Social Business Intelligence

- OLAP Applied to User Generated Contents –

Matteo Golfarelli

University of Bologna - Italy

Data Conference – Vienna, 30 August 2014

Summary

- Introduction to Social BI
- An architecture for SBI
- Data Modeling in SBI
  - ✓ MetaStar
- Our prototype
- Conclusions

Introduction to SBI

User Generated Contents

- **User-generated content (UGC)** refers to a variety of media content available in a range of modern communications technologies. UGC is often produced through open collaboration [wikipedia].
- Although, a UGC can be an audio or a video in the following we will refer to textual ones, that can be classified according to the media that generated it:
  - ✓ Internet forums
  - ✓ Blogs
  - ✓ Wikis
  - ✓ Social networks
  - ✓ Customer review sites

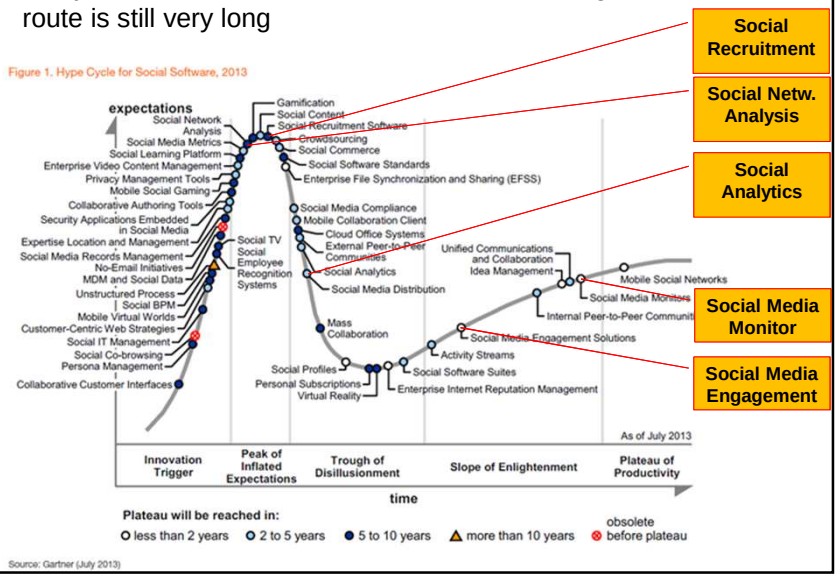
- UGCs do not carry only news, preferences, opinions, news, etc. but they are also rich of meta-data such as:
  - ✓ Geo-localization
  - ✓ Information about the authors
  - ✓ Information about the media

UGC Exploitation

- UGC is raising an increasing interest from decision makers because it can give them a fresh and timely perception of the market mood (inbound) and can be used to deliver important messages to potential customers (outbound)
  - ✓ Social events are perceived by traditional information systems **when they impact** on the company processes (e.g. sales reduction). Social events are perceived by SBI systems **when they start happening**, that can be several days/weeks/months before their effects impact the company information system
- Exploiting such opportunities requires the companies to adapt their business model to the new market that implies
  - ✓ new ways to communicate
  - ✓ new competitors
  - ✓ new consumers
- Such model is often called **Social Business Model** (SBM) since business processes are influenced by the internet user behaviors that can be captured and influenced through the analysis and production UGCs.

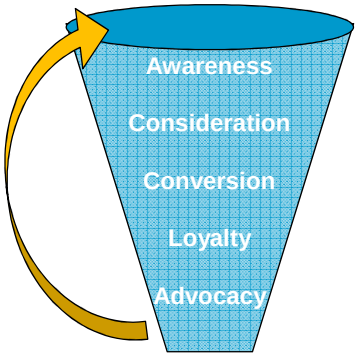
UGC Exploitation

- Many Social Business related software are sailing the wave but the route is still very long



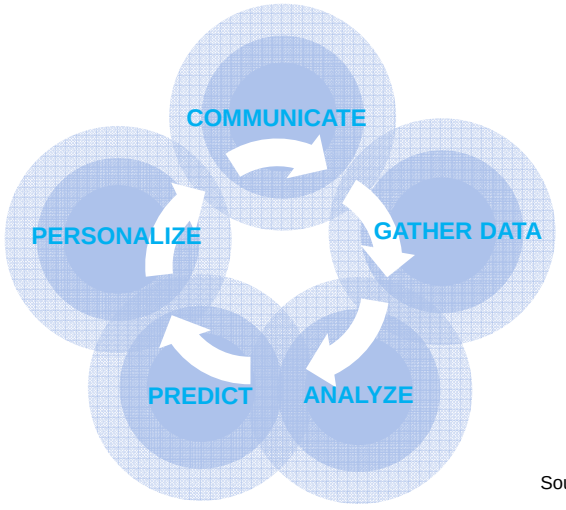
UGC Exploitation: Digital Marketing

- The division that is more affected by UGC is often the marketing one
  - ✓ Digital marketing divisions are growing their budgets and their relevance within the company strategy
- Marketing in the era of social network is based on **Word of Mouth** and is aimed at making the customer the main actor in communicating the value of a product or service



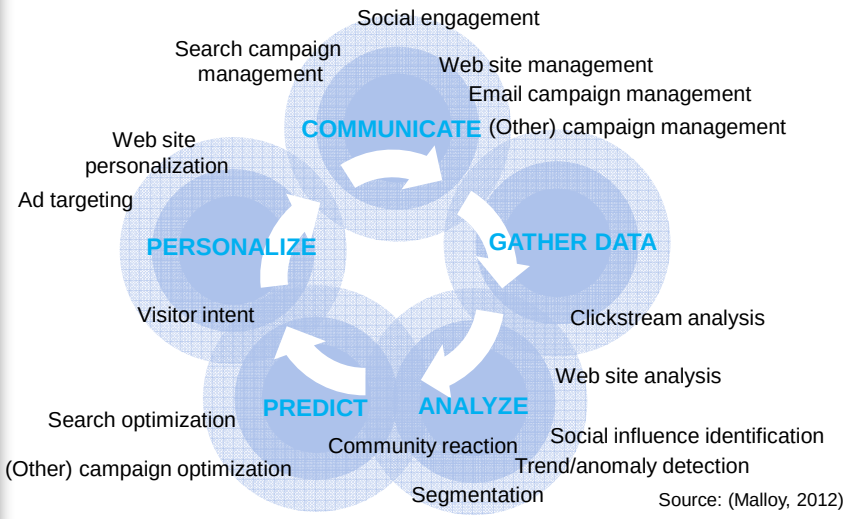
Source: (Malloy, 2012)

Digital Marketing: activities breakdown



Source: (Malloy, 2012)

Digital Marketing: Activities Breakdown



Source: (Malloy, 2012)

Digital Marketing: Technology Building Blocks

Source: (Malloy, 2012)

Social-Media Monitoring tools

- Many commercial tools and platforms are available for analyzing the UGC
  - ✓ Brandwatch
  - ✓ Tracx
  - ✓ Clarabridge
- They typically rely on a large but fix set of glitzy dashboards that analyze the data from set of points of view…
  - ✓ Topic usage
  - ✓ Topic correlation
  - ✓ Brand reputation
- … and using some ad-hoc KPIs
  - ✓ Topic counting (e.g. Top topic, Trending topic)
  - ✓ Sentiment and polarization
- Rely on a cloud architecture and are oriented to business users with limited capabilities in managing data

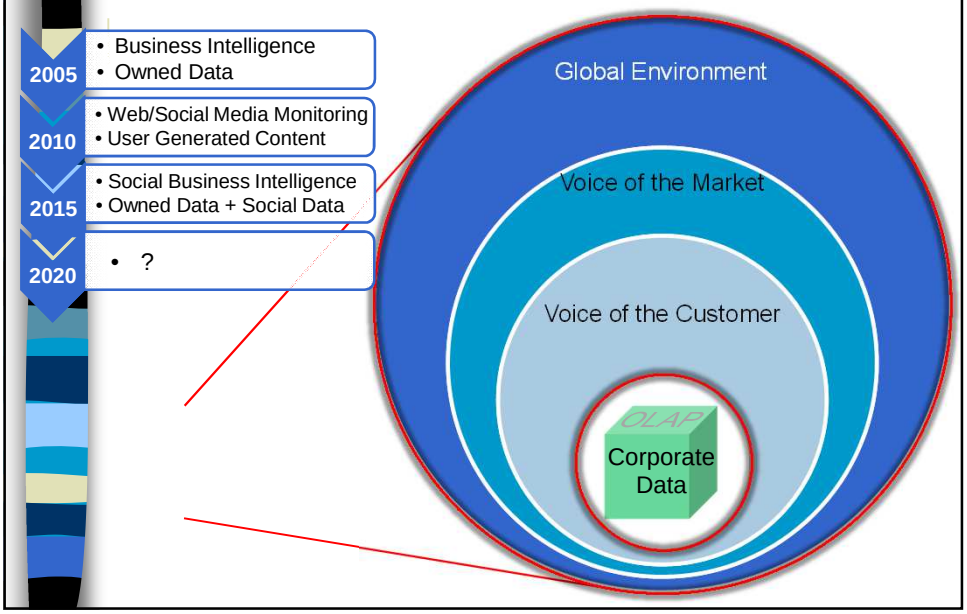
Social-Media Monitoring tools

- Social-Media Monitoring tools are often offered as-a-service
  - ✓ Are project-oriented (typically with a narrow time-horizon)
  - ✓ Lack in providing a sufficient verticalization/personalization of the system in term of dictionaries, rules, etc..
  - ✓ Provide limited capabilities for data cleaning and data enrichment
  - ✓ The historical depth of data is limited or expansive
  - ✓ Data reworking in presence of new requirements is unfeasible
  - ✓ Are perceived by companies as self-standing applications, so UGC-related analyses are run separately from those strictly related to business
  - ✓ Does not allow integration with corporate data (Grimes, 2014)
  - ✓ Lack in providing flexible and user-driven analysis (Grimes, 2014)

From Social-Media Monitoring to Social Business Intelligence

- Social-Media Monitoring process can be seen as a DW process
  - ✓ Extract semi-structured data from the web/data provider/CRM
  - ✓ Transform, enrich and clean data
  - ✓ Load data in a system oriented to data analysis
- The process is much more complex since
  - ✓ Crawling the web is not as easy as accessing the enterprise DBs
  - ✓ Data are semi-structured
  - ✓ Enrichment is based on text-mining and NLP techniques
  - ✓ Data volume could be huge

The Trend



Social BI: a Definition

- **Social Business Intelligence** is the discipline that applies DW and OLAP approaches to the analysis of user-generated content to let decision-makers improve their business based on the trends perceived from the environment.
- As in traditional BI the goal of SBI is to enable powerful and flexible analysis even for decision makers with limited technical skills.
- In a SBI system
  - ✓ OLAP-like operators allow flexible, detailed and user-driven analysis
  - ✓ Social data becomes an asset of the company
    - **Verticalization/Personalization** improves analysis effectiveness
    - Data can be **reworked** in order to clean and enrich data as much as needed
    - Social data can be **integrated** with corporate data in order to better analyze the effect of social behaviors on the enterprise

Verticalization

- Verticalization of an SBI system refers to the possibility of tuning the system based on the specific domain of listening the application is running on:
  - ✓ Dictionary enrichment, words specific of the domain of listening are added to the system dictionary in order to make it able to recognize and analyze them
  - ✓ Polarization changes, refer to changing the general polarization of a word (positive or negative) in order to better capture its understanding in the specific domain of listening
  - ✓ Semantic Enrichment tuning: is aimed at improving the effectiveness of semantic enrichment phase helping the system to understand more text

Integration with Enterprise Data

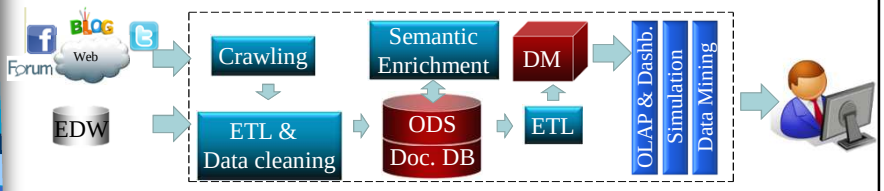
- Carrying out cross-analysis between enterprise and social data is fundamental to properly understand the impact of social events on the enterprise
  - ✓ Ex-post analysis
    - Coupling the trend of a product sentiment with its sales
    - Coupling customer complaints with the churn rate
    - Coupling the level of appreciation of a marketing campaign with the sales increase
  - ✓ Ex-ante (real time)
    - Situational Awareness: understand which enterprise facts (contract, plant, delivery) are affected by an external event (strike, earthquake, accident)

Some Non-Technical Remarks

- The interest around Social-Media Monitoring and Social BI is high since companies cannot ignore UGC-derived information
- Nowadays project costs are largely determined by the licenses of commercial tools
  - ✓ Crawling engines for gathering data from the web
  - ✓ Semantic engines for data enrichment
- The future of Social BI will be determined by:
  - ✓ The spreading of open source software or with more affordable prices (e.g. Google Prediction API?)
  - ✓ A more clear understanding of the impact of the social environment on the performance of real enterprise

An Architecture for SBI

Functional Modules



- ODS (Operational Data Store) that stores all the relevant data about clips, their topics, their authors, and their source channels to this end, a relational database is coupled with a document-oriented database that can efficiently store and search the text of the clips and with a triple store to represent the topic ontology.
- A semantic enrichment component extracts information from raw text applying information retrieval, text mining, NLP techniques.
- A crawling component that runs a set of keyword-based queries to retrieve the clips (and the related meta-data) that lie within the subject area.

Project types

- The components mentioned above are normally present, though with different levels of sophistication, in most current commercial solutions for SBI

- Level 1: Best-of-Breed A best-of-breed policy is followed to acquire tools specialized in one of the steps necessary to transform raw clips in semantically-rich information.
  - ✓ Followed by those who run a medium to long-term project to get full control of the SBI process by finely tuning all its critical parameters
  - ✓ Typically aimed at implementing ad-hoc reports and dashboards to enable sophisticated analyses of the UGC.

