1 The emergence of neo-renaissance paradigm

Looking to next years for post-industrial, ‘techno-elite’ cultures (i.e. US, European Union, Japan, Australia), the primary societal change driver is expected to be a paradigm shift from an organization-centric society revolving around an external workplace to an individual-centric society centered around the home as well as the local and virtual communities.

An expected paradigm shift from an organization-driven societal architecture (i.e. communities and social activities focused around large companies or organizations) to an individual-centric societal architecture (increased choice and availability of opportunities focused around individual wants and desires) is referred to as a neo-renaissance [1]. A neo-renaissance lifestyle is a home, local community, and virtual-community-centric societal architecture. This new lifestyle is due to a growing workforce composed of digitally connected free-agent workers able to work from any location. Boundaries between work and personal life will become increasingly blurred [2]. The neo-renaissance, decentralized business culture is dominated by a virtual workforce that can be engaged anywhere on the globe, and business will be conducted from any location by a virtual workforce of free agents. Focus shifts away from the corporate site as the primary workplace to the virtual workspace of the virtual office, collaboration centers, and virtual communities.

Due to this growing of virtual communities, a strong interest towards the automatic evaluation of intellectual and human capital as expressed in virtual communities is also growing [3].
2 Intellectual and human capital

2.1 The real wealth

Margaret Blair of the Brookings Institution has calculated the relationship between tangible assets (property, plant, and equipment) and total market value for every US manufacturing and mining company in the Compustat database. In 1982, she found, those assets accounted for 62.3 percent of the companies’ market value; ten years later, they made up only 37.9 percent of the value of the whole [4]. This was the proof that the hard assets of a knowledge company contribute far less to the value of its ultimate product (or service) than the intangible assets – the talents of its people, the efficacy of its management systems, the character of its relationships to its customers – that together are its intellectual capital.

2.2 Intellectual capital taxonomy

A structured way of presenting Intellectual Capital is the following one, proposed by Gartner Group [5].

*Intellectual capital* is composed of tangible and intangible resources, including human capital, intellectual assets and intellectual property. *Human capital*, the expertise and know-how of a company employees, the relationships with a company’s customers and prospects, their opinion and feelings are the primary form of Intellectual Capital.

![Figure 1: Example of a Human Capital Index dashboard, based on W. Wyatt theory [17] and text mining analysis of textual interactions (email, chatlines, instant messages) as represented inside a large company (Conoco, Inc.).](image-url)
*Intellectual assets* include databases of customer information, installed applications, engineering drawings, business documents, business processes and other work products.

*Intellectual property* is intellectual capital realized as a product or service, process or patent, for which the enterprise has sought legal protection and that can be bought, sold or stolen.

Text mining technology has already proved to be the correct way to analyze intellectual property [6] and intellectual assets [7]. Here we address the human capital aspect of intellectual capital. And, particularly, the opinions and feelings as expressed in virtual communities.

Where to look to find human capital? In customer and employee communities and, generally, in all the places where it is possible to recognize a social context. Here we analyse human capital as it appears in the so-called ‘virtual communities’

3 Virtual communities: where text mining is applied

A virtual community is a community of people sharing and communicating common interests, ideas, and feelings over the Internet or other collaborative networks. The possible inventor of this term was Howard Rheingold, who defines virtual communities as social aggregations that emerge from the Internet when enough people carry on public discussions long enough and with sufficient human feeling to form webs of personal relationships in cyberspace [8].

Full-function technology for virtual communities is comprised of four functional segments, and all of them profit highly of a text mining capacity to extract human capital:

- Community structuring
- Participant interaction and access
- Content management
- Community leveraging

3.1 Community structuring

Structuring includes capturing and responding to participant feedback and monitoring activities. It will be critical to define the rules for monitoring. If it is acceptable in a customer community, in an employee community the use of a moderator can be a double-edged sword: management does not want malcontents to openly degrade the company or its management team, but a moderator who ensures that only ‘safe’ opinions or comments are posted will prevent community postings from truly reflecting the community members feelings [9].

Text mining, with its capacity of analyzing the language and of extracting feelings allows to eliminate the moderator role and to alert the company if offensive or defamatory material appears on the community sites.
3.2 Participant interaction and access

Participant interaction needs are broad and diverse. Most community members need to interact with a single individual in a one-to-one conversation or participate and collaborate in idea development via threaded conversations with multiple people or groups.

