Social Media Analytics

Making sense of Big Data

Henrik Hammer Eliassen
Predictive Analytics Specialist
IBM Business Analytics
E-mail: henrik.hammer@dk.ibm.com

http://www.youtube.com/user/HammerDemos/videos

Statistics play a big role in big data and data science…and it is hot

How would you present the meaning of all this?

More than **1 billion unique users visit Youtube each month** watching over 6 billion hours of video

More than 388 million people view more than **12.7 billion blog pages each month**

There are 500 million tweets daily – that's **5,700 per second**

50% of Facebook users check it daily – there are **more than 1 billion users world wide**

The Big Picture
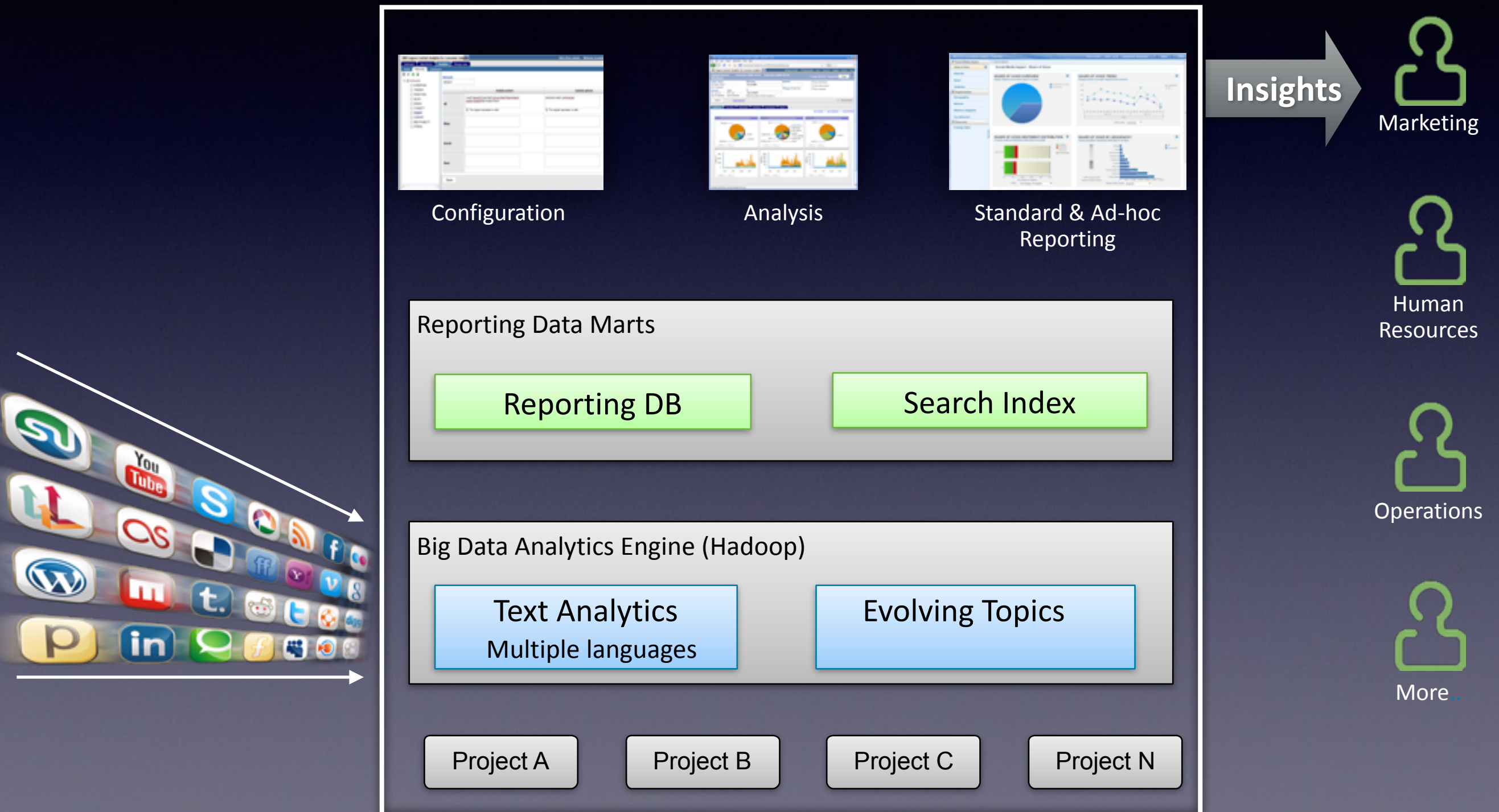
Configuration

Analysis

Standard & Ad-hoc Reporting

Reporting Data Marts

Reporting DB

Search Index

Big Data Analytics Engine (Hadoop)