- Level 2: End-to-End A single software/service is acquired and tuned.
  - ✓ Customers only need to carry out a limited set of tuning activities that are typically related to the subject area, while a service provider or a system integrator ensures the effectiveness of the technical (and domain-independent) phases of the SBI process.

Project types

- Level 3: Off-the-Shelf Consists in adopting, typically in a as-a-service manner, an off-the-shelf solution supporting a set of reports and dashboards that can satisfy the most frequent user needs in the SBI area (e.g., average sentiment, top topics, trending topics, and their breakdown by source/author /sex).
  - ✓ With this approach the customer has a very limited view of the single activities that constitute the SBI process, so she has little or no chance of positively impacting on activities that are not directly related to the analysis of the final results.

The Crawling module

- This module is in charge of capturing UGCs through a set of keyword-based queries
- While this task is quite easy when the UGC sources are controlled by the enterprise (e.g. Enterprise CRM), it becomes very hard when the listening domain is the web
  - ✓ Parsing XML or JSON data
  - ✓ Split content from advertising
  - ✓ Discard duplicate contents/clips
  - ✓ Collect meta-information about sources, authors, etc.
- Approaches are mainly based on
  - ✓ Templates of the web-source pages
  - ✓ API provided by the web-source

The Semantic Enrichment module

- This module is in charge of extracting from the raw text as many information as possible
- A large amount of research has been carried out on this issue. The main tasks carried out are:
  - ✓ **Entity extraction:** locates and classifies elements in text into pre-defined categories (e.g. names of persons, organizations, locations)
  - ✓ **Relation extraction:** identifies relations between named-entities
  - ✓ **Sentiment analysis:** identifies the positive, negative or neutral polarization text. Depending on the adopted technique can be roughly computed at the clip level, or can be detailed for each words group. (Liu, 2012)
  - ✓ **Clip clustering**: identifies clusters of clips related to the same topics.
- Approaches range in
  - ✓ Statistical and Text mining
  - ✓ Machine-learning
  - ✓ Natural Language Processing

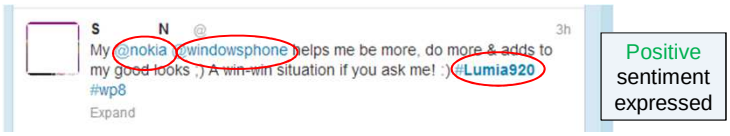
Data Modeling in SBI

Analysis of textual UGC through relevant topics

- A key role in the analysis is played by topics, meant as specific concepts of interest within the subject area

Analysis of textual UGC through relevant topics

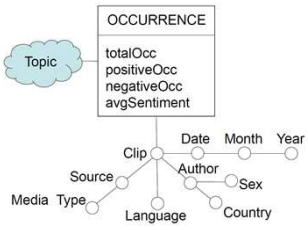
- A key role in the analysis is played by topics, meant as specific concepts of interest within the subject area



Analysis of textual UGC through relevant topics

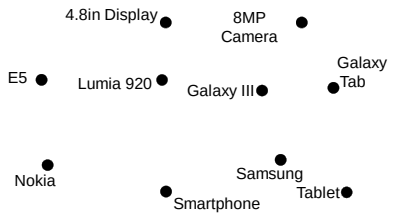
- A key role in the analysis is played by topics, meant as specific concepts of interest within the subject area

- Topics are an obvious candidate dimension of the cubes for Social BI, but:
  - ✓ Trending topic are heterogeneous and change quickly over time
  - ✓ A classical dimension table with a static hierarchy is not suitable



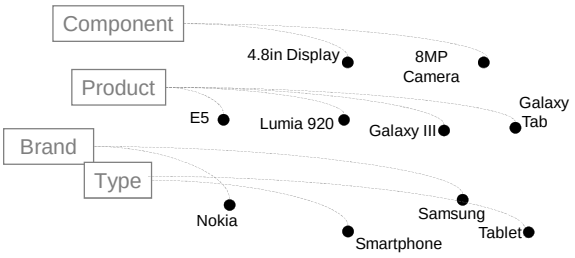
Topic hierarchy schema

- Consider a mobile-oriented scenario



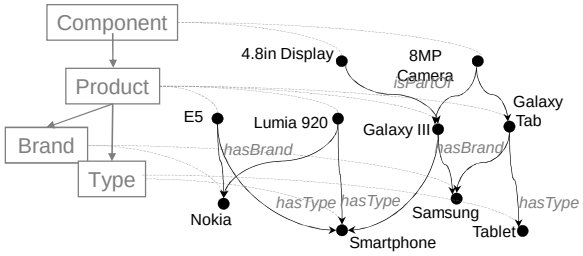
Topic hierarchy schema

- Consider a mobile-oriented scenario
  - ✓ Most topics can be classified into levels, that correspond to aggregation levels in traditional hierarchies



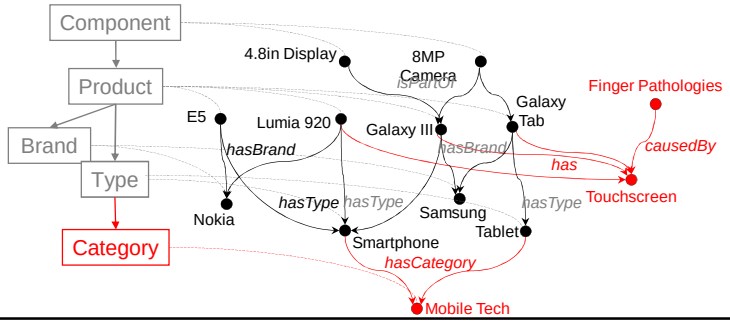
Topic hierarchy schema

- Consider a mobile-oriented scenario
  - ✓ Most topics can be classified into levels, that correspond to aggregation levels in traditional hierarchies
  - ✓ Relationships between topics highlight roll-up relationships



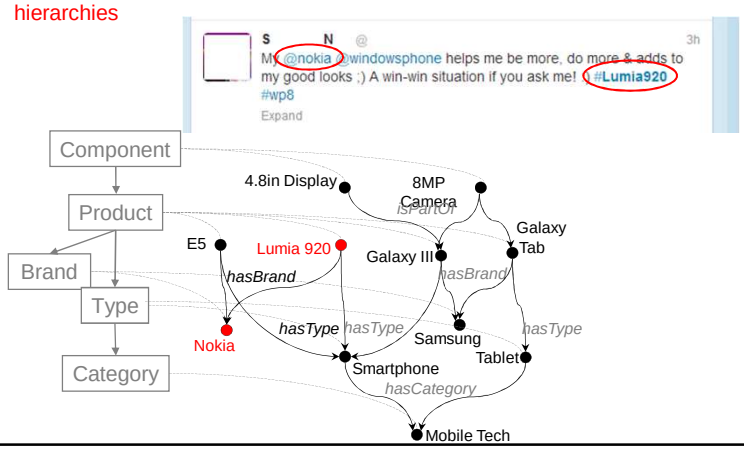
Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  1. Dynamicity: new topics, relationships and aggregation levels might be added at any time



