# WeCIM – Web Competitive Intelligence Methodology

Joaquim Fonseca and António Grilo

***Abstract*—Disruptive change causes value displacements in today's competitive economic environment. To enhance survival capabilities organizations are increasing efforts in more untraditional business value assets such the intellectual capital and intelligence. Dynamic capabilities theory states that companies have to develop adaptive capabilities to survive disruptive change and increase competitive advantage in incremental change phases. Taking advantage of the large amount of information in the Web we proposed a methodology to develop an application to gather, filter and analyze web data and turn it into usable intelligence (WeCIM). In order to enhance information search and management quality it is proposed to use ontologies that allow reasoning over knowledge concepts without human intervention. Two case studies were conducted with satisfactory results. Two software prototypes were developed according to the proposed methodology that validated the approach in two specific real-life situations.**

***Index Terms*—Competitive intelligence, ontologies, WeCIM, World Wide Web.**

## I. Introduction

Companies face an environment of increasing competitiveness. Many theories try to characterize organization strategy in their market and their ability to survive. It is accepted that changes in technology are a serious threat. The most acknowledged theory in strategizing has been proposed by Porter [1]. The author identified five factors from which companies should shape their strategy that would yield the best competitive advantage.

Why companies fail to survive is also a matter of great interest in today's academic research. Authors have classified big technology changes and general innovations has a disruptive change, capable of changing market equilibrium, displacing product value and contributing to companies rise and fall.

These two subjects collide in a way that good strategizing should prevent failure in a presence of a disruptive change. What should be the ideal organization's capabilities that prevent failure and yield great competitiveness? Studying companies' survival capabilities some authors' proposed different strategizing theories. One of the most recent and most complete theories considers the capability of a company to adapt to a new environment and create new sources of value. This adaptation capability is surely necessary when disruptive change occurs. Since the pace of changing is increasing [2], [3], those capabilities have to be dynamic and therefore, one company's survival capability is related to their dynamic capabilities [4]. The theory addresses why and how companies should adapt.

Aiming to study one of the many factors of why a company fails to adapt or even detect the occurrence of disruptive change that may determine that a strategic move is necessary, our work focus on data analysis in the biggest source of information, the Internet. We propose that companies could generate better insights on strategic position by "listening" to information available on web. This information seeking should be part of an active Competitive Intelligence and Knowledge Management strategy. Companies could better detect strong and weak signals that may support decision-making, allowing then to gain competitive advantage and survive in the competitive global economy.

Manuscript received September 14, 2012; revised January 23, 2013.

The authors are with UNIDEMI, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (e-mail: jpn.fonseca@campus.fct.unl.pt).

## II. Objectives

The present work has the objective to conceptualize a methodology for the development of decision making aiding tools based on web collected data. This framework focuses on Competitive Advantage and Knowledge Management subjects based on semantic search of information sources.

The challenges for this task can be enumerated as:

- To conceptualize a framework which can manage information gather from the web and store on an ontology based data set.
- Integrate tools responsible for simple tasks regarding extraction, managing, analyzing and presentation of data.
- Guarantee interoperability capability to ensure that multiple tools can operate over the same library of ontologies.

The value of the framework is to provide a superior information management with automated and semantic capabilities that enhances the organization ability of gathering and process contextualized information. These could play and determinant role to enriched dynamic capabilities and survival abilities of the company in the present of disruptive change. During incremental change the framework would enhance competitive capability maintaining automated set of information management processes.

## III. Literature Review

### A. Disruptive Change vs. Companies

In 1980 digital cameras appear in the market and Kodak had already invented their product. A tremendous change had to be made on Kodak's business orientation. Kodak didn't pursue its invention because it was necessary to shift his high profitable film industry to a new digital one. Kodak had to cannibalize their human resources and R&D structure from a chemistry based operation to electrical and physicist research

DOI: 10.7763/JOEBM.2013.V1.26

areas [5]. The internal restructuring Kodak made as a response to the emergent competition in digital photography lead Kodak's workforce to fall from 145.000 to 27.000 employees between the 90's and 2007.

This subject is well known and many researches pointed different reasons for Kodak's failure. Regardless what was the key reason or bundle of factors for Kodak's downsizing and profit lost the fundamental cause is clear: a disruptive change.

Although changing factors can be very industry specific a broad definition of disruptive change can be export from the Christense [6] view on disruptive technology as the appearance of products with a very different value proposition which often are *cheaper*, *simpler*, *smaller* and *more convenient* to use.