Text Analytics
Multiple languages

Evolving Topics

Project A

Project B

Project C

Project N

Insights

Marketing

Human Resources

Operations

More

Aggregation
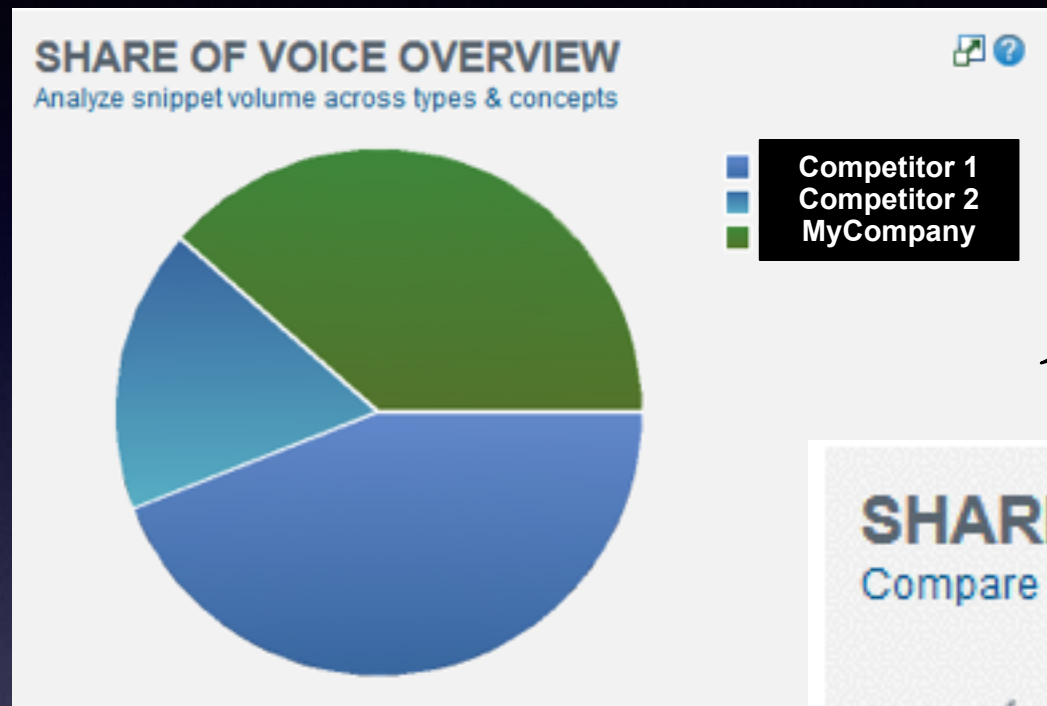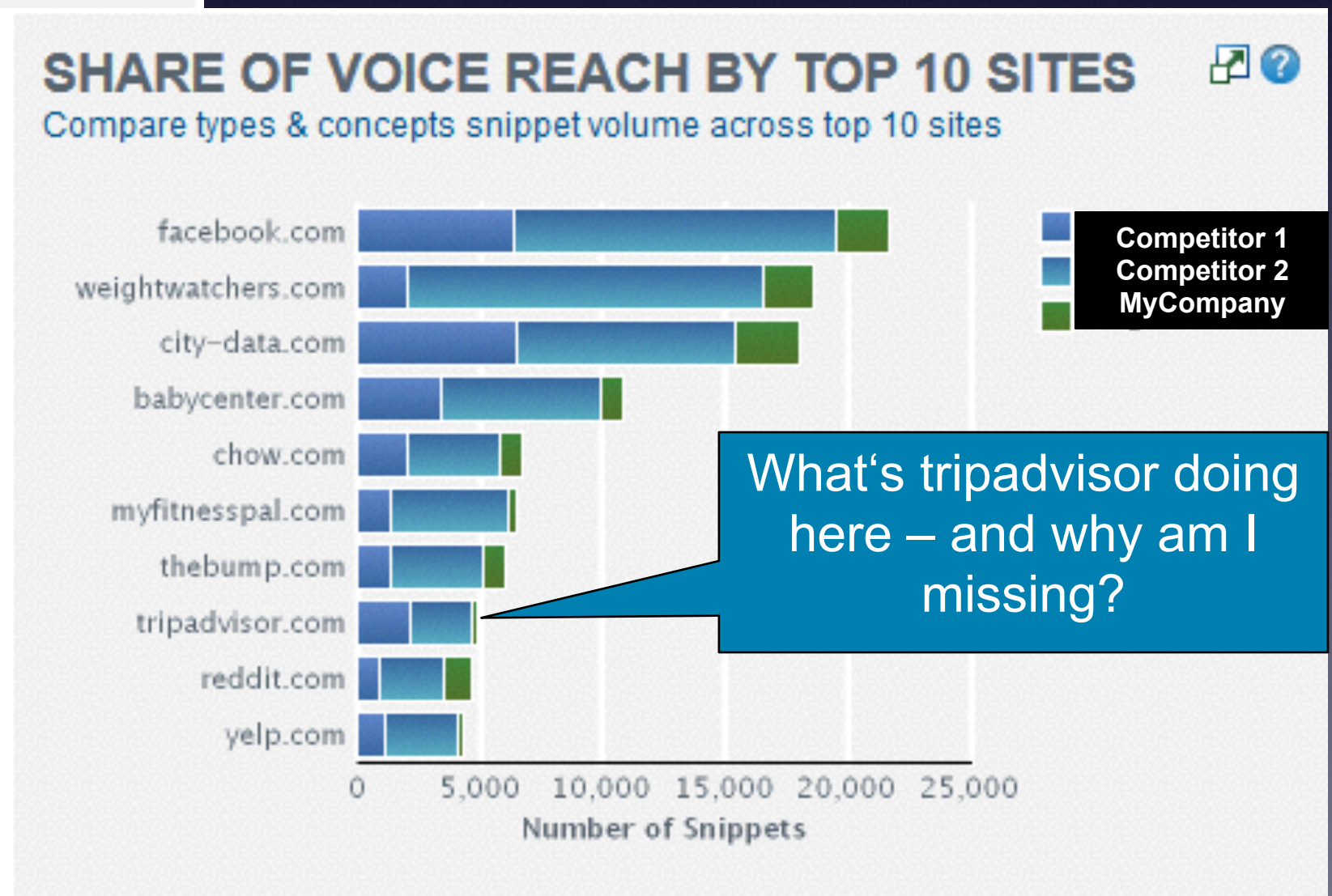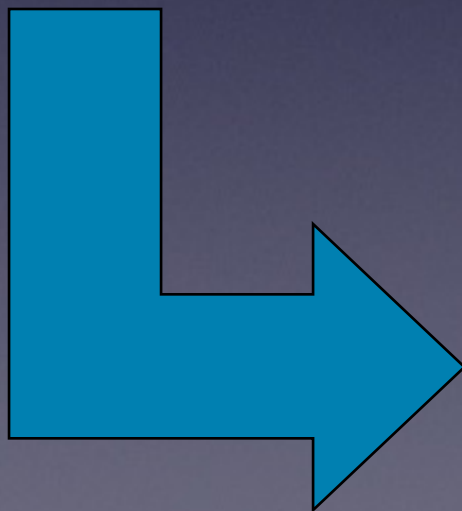
Behavior & Evidence

