

Social Business Intelligence in Action

Matteo Francia, **Enrico Gallinucci**, Matteo Golfarelli, Stefano Rizzi

DISI – University of Bologna, Italy



15/06/2016

Introduction

Several Social-Media Monitoring tools are available for UGC analysis

- Perceived as **self-standing apps**: no integration with corporate data
- Offered **as-a-service**: lack in sufficient verticalization / personalization
- **Project-oriented**: narrow time-horizon, limited historical depth



*"**Social Business Intelligence** is the discipline that aims at **combining corporate data with user-generated content (UGC)** to let decision-makers analyze and improve their business based on the **trends and moods** perceived from the environment."*

15/06/2016

Introduction

In SBI, social media monitoring becomes an integrated **DW process**

- **Cross-analysis** between enterprise and social data is **fundamental** to properly understand the impact of social events on the enterprise
- **Social data become an asset** of the company

The goal of our approach is to **discuss the architectural options available for an SBI project**

- Help designers in finding the right cost-benefit compromise

Based on the experience of **several real world projects**, including:

- Collaboration with **Amadori**, Italian leader in poultry industry
- Regional project on monitoring **vaccine-related discussions** and fears
- Ministerial project on the analysis of the **2014 European elections**

15/06/2016

Introduction: the project

The WebPoLEU Project aimed at studying the connection between **politics and social media**

- <http://webpoleu.net>

SBI is used as an enabling technology for analyzing the UGC

- When: **March 2014 to May 2014** (elections held on May 22-25, 2014)
- Where: **Germany, Italy, United Kingdom**

Around **10 millions clips** collected

- Facebook posts, tweets, blogs and forum posts, news feeds and comments, etc.



15/06/2016

SBI architecture: classification

In an SBI architecture, the roles of each component may vary from project to project

- Design complexity and control level by the user may vary

Off-the-Shelf solution

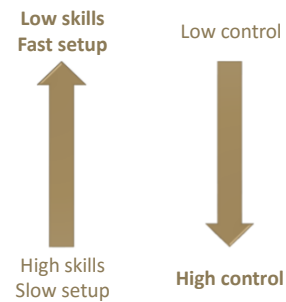
- Adopt a full solution, supporting a set of standard reports and dashboards

End-to-End solution

- Acquire and tune an end-to-end software / service

Best-of-Breed solution

- Acquire specialized tools in one or more parts of the process



15/06/2016

SBI architecture: reference

A reference architecture for SBI has been proposed in [2]

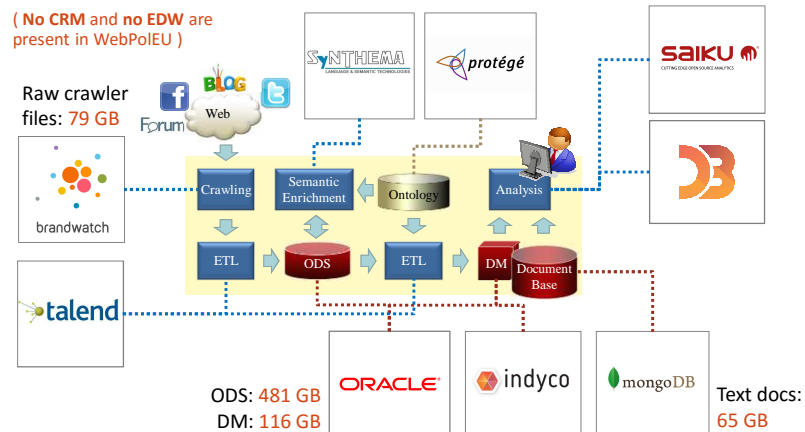


- [2] M. Francia, M. Golfarelli, S. Rizzi. A methodology for social BI. In Proc. IDEAS, pp 207-216, 2014

15/06/2016

SBI architecture: WebPolEU

(No CRM and no EDW are present in WebPolEU)



15/06/2016

SBI architectural options

Crawling



Context

- The main burden is in ensuring a good compromise between too much / too little content

Considerations

- **Off-the-shelf**: the designer only carries out macro analysis
- **End-to-end**: clipping is guaranteed by the service provider, querying is controlled by the designer
- **Best-of-breed**: all technical activities are in charge of the designer; potentially burdensome and very time-consuming