Rohrbeck [7], on the other hand, summarized why companies had difficulties in their ability to adapt to changes in the environment. There were three main categories for failure:

- High rate of speed
- Ignorance
- Inertia

Disruptive changes lead to business failure because they are not able to catch up with innovation speed, don't capture signals of an imminent change or are enable to adapt due to organizational complexity. To change the company's ability to adapt, top level decisions have to be made, in other words, strategizing. Organizations need to focus in their expertise and develop difficult to imitate competencies.

### B. Dynamic Capabilities for a Successful Strategy

The most successful framework in the last decades to shape the organizations strategy and achieve competitive advantage was Porter's five competitive forces [1]. The framework stated that companies should analyze five different forces that shaped the industry and defined how they could integrate and gain competitive advantage.

Teece et al [4] discussed other strategizing theories to conclude that they all lack competitive factors that shape successful organizations. Strategizing according to competitive forces and industry specific assets fail to pursue and create new sources of value. They propose Dynamic Capabilities as the:

> *"(…) role of strategic management in appropriately adapting, integrating, and reconfiguring internal and external organizational skills, resources, and functional competences to match the requirements of a changing environment."*

The main point in dynamic capabilities is that companies should develop a set of assets combined with knowledge and routines creation to produce a difficult to replicate or imitate value proposition.

The linkage between dynamic capabilities and knowledge management is demonstrated by Eisenhardt and Martin [8]. Their view of dynamic capabilities reinforced the idea that the concept is based on well-known processes, product development and strategic decision making. This dynamic capabilities and knowledge management relation is even more evident in Cepeda and Vera [9] view. They claim that with knowledge-based transformation processes, managers are able to create, join and codify knowledge configuration which allow new improvements in the organizations routines.

### C. Information in Dynamic Capabilities

Dynamic capabilities have a core competency that should drive management practices, but base on what? How can managers and decision makers base their policies? How can they even assess the need for change? What feeds their decision actions in order to achieve procedure changes? Strategic foresight answers those questions in the management perspective: where to look, what to retain, how deeply, how often, but what is the pure raw material?

Three decades ago Kantrow [10] mentioned the need to incorporate the technology information discipline within the strategic decision making. Strategy and technology were unsociable. Just a couple years later, Porter and Millar [11] explained how information could yield a competitive advantage to any organization with IT adoption and how game changing this technology would be.

Information can be a characterized by the way company uses it. Rohrbeck [7] described as the elements of information usage and it plays important roles in his propose for a strategic foresight framework:

- Reach
- Scope
- Time Horizon
- Sources

It is important that any improvement in intelligence have concerns over the four information elements.

### D. Competitive Intelligence for Competitive Advantage

Chawner [12] exported the Society of Competitive Intelligence Professionals to define CI as the

> *"the process of ethically colleting, analyzing and disseminating, accurate, relevant, specific, timely , foresighted and actionable intelligence regarding the implications of the business environment, competitors, and the organization itself."*

Authors had the necessity to clearly distinguish information from intelligence. Rouach and Santi [13] review that information is factual data and needs to undergo a process of filtering and treatment to become usable intelligence. The authors also classified a process to accomplish accurate CI. It starts by planning and defining a direction for a course of action. Collection of information and the correspond analysis to turn it to intelligence. Dissemination is the end activity where intelligence should reach the appropriate management level.

## IV. Web Competitive Intelligence Methodology

The basic objective is the creation of a simple methodology to develop, implement and manage Competitive Intelligence tools based on information collected from the web.

The tools consist in the interaction of four elements organized in a three-component framework illustrated in Fig. 1. These concepts are correspondent with Lixto classification

of data course: *sources*, *integration*, *storage* and *usage* [14]. The crawler component deals with information *sources*; information is *integrated* by wrappers in an ontology based *store* system from where data analysis tools *use* the collected information.

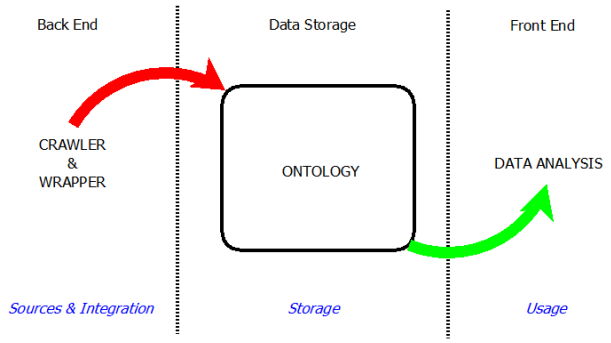


Fig. 1. Software elements with data classification.

Phases of development follow the four component of the framework.