Author behavior

Users of a certain product or service
    What product features are relevant for them?

Recommenders
    E.g., authors mentioning „you should use X"

Detractors
    e.g. authors mentioning „stay away from X"

Prospective users
    Potential sales leads for 1:1 engagement
    Identify sites where prospective users congregate

HOTWORDS BY AUTHOR BEHAVIOR

Analyze author distribution by hotwords for each behavior attribute

User: S3
User: Lumia 920
User: Lumia 820

0%   20%   40%   60%   80%   100%
Percentage of Authors

Legend:
- Screen
- Music
- Memory
- Video
- Maps
- Networking
- Operating System
- Camera
- Battery
- Apps
- GPS
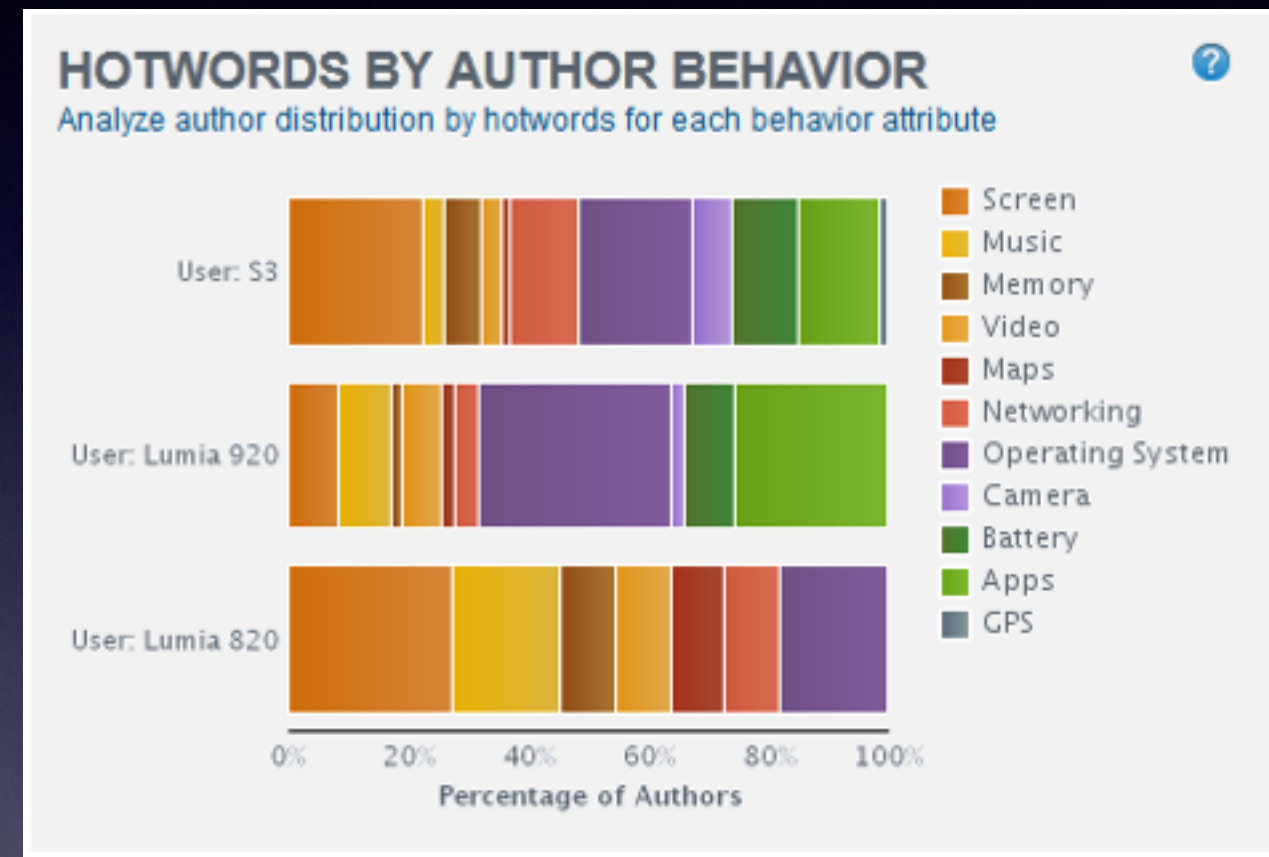
Evidence

Author Name	Site URL	Number Of Snippets	Gender	Is Married	Has Children	Author Location	Behavior: Concept	Evidence Text
Blue_Tooth	http://www.ign.com	2	Unknown	Unknown	Unknown	Canada EDMONTON	Prospective User: S3	getting an S3
Ju5tin	http://www.golivewire.com	2	Male	Unknown	Unknown	Canada SASKATCHEWAN	Prospective User: S3	will be getting an S3