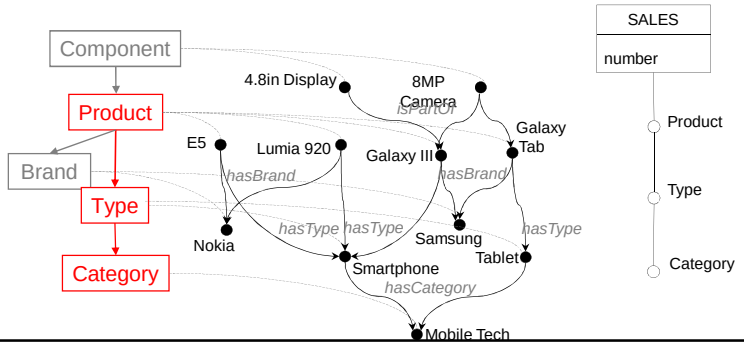
Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  1. Dynamicity: new topics, relationships and aggregation levels might be added at any time
  2. Mixed granularity (facts associated to non leaf-topics) and unbalanced hierarchies



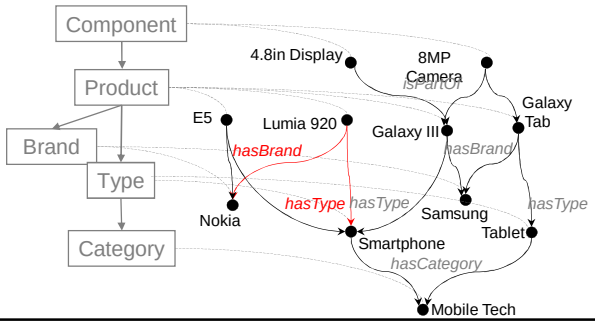
Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  1. Dynamicity: new topics, relationships and aggregation levels might be added at any time
  2. Mixed granularity (facts associated to non leaf-topics) and unbalanced hierarchies
  3. Integration: some topics are also part of business hierarchies and require a direct connection with the enterprise cube



Topic hierarchy schema

- How is a topic hierarchy different from a traditional hierarchy?
  2. Mixed granularity (facts associated to non leaf-topics) and unbalanced hierarchies
  3. Integration: some topics are also part of business hierarchies and require a direct connection with the enterprise cube
  4. Semantics: roll-up relationships between topics can have different semantics



The Meta-Star approach

- Meta-Stars overcome these issues by using a combination of modeling strategies (Gallinucci, 2013)

- Navigation tables
  - ✓ Support hierarchy instances with different lengths and non-leaf facts
  - ✓ Allow different roll-up semantics to be explicitly annotated

- Meta-modeling
  - ✓ Enable hierarchy heterogeneity and dynamicity to be accommodated

- Traditional dimension tables
  - ✓ Easy integration with standard business hierarchies

The Meta-Star approach

- Implementation of a Meta-Star requires two components:
- A Topic Table
  - ✓ Stores all the topics of the hierarchy
  - ✓ Topic levels can be modeled in a static way (i.e., like in a classical dimension table)
- A Roll-up Table
  - ✓ Stores every relationship between two topics in the transitive closure

The Meta-Star approach

- Implementation of a Meta-Star: the **topic table**
  - ✓ One row for each topic

TOPIC_T

IdT	Topic	Level
1	8MP Camera	Component
2	Galaxy III	Product
3	Galaxy Tab	Product
4	Smartphone	Type
5	Tablet	Type
6	Mobile Tech	Category
7	Samsung	Brand
8	Finger Path.	-
9	Touchscreen	-
...	...	...

The Meta-Star approach

- Implementation of a Meta-Star: the **topic table**
  - ✓ One row for each topic
  - ✓ With respect to start-schema levels are meta-modeled and become instance rather than attributes

TOPIC_T

IdT	Topic	Level
1	8MP Camera	**Component**
2	Galaxy III	**Product**
3	Galaxy Tab	**Product**
4	Smartphone	**Type**
5	Tablet	**Type**
6	Mobile Tech	**Category**
7	Samsung	**Brand**
8	Finger Path.	-
9	Touchscreen	-
...	...	...

The Meta-Star approach

- Implementation of a Meta-Star: the **topic table**
  - ✓ One row for each topic
  - ✓ With respect to start-schema levels are meta-modeled and become instance rather than attributes
  - ✓ Columns for each static level, like in a classical dimension table

TOPIC_T

IdT	Topic	Level	**Product**	**Type**	**Category**
1	8MP Camera	Component	-	-	-
2	Galaxy III	Product	Galaxy III	Smartphone	Mobile Tech
3	Galaxy Tab	Product	Galaxy Tab	Tablet	Mobile Tech
4	Smartphone	Type	-	Smartphone	Mobile Tech
5	Tablet	Type	-	Tablet	Mobile Tech
6	Mobile Tech	Category	-	-	Mobile Tech
7	Samsung	Brand	-	-	-
8	Finger Path.	-	-	-	-
9	Touchscreen	-	-	-	-
...	...	...	...	...	...

The Meta-Star approach

- Implementation of a Meta-Star: the **roll-up table**
  - ✓ One row for each arc in the transitive closure of the hierarchy

ROLLUP_T

ChildId	RollUpSignature	FatherId
1	000000	1
2	000000	2
…	000000	…
1	100000	2
1	100000	3
2	010000	4
2	001000	7
4	000100	6
8	000001	9
2	000010	9
…	…	…
1	110000	4
1	110000	5
1	101000	7
1	100010	9
2	010100	6
3	010100	6
…	…	…
1	110100	6
…	…	…

The Meta-Star approach

- Implementation of a Meta-Star: the **roll-up table**
  - ✓ One row for each arc in the transitive closure of the hierarchy

ROLLUP_T

ChildId	RollUpSignature	FatherId
1	000000	1
2	000000	2
…	000000	…
1	100000	2
1	100000	3
2	010000	4
2	001000	7
4	000100	6
8	000001	9
2	000010	9
…	…	…
1	110000	4
1	110000	5
1	101000	7
1	100010	9
2	010100	6
3	010100	6
…	…	…
1	110100	6
…	…	…

- Each bit of the *roll-up signature* corresponds to one roll-up semantics
- If the hierarchy includes a directed path from t_1 to t_2, the bits corresponding to the involved roll-up semantics are set to 1

t_1	isPartOf	hasType	hasBrand	hasCategory	has	causedBy	t_2
8MP Camera	1	1	0	1	0	0	Mobile Tech

The Meta-Star approach

- The combination of meta-modeling with the roll-up table accommodates the dynamicity of the topic hierarchy

TOPIC_T

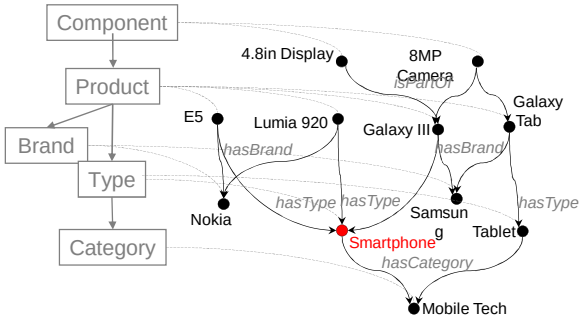
IdT	Topic	Level
1	8MP Camera	Component
2	Galaxy III	Product
3	Galaxy Tab	Product
4	Smartphone	Type
5	Tablet	Type
6	**Mobile Tech**	**Category**
7	Samsung	Brand
8	Finger Path.	-
9	Touchscreen	-
...	...	...

