Expert Systems with Applications 41 (2014) 1315–1325

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Knowledge-based extraction of intellectual capital-related information from unstructured data

Eric Tsui *, W.M. Wang, Linlin Cai, C.F. Cheung, W.B. Lee

Knowledge Management and Innovation Research Centre, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Intellectual capital management
Unstructured information management
Information extraction
Knowledge-based systems
Case-based reasoning

ABSTRACT

Nowadays, there is an increasing demand for the identification of an organization's intellectual capital (IC) for decision support and providing important managerial insights in knowledge-intensive industries. In traditional approaches, identification of an organization's IC is usually done manually through interviews, surveys, workshops, etc. These methods are labor and time intensive and the quality of the results is highly dependent on, among other things, the experience of the investigators. This paper presents a Knowledge-based Intellectual Capital Extraction (KBICE) algorithm which incorporates the technologies of computational linguistics and artificial intelligence (AI) for automatic processing of unstructured data and extraction of important IC-related information. The performance of KBICE was assessed through a series of experiments conducted by using publicly available financial reports from the banking industry as the testing batch and encouraging results have been obtained. The results showed that, through the use of hybrid intelligent matching strategies, it is possible to extract commonly referred IC-related information from unstructured data automatically. IC information analyst can rely on this method as an additional mean to identify and extract the commonly sought IC information from financial reports in a fast, systematic and reliable manner.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Intellectual capital (IC) as a whole refers to the total resources and potential that determines the value and competitiveness of an enterprise (Magrassi, 2002). The concept of IC has been associated with organizational knowledge and intangible resources. Edvinsson and Malone (1997) refer IC as the cumulative value of an organization's intangible assets. The intangible assets are important today as knowledge and innovation are the key drivers to long-term business competitiveness. Current research has found that there are significant relationships between the IC and value-added productivity, measuring IC can therefore strengthen ongoing productivity measurement efforts on a firm's intangible assets (Phusavat, Comepa, & Sitko-Lutek, 2013). It has also been shown that IC is significantly positively associated with firm operating efficiency hence companies should invest and fully utilize IC to gain competitive advantage (Lua, Wang, & Kweh, 2014). According to the knowledge-based view of the firm (Grant, 1996; Spender, 1996), all intangible assets can be categorized into different types of knowledge. Similarly, Brooking (1996) conceptualizes IC as combined intangible assets of market, intellectual property, human capital, and firm's infrastructure that all together enable a company to function. Prior research suggests that the development of IC resources creates value for organizations, especially since the majority of an organization's assets are intangibles that are not shown on the balance sheet (Stewart, 1997). IC is rapidly becoming a new instrument for gauging organizational hidden values; measuring the real value and the total performance of IC are essential to any corporate heads who know how high the stakes have become for corporate survival in the knowledge and information age (Khavandkar & Khavandkar, 2009). As a result, the identification of an organization's IC is important as this provides insights on management action. Such actions often relates to the goal of enhancing the transparency of the concerned organization and benefits both internal stakeholders and external investors and beneficiaries.

However, the identification of an organization's IC is intrinsically difficult and is often subjective and inaccurate. Such inaccuracies not only mislead the market's observation over a corporation's consistency in performance but may also cause legal issues. This in turn leads to low business transparency in a knowledge-economy with a huge service-sector from which the most valuable assets are in fact intangibles. The fundamental challenge for these quality variations is the knowledge related to intangible assets which is mostly represented in unstructured or semi-structured formats. Due to the inherent nature of information, intangible assets account for a big proportion of a firm's capital that are neither has properly refined

* Corresponding author. Tel.: +852 2766 6609.
E-mail address: Eric.Tsui@polyu.edu.hk (E. Tsui).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.08.029

nor structured. In traditional approaches, the identification of an organization's IC is done manually through interviews, surveys, workshops, etc. (Yin, 2003). These methods are not only labor and time intensive but the quality of the results is highly dependent on the experience of the investigators. Up to now, there is still no uniform architecture available for intangible knowledge acquisition and elicitation. As a result, there exist significant variations in the quality of reported intangibles by organizations.

On the other hand, the established eXtensible Business Reporting Language (XBRL) attempts to standardize financial reporting with a machine-interpretable format that makes corporate reports easier to consume (interpret) and integrate data (O'Riaina, Currya, & Harth, 2012). Since its inception, XBRL has become an important element of the financial reporting landscape (Vasarhelyi, Chan, & Krahel, 2012). However, there are also problems with the use of XBRL. As XBRL has a highly structured format for exchanging business information, tedious initial manual efforts are required in filing XBRL-compliant documents. Furthermore, prior research has revealed that XBRL documents often contain multiple errors in signage, amounts, labeling, and classification (Bartley, Chen, & Taylor, 2010). These are serious errors, since XBRL data is computer-readable and users are generally unable to visually recognize and correct these errors.

Although the value of manual approaches (e.g. surveys and interviews) in identifying and collecting IC information is not to be under-estimated, there is also a considerable number of existing and alternative sources whereby an organization can tap into in order to retrieve IC-related information. Some of these sources indeed are explicit knowledge assets which are routinely produced by an organization (e.g. annual report) as part of its normal operation. It is not uncommon that a lot of efforts has been expended on the compilation of such explicit assets though such efforts are generally not IC-directed.