Sentiment Analysis
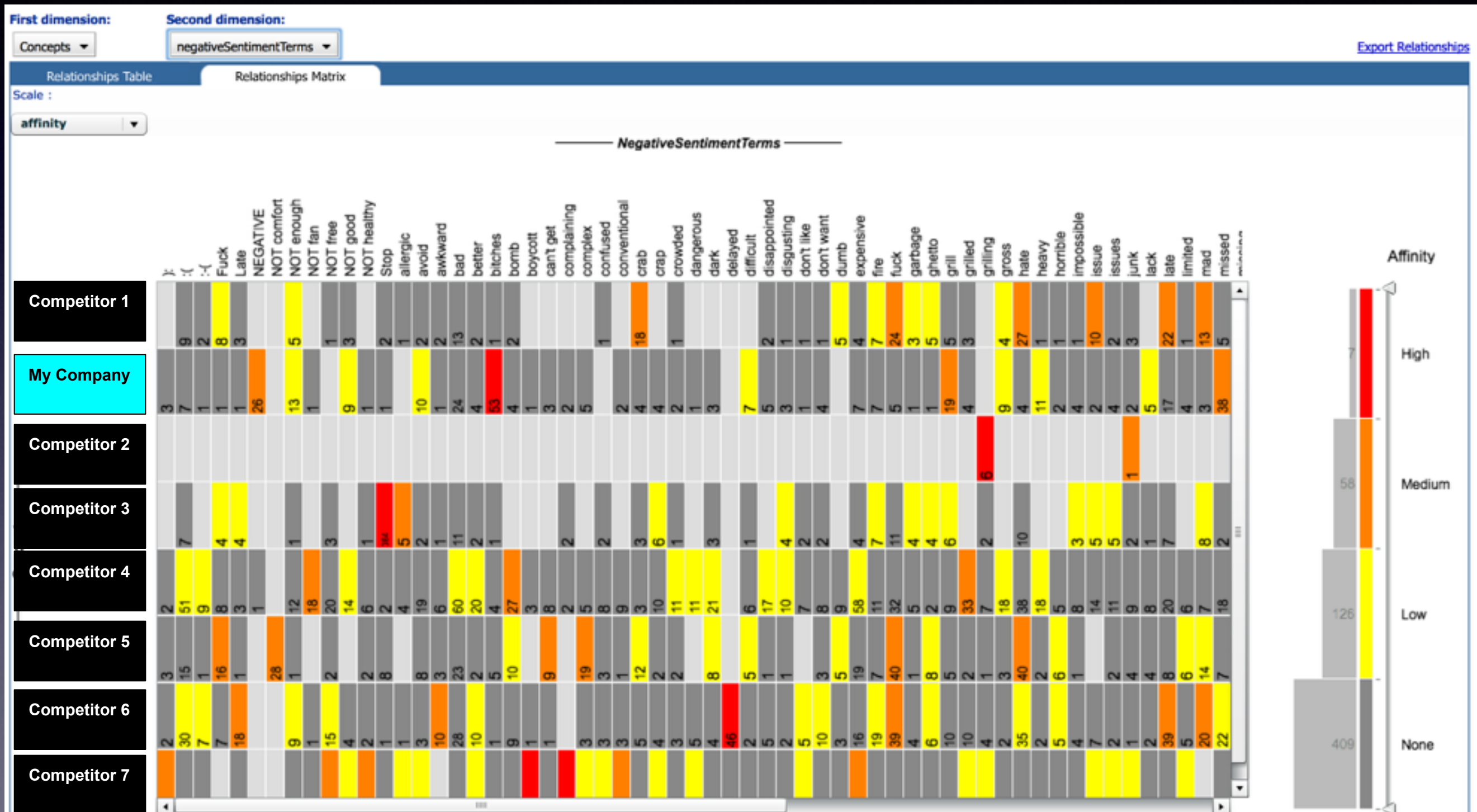
Colors show "correlation" between a concept and a topic

Sentiment must be actionable

What is the sentiment of the following tweet?

"Bank X sucks, switched to Y now, and they're great!"

We think this tweet does not have one **single** sentiment value. It's a positive sentiment for Y, and a negative one for X

To be **actionable**, sentiment needs to be determined for a **particular concept** within a document. Document-level sentiment is not enough

Must be concept related

What's the sentiment of this facebook post?

Lmao everyone I ask says my galaxy s5 is way better than my brothers iphone 5s

Like · Comment · Share

Neutral or... positive with a 0.5 probability ?

We think it's...

Positive for the Galaxy S5

... Lmao everyone I ask says my galaxy s5 is way better than my brothers iphone 5s ...

No further content   Language: English   Author Nathaniel R_____ (_____)
Sources: boards (Facebook / Nathaniel ._____)

Negative for the iPhone

... Lmao everyone I ask says my galaxy s5 is way better than my brothers iphone 5s ...