Technologies for participant interaction and access include: chat rooms, conferencing, e-mail, instant messaging, portals (provide some functionality for access, interaction and content management) whose content is easily extracted through text mining [10].

3.3 Content management

Content management includes capturing community conversations and leveraging its intellectual capital. Community activity is about sharing information and knowledge, thus effective management of community content is critical. Content management technologies include those for organizing content or members (i.e. profiling them) by interest area, building ontologies, managing repositories for storage and retrieval, and managing subscription services: all these technologies are strongly boosted up by text mining technology. In fact content can refer to any digitized information – structured data files, unstructured text files, formatted documents, messages, metadata about video or audio files and so on and these media contain at least a textual part.

The language of the community must be understood, and terms or synonyms relevant to the community should be incorporated thanks to text mining and its capacity of categorizing using the hands-on training of an ontology or library sciences expert.

Text mining technologies for content management include: extraction, to allow members profiling; clustering, to build dynamic taxonomies of content; ontologies, to understand the community language; automated categorization, to allow organization of content.

3.4 Community leveraging

These technologies and processes enable the community analyst to discover and exploit the knowledge and information embedded in the community interactions (e.g. to detect customer feelings, as satisfaction, in the email or opinion surveys sent to a company). Leveraging activities comprehend mining e-mail and other text to discover unstated needs and behavioral trends. The knowledge and feedback obtained from a robust, active community is invaluable to the analyst and should be viewed as part of its overall intelligence and decision-making capabilities. For customer communities, this knowledge could affect support, pricing, channels or even product design [11]. For employee communities, it may affect working conditions, compensation and employee morale programs [10]. In generic social contexts, it may boost up the investigator activity in search of criminals and/or terrorists [12].
Privacy and trust in the community are critical: before entering the community, members must know how their discussions and interactions will be used, whether they will be monitored, and any potential negative ramifications of participating.

Engineered communities can be designed for mining knowledge, but organic communities are more difficult seen that members of an organic community are more likely to revolt against monitoring [9], anyway anonymity may be preserved.

4 Human capital in customer communities

The principal business drivers of this interest for language management capability to better understand customers are: (1) the rapidly increasing interest in customer behavior and satisfaction expressed through CRM (Customer Relationship Management), e-commerce markets, specialized portals, and intranets; (2) the focus on rich textual data flood (e.g. e-mail messages, memos, reviews, web pages, chat lines, public forums, and mailing lists), no longer seen as a problem but as an asset; and (3) the focus on the market, with the need of discovering new business opportunities and potential competitors strategies.

Customer capital is the value of an organization’s relationships with the people with whom it does business.

CRM (Customer Relationship Management) systems are getting more and more interested in extraction of customers and prospects feelings, opinions, tastes, cultural behaviors of the individuals [13]. In fact those feelings help in forecasting their future behaviours much more than studying old behaviour trends.

For CRM systems it is more and more important to extract from customer interactions relevant information, as satisfaction, unsatisfaction and complex, contradictory opinions, and automatically structure them according to customized semantic models and dictionaries enabling users to define precisely about which facts customer express themselves and which sentiments, judgments and needs they express [14].

Semantic models are defined according to the user requirements, but allow to extract the principal topics of interest to the user. These semantic models are contained in specific ontologies, engineering artifacts which contain a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words. An example of them, regarding Temis Skill Cartridges, is discussed in [11].

The most common example of data flood regarding customers is their e-mail that arrives daily to customer service office of a corporation. Text mining enables end users to order and classify them before reading them.

5 Human capital in employee community

5.1 An example of human capital: Employee attitudes

Human capital matters because it is the source of innovation and renewal, whether from brainstorming in a lab or new leads in a sales rep’s little black book
Studies consistently show a correlation between high morale and superior financial performance [16]. The mere fact that financial excellence cohabits with shiny, happy people does not establish causality, of course. But we know that people who feel as if they are learning, needed, and useful will be more productive than people who are idle and uncertain of their role in the company’s success; they are also likely to treat suppliers, customers, and each other better. Studies show a strong relationship between employee attitudes and customer attitudes – evidence of the interplay between human and customer capital.

