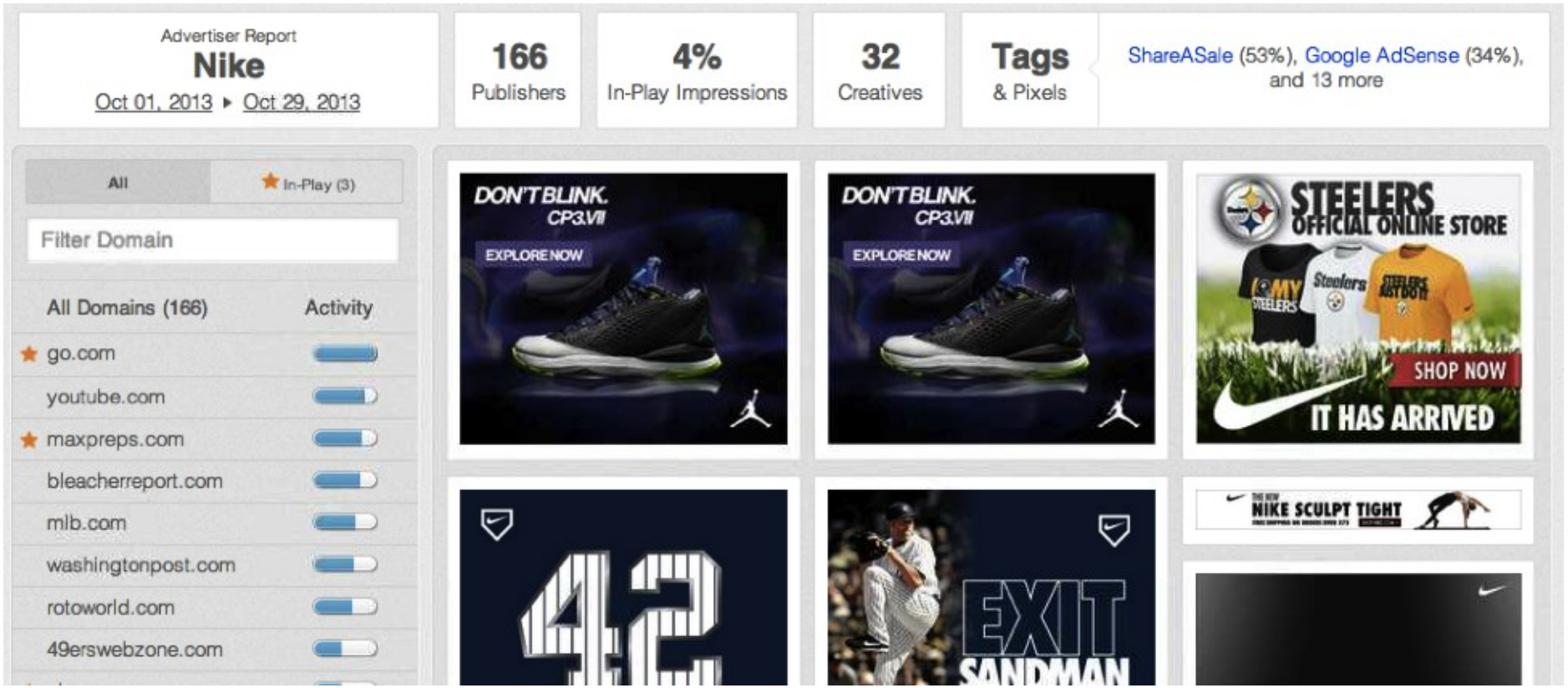


Example 1: Visa Applications

Wow, my competitor is aggressively hiring new engineers.
Should we be investing more in our product?

