NORMA SEARCH: a Big Data application for financial services.

Ylenia Maruccia
Gloria Polimeno
Giovanni B. Pansini
Felice Vitulano

Bari, June 16th 2016
1. **CONTEXT ANALYSIS**

2. **PROPOSED SOLUTION**

3. **FINANCIAL SCENARIOS**

4. **OUR SOLUTION: NORMA SEARCH**

5. **NORMA SEARCH: FUNCTIONALITIES**

6. **NORMA SEARCH: A BIG DATA ARCHITECTURE**

7. **APPLICATION IN THE BANKING SYSTEM**

8. **IMPLICATIONS & CONCLUSIONS**
With the advent of Big Data, several approaches have been proposed to manage and analyse them, at the aim of finding added value from this huge amount of data and supporting the business ecosystems.

One of the main difficulties lies in the unusual format in which data comes from.

In the banking system, all the legislations are in an unstructured format and it is not easy to automatically manage them.

**Challenge:** How to automatically extract knowledge from them?
NORMA SEARCH is an application with a Big Data architecture that helps finding and analysing new legislations and giving support in a strategic decision making process.
The financial crisis and the speculative use of the derivative instruments has placed the reform of the derivative markets “Over The Counter” among the priorities of the legislature in terms of standard negotiation procedures, as well as more stringent rules pertaining to the capitalization of financial intermediaries.
In terms of rules designed to standardise the trading of OTC derivatives, it has been promulgated different regulations, such as Dodd Frank Act (European Market Infrastructure Regulation), that revived the role of the Central Counter-Parties (CCPs), with the aim of increasing transparency and reduce both the counterpart risk and the operational one.

To ensure the soundness of the banking system, the Basel agreements require the banks of the leading world countries some limits about their operational activities, especially regarding the amount of assets which they have to equip themselves for their clients’ protection, thus allowing the capitalization of banks (and, consequently, liquidity guarantees), to guarantee the operations - collection, financing and investment - put in place with customers.
In the financial system the granularity and, at the same time, the complexity of all the regulations necessitate a constant attention and the monitoring of them, in order to anticipate future changes, integrations or evolutions.

NS is a machine learning tool that:
• may provide guidance on the involved bank process
• indicate with almost predictive function the impacts on the IT applications, in terms of changes and/or new implementations through the analysis of the newly introduced legislation (or that are in its approval process) and/or the changes in the requirements previously promulgated.
Norma Search: Functionalities

- Refer autonomously a set of sites, blogs, forums and so on, looking for info about a set of concepts of interests identified by the machine training activity by examples (weakly supervised training); the system is able to consult a set of predetermined sites (authoritative sites) or even the whole www.

- Identify, in every web page retrieved, the individual portions of text (HTML page section) in which are expressed the sophisticated concepts, by associating a percentage indicator of relevance to such concepts with each section identified.

- Automatically classify and organize web sites and pages that belong to them according to a predetermined conceptual taxonomy or derivable during the training phase machine.
• Filter, as needed, specific types of web sites that tend to generate noise, such as for example search engines based on search engine spamming techniques.

• Identify only new content found on each new consultation.

• Present the results through a simple web interface or as a report directly downloadable from the interface; reports can also be sent from the application via e-mail.

• Independently identify potentially authoritative sites and recognise inactivity of authoritative sites.
**CLOUDERA SEARCH**

It works as:
- **DATABASE** for structured data.
- **SEARCH MODULE** for unstructured data.

**SPARK**

It is useful for:
- **DISTRIBUTED COMPUTING**.
- Its libraries for the application of different **MACHINE LEARNING TECHNIQUES**.
- In particular: **LATENT SEMANTIC ANALYSIS, MARKOV CLUSTERING ALGORITHM**.
Norma Search: Architecture

**NormaSearch Client**

- Administration Console
  - Security rules
  - Web monitoring
  - Retrieval of docs of interest
  - Defining categories/sub-categories
  - Training of security rules
  - Defining new projects using Cool Stuff as starting point to expand research

- Fruition Console
  - Manage docs
  - What docs to be processed?
  - What docs are useful for new experimental projects?
  - What docs to be dismissed?

Bari, 16/06/2016

Y. Maruccia, G. Polimeno, G. B. Pansini, F. Vitulano
Application in the banking system

• NS simulates the human activities: browsing the web, searching on specific sites for gathering specific information, and so on, providing a major number of sites analysed.

• Once the news had, NS automatically classifies them, trying to understand the written text.

• Next step: correlation of the heterogeneous sources based on content, profiles entities (people or companies) on the basis of their behaviour (log analysis) or information, and finally georeferences unstructured information.

**RESULTS:** FIND NEW LEGISLATIONS, UNDERSTAND WHAT ARE THE DIFFERENCES WITH RESPECT TO YET EXISTING LAWS OR IF THEY INTRODUCE NEW ASPECTS OR CONSTRAINTS, BY EVALUATING WHAT COULD BE THE IMPACTS ON BOTH MANAGEMENT AND STRATEGIC ACTIVITIES.
If well trained, through the analysis of legislation (that are newly introduced or in the approval process), or of the changes in the requirements previously promulgated, NS could therefore:

• **Provide guidance** on the impacted bank processes

• Indicate with almost predictive feature what are the impacts on the IT applications (in terms of changes or new implementations), enabling to take strategic decisions in order to minimise potential costs in the banking system.