Submission to BUSINESS STRATEGY AND THE ENVIRONMENT

Title:

Web Content Mining for Comparing Corporate and Third-Party Online Reporting: A Case Study on Solid Waste Management

Authors:

Irene Pollach, Aarhus University, Denmark
Arno Scharl, MODUL University Vienna, Austria
Albert Weichselbraun, Vienna University of Economics, Austria

Address for Correspondence:

Irene Pollach
Aarhus University, Aarhus School of Business
Fuglesangs Allé 4
8210 Aarhus
irpo@asb.dk

Web Content Mining for Comparing Corporate and Third-Party Online Reporting: A Case Study on Solid Waste Management

ABSTRACT

This study investigates the coverage of solid waste management on 1,142 websites maintained by companies, news media, and non-governmental organizations to validate an automated approach to content and language analysis. First, a frequency analysis of waste management terms is used to shed light on the breadth and depth of their environmental discourses, revealing that corporate and media attention to waste management is small compared to that of non-governmental organizations. Second, an investigation of their attitudes toward waste management suggests that companies avoid negative information in environmental communication, unlike news media or non-governmental organizations. Ultimately, an automated tool for ontology building is employed to gain insights into companies' shared understanding of waste management. The ontology obtained indicates that companies conceptualize waste management as a business processes rather than framing it from an ecological perspective, which is in line with findings form previous research.

KEYWORDS

Environmental reporting, websites, content analysis, quantitative linguistics, ontology, waste management

INTRODUCTION

Three interrelated forces drive worldwide demand for public access to environmental information: the urgency and scope of environmental problems, the increasingly active character of civil society, and the revolution in information and communication technologies (Scharl, 2004). Disseminating environmental information via the World Wide Web, directly or via online media as intermediaries, helps enhance public knowledge and build awareness of the interdependency of ecological, economic, and social issues. The importance of environmental communication through scientific exchange, educational programs, and the media has been recognized internationally at least since the United Nations 1972 Conference on the Human Environment in Stockholm (UNEP, 1972). Given that the Earth's growing population places greater and greater pressure on already dwindling resources, environmental communication and education may slow down this process and pave the way for a more sustainable future (Kunst and Witlox, 1993).

Waste management is a key element of strategies aiming for environmental sustainability. Countries that support recycling not only reduce resource consumption and discharge less waste into the environment, but also achieve higher total output rates (Di Vita, 2004). Integrated waste management, and in particular composting, also eases climate-change problems by reducing the net flux of greenhouse gases (BioCycle, 2004). Although most OECD countries have committed themselves to the objectives of waste prevention and minimization, the waste volumes they produce have increased substantially over the past decades. The attitudes and behaviors of the public at large play a key role in waste reduction and recycling, but sustainable waste management is not in their hands alone (OECD, 2004). Rather, producers need to recognize that their

responsibility extends to the post-consumer stage of a product's life cycle. They can prevent waste and limit the use of resources by offering eco-friendly designs, reducing packaging waste, or setting up take-back systems (Runkel, 2003).

The purpose of this study is to present an automated approach for analyzing environmental online communication using web-mining technology. To demonstrate the robustness of this approach, we compare environmental reporting on websites of organizations that disseminate environmental information for entirely different reasons. These organizations include corporations, news media, and non-governmental organizations in the environmental domain. While companies make environmental information available on their websites to satisfy stakeholder demands (Cormier *et al*., 2004), news media do so only if they consider environmental issues newsworthy, and environmental NGOs do so as their *raison d'être*. Based on their web coverage of solid waste management, a narrow and thus manageable topic, we examine the attention they pay to waste management, gauge their attitudes toward this issue, and determine whether they share the same conceptualization of waste management. In the following sections, we first review the literature on corporate environmental reporting and describe the methodology of data collection and analysis before presenting the findings of our analyses and discussing the results.

CORPORATE ENVIRONMENTAL REPORTING

We are witnessing rapid advances in gathering and processing information. These advances are transforming the way society handles the explosive growth and reduced lifespan of information (Bell, 1973; Haddad and Draxler, 2002; Weingart, 2002), particularly in highly dynamic domains such as the environment (Pick *et al*., 2000). The

past decade has also seen enormous growth in corporate environmental communication with external stakeholders (Elkington, 1994). These efforts have been spurred by companies' realization that environmental communication closes the information gap between companies and their external stakeholder groups (Lev, 1992), demonstrates environmental leadership (Dando *et al*., 2003), and may even steer public attention away from actual problems (Cerin, 2002). Voluntary reporting of corporations is thus used both as a response to public pressure and as a proactive attempt to shape perceptions and enhance the corporate image (Hooghiemstra, 2000). Clearly, the voluntary nature of environmental disclosures makes it tempting for businesses to skew the environmental information they disseminate (Watson *et al*., 2003), which may be the reason why corporate environmental reports tend to contain good intentions rather than measurable objectives ("Corporate storytelling", 2004). The danger inherent in such practices is that they may unduly raise the public's expectations of a company's performance and create a *façade* of environmental concern, which is hardly sustainable in the longer term (Schlegelmilch and Pollach, 2005).

Broadly, previous research on corporate environmental reporting has focused on the prevalence of environmental disclosures, attitudes of environmental managers, company performance, target audiences, and the content of environmental reports. The prevalence of corporate environmental disclosures in different industries or countries has been examined in both snapshot and longitudinal analyses, including studies on Australia (Deegan and Gordon, 1996), Finland (Niskanan and Neiminen, 2001), France (Cormier and Magnan, 2003), the United Kingdom (Campbell, 2003), Europe, Japan and the US (Kolk, 2005). Frequently, an industry's environmental impact positively correlates with

its degree of environmental disclosure (Deegan and Rankin, 1996; Kolk *et al*., 2001; Gray *et al*., 2001; Cormier and Magnan, 2004).

Corporate environmental managers have been surveyed on environmental reporting practices, focusing for example on the determinants of corporate environmental disclosures (Cormier *et al*., 2004), on the influence of such disclosures on corporate environmental performance (Annandale *et al*., 2004), on the development of environmental policies in companies (Tilt, 1997), and on the media channels used to disseminate environmental information (Stray and Ballantine, 2000). Other studies have investigated financial determinants of reporting activities (Cormier and Magnan, 2003) and the relationship between corporate financial performance and environmental reporting efforts (Connelly and Limpaphayom, 2004; Stanwick and Stanwick, 2000).

Another strand of research has focused on the target audience of corporate environmental reporting, investigating whether corporate environmental reports meet the information needs of their readers (Noci, 2000), whether readers perceive them as comprehensible (Rockness, 1985), how particