INTELLIGRATE
COMPETITIVE INTELLIGENCE AND DATA INTEGRATION

Text Mining in Science and Technology

Leveraging unstructured data to provide knowledge discovery and analysis

"Finding and analyzing relevant textual data could be a gigantic work: Text Mining can help you save time and money"

The information is there. Just find it!

Although information access has been widely simplified by the Internet era, managing **unstructured data** still remains very complex and time consuming in the science and technology industry.
Large online databases, patents, news flows, competitors' websites, financial reports, document spread all across the offices, etc. are holding critical information that can give your company a serious **competitive advantage**.
Text Mining can unlock those valuable contents by combining state-of-the-art linguistic and statistic technologies, both for Corporate and for Business Unit issues.

Background

Science and Technology (S&T) knowledge is now shared among several different and heterogeneous sources: technical literature, patents, Large online databases, news flows, competitors' websites, financial reports.

Rapid advancement of S&T depends on the efficiency of knowledge extraction from this sources, including both infrastructure (authors, journals, institutions) and thematic (technical thrusts, relationships) information. Relative to global S&T, questions of interest center around what S&T is being performed, who is performing the S&T, where is it being performed, and what messages and heretofore undiscovered information can be extracted from the global literature. The expert analysts can then judge what is not being done, and recommend what should be done differently.

In the past, the technical community used the thorough but inefficient approach of visually scanning printed and electronic technical literature to identify relevant documents, then reading the relevant documents (with no decision aids) to extract the information. Now, techniques have been developed to perform the **selection of relevant literature** semi-automatically, and to order the intrinsic technical concepts and their relationships to provide a framework for an **integrated analysis**. These techniques are encompassed under the umbrella of S&T **text mining**.

Contatti: Intelligrate srl
Via XII Ottobre, 2/92 16121 Genova
e-mail: info@intelligrate.it

INTELLIGRATE
COMPETITIVE INTELLIGENCE AND DATA INTEGRATION

"Text Mining help you to mine technical literature, patents, large online databases, news flows, competitors' websites, financial reports, find what you need and report the way you need it"

Text-mining can benefit S&T performers, managers, sponsors, administrators, evaluators, and oversight organizations. It can serve as a catalyst to enhance peer review, metrics, road-mapping, and other decision aids. It could allow comprehensive roadmaps for strategic planning to be constructed, and thereby serve as a foundation for international policy assessment. Text Mining can support workshops and S&T reviews by identifying the key performers in disciplines related to those being evaluated. It can identify productive sites to be visited in global S&T evaluations. It can identify new information groupings, to provide novel technical insights that could lead to discovery and innovation. In parallel, this could lead to promising new S&T opportunities, and new research directions.

Definition

Text mining should not be confused with the better known Internet search engine tools or database management capabilities. Analogous to data mining, which extracts useful information from any type of data with large quantities, text mining is a procedure applied to large volumes of free unstructured text. After a traditional search for documents is completed, such as in format of full text, abstracts, or indexed terms, text mining explores the complex relationship among documents.
Science & Technology (S&T) text mining is the application of text mining to highly detailed technical material. There are three major components of S&T text mining.

(1) **Information Retrieval**, the foundational step of text mining. It is the extraction of relevant records from the source technical literatures or text databases for further processing.
(2) **Information Processing**, the extraction of patterns from the retrieved data obtained in the previous step. It has three components: bibliometrics, computational linguistics and clustering techniques. This step typically provides ordering, classification and quantification to the formerly unstructured material.
(3) **Information Integration**. It is the combination of the information processing computer output with the human cognitive processes.

Contatti: Intelligrate srl
Via XII Ottobre, 2/92 16121 Genova
e-mail: info@intelligrate.it

INTELLIGRATE
COMPETITIVE INTELLIGENCE AND DATA INTEGRATION

S&T Text Mining Applications

There are several existing and potential text mining applications.

Retrieving Documents

Text mining can be used to improve the comprehensiveness and relevance of information retrieved from databases. Today, high quality methods use some type of iterative method with relevance feedback to modify the initial test query for increased comprehensiveness and precision of the records retrieved.

Identify Infrastructure

Text mining can be used to identify the elements of the infrastructure of a technical discipline. These infrastructure elements are the authors, journals, organizations and other group or facilities that contribute to the advancement and maintenance of the discipline. Additionally, text mining can provide their specific relationships to the total technical discipline or to sub-discipline areas.