In order to provide a fast, systematic, consistent and reliable way to identify IC, this paper illustrates how to extract the related information of organization's IC from unstructured documents of an organization using an automatic and knowledge-based information extraction approach.

The amount of available electronic data of all kinds is increasing dramatically (Abiteboul, 1997). It was found that most of the information or knowledge in an organization is unstructured or semi-structured (Waters, 2005) such as e-mails, office documents, PDF documents and many other text-based documents, which contain much human knowledge and details of customer relationships related to daily operations. Many companies realize the value of the knowledge inherent in unstructured information which constitutes up to 80–98% of all the data, information and knowledge in an organization (Cheung, Lee, & Wang, 2005). In this paper, to address the challenge, a study has been conducted for efficient knowledge discovery and the extraction of intangibles by revealing IC-related information that are embedded mostly in unstructured data and partly in semi-structured data. A Knowledge-based Intellectual Capital Extraction (KBICR) algorithm is presented; this algorithm incorporates a 2-tier filtration by applying Rule-Base Reasoning (RBR) and Case-Based Reasoning (CBR). The KBICE algorithm has been evaluated by applying it to several publicly available financial reports from the banking industry.

2. Related work

Maeques, Simon, and Caranana (2006) divide IC into three dimensions including human capital, structural capital, and relational capital, based on the knowledge source and structure. Subramaniam and Youndt (2005) believe that IC consists of three highly interdependent facets of human capital, organizational capital, and social capital. Human capital comprises of all individual knowledge, both tacit knowledge (knowing how) and explicit knowledge (knowing what). Recently, Joshi, Cahill, Sidhu, and Kansal (2013) discovered that the value creation capability is highly influenced by human capital. The paper addressed those factors affecting IC performance in to the process of maximizing value creation. Structural capital composes of organization's routines, procedures, strategies, and policies that are in charge of organization's daily operations whereas organizational capital is the collective and institutionalized knowledge and experience residing within and utilizing through databases, patents, manuals, structures, systems, and processes of an organization. In Pandey and Dutta (2013), they found that there is relevance between knowledge infrastructure capability and KM excellence. They highlighted that the important role of a knowledge-sharing culture throughout management systems and routines. Their findings also suggested that organizational structure (a principal part of an organization's structural capital) plays both facilitating and steering roles in developing the culture of knowledge. Relational capital refers to all knowledge acquired by organizations because of their interaction with the external environment such as competitors, partners, customers, regulators, etc. Social capital, on the other hand, is defined as knowledge embedded within, available through and utilized by interactions among individuals and their social networks. In particular, empirical results have revealed that the social capital has significant effects, directly or indirectly, on supply chain integration and performance (Yim & Leem, 2013); it is suggested that supply chain integration among partners in the value chain can be improved by building up social capital.

Content analysis is the most popular method adopted to identify intellectual capital-related information (Guthrie, Petty, Yongvanich, & Ricceri, 2004). It is a manual method that involves in codifying qualitative and quantified IC-related information into the pre-defined IC indicator categories. A list of IC indicators is prerequisite, which was first compiled by Guthrie, Petty, and Wells (1999) based on the literatures on government policy and professional policy pronouncements. According to the context, culture as well as the environment changing, the list are modified by various scholars based on the various materials, such as the project results of US Financial Accounting Standards Board (FASB) (Bozzolan, Favotto, & Ricceri, 2003), the extant IC academic articles (Abdolmohammadi, 2005), stakeholder consultation principles (Schneider & Samkin, 2008), etc. Then these indicators are put into different categories for coding IC-related information. One of the most commonly used frameworks is derived from the Sveiby (1997) IC framework: internal structures, external structures; and employee competence. Coders record the IC data in the materials such as annual reports (Bontis, 2003; Guthrie & Petty, 2000; Guthrie et al., 1999), IPO prospectuses (Bukh, Nielsen, Gormsen, & Mouritsen, 2005), sustainability reports (Cinquini, Passetti, Tenucci, & Frey, 2012), etc. Finally, other coders put the IC data that was recorded into the designated classification. However, a completely manual method greatly limits the volume of process-able texts due to the labor-intensive data collection process (Beattie & Thomson, 2007; Oliverira, Gowthorpe, Kasperskaya, & Perramon, 2008). Even though at least two researchers participate in the assessing process, the subjectivity is inevitably involved, thus the reliability of the extracted data is also affected by personal bias (Abeysekera, 2006; Beattie & Thomson, 2007; Guthrie et al., 2004; Lee & Guthrie, 2010). Furthermore, multiple coders increase the risk of inconsistency due to the different coding rules (Abeysekera, 2006; Beattie & Thomson, 2007; Lee & Guthrie, 2010) being applied/interpreted.

Considering these disadvantages, some researchers turn to computers to solve the problem. Bontis (2003) applied the electronic search to identify the IC-related information in the electronic database which contains approximately 11,000 Canadian Corporations annual reports. Then a list of IC terminology was used

Download English Version:

https://daneshyari.com/en/article/386803

Download Persian Version:

https://daneshyari.com/article/386803

Daneshyari.com