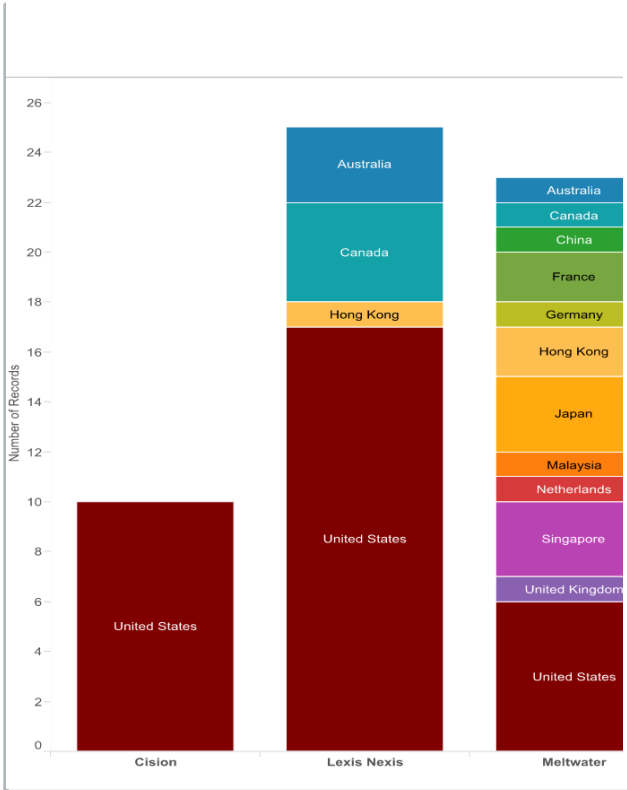
Example 2: Online ads

Hmm, Nike is getting a great ROI on Go.com.
Should we should start advertising there too?

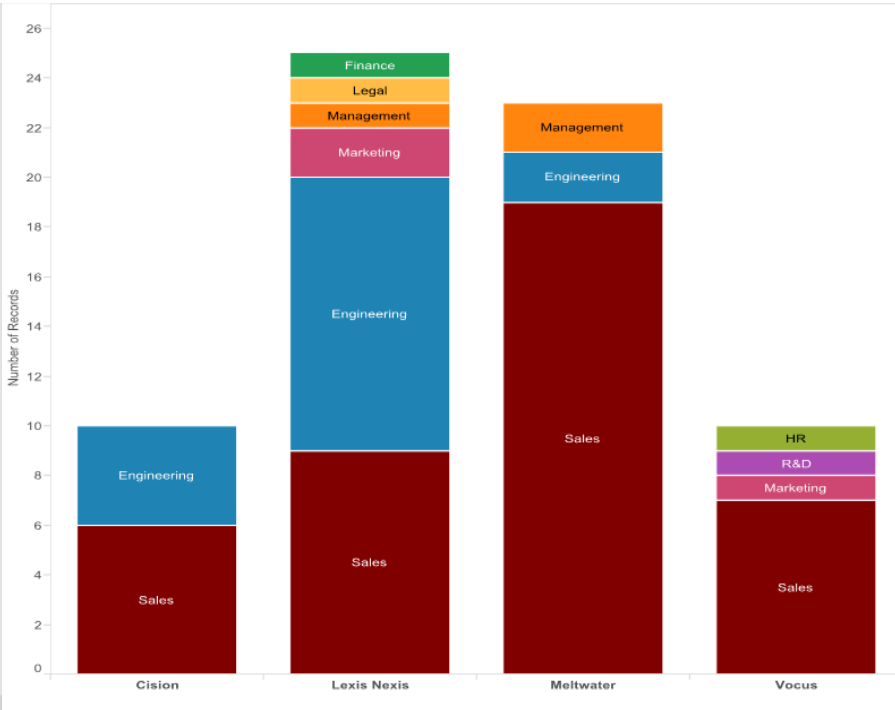


Example 3: Jobs

Woah, My biggest competitor just started outsourcing
Should we consider doing that?



