Chapter II
Creating Strategic Information for Organizations with Structured Text

Roberto Penteado
Brazilian Agricultural Research Corporation, Brazil

Eric Boutin
University du Sud Toulon-Var, France

ABSTRACT
The information overload demands that organizations set up new capabilities concerning the analysis of data and texts to create the necessary information. This chapter presents a bibliometrical approach for mining on structured text and data tuned to the French school of information science. These methodologies and techniques allow organizations to identify the valuable information that will generate better decisions, enabling and capacitating them to accomplish their mission and attain competitive advantages over the competition. The authors think that information treatment and analysis is the most critical organizational competence on our information society and that organizations and universities should take measures to develop this new field of research.

INTRODUCTION
There is a universal tendency for an “information overload.” Naisbitt (1996) best summarized this situation: “we are drowned in information but thirsty for knowledge.” Human knowledge more than duplicates each year. This quantitative evolution is due to the convergence of several phenomena: technological development, the emergence of the Web, and the widening of the economic actors action capacity.

Fayyad and Stolorz (1997, p.101) indicated that since “the rates of growth of datasets are exceeding by far any rates that traditional ‘manual’ analysis...
techniques can cope with,” the option available to organizations is to leave most of the data unused. Thus, they state that such a scenario is not realistic in any competitive environment where others who do utilize data resources better will gain a distinct advantage,” and that all kinds of organizations, “spanning from business, to science, to government” are concerned by this sort of pressure (Fayyad & Stolorz, p. 101).

Added to this quantitative increase in information available is a qualitative problem corresponding to the shift that exists between supply of information available and demand of information expressed by the economic actors.

“While available information is abundant, incomplete, doubtful, public and confused, necessary information is synthetic, complete, reliable, confidential and precise” (Boutin, 1999), the use of available information is limited. The more unrefined the information, the slower it takes to the decision maker to act on it. While making sense of raw information can take more than one day, the strategic information that is processed and validated could be understood and adapted for action by the decision maker in just one minute. Quoniam (2001) affirms that it should be “the right information, provided at the right time, in the right form, to the right person and should result in the right decision.” The usage determines the true value of the information. If it cannot be used, it has zero value. This shift between supply and demand in the market of information requires new capabilities, which ask for the development of analysis and mining of texts and data inside the organizations. In order to remain viable and robust, organizations should dominate methodologies and techniques with human and material resources, able to accomplish the operations demanded to create intelligence, in other words, the necessary information.

Levet (2001) affirms that “it is no longer access to information that provides the wheel for economic growth and employment but the actor’s aptitude to transform, understand, interpret and use information” (p. 38). This was recognized by the international community in the environmental management standards known as the ISO 14.000 family of international standards. They enforce a system of environmental administration. It requires, among other relevant environment aspects, to register and to analyze information on legislation, processes, products, sub-contractors, suppliers, management, prevention of crises and response capacity.

Dou (2003) recognizes that organizations need to anticipate on competitors’ actions. They need present or near future information and should use it as soon as possible. What is at stake is the organization capacity to analyze, in real time, the maximum possible amount of information and to establish a significant number of economic, financial, juridical, diplomatic, cultural, social, scientific, and political intelligences. To perceive, before the competitors, some week signs, indicating a business opportunity or a new technology, and to act in this sense can result in competitive advantages.

These methodologies and techniques to “create, associate, treat and use information to answer specific questions concerning current operations, research, product improvement or commercial conquest” (Dou, 2003) are the main objectives of this chapter. They refer to a well-defined process that has its productivity expanded by new data storage technologies that have emerged over the last decades, and does not spare data preparation and reformatting before analysis. First, we will provide a synthetic vision of the different stages of the information process that are becoming vital to every organization; then we will describe some ways on accomplishing the information analysis generating information ready for action. We make two restrictions on the many techniques available for data mining and text mining: first, we favor those methodologies that are tuned to the French School of Information Science and second, we use a bibliometrical approach for mining on structured text and data.