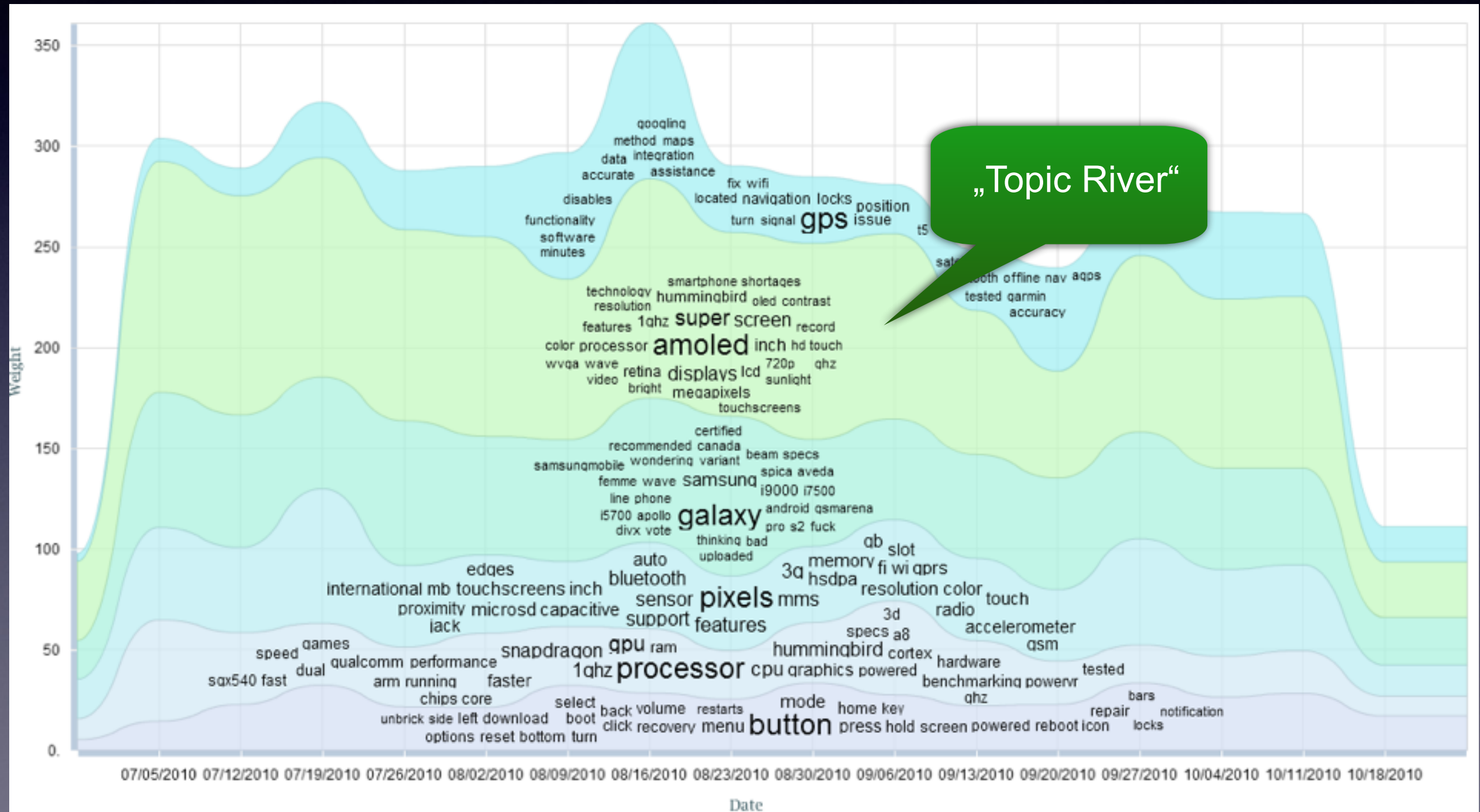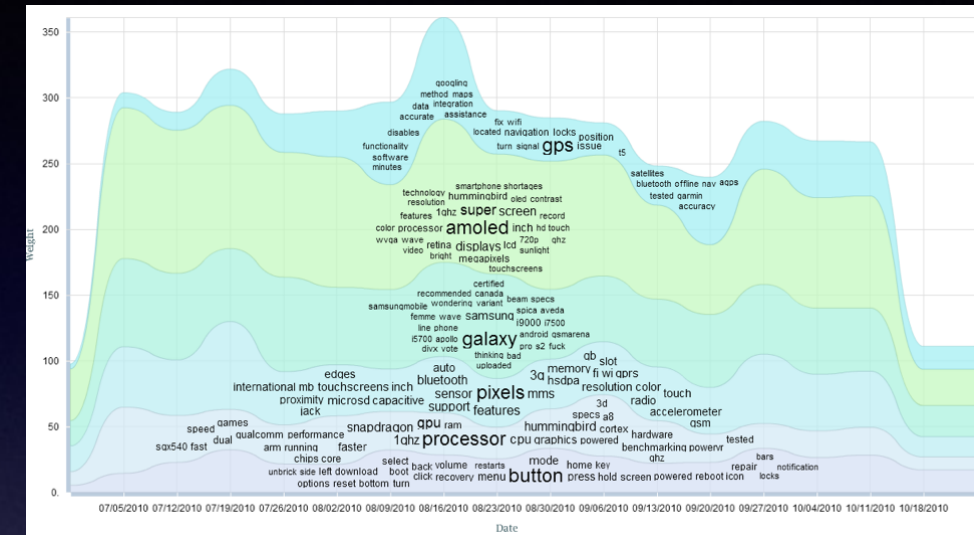No further content   Language: English   Author Nathaniel R _____)
Sources: boards (Facebook / Nathaniel ._____)

Topic Rivers

## Evolving Topics around a particular smartphone

How is it done?