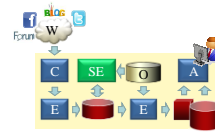
In WebPolEU

- Crawling process relies on **Brandwatch**, a third-party service (end-to-end)
- It also extracts metadata (source, author, etc.) and derives clip sentiment

15/06/2016

SBI architectural options

Semantic enrichment



Context

- A wide spectrum of technological alternatives

Considerations

- Basic techniques may be sufficient to analyze raw data
 - e.g., count topic occurrences
- NLP analysis techniques are powerful but potentially expensive
 - e.g., extract lemmas, semantic relationships between lemmas, more detailed sentiment
- For inherently complex languages (e.g., German), automated analysis and interpretation of sentences is discouraged

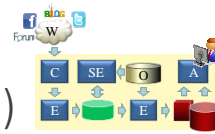
In WebPoIEU

- Semantic enrichment is achieved as the combination of both basic and advanced techniques
- NLP analysis is carried out by the commercial system SyN Semantic Center

15/06/2016

SBI architectural options

ODS (Operational Data Store)



Context

- The ODS component is not strictly necessary

Considerations

- ODS is actually recommended
 - Buffering and early analysis (separate crawling from semantic enrichment)
 - Clip reprocessing (semantic enrichment is an iterative process)
 - Data cleaning (more effective on materialized data rather than on-the-fly)
- Relational or NoSQL?
 - NoSQL guarantees scalability
 - Transactional workload is better handled with ACID properties
 - Metadata processing also favors a well-defined, normalized schema

In WebPoIEU

- An RDBMS is used; a NoSQL repository is only used to enable text search

15/06/2016

SBI architectural options Analysis



Context

- Analysis is a **key component** of SBI architectures; it can take a variety of shapes, with quite different capabilities
- Dashboards**: small number of predefined views and navigations
- Text search**: detailed analysis up to the single UGC level
- Text mining**: advanced analyses (e.g., clip clustering, topic discovery)
- OLAP**: flexible analysis on the multidimensional metaphor

Considerations

- Off-the-shelf** solutions: dashboards and text search, rarely text mining
- OLAP capabilities** are clearly more powerful, but also complex to provide

In WebPoIEU

- Dashboards, text search and OLAP capabilities are enabled

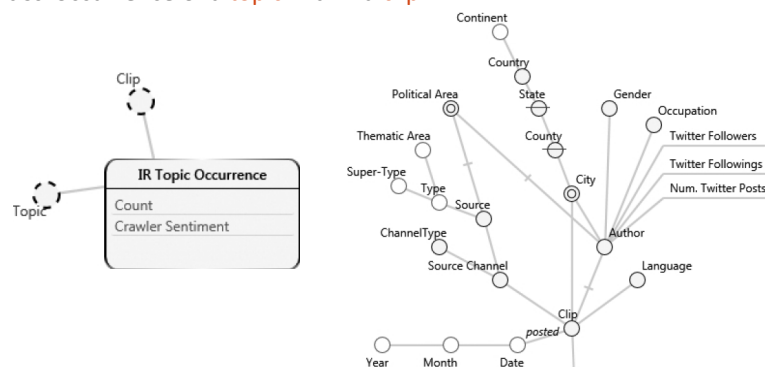
15/06/2016

SBI architectural options Analysis (WebPoIEU)



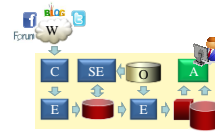
Clip hierarchy: built with **metadata** from the crawling service

Fact: occurrence of a **topic** within a **clip**

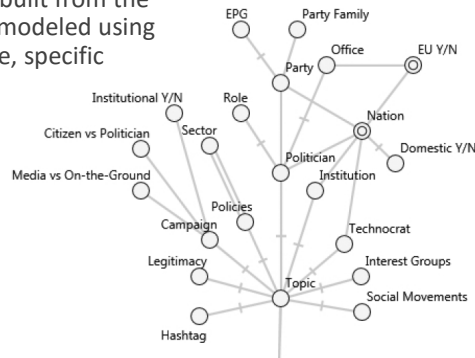


15/06/2016

SBI architectural options Analysis (WebPolEU)