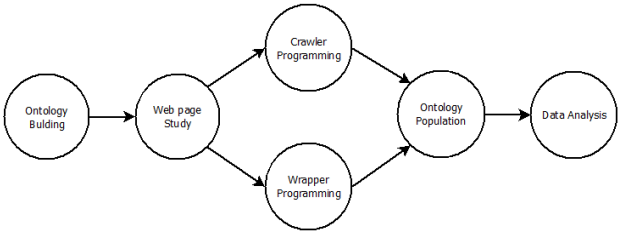


Fig. 2. WeCIM's tasks precedencies.

### A. Ontology Building

The ontology building methodology states that it should start with the requirements definition. They inherit from the competitive objectives that the software proposes to accomplish. Simple and focused objectives need simple ontologies schemas. Simple schemas are favorable to integrity control.

Some typical objectives can be proposed provided that they are published on the web:

- Competitors products price monitoring.
- Raw materials price in stock markets.
- Public calls from international organizations.
- Transportations and other services strike warnings.
- Business related news.
- Exchange rates
- Competitors stock value
- Business related scientific papers.

With the requirements set the ontology definition should start. In this matter an iterative procedure is done until the final ontology is approved. The final result as to be capable of:

- Represent the knowledge field of the target information.
- Allow all data analysis proposed to gained competitive advantage.

### B. Web Page Study

The web page study serves two proposes. How the crawler algorithm is going to be program to access the target web page where information or documents are being hold, and how the information in the HTML or documents is organized for the wrapping procedures.

The first step is to understand where the target information is kept on the web page and how it has to be accessed. Information can be hold in various formats, can be structure or unstructured. Some sites already provide semi-structure formats like CSV or XL documents or even complete structure formats life XML. Information has to be collected in order to program the algorithm. Examples of pertinent information are the URL structure, key strings to identify second level URL in the HTML document, variables that control the algorithm behavior, textbox and buttons names for automated procedures and other information to control errors and event exceptions.

In order to correctly program wrapper function a different set of information is needed to identify all key elements that guarantee the correct collection of data and the text structure in order to clean unnecessary substrings. The crawler output may be unstructured or structure file formats. Some file formats may require converting to readable formats. Different procedures have to be idealized according to file converting and reading requirements.

### C. Crawler Programming

We defined crawling procedures as the operation of interpreting the target web site URL and HTML structure to correctly access the target web page or documents from where information is to be collected.

The crawling structure may also depend on the periodicity of information publishing. In the first case study presented, the information was gathered in a daily basis and the URL was structured accordingly. The output of the crawling activity would be a list of URLs that represent either web pages or documents to be downloaded, such as PDF's, containing desired information. These URLs can be saved in a file for future use or saved internally in the program.

Some concerns while programing crawler features are: accessing the first URL, retrieve second level URLs, check string encoding, adding the root element, undergo log-in search and log-in procedures, and navigate through multiple pages.

### D. Wrapper Programing

Wrapper programing performs the key element of collecting strings containing the target data. Those strings may represent entity names or values. Three possible scenarios may occur depending on the crawler output, its formats and the methods chosen to reach information.

- Extract data from file formats that can be directly read with no necessary conversion such as TXT, HTML, XML, CSV and other Excel formats.
- Extract after converting formats into readable files using specific add-ins or methods libraries. Complete Excel or Word formats have to be managed to extract the text content. XML or TXT are possible outputs of these file conversions.
- A third case can take place. Structure formats can be queried such as XML.

Therefore the main concerns in wrapping programing are: determine is conversion methods are necessary, plan extraction according to structure or unstructured formats and

perform consistency checks to determine the output integrity.

### E. Ontology Population

Ontology population defines how gathered information will be introduced in the ontology file. Automatic procedures should be capable of dealing with the wrapper output and compile it in the ontology.

Our ontology file has a RDF/XML format. During development some problems arise when combining schema file with files containing individuals. The solution approach was a file with classes and relations (schema) that is kept separated from files with information. Another reason to do so is information size. To mitigate processing time problems, multiple files can be created according to time parameters.

### F. Data Analysis

The key element of data analysis is to plan relevant data configurations that represent a support for decision making. According to the CI objectives planned in the first step, data analysis can be program. They should be design in a way that answer strategic questions and support market and competitors characterization.

Data analysis uses queried information from the ontology file and further processing. Data is collected according to different parameters or logical relations in order to capture pertinent intelligence.

