Chapter III
Automatic NLP for Competitive Intelligence

Christian Aranha
Pontifical Catholic University of Rio de Janeiro, Brazil

Emmanuel Passos
Pontifical Catholic University of Rio de Janeiro, Brazil

ABSTRACT
This chapter integrates elements from natural language processing, information retrieval, data mining and text mining to support competitive intelligence. It shows how text mining algorithms can attend to three important functionalities of CI: filtering, event alerts and search. Each of them can be mapped as a different pipeline of NLP tasks. The chapter goes in-depth in NLP techniques like spelling correction, stemming, augmenting, normalization, entity recognition, entity classification, acronyms and co-reference process. Each of them must be used in a specific moment to do a specific job. All these jobs will be integrated in a whole system. These will be ‘assembled’ in a manner specific to each application. The reader’s better understanding of the theories of NLP provided herein will result in a better ‘assembly’.

COMPETITIVE INTELLIGENCE
There has been a huge computer technology development and an accelerated growth in information quantity produced in the last two decades of the 20th Century. But how do companies use these published data, mainly in the digital media? What do they use to increase their competitive advantages? It is true that most companies recognize information as an asset and believe in its value for strategic planning. The big difficulty however is to deal with information in a changing environment. To plan on an annual basis seems not to be sufficient anymore. The dynamism of the environment is faster. The temporal character of information is becoming more and more critical. Information valid today may not be valid tomorrow anymore. Data are not static blocks to become a building block of a temporal reality. Information analysis is no longer an action, it has become a process.
If in the past there was an information tap that could be opened and closed, today there is a river of information whose flow never ends.

Leaders can go out of business in a matter of months and challengers can assume their thrones. So companies wanting to succeed need to adopt a continuous and systematic approach in order to stay ahead of competition and be prepared for upcomers, technological changes or economic turmoil. That is exactly the role that competitive intelligence (CI) promises to play. By definition, competitive intelligence is a continuous and systematic process of gathering information regarding the external environment, analyzing, and disseminating it to the top management. In thesis, the more systematic the approach, the more responsive the company will be to potential outside changes. What is the value of implementing a formalized CI process? It’s hard to quantify, but quantification is probably unnecessary. How much is it worth if your company launches a new product before your main competitor? What about anticipating a major technological change and acting ahead of the industry to take advantage?

Furthermore, facts such as the Enron and WorldCom scandals have made shareholders more apprehensive about administrating a company. That is why, nowadays, we can observe a disclosure trend, i.e., investors and market regulation agencies pressure companies, more and more, to operate with increased transparency. If the tendency is a bigger amount of published information, those who handle it better will be the most competitive, heightening the importance of the CI technology market.

Additionally, the volume of information also grows day by day, calling for more analysts to process all the available content. Meanwhile, CI departments tend to be too small when compared to the volume of information that needs to be processed. This increases the demand a technology that would allow for a reduced number of high quality analysts to handle the big volume of information present inside, in the company, and outside, in the environment.

In this chapter we will cover natural language processing (NLP) techniques for text mining (TM) in order to make information analysis more efficient.

**SUPPORTING FUNCTIONALITIES TO CI**

A CI analyst should basically be aware of information surrounding him/her in order to be able to evaluate its importance and to prepare reports for decision making. As most data are presented in the form of texts, we will consider, in this chapter, three text mining functionalities to help this process. We should have in mind that the central objective is to keep the analyst informed. Besides TM, we should add CI network management, discussion forum and business intelligence as other examples of supporting functionalities to CI.

The first supporting functionality to the analyst comes from the limited daily reading capacity of a human being. The Filtering functionality has been created with the purpose of allowing pre-selection of reading contents, assuming that the important information is, very likely, within the filtered subset. These data are read progressively by the analyst in order to extract relevant facts.

This reading process, however, can still be time consuming. When dealing with information, timing is a crucial factor. The technological Event Alert functionality was developed with the objective of advising the analyst as soon as possible of some pre-specified events as being important to his business.

The third and last functionality refers to a Semantic Search tool that becomes necessary for ad-hoc demanded information. This demand comes from the fact that both Filtering and Event Alert are planned and predefined. The objective of this tool is, therefore, to allow the analyst to reach the information required in a particular instance, as soon as possible. Finally, the intersection of three sets of information, resulting from three functionalities of text mining, minimizes the in-
Related Content

Sequential Patterns Postprocessing for Structural Relation Patterns Mining
www.igi-global.com/chapter/sequential-patterns-postprocessing-structural-relation/40406?camid=4v1a

ASCCN: Arbitrary Shaped Clustering Method with Compatible Nucleoids
www.igi-global.com/article/asccn-arbitrary-shaped-clustering-method/46940?camid=4v1a

A Survey of Extract–Transform–Load Technology
www.igi-global.com/article/survey-extract-transform-load-technology/3894?camid=4v1a

Finding Associations in Composite Data Sets: The CFARM Algorithm
M. Sulaiman Khan, Maybin Muyeba, Frans Coenen, David Reid and Hissam Tawfik (2013). Developments in Data Extraction, Management, and Analysis (pp. 162-186).
www.igi-global.com/chapter/finding-associations-composite-data-sets/70797?camid=4v1a