ROLLUP_T

ChildId	RollUpSignature	FatherId
1	000 **0** 00	1
2	000 **0** 00	2
...	000 **0** 00	...
1	100 **0** 00	2
1	100 **0** 00	3
2	010 **0** 00	4
2	001 **0** 00	7
4	**000 1 00**	**6**
8	000 **0** 01	9
2	000 **0** 10	9
...	...	...
1	110 **0** 00	4
1	110 **0** 00	5
1	101 **0** 00	7
1	100 **0** 10	9
2	**010 1 00**	**6**
3	**010 1 00**	**6**
...	...	...
1	**110 1 00**	**6**
...	...	...

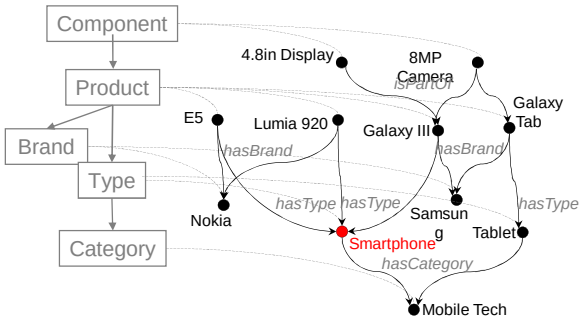
Querying Meta-Stars

- We distinguish two main type of queries
  - ✓ **Schema-free topic query**: the GP component is defined by topics
    - ✓ For each of the given topics a group is defined
  - ✓ **Schema-aware topic query**: the GP component is defined by levels
    - ✓ As in traditional OLAP query a group is created for each topic of the given level
- The semantic filters determines the composition of each group
  - ✓ **Queries without topic aggregation**: only facts related to the specific topic are considered



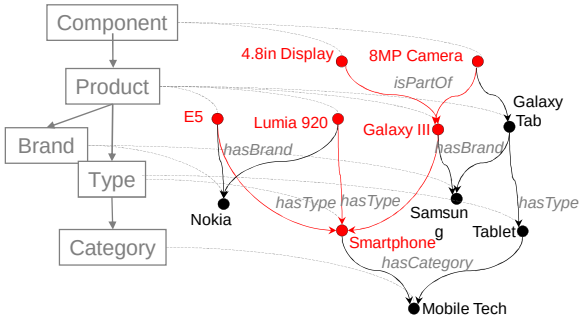
Querying Meta-Stars

- We distinguish two main type of queries
  - ✓ **Schema-free topic query**: the GP component is defined by topics
    - ✓ For each of the given topics a group is defined
  - ✓ **Schema-aware topic query**: the GP component is defined by levels
    - ✓ As in traditional OLAP query a group is created for each topic of the given level
- The semantic filters determines the composition of each group
  - ✓ **Queries without topic aggregation**: only facts related to the specific topic are considered



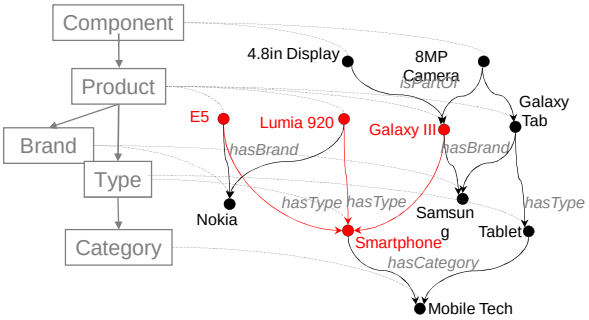
Querying Meta-Stars

- We distinguish two main type of queries
  - ✓ **Schema-free topic query**: the GP component is defined by topics
    - ✓ For each of the given topics a group is defined
  - ✓ **Schema-aware topic query**: the GP component is defined by levels
    - ✓ As in traditional OLAP query a group is created for each topic of the given level
- The semantic filters determines the composition of each group
  - ✓ **Queries with full topic aggregation**: no filter on semantics is applied
    - ✓ It is the standard OLAP semantic



Querying Meta-Stars

- We distinguish two main type of queries
  - ✓ **Schema-free topic query**: the GP component is defined by topics
    - ✓ For each of the given topics a group is defined
  - ✓ **Schema-aware topic query**: the GP component is defined by levels
    - ✓ As in traditional OLAP query a group is created for each topic of the given level
- The semantic filters determines the composition of each group
  - ✓ **Queries with semantic topic aggregation**: semantic filter is user-defined



Querying Meta-Stars

- Question: what is the current average sentiment over smartphones?
  - Example of query with full-topic aggregation

FT

AvgSentim	IdT
+5	1
-3	1
+2	1
+15	2
-20	2
…	3
-15	4
+20	4
…	5
…	6
…	7
…	7
…	7
…	8
…	…

ROLLUP_T

ChildId	RU_Sign	FatherId
1	000000	1
2	000000	2
4	**000000**	4
7	000000	7
1	100000	2
1	100000	3
2	**010000**	4
2	001000	7
4	000100	6
8	000001	9
2	000010	9
1	**110000**	4
1	110000	5
3	001000	7
1	100010	9
2	010100	6
3	010100	6
1	110100	6

Topic = 'Smartphone'

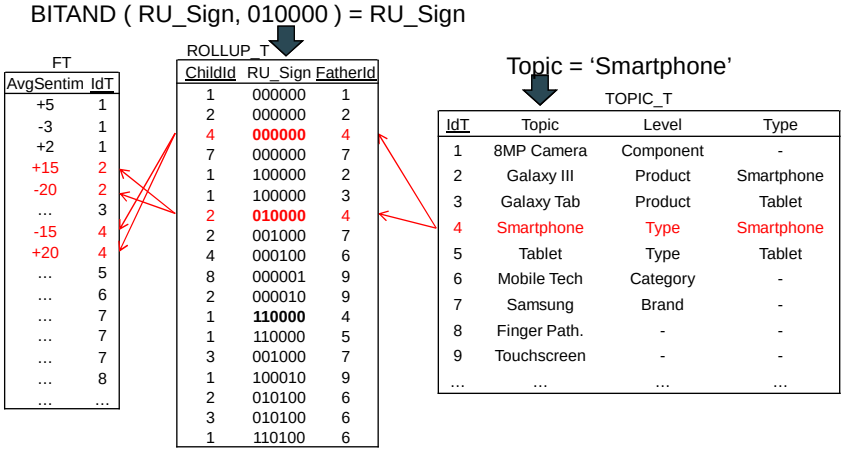
TOPIC_T

IdT	Topic	Level	Type
1	8MP Camera	Component	-
2	Galaxy III	Product	Smartphone
3	Galaxy Tab	Product	Tablet
4	Smartphone	Type	Smartphone
5	Tablet	Type	Tablet
6	Mobile Tech	Category	-
7	Samsung	Brand	-
8	Finger Path.	-	-
9	Touchscreen	-	-
…	…	…	…