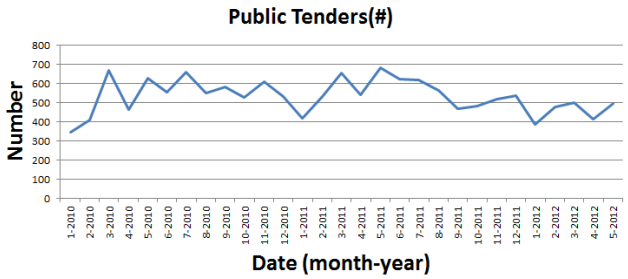


Fig. 3. Number of PT from 2010 to Mat 2012.

A user interface can be design not only to visualize analysis results but also to search and query specific information. Filters should be created to limit the scope and amount of data. Often, ontology classes are a clear suggestion to relevant filters.

In our case studies, however, we only exported information to an excel spreadsheet to demonstrate possible analysis. This proof of concept was necessary since time constrains wouldn't allow full development of a user interface and correspondent analysis procedures.

## V. Case Studies

### A. Open Public Tenders from Diário da República

The first case study arises from the need for companies to assess public tenders (PT) they may be interested applying to. In the business competitive intelligence point of view we want to prove that our software could automatically gather web related information and characterize the related business market to support top level decision making.

The target page was "dre.pt" the official Portuguese Government Publication. The developed software has a crawler and wrapper capabilities, and stored the information in a specifically created ontology and data analysis where made based on 2010, 2011 and 2012 data.

The crawler was able to know the last successful run. It repeats that run and continues to the next day until no more information is found. This prevented that information could be missed if late documents were published.

Wrapping procedures are triggered in between successful crawling activities. The information is gathered, consistency checks are made to ensure no error occurred and each PT's information is inserted in the ontology.

To gather past information, the software was forced to use variables corresponding to data from 2010, 2011, to May of 2012. After past data collection, the programed is capable of autonomously collect up to date information.

From the populated ontology, information from 2010, 2011 and 2012 was massively gathered and data analyses were made. 15432 valid instances were collected.

The results were very interesting since a high level of efficiency was obtaining will minimal fails. For consistency reasons, the software creates reports over all information collection and missed data errors for a continuous improvement of the algorithm's.

### B. Closed Public Contracts from Base

The second case study relates to the first one but an independent approach was intended. In this case we gathered information about all public contracts, including public tenders. Again the software must have to retrieve all related information in an autonomous manner. Public tenders from private companies are outside reach because only public institutions have to present sealed contracts information.

The target web page is a Portuguese governmental site called "base.gov.pt". The crawler capabilities were able to collect daily data from sealed public contracts. The wrapper features, made sense of the output document to retrieve data and store in a designed ontology.

Similar procedures were adapted from the first case study that showed that a base structure can be designed for this kind of software. Therefore not only a methodology can be proposed but a solid scheme foundation could be created.

To gather past information, the software allows the introduction of an initial and final date inputs from which it collects all data in-between. Using those input fields, data from January 2012 to July 2012 was collected.

Due to a bigger amount of information, a different ontology approach was developed. Instead of having one central ontology file, each month has a correspondent file. All ontology files share the same base schema.

The results allow us to study public organizations buying behavior. Critical business areas, most active organizations, and spending values can be characterized. Competitor's contracts can be tracked to gain competitive intelligence that allows a comparison between the organization and its competitor's market share.

## VI. Conclusions and Future Work

In current days, gaining competitive advantage depends much on the ability to filter and be able to rapidly analyze

available information on the web. The case studies have preliminary validated the applicability of the semantic web–based competitive intelligence approach, though much research work still needs to be developed.

An interesting future work will be to analyze the use and reuse of already formulated ontologies to accelerate the process of software development and enrich capabilities. The global acceptance of the Semantic Web practices could also improve the development of such applications.

It is also of great importance the development of a high level ontology that could represent the enterprise knowledge field. That ontology should be integrated in a base framework upon small application modules could be implemented. This complete approach sustains not only competitive intelligence objectives in different sectors of organizations but to also internal processes according to business models and strategic premises.

## VII. Conclusion

The development of the software prototypes showed it is possible two create simple, efficient and reliable applications responsible for gathering, filtering and store web information for data analysis with competitive intelligence value.

Both cases demonstrated a high level of effectiveness with minimal errors, most associated with problems of information structuring from the web site side. Although no user interface was developed due to time constrains, data analysis are viable and reflect the main value component if they are capable of transforming information into intelligence.

The development of the second case study showed a high level of similarity between software architecture suggesting not only that the methodology is solid and replicable but also that a common software structure can be proposed. For instance, crawler and wrapper are highly integrated; common phases of information treatment and document handling are present.

