Chapter 36
Corporate Disclosure Measurement

Md. Salah Uddin Rajib
Jahangirnagar University, Bangladesh

Md. Qutub Uddin Sajib
China University of Geosciences (CUG), China

ABSTRACT

This chapter conducts the study on the use of text mining for corporate disclosure measurement. During the last few decades, research on corporate disclosure has received significant concentration from the researchers. Corporate disclosure measurement has been used by researchers to reveal the real-world scenario of corporate disclosure practice. Very often, the researchers have used content analysis for making comments on the corporate disclosure practice. This chapter investigates the usability of text mining approaches for measuring the corporate disclosure. To do this, stages and general techniques of text mining have been discussed. The usability and advantages of text mining approaches for corporate disclosure measurement have been investigated. And prospective stages of text mining have been proposed to measure the corporate disclosure. The investigation and study indicates that text mining can contribute significantly to measure the corporate disclosure.

INTRODUCTION

Researchers have mentioned that about 90% of the entire world's data has been created in the past 2 years (e.g., Kolb & Kolb, 2013). Most of these data are made in electronic format. Corporate disclosures (CDs) are available now in electronic formats through different documents also. As the organizations are disclosing the disclosure annually or in the interim period, the numbers of volumes are increasing gradually. Organizations disclose disclosure through various media, e.g., annual reports, websites, supplementary of financial statements and others.

A number of researchers have investigated the corporate disclosure (CD) scenario in the last decades as it is assumed to have an impact on the capital market (Healy & Palepu, 2001). CD can be categorized into two broad groups – Mandatory disclosures and Voluntary disclosures. Mandatory disclosures are obligatory by the concern organization of a country (e.g., Securities and Exchange Commission). In

contrast, voluntary disclosures are made by the firms voluntarily considering the interest of stakeholders. A number of topics on CD have been investigated e.g., financial disclosure (e.g., Malone et al., 1993), human resource capital (e.g., Bontis, 2003), environmental disclosure (e.g., Gamble et al., 1996), strategic disclosure (e.g., Wagenhofer, 1990) and others.

Very often, researchers have used content analysis to explore the scenario of CD practice (e.g., Bontis, 2003; Khan et al., 2011). In content analysis they normally count the frequency of words to explore the scenario. Researchers normally select a list of related terms or keywords to the concern topic and search the terms in the annual reports or other related supplementary where disclosure has been presented by the organizations (e.g., Bontis, 2003).

Researchers have urged for more reliable measurement of CD practice (e.g., Rahman & Post, 2011). Consistent with that, this chapter proposes the use of text mining (TM) approach for measuring the CD. Currently, different areas of researches are applying the approaches and techniques of TM. Researcher of medical science, politics, business and others have used the techniques of TM to extract the information (Moohebat et al., 2015). The purpose of TM is to process unstructured textual information. In large text collection, TM works as a tool for knowledge discovery (Gomez et al., 2002). TM uses several techniques to extract the information, e.g., text classification, text clustering, text summarization, sentiment analysis and others. This chapter investigates the usability and advantages of TM approaches in the measurement of CD. Therefore, the aim of this paper is fourfold. First, it reviews the background of CDs and their measurement and background of TM. Second, it discusses the TM approaches, including general model of TM and methodology. Third, the usability of TM in measuring the CD has been investigated. And finally, it discusses the prospects of research in the area of using the TM in CD measurement.

BACKGROUND

As mentioned earlier, CD can be categorized into two broad groups - mandatory and voluntary disclosure. The demand for CD primarily arises from the perspective of regulation. Therefore, Securities and Exchange Commission (SEC) played an important role in the CD, made by the organizations. History states that because of various crises in US e.g., war, unemployment, racial conflict, environmental degradation and others, authority tempts to examine the fundamental principles underlying the body of corporate law for policy making (Schoenbaum, 1972). Sarbanes-Oxley (SOX) requires more and better information also. Consisting with the situation The Securities Exchange Act of 1934 imposes disclosure requirements on corporations in the US (Schoenbaum, 1972). With time, Generally Accepted Accounting Principles (GAAP) imposes some disclosures on corporations. Voluntary disclosure drives by various factors, e.g., manager-specific influences (e.g., Bamber et al., 2010), corporate governance (e.g., Hermalin & Weisbach, 2012), external pressure groups (e.g., Islam & Deegan, 2010) and others. Therefore, voluntary disclosures are associated with various areas and it is unstructured in nature. Researchers have tempted to find out the relationship of CD with the organization performances (Healy & Palepu, 2001). Therefore, measurement of CD emerges as a research area for the researchers.

This chapter conducts studies on the usability of text mining (TM) for CD measurement. TM is a discipline of natural language processing. The approach of TM has evolved from the simple word processing at the end of the 1990s to now (Antonio & Ferneda, 2008). TM derives much of its inspiration and direction from seminal research on data mining (Feldman & Sanger, 2007). The increased numbers of electronic documents have triggered the use of TM to extract the knowledge from textual data.

Related Content

Intelligent Medium Access Control Protocol for WSN
H. Malik, E. Shakshukiand M. Denko (2007). *Encyclopedia of Mobile Computing and Commerce (pp. 328-333).*
www.irma-international.org/chapter/intelligent-medium-access-control-protocol/17096

Clustering-Based Optimal Relay Vehicle Selection Scheme for Vehicular Adhoc Networks (VANETs)
Virender Kumarand Pawan Kumar Dahiya (2020). *International Journal of Mobile Computing and Multimedia Communications (pp. 67-83).*
www.irma-international.org/article/clustering-based-optimal-relay-vehicle-selection-scheme-for-vehicular-adhoc-networks-vanets/273169

Design Methodology for Mobile Information Systems
Z. Maamarand Q. Mahmoud (2007). *Encyclopedia of Mobile Computing and Commerce (pp. 190-194).*
www.irma-international.org/chapter/design-methodology-mobile-information-systems/17075

From Ethnography to Interface Design
Jeni Paay (2009). *Mobile Computing: Concepts, Methodologies, Tools, and Applications (pp. 3333-3348).*
www.irma-international.org/chapter/ethnography-interface-design/26727

Context-Aware and Adaptive Units of Learning in mLearning
Sergio Gómezand Ramón Fabregat (2012). *International Journal of Handheld Computing Research (pp. 1-25).*
www.irma-international.org/article/context-aware-adaptive-units-learning/67094