The **topic hierarchy** is built from the **domain ontology** and modeled using an advanced technique, specific for topic hierarchies (*meta-stars*) [4]



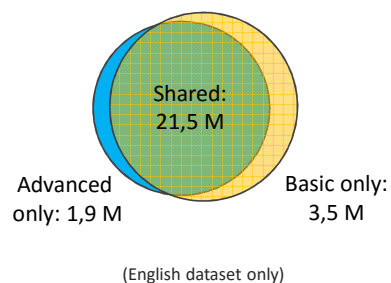
- [4] E. Gallinucci, M. Golfarelli, S. Rizzi. *Advanced topic modeling for social business intelligence*. Information Systems, pp. 53:87-106, 2015.

15/06/2016

Case study: effectiveness

Comparison of the two semantic enrichment techniques

Topic occurrences in the clips



Advanced: SyN semantic engine
Basic: in-house procedure

In most cases, the two techniques find the same topic occurrences

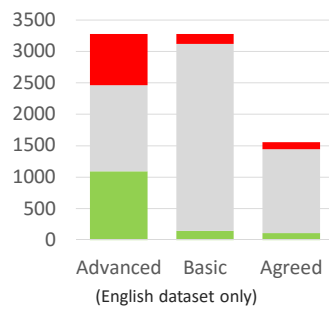
- Basic techniques could be sufficient for KPIs based on topic counting
- Sophisticated ontology-based techniques are required for deeper analyses (e.g., semantic co-occurrences)

15/06/2016

Case study: effectiveness

Comparison of the two semantic enrichment techniques

Clip sentiment comparison



Advanced: SyN semantic engine (lexical analysis of the sentences)

Basic: Brandwatch service (rule-based technique)

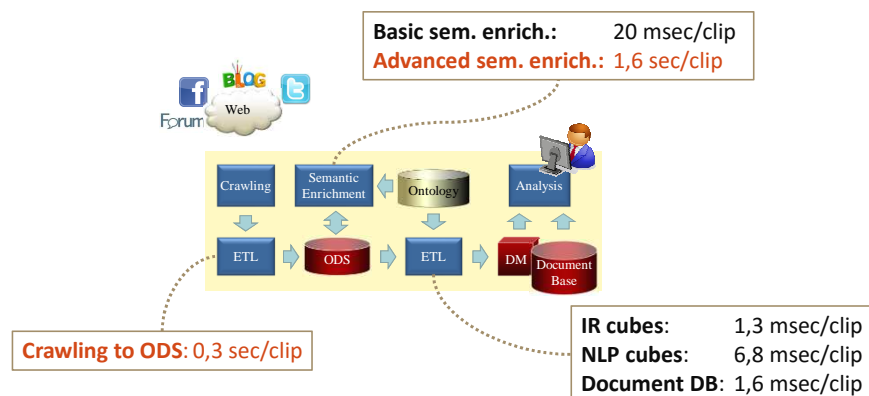
Brandwatch hardly assigns a non-neutral sentiment to a clip

- Due to its inability / caution in assigning a non-neutral sentiment

15/06/2016

Case study: efficiency

Average ETL times



15/06/2016

Case study: sustainability

The design of the architecture is an iterative task

- First design iteration: 84 person-days
- Second design iteration: 30 person-days

Main critical issues:

- **Ontology design:** correctness of the results is deeply affected by the completeness of the domain ontology
- **Crawling setup:** proper formulation within the boundaries set by the service provider (e.g., number of queries, query length) may become a nightmare
- **ETL & OLAP design:** continuous tuning required to handle all possible unexpected results due to bad clipping

15/06/2016

Conclusion

Rules of thumb

- **The adoption of sentiment analysis** should be carefully evaluated
 - Aside from specific sources/closed domains, sentiment accuracy **easily drops**
- **Twitter** is possibly the best source for sentiment analysis
 - Due to the **shortness** of tweets and the **high percentage of opinions**
- **Off-the-shelf solutions** only provide *quick-and-dirty* answers
 - To pursued only with limited available resources or at early stages
- OLAP analysis has been recognized as **truly valuable** by the WebPoIEU users
 - **Full OLAP capabilities** will increasingly be provided as SBI gradually gain importance

WebPoIEU data is going to be **released as a benchmark** for SBI

- The goal is to enable the possibility to test every task of the SBI process, thanks to expert-validated ground truth