- Goal: find „lists of keywords" (=topics) that allow to „reconstruct" a social media post through a combination of topics

- Approach: Non-Negative Matrix Factorization

- Advantage over document clustering: focus is on getting representative topic keywords, which helps the user to understand what the documents „are about", not „perfect" document clusters

- Factorization problem:
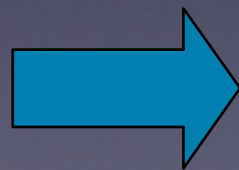
term

Document-Term-Matrix
X

K topics

Document-Topic-Matrix
W

term

Topic-Term-Matrix
H

Topic Rivers: Example Use

EVOLVING TOPICS TREND
Analyze topics weight over time

Legend:
- recalls, butter, peanut, ▓▓▓ salmonella
- companies, marketing, custom, business, product
- food, marketing, emporium, stores, eating
- grocery, stores, shop▓▓▓▓▓▓▓s
- cooking, cup, oil, eating, salt
- opened, stores, ▓▓▓▓▓▓▓▓▓
- develop, square, build, projects, retailer
- foods, stores, marketing, recalls, listeria
- feel, life, things, people, time
- label, genetically, gmo, modified, product
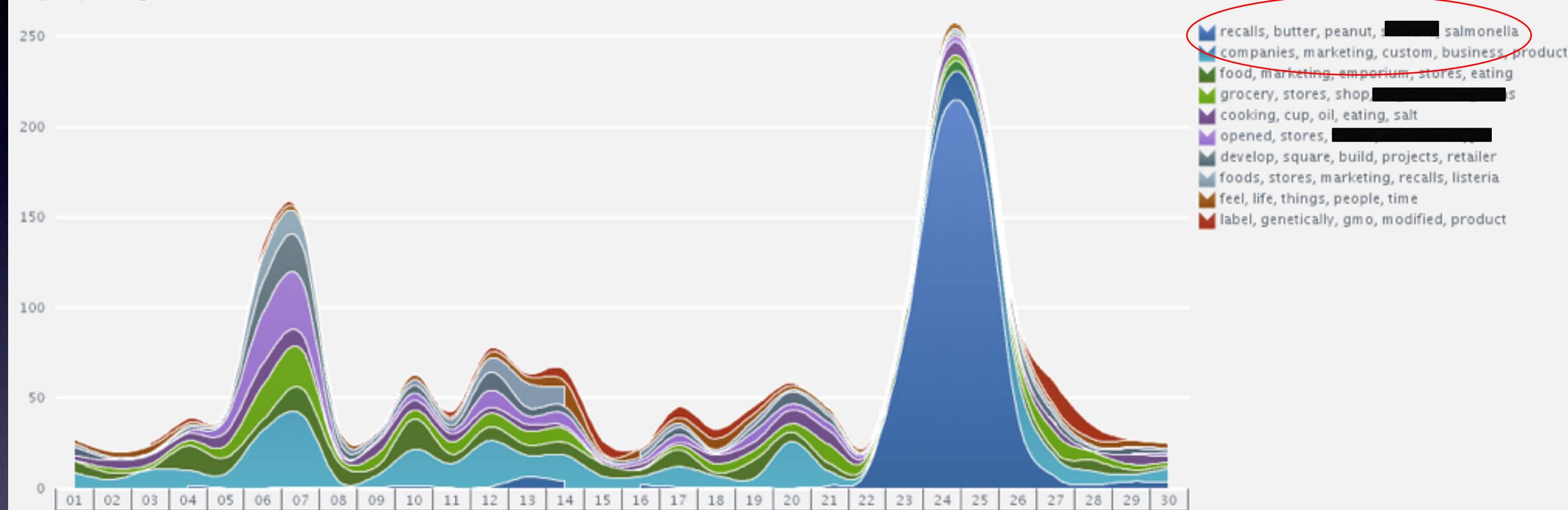
AuthorName: ▓▓▓▓▓▓▓▓▓▓
Snippet: Peanut butter recall expands beyond ▓▓▓▓▓▓▓ By ▓▓▓▓▓▓▓ Associated Press WASHINGTON - A New Mexico-based company is recalling 76 types of peanut butter and almond butter after one of its products was linked to a salmonella outbreak ▓▓▓▓▓▓▓▓▓▓▓▓ Inc. recalled the products under multiple brand names after the Food and Drug Administration and the federal Centers for Disease Control and Prevention linked 29 salmonella illnesses in 18 states ▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓ and manufactures and packages the ▓▓▓▓▓▓ product.

