Computers in Industry 65 (2014) 393–400

Contents lists available at ScienceDirect

Computers in Industry

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

Sustainability trends in the process industries: A text mining-based analysis

Wan Te Liew [a], Arief Adhitya [b], Rajagopalan Srinivasan [a,*]

[a] Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576, Singapore
[b] Institute of Chemical and Engineering Sciences, A*STAR (Agency for Science, Technology and Research), 1 Pesek Road, Jurong Island, Singapore 627833, Singapore

ARTICLE INFO

Article history:
Received 4 March 2013
Received in revised form 12 November 2013
Accepted 13 January 2014
Available online 16 February 2014

Keywords:
Sustainability trends
Text mining
TF-IDF
Ontology
Chemical industry
Corporate Social Responsibility Report

ABSTRACT

Sustainability is widely recognized as one of the most important challenges facing the world today. Companies publish sustainability reports that present their efforts and achievements in meeting sustainability goals and targets. In this paper, text mining is used to identify sustainability trends and practices in the process industries. Four main sectors of the industry are studied: oil/petrochemicals, bulk/specialty chemicals, pharmaceuticals, and consumer products. Our study reveals that the top sustainability focuses of the four sectors are very similar: health and safety, human rights, reducing GHG, conserving energy/energy efficiency, and community investment. Sector-specific sustainability issues have also been identified, for example oil spill prevention in the oil/petrochemicals sector and access to medicine in the pharmaceuticals sector. Environment is identified to be the predominant sustainability aspect in the process industries. The text mining methodology, results, and findings are detailed in the paper.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In the face of mounting environmental challenges and shortages of natural resources, growing awareness of social responsibility and the need to remain profitable, sustainability has become a key agenda for every industry. The most commonly used definition of sustainability or sustainable development is from the Brundtland Commission: development that meets the needs of the present without compromising the ability of future generations to meet their own needs [1]. While this definition reflects the philosophy of sustainability, it does not provide any specific directions or action items that companies can readily adopt. They may interpret sustainability differently and pursue it through different measures. This paper therefore seeks to understand what sustainability means to companies and how it is translated into practice.

Sustainability practices are industry-specific as different industries face different challenges. For example, sustainability issues faced by the electronics industry are human rights labor violations and product recycling or take back [2], whereas the automotive industry is more focused on fuel consumption and emissions [3]. In this paper, we are interested in the chemical process industry, which in particular has a strong emphasis on sustainability due to its nature of business. The industry involves extraction of raw materials such as crude oil, gas and minerals and processes which are highly energy intensive. The industry also handles large volume of toxic, flammable, and hazardous chemicals. Different sectors within the chemical process industry emphasize different aspects of sustainability. This paper focuses on four main sectors of the chemical process industry: oil/petrochemicals, bulk/specialty chemicals, pharmaceuticals, and consumer products.

There are many ways that companies can contribute to sustainability. A study by McKinsey shows that sustainability agendas of companies cover the social, environmental and economic aspects – the triple bottom line of sustainability [4]. These include reducing energy and water input in operations, reducing emissions and waste, improving employee retention and motivation, compliance to regulations, and reducing operational risk. Other measures include forging stronger relationships with communities, NGOs, governments, and industry peers.

Different companies even within the same sector undertake different measures. These are typically highlighted in sustainability reports that many companies voluntarily publish to describe their sustainability goals, efforts, and achievements. These reports are usually published on an annual basis, aimed at demonstrating transparency and accountability, as well as to engage stakeholders and governments and to attract investors. It has been proposed

* Corresponding author. Tel.: +65 65168041; fax: +65 67791936.
E-mail address: chergs@nus.edu.sg (R. Srinivasan).

0166-3615/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compind.2014.01.004

that accountability for sustainability impacts can lead to structures and models that improve sustainability performance of companies [5]. These reports also reflect the growing importance of stakeholders' perception [6]. Sustainability has also been linked with increased competitiveness and economic performance [7].

In this paper, we analyze sustainability trends and practices in the process industries through text mining of sustainability reports. Text mining provides the capability to extract information and trends from large amount of textual data, giving an overview of the main sustainability issues and practices discussed in the reports. To our knowledge, there has not been any reported work on identifying sustainability trends in the industry. Modapothala and Issac [8] used a data mining approach to evaluate corporate environmental reports based on 10 sustainability criteria (variables) such as CEO statement on sustainability, carbon dioxide emissions, human rights, etc. They then analyzed the correlation between the term frequencies of the 10 variables in the reports to identify close variables that can be grouped together. Modapothala et al. [6] used text mining and multi-discriminatory analysis to appraise corporate sustainability reports and found that disclosures made by the companies differ across industrial sectors and that the environmental variable is the main contributing factor in the reports. Shahi et al. [9–11] developed an intelligent software system to analyze and score sustainability reports within the Global Reporting Initiative framework using machine learning and text categorization. Text mining has also been used to identify research trends and clusters by analyzing published journal papers, the objective being to identify fading or emerging research topics, lifetime of the topics, and rising or declining theories [12,13]. All of these demonstrate the efficacy of text mining in generating knowledge from large amount of textual data.