But how to measure employee attitude? How to use an ERM (Employee Relationship Management) system? A new area of interest for text mining applications is just Human Resources, where it is used to match employees’ and candidates’ resumes and enterprise needs, as well as to measure their skills and motivations.

5.2 Vital signs monitor

Conoco, Inc. developed a Vital Signs Monitor, a tool for monitoring its intangible assets as the corporate human energy level and health. The concept is to create probes to assimilate data about various elements in an organization such as corporate culture, people practices, workforce motivation, level of innovation,
knowledge sharing, customer satisfaction and business results. The first premise of the VSM concept is that the total of these probes provides a more complete picture of organizational health, and therefore enhances the ability to make better decisions on unleashing the energy of the organization. The second one is that company vital signs, expressed in textual interactions (e-mails, opinion surveys, management declarations, internal chat lines, etc.), may be extracted and measured through text mining techniques.

5.3 VSM key concepts definition

The management theoretical framework which provided the guidelines to define the Vital Signs as appear in textual interactions to allow text mining technology to extract them was based on Individualized Corporation theory, developed by Sumantra Ghoshal [17]. This theory plans a transformation from a structure and strategy oriented leadership model to a new model encompassing purpose, process and people.

This theory, for example, establishes that the ability and willingness of people to take initiative is rooted in the tension between stretch and discipline: the former serving as the source of energy and the latter converting that energy into tangible and time bound action. Stretch without discipline leads to daydreaming, while discipline without stretch locks the company into an ever narrowing spiral of refining existing operations without the courage to make a creative leap.

Similarly, it is the combination of trust and support that motivates cooperation and collaboration. Trust makes cooperation desirable; support enables individuals to convert that desire into action. Each is a necessary element of the organizational glue, but only in combination do they create the sufficient condition for integrating the disparate actions of dispersed people.

Utilizing this theory, and the text mining technology to extract the previous key concepts, Conoco detected and measured the employee morale and management style, as appeared in Conoco textual interactions (i.e. email, chat lines, newsgroups and forums) [10].

Figure 3: Conoco passed from an ‘old’ to a ‘new’ management model, both of them defined following Ghoshal’s theory, by a set of different concepts.
6 Human capital in social contexts

After the description of human capital from the point of view of the extraction of its contents (opinions, feelings, know how) we now present it as it appears in its relational structures (identification of the author of an action, his personal characteristics, his social relations).

6.1 Defining anonymous terrorist authorship

Frequently the only available traces after a terrorist action are letters or emails which revendicate the action. In such cases not even the paper is available to be utilized in the judiciary police investigations but only the electronic text. On the contrary, the style by which these documents have been written, the concepts that appear and the links among them, and their writing structure, are the only useful traces.

These documents may be compared automatically to the typical production of, for instance, thousands of possible suspects to detect the author or, at least, its socio cultural origin. Famous attackers (Unabomber was the most famous one) were precisely described, before being really detected, using this type of analysis [18].

Theodore Kaczynsi was identified as Unabomber after his manifesto, composed of 35,000 words was compared with his writings, provided to the police by his brother David.

6.2 Digital signatures

Human beings are habit beings and have some personal characteristics that are inclined to persist. Some of them, as associations between words, their frequencies, the richness of the vocabulary, the average length of the words and of the sentences, percentages of foreign words, will remain constant. Totally more than 1000 ‘style markers’ have been quoted in literature [19–20].

Thanks to analysis based on these style markers and other elements text mining technology may help in answering questions as:

1. is the author a man or a woman?
2. young or old?
3. cultivated or not?
4. unique or multiple? i.e., are there sudden style changes inside the same text, that may be explained by a change of the author?
5. the document is an original one or has been copied?
6. are there unexpected analogies among documents (or at least parts of them), written in different periods by persons expected to be different but that, in reality, are the same one?
7. the document is homogeneous or not?
8. the total document has been written in the same moment or in a different time?
9. part of the document has been written quicker than another one?
10. and, most important question of all the previous ones, the author has a writing style which is recognizable and that may help us in identifying him?