Querying Meta-Stars

- Question: what is the current average sentiment over smartphones?
  - Example of query with semantic topic aggregation

BITAND (RU_Sign, 010000) = RU_Sign



FT

AvgSentim	IdT
+5	1
-3	1
+2	1
+15	2
-20	2
…	3
-15	4
+20	4
…	5
…	6
…	7
…	7
…	7
…	8
…	…

ROLLUP_T

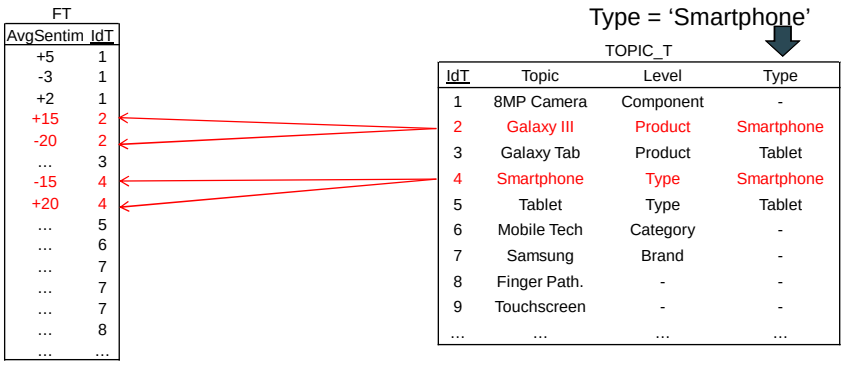
ChildId	RU_Sign	FatherId
1	000000	1
2	000000	2
4	**000000**	4
7	000000	7
1	100000	2
1	100000	3
2	**010000**	4
2	001000	7
4	000100	6
8	000001	9
2	000010	9
1	**110000**	4
1	110000	5
3	001000	7
1	100010	9
2	010100	6
3	010100	6
1	110100	6

TOPIC_T

IdT	Topic	Level	Type
1	8MP Camera	Component	-
2	Galaxy III	Product	Smartphone
3	Galaxy Tab	Product	Tablet
4	Smartphone	Type	Smartphone
5	Tablet	Type	Tablet
6	Mobile Tech	Category	-
7	Samsung	Brand	-
8	Finger Path.	-	-
9	Touchscreen	-	-
…	…	…	…

Querying Meta-Stars

- Question: what is the current average sentiment over smartphones?
  - Example of query with full topic aggregation using static levels



FT

AvgSentim	IdT
+5	1
-3	1
+2	1
+15	2
-20	2
…	3
-15	4
+20	4
…	5
…	6
…	7
…	7
…	7
…	8
…	…

TOPIC_T

IdT	Topic	Level	Type
1	8MP Camera	Component	-
2	Galaxy III	Product	Smartphone
3	Galaxy Tab	Product	Tablet
4	Smartphone	Type	Smartphone
5	Tablet	Type	Tablet
6	Mobile Tech	Category	-
7	Samsung	Brand	-
8	Finger Path.	-	-
9	Touchscreen	-	-
…	…	…	…

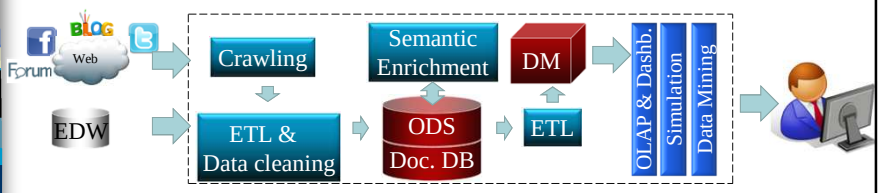
A Case Study on 2014 European Election

Our projects

- We collaborated with the following projects (carried out without adopting an ad-hoc methodology)
  - ✓ DOXA – the widest Italian Market Analysis Company
  - ✓ Amadori - the italian leader in the poultry industry
  - ✓ *City mood - Bologna*
- We are now running a large Social BI project within the WebPolEU – FIRB project (http://webpoleu.altervista.org/)
  - ✓ *The project aims at studying the nexus between politics and social media in comparative perspective from the viewpoint of both citizens and political actors*
- We monitor the European Election 2014 over three different countries (Italy, England, Germany)
  - 2 months of listening
  - 20 milions of raw clips
  - 60,000 web sources
  - 3 different languages
  - 2 full-time + 2 half-time + 5 *end user*