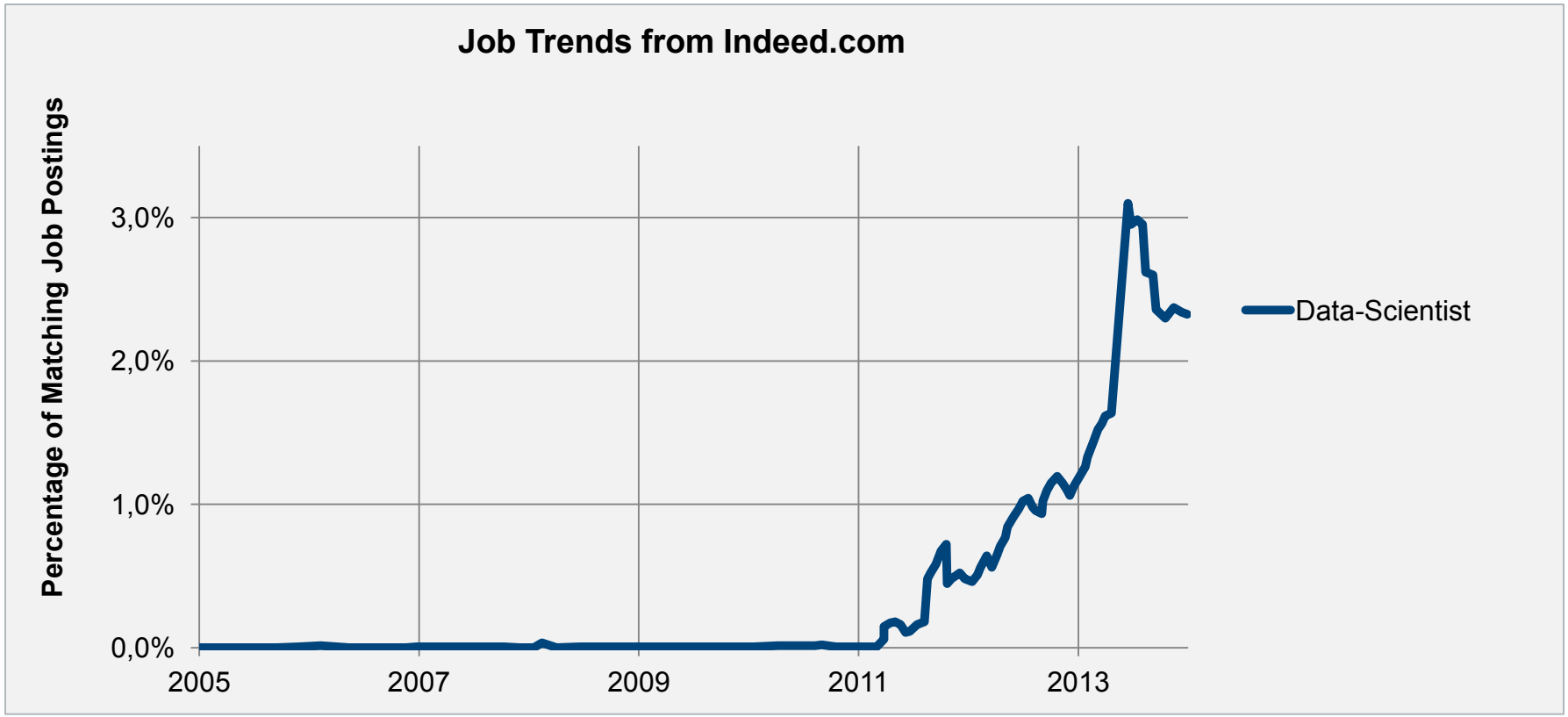
Hiring by Country



Hiring by Function

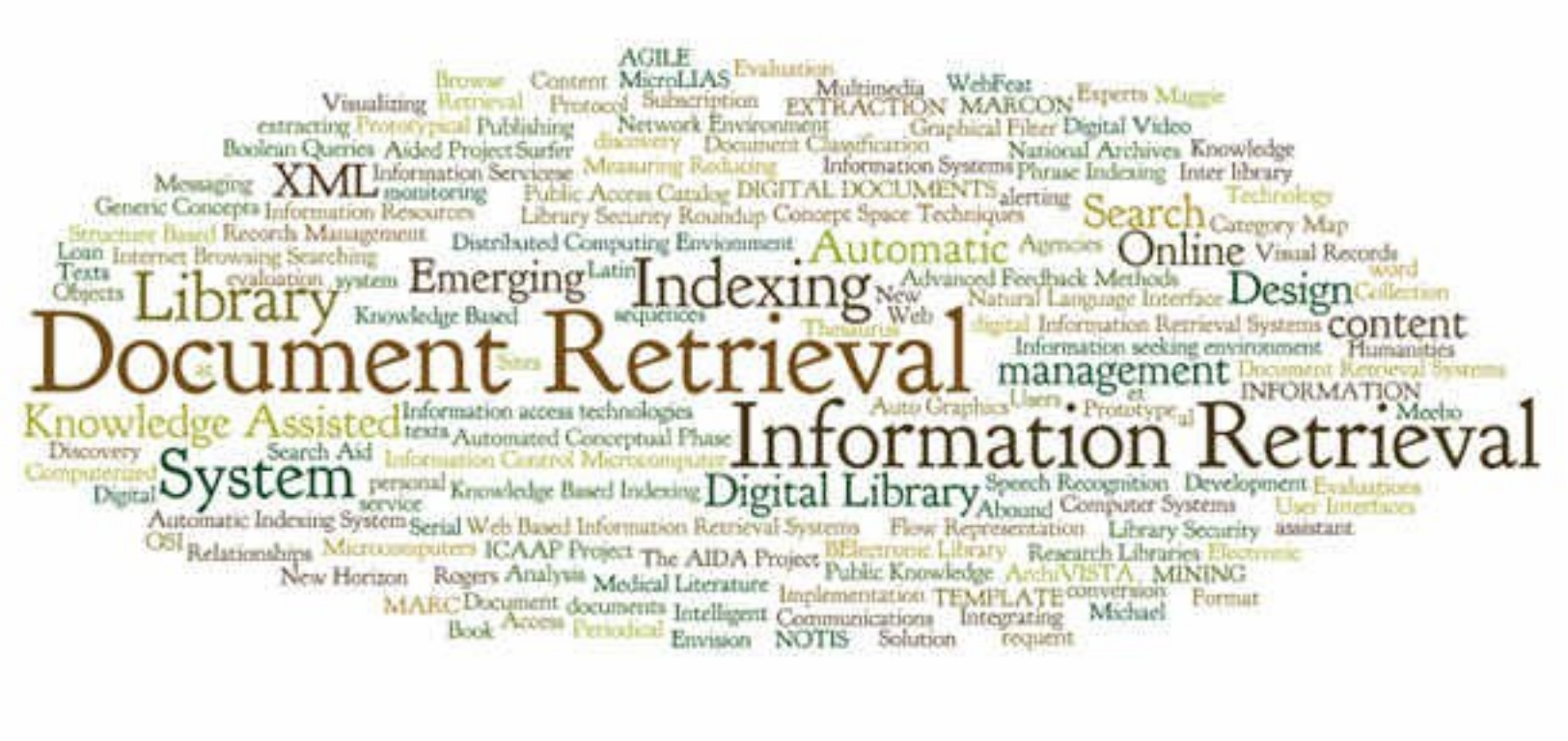
Example 3: Jobs (cont'd)

Kablam! I guess this Big Data trend is real…
- I wish there was a solution to help us harvest those insights!



Example 4: Patents

Our competitor is building a big portfolio of information extraction patents.
Do they plan to expand their business?



Data Analytics, Monitoring and Beyond

- Processing close to real-time
  - Alerting – quick reactions
- Data segmented by
  - Geographic location
  - Demographics (gender, age, level of education, social status etc.)
- Correlations – even across the firewall
  - Integration to traditional BI systems

Summary

- Information Explosion Creates a whole New Industry – Open Data Intelligence
- Classical BI Technologies are not applicable
- Natural Language Processing, Machine Learning, Big Data analytics help “connecting the dots”
- Companies of the 21st century cannot afford not using such information.

Thank you!

Gábor Pécsy (gabor.pecsy@meltwater.com)