Identify Technical Themes / Relationships

Text mining can be used to identify technical themes, their inter-relationships, their relationships with the infrastructure and technical taxonomies through computational linguistics. By categorizing phrases and counting frequencies, S&T text mining can estimate adequacies and deficiencies of S&T in sub-technology areas.

Discovery from Literature

There are different kinds of literature-based discovery: examining relationship between liked, overlapping literatures, and discovering relationships or promising opportunities that wouldn't be found when read separately. Successful performance of literature discoveries can lead to identification of promising new technology opportunities and research directions, such as extrapolation of ideas from on discipline to a disparately related discipline.

Technology Forecasting

In the process of retrieving and relating useful text data, text mining can also provide the time series for trend extrapolation. As an extension of the process, text mining can be used to identify state-of-the-art Research & Development (R&D) emphases and portend future development.

Contatti: Intelligrate srl
Via XII Ottobre, 2/92 16121 Genova
e-mail: info@intelligrate.it

INTELLIGRATE
COMPETITIVE INTELLIGENCE AND DATA INTEGRATION

"Analyzing competitors' information can be an easy task with the proper tools"

Text Mining Applications

Competitive Intelligence

Competitors strategies, product launches, lawsuits, product withdrawals, FDA approvals, thought leaders' concerns, etc. are critical information for your team. Text Mining solutions unlock this knowledge and help you to monitor your competitors' activities and your market environment.

Intellectual Property Management

Patent protection is central to the preservation of your R&D investments.
Text Mining provides you with patents and portfolios analysis to track your competitors strategies.

Innovation Indicators

As mentioned before, bibliometrics is one of the text mining techniques to capture desired information. It uses counts of publications, patents or citations to measure and interpret the advance of technologies. These counts can then be reasoned as innovation indicators of a certain technology.
Innovation indicators collect information on technology life cycle status, innovation contextual influence, and product market potential concepts.
These innovation indicators can be defined along with the identification of technical themes, where co-occurrence of clustering analysis is required. Therefore, combined with text mining technique, innovation indicators can be generated as aids to show the maturity level of technology, similar in concept to Technology Readiness Levels (TRLs).

Innovation Flow Mapping

Innovation flow mapping is a technique to model the influences on or drivers of technology development in a graphical manner. It can be used as a brainstorming tool in the early stage of planning or examining the prospects for a technology and whether the institutions and the organizational capability exist to complete the development. Similar to an *Interrelationship Diagraph*, an innovation flow map consists blocks of identified technologies/sources and cause/influence relationship arrows in between blocks. It can depict the location of research domain relative to each other and institutional interest and overlaps.

Technology Opportunities Analysis

Technology Opportunities Analysis (TOA) exploits electronic information resources to provide technology foresight. The TOA approach proceeds as follows:

Contatti: Intelligrate srl
Via XII Ottobre, 2/92 16121 Genova
e-mail: info@intelligrate.it

INTELLIGRATE
COMPETITIVE INTELLIGENCE AND DATA INTEGRATION

1. Search on a topic of interest in one or more databases (e.g., *Engineering Index, U.S. Patents)* and retrieve the resulting electronic records (typically publication, patent, or project abstracts)
2. Apply software programs to profile the content of those records and find relationships of interest
3. Represent the activity and relationships in informative ways
4. Combine this information with expert opinion to generate valuable technological intelligence.

Sales Support

Sales effectiveness evaluation requires qualitative indicators to be significant.
Through Text Mining you will be able to analyze free text contained in daily reports from your representatives.

Contatti: Intelligrate srl
Via XII Ottobre, 2/92 16121 Genova
e-mail: info@intelligrate.it

INTELLIGRATE
COMPETITIVE INTELLIGENCE AND DATA INTEGRATION

About Intelligrate

Intelligrate provides Competitive Intelligence consultancy services in several business areas, along with Intellectual Property management assessment.
Further, by leveraging the power of Data Mining and Text Mining technologies, Intelligrate is able to provide Data Integration solutions, to discover and analyze the knowledge both in structured and non-structured format (text).

Offices

Company Headquarter

Intelligrate srl
Via XII Ottobre, 2/92
16121 Genova

Branch Office

Via Marazzani, 9
20132 Milano
Tel. 02 36554259

Contatti: Intelligrate srl
Via XII Ottobre, 2/92 16121 Genova
e-mail: info@intelligrate.it