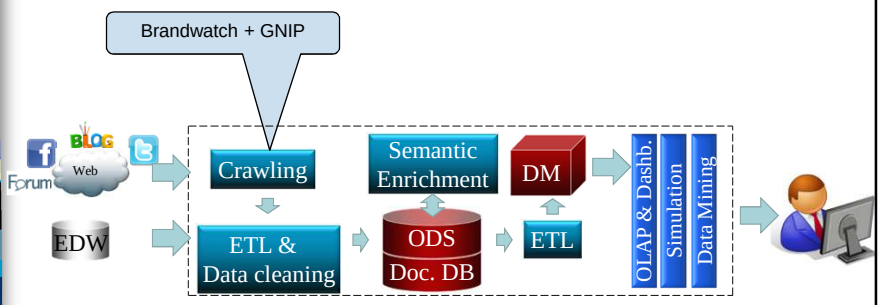
Our architecture

- We adopted an “Best-of-breed” solution



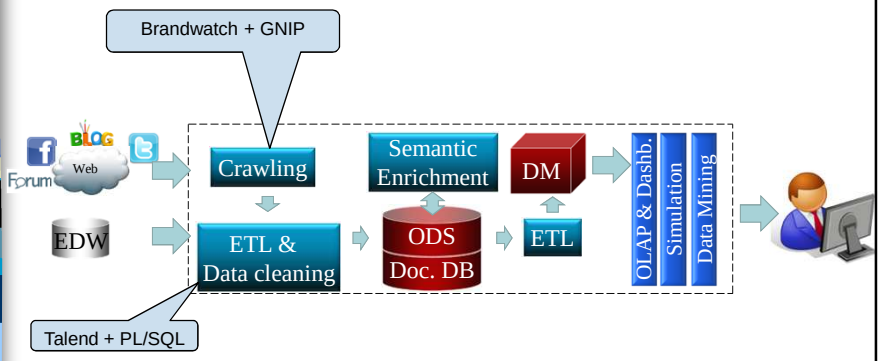
Our architecture

- We adopted an “Best-of-breed” solution



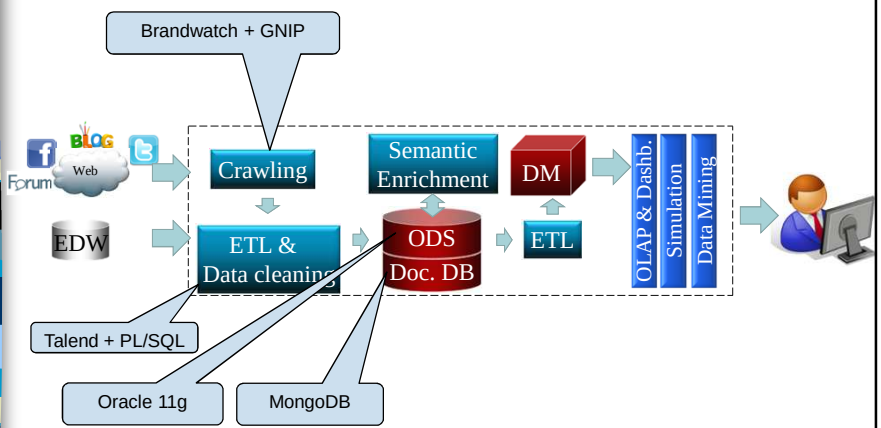
Our architecture

- We adopted an “Best-of-breed” solution



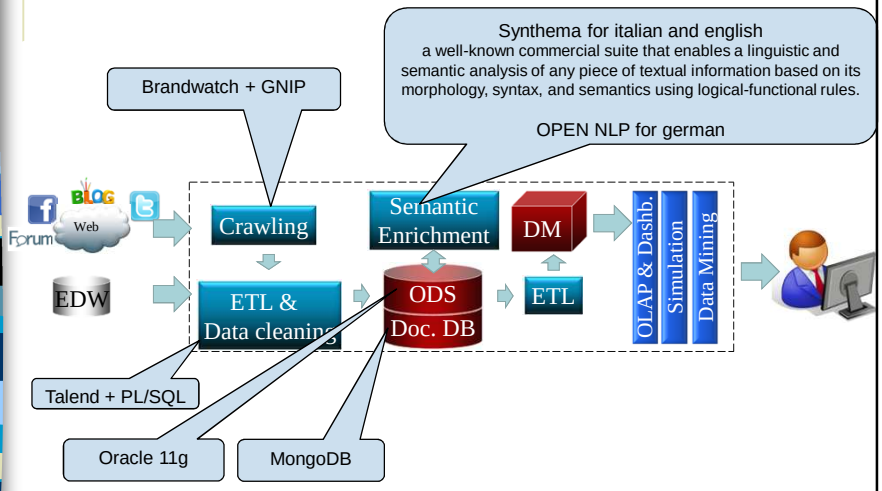
Our architecture

- We adopted an “Best-of-breed” solution



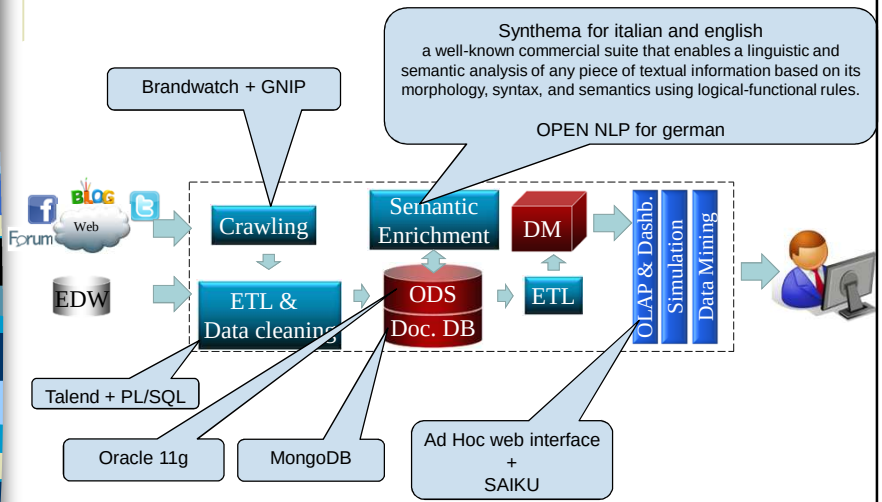
Our architecture

- We adopted an “Best-of-breed” solution



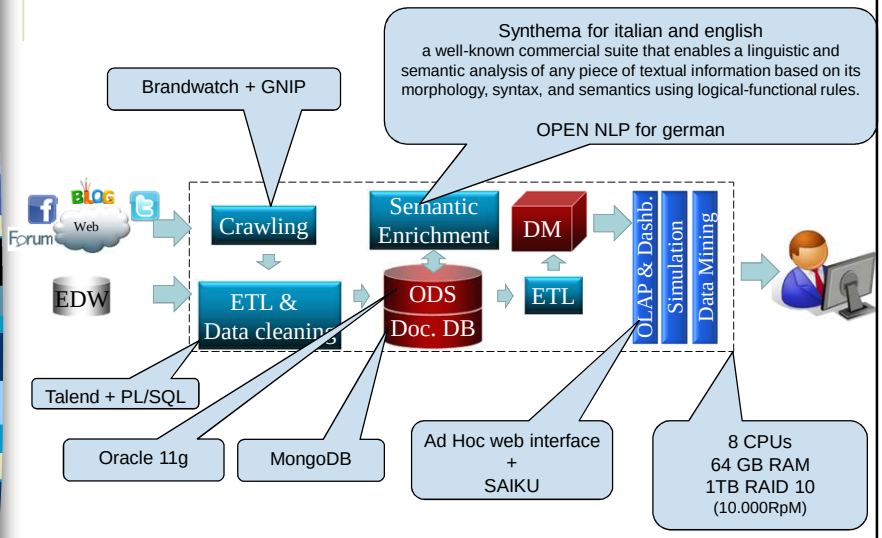
Our architecture

- We adopted an “Best-of-breed” solution

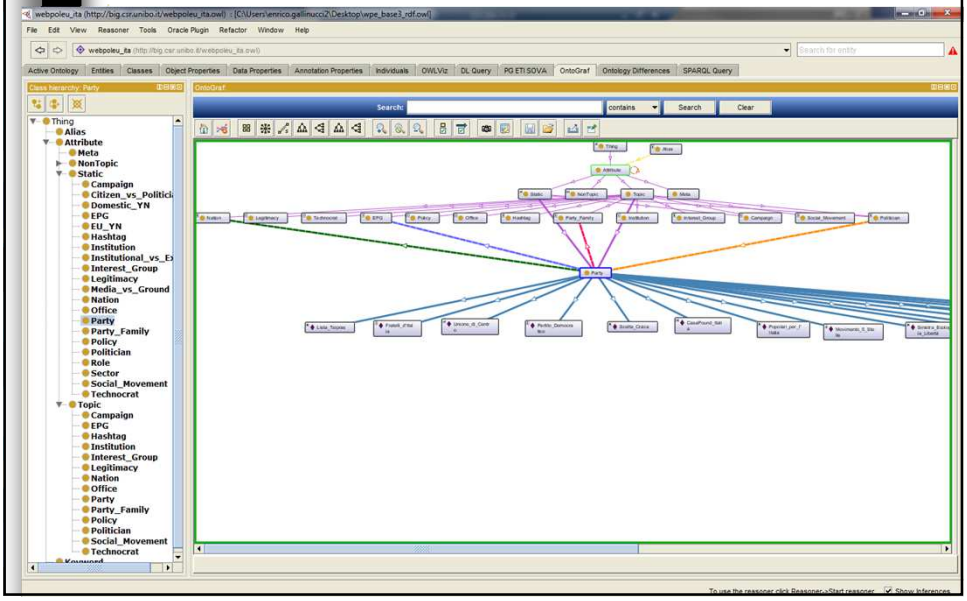


Our architecture

- We adopted an “Best-of-breed” solution

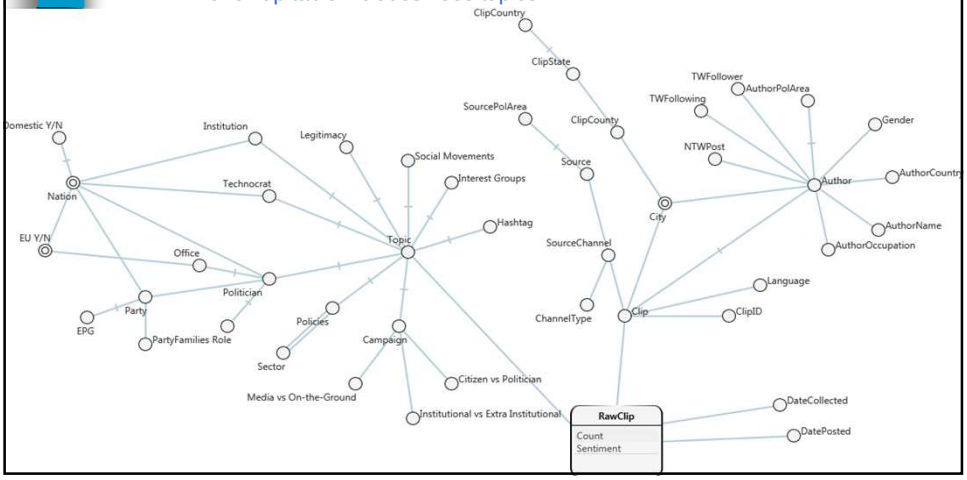


Building the topic hierarchy



Building the topic hierarchy

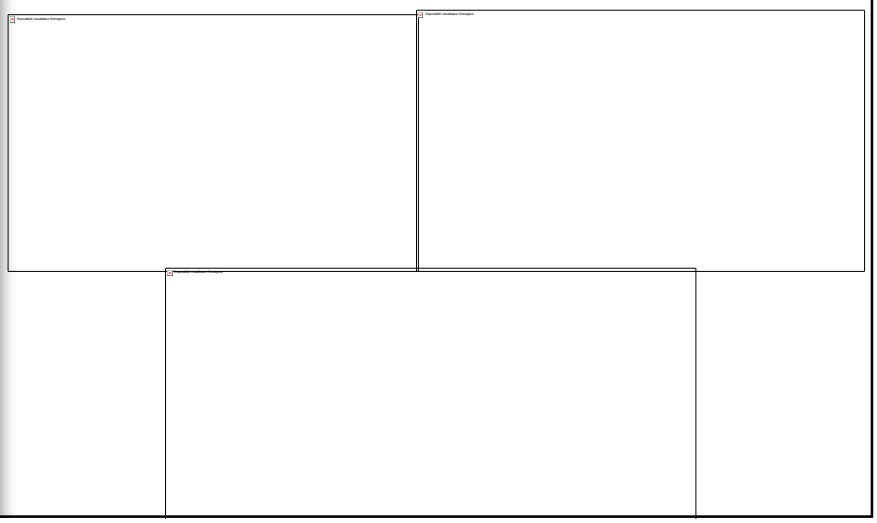
- Working with our users we detected
  - ✓ 669 alias
  - ✓ 396 topic
  - ✓ The roll-up table includes 1588 tuples



Demo time!

Some comments

- Twitter is largely the main (viral) clip source



Some comments

- Twitter is largely the main (viral) clip source
- The project scope determines the quantity of data to be handled
  - ✓ In many cases storing can be handled with traditional technologies but in many others a Big Data approach must be followed
  - ✓ OLAP with Big Data is far to be an explored topic
- Deep semantic analysis may largely increase the size of the data to be handled
  - ✓ It is not clear (*at least to me*) if this extra cost has a value for customers
  - ✓ The polarization correctness has still a statistic value
- Responsiveness in an SBI project is not a choice but rather a necessity, since the frequency of changes requires a tight involvement of domain experts to detect these changes and rapid iterations to keep the process well-tuned.