In this work, sustainability reports of major companies in the four sectors of the chemical process industry are obtained and used as the basis for text mining analysis. The companies chosen are the largest companies listed in the Forbes ranking of the global top 2000 companies in 2011 [14]. The Forbes ranking takes into account sales, profits, assets, and market values of the companies and provides a useful indicator of the biggest public companies in the world. These prominent companies should provide meaningful and representative insights on the sustainability trends in the process industries. The rest of the paper is organized as follows. The text mining methodology is described in Section 2. The results are discussed in Section 3 and Section 4 presents the conclusion.

2. Methodology

2.1. Key concepts

In this age of rapidly proliferating documents on the Internet, the amount of data and information available is overwhelming and clearly too much to be processed on a manual basis. Text mining attempts to uncover patterns and non-trivial information and knowledge from unstructured text. It is a relatively new research area that attempts to resolve the information overload problem by using techniques of data mining, machine learning, natural language processing, information retrieval, and knowledge management [15]. Since text is the most readily available form of information around, text mining has a high potential value and wide applications. Some daily applications that make use of text mining are spam filtering, sentiment analysis, and measuring customer preferences.

In this paper, analysis is conducted on the basis that frequent use of sustainability-related terms in a company's sustainability report reflects its main concern and emphasis. The higher the frequency or occurrence of a certain term, the greater is the emphasis that companies place on them and hence the more important they are to the companies. Thus through text mining and term frequency analysis, the main sustainability trends and focuses in each of the four sectors and the process industry as a whole can be identified. Text mining consists of two main phases: text refining and knowledge distillation [16]. Text refining refers to preprocessing of the text inputs to the system to prepare the data for knowledge distillation. Text refining converts the original, unstructured text data input into a standard form for analysis. The knowledge distillation phase focuses on deducing patterns and identifying information. This reveals implicit knowledge that is not noticeable from just one document but perceptible from a series of documents.

2.2. Overall procedure for text refining

2.2.1. Gathering of files and conversion

The reports were first converted into plain text format to facilitate processing in the subsequent steps. During the conversion, irrelevant content such as content pages, graphs, tables, headings, and footers were removed as they are mostly repetitions and could introduce bias into the analysis.

2.2.2. Basic text processing

We used the open-source RapidMiner [17] with its Text Processing Extension for basic text processing. The steps involved are detailed in Table 1. The output from this step is a filtered list of terms, which include single words and phrases of two (bigrams) and three words (trigrams), from all reports in each sector group.

2.2.3. Generating frequency statistics

This step generates the frequency statistics that will form the basis for the analysis. They are in the form of *term occurrence*, *term frequency-inverse document frequency* (TF-IDF), and *involvement* for all terms in all documents (reports).

Term occurrence (also called term frequency or TF) is the count of a term in a document, i.e. the number of times that term appears in a document.

TF-IDF is calculated as follows [18]:

$$\text{TF-IDF} = \text{TF} \times \log \frac{\text{Total number of reports in the sector}}{\text{Number of reports in the sector that contain a term `}x\text{'}}$$

TF-IDF is used to identify terms that are unique to a particular document in a large sample space. The TF-IDF score will be highest when a term occurs many times within a small number of reports in the particular sector. This helps to identify rare terms that occur in an individual company report in a sector. The TF-IDF score will be lowest when the term occurs in almost all reports.

Involvement measures how widespread a particular issue is among companies in a sector and is calculated as:

$$\text{Involvement} = \frac{\text{Number of reports in the sector that contain a term `}x\text{'}}{\text{Total number of reports in the sector}} \times 100\%$$

The output from this step is four matrices, one for each sector, containing all the terms in the sector and the statistics values for each of these terms.

2.2.4. Pruning

The matrices are refined by pruning those terms that occur in less than a lower limit number of documents. In this work, terms have to appear in at least two documents to be included; terms that

Download English Version:

https://daneshyari.com/en/article/508956

Download Persian Version:

https://daneshyari.com/article/508956

Daneshyari.com