Competitor 1

Action → Check own supply chain to pro-actively avoid this problem
Prepare statement to clarify that your brand is not affected

# Listen to Social Media to design better products

Ford considers social media to be a boundary-less group and their largest focus group. Here they listen to what is being said about Ford and competitive products.

This information is then used in the development of future models (fx. The 2016 Ford Explorer) focusing on what people like, don't like, need to be changed, need to be added etc.

IBM Social Media Analytics used as the solution
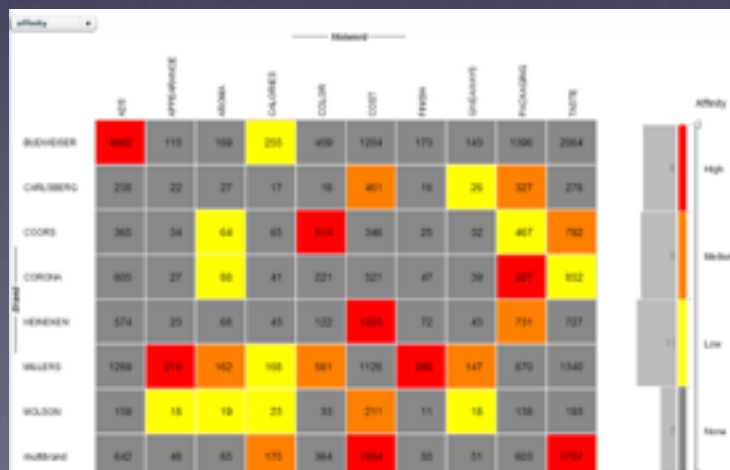
Want to now more ?
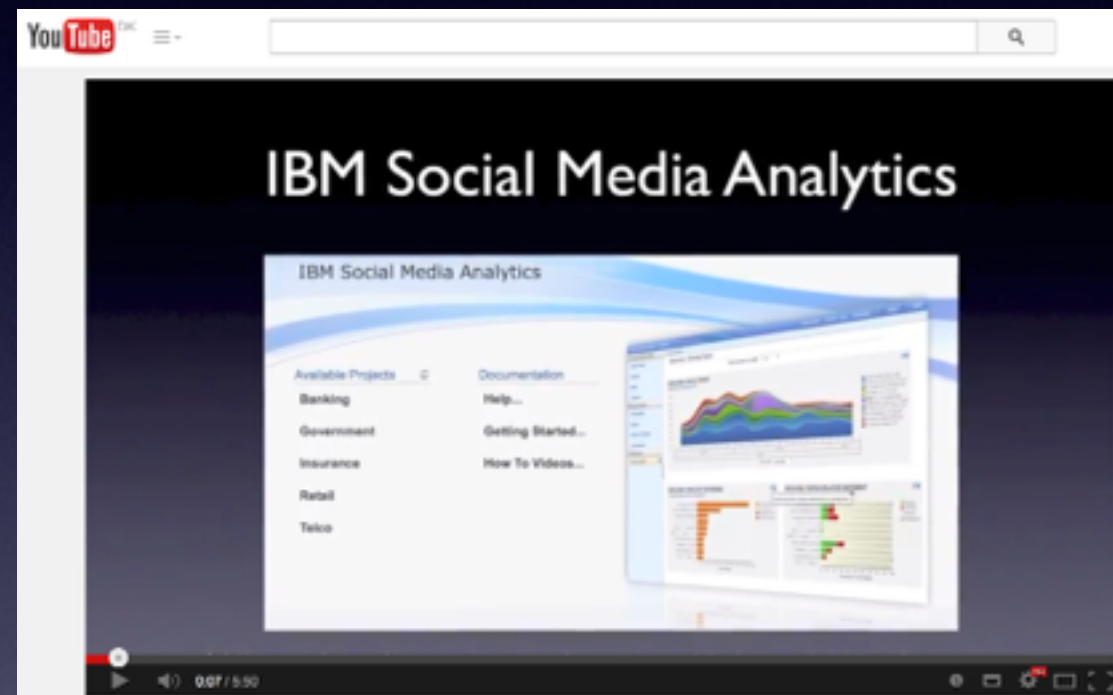
https://www.youtube.com/watch?v=vAcqZ_THHwI

Henrik Hammer Eliassen
Predictive Analytics Specialist
IBM Business Analytics
E-mail: henrik.hammer@dk.ibm.com

http://www.youtube.com/user/HammerDemos/videos