6.3 Lobby detection

Text mining helps in grouping, automatically, and in function of the opinions there expressed, the writings or the declarations about a series of topics. This allows the recognition of unexpected associations (‘lobbies’) of authors (as journalists, interest groups, newspapers, media groups, politicians) detecting whom, among them, is practically forming an alliance.

If these alliances are detected more often than we expect from a statistical point of view, we may suppose possible collusions.

6.5 Monitoring of specific areas/sectors

One of the tasks of the intelligence services consists in monitoring various documentation to detect interesting concepts and/or topics (economical, political, social ones) once we have defined our interests.

Text mining technology detects these concepts [21] inside large volumes of documents coming from heterogeneous sources, and outline, count, summarize them, recognizing, for example, how many documents, out of 10,000, forecast a ‘political crisis in Brazil’, ‘a worsening of the world economy’, ‘a return of Islamic terrorism in Russia’...

Text mining technology may also help investigators to analyze social communities to extract member feelings and opinions, as shown in the paragraphs dedicated to customer and employee communities

6.6 Chatlines and other open sources analysis

May we detect the authors of some popular theories/ideas/concepts to detect if the current common opinions were originated by only one person or a group? Are appearing theories/ideas/concepts which will produce risks for the national security? Dr.X in his declarations appears as changing his opinion? The opinions of Dr.X are approaching those of Mr.Y?

The first enemy of intelligence activity is the ‘avalanche’ of information that daily the analysts must retrieve, read, filter and summarize. The Al Qaeda terrorists interact through chatlines to avoid being intercepted [22], without knowing that interception and analysis of chatlines content is possible and frequently done in commercial situations [10].

Using different text mining techniques it is possible to identify the context of the communication and the relationships among documents. This approach not only allows to capture the opinions contained in the data but allows to monitor
trends and important patterns inside the information process, retrieving and analyzing sources as different as:

- Databases, CD Rom
- Websites
- News agencies, Press reports
- E-mails, Chat rooms, Forums, Newsgroups
- Sms, Transcribed calls, Instant messages
- ………………………..

detecting the references to the interesting topics.

7 Social network links detection

7.1 Social structure

‘Social structure’ has long been an important concept in sociology. Network analysis is a recent set of methods for the systematic study of social structure and offers a new standpoint from which to judge social structures [23].

Text mining is giving an important help in detection of social network hidden inside large volumes of text. Thanks to it we can easily extract concepts such as names, roles, organizations, departments from large volumes of messages, emails, documents, trials/processes, texts, and also the links among them (not only in the same document but also outside it).

Text mining may also detect the simultaneous appearance of names, events and concepts. For instance it can detect all of the times which the name of a person (or an event or a concept) appears, also the name of another person (or an event or a concept) appears.

The investigator is now able to detect the existence of hidden social structures otherwise impossible to be detected inside, for example, the proceedings of a long criminal trial. It is easy to infer a relation between two companies which share the same person on their board, also if the information is not in the same document and the available documents number tens of thousands.

7.2 Graphical representation of connections

Temis developed a graphical representation of the network as it appears in text, to put in evidence concepts and their links, detectable through its Insight Discoverer Extractor (IDE). These concepts are extracted by IDE through Skills Cartridges (which contain the specific language used in the text). They may regard the social and economic language: ‘Mr. Baranov is on the Board of Gazprom and Yukos’. Thanks to this representation we quickly detect who is the link between Gazprom and Yukos. But these concepts may be extracted through other cartridges regarding, for example, human relationships language as well: ‘Mr. Baranov doesn’t like the President of IBM but loves Ms. Linda’ recognizing the type of sentiment.
This construction is a development of a very well known approach to language analysis, the so-called ‘semantic net’ approach. Semantic net idea was developed firstly by Ross Quillian in 1968 [24] to connect objects, concepts (as meaning of words) or events through their relationships and was used to represent the knowledge base of an expert system or of a rules system (for example semantic nets are useful to detect mistakes in the use of a language).

The Skills Cartridge approach used by Temis overcomes some limits of semantic nets as the lack of expressiveness (negative, disjunctions, quantification) and the rigidity of the meaning of a concept, extending their applicability to other, more complex concepts.

Thanks to skills cartridge flexibility, concepts, feelings and their complex relationships in all the specific languages (e.g. military, marketing, scientific ones) can be detected and visualized.

References