- If a proper methodology is not adopted the main problems are:
  - ✓ a lack of synchronization between the activities, that reduced their effectiveness
  - ✓ an insufficient control on the effects of changes

Questions?

Related Readings

(Malloy, 2012) Tom Malloy. *Revolutionizing Digital Marketing with Big Data*, In CIKM 2012. Hawaii (USA), 2012.

(Grimes, 2014) Seth Grimes. *Sentiment Analysis and Business Sense*. Retrieved on 30th April 2014 from clarabridge.com.

(Lee, 2000) J. Lee, D. Grossman, O. Frieder, M.C. McCabe. *Integrating structured data and text: a multi-dimensional approach*, in Int. Conf. on Information Technology: Coding and Computing, Las Vegas, 2000.

(Ravat, 2008) F. Ravat, O. Teste, R. Tournier, G. Zurfluh *Top_Keyword: an Aggregation Function for Textual Document OLAP*. In DaWaK 2008 Turin, Italy, 2008.

(Rehman, 2012) N. Rehman, S. Mansmann, A. Weiler, M.H. Scholl. *Building a Data Warehouse for Twitter Stream Exploration*. In Int.Conf. on Advances in Social Networks Analysis and Mining (ASONAM), Istanbul, Turkey, 2012

Related Readings

(García-Moya, 2013) L. García-Moya, S. Kudama,M. J. Aramburu, R. Berlanga. *Storing and analyzing voice of the market data in the corporate data warehouse.* In Information Systems Frontier2012. Vol. 15(3), pp. 331-349, 2013.

(Gallinucci, 2013) E. Gallinucci, M. Golfarelli, S. Rizzi. Meta-Stars: Multidimensional Modeling for Social Business Intelligence. In *Proc. DOLAP 2013*, San Francisco, USA, 2013.

(Francia, 2014) M. Francia, M. Golfarelli, S. Rizzi. A Methodology for Social BI. In *Proc. IDEAS2014*, Porto, Portugal, 2014.

(Liu, 2012) Bing Liu. Sentiment Analysis and Opinion Mining. Morgan & Claypool, 2012.