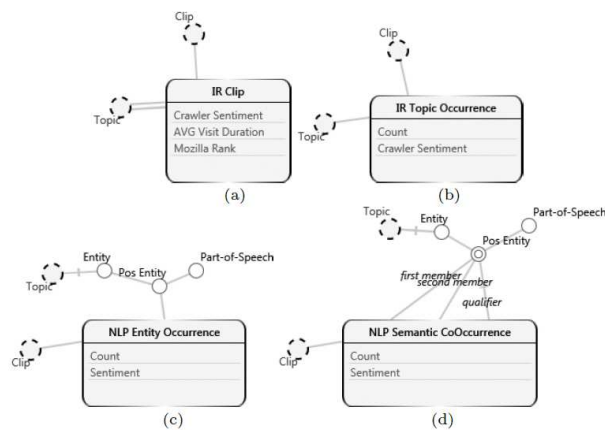
15/06/2016

Thanks



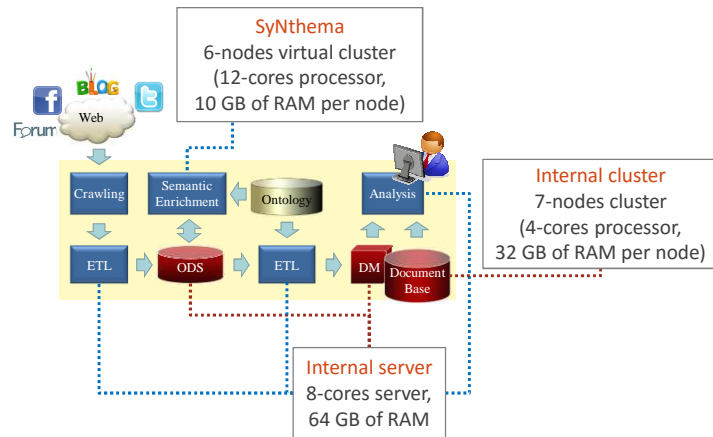
15/06/2016

Architectural options Analysis (WebPoLEU)



15/06/2016

Architecture: specifications

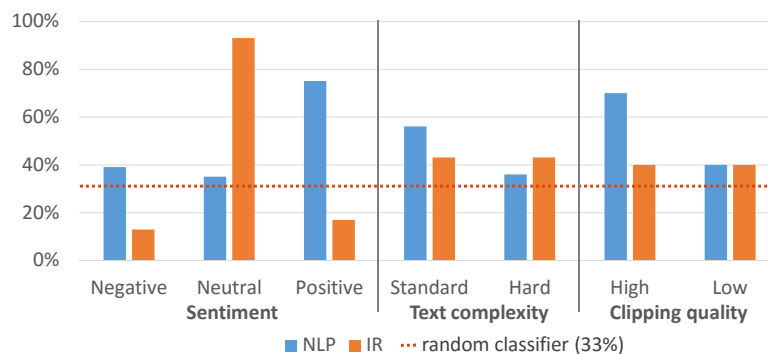


15/06/2016

Case study: effectiveness

Clip sentiment accuracy

- Sample of 600 English clips, manually tagged by domain experts

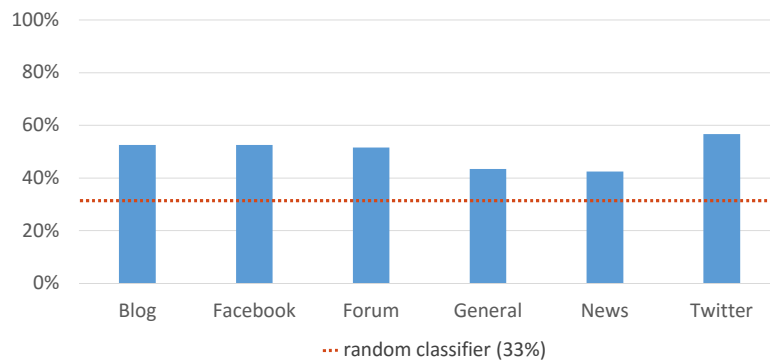


15/06/2016

Case study: effectiveness

Clip sentiment accuracy

- Sample of 600 English clips, manually tagged by domain experts



15/06/2016