Regarding the value of the two applications we think we demonstrated two value components. The first is the competitive advantage that one organization can withdraw from data analysis and how it can support strategic decision making. The second has functional value in internal procedures if the application itself has capabilities already developed in the organization process. Information search by human work can be automatically processed by the application and workforce can focus on other tasks.

The application helps achieving superior competitive advantage if it is capable of support strategic decision that enhances survival capabilities. All four information gaps suggested by Rohrbeck [7] are somehow filled with the development of such applications. They enhance *reach* capability since they can be responsible of gathering information from otherwise difficult to intensively search web sites. *Scope* if they intend to widen the business related information. *Time Horizon* element is present by the storage of all gathered data allowing tendencies to by study. Multiple configurations of software modules can decrease the *Source* gap. Humanly impossible information search can now be easily achieved with the right deployment and with pertinent intelligence objectives.

For all the presented work we considered of great value the presented methodology.

## Acknowledgment

The authors would like to thank Miguel Soares of Vortal for the pertinent suggestions to develop the case studies.

Marco Delgado and Sudeep Ghimire helped us to address the technological dimensions of the research and we would like to express our gratitude to both.

## References

[1] M. E. Porter, "How competitive forces shape strategy," *Harvard Business Review 57*, vol. 2, pp. 149-160, 1985.
[2] E. H. Kessler and A. K. Chakrabarti, "Innovation speed: a conceptual model of context, antecedents, and outcomes," *Academy of Management Review 21*, vol. 4, pp. 1143-1191, October 1, 1996.
[3] A. Sood and G. J. Tellis, "Technological evolution and radical innovation," *Journal of Marketing*, vol. 69, pp. 152-168, July 2005.
[4] D. J. Teece, G. Pisano and A. Shuem, "Dynamic Capabilities and Strategic Management," *Strategic Management Journal 18*, vol. 7, pp. 509-533, August 1997.
[5] C. H. Deutsch. (May 2008). At kodak, some old things are new again. *The New York Times*. [Online]. Available: http://www.nytimes.com/2008/05/02/technology/02kodak.html
[6] C. M. Christense, *The Innovator'S Dilemma: When New Technologies cause Great Firms to Fail*, 1st ed. Harvard College, U.S.A, 1997, ch.1 pp. 30-61.
[7] R. Rohrbeck, *Corporate Foresight: Towards a Maturity Model for the Future Orientation of a Firm*, 1st ed., Springer, 2010, ch.1, pp. 10-16.
[8] K. M. Eisenhardt and J. A. Martin, "Dynamic Capabilities: What Are They?" *Strategic Management Journal*, vol. 21, pp. 1105-1121, October 2000.
[9] G. Cepeda and D. Vera, "Dynamic capabilities and operational capabilities: a knowledge management perspective," *Journal of Business Research* 60, vol. 5, pp. 426-437, 2007.
[10] A. M. Kantrow, "The strategy-technology connection," *Harvard Business Review 58*, vol. 4, pp. 6-21, 1980.
[11] M. E. Porter and V. E. Miller. "How information gives you competitive advantage." *Harvard Business Review 63*, vol. 4, pp. 149-160, 1985.
[12] B. Chawner, "Millennium intelligence: Understanding and conducting competitive intelligence in the digital age," *Online Information Review*, vol. 25, pp.131-141, 2001.
[13] Rouach, Daniel, and Patrice Santi. "Competitive intelligence adds value: five intelligence attitudes," *European Management Journal 19*, vol. 5, pp. 552–-59, October 2001.
[14] R. Baumgartner, O. Fröhlich, G. Gottlob, P. Harz, M. Herzog, P. Lehmann, and T. Wien, "Web data extraction for business intelligence: The Lixto approach," in *Proc. BTW*, 2005, pp. 48–65.

**Joaquim Fonseca** is a PhD student. Recently finished his Master Degree in Management and Industrial Engineering, conducted research in the research lab at the Mechanical and Industrial Engineering Department (UNIDEMI).

**António Grilo** holds a PhD degree in Industrial Management by the University of Salford, UK. He is Auxiliary Professor of Industrial Engineering and Management at the Faculdade Ciências e Tecnologia da Universidade Nova de Lisboa, in doctoral, master and undergraduate degrees. He is also a member of the board of director of the research centre UNIDEMI. He has over 50 papers published in international conferences and scientific journals, and he is an expert for the European Commission DG-INFSO. Besides academia, he has been working in the last 10 years as a management information systems consultant, particularly in e-business, e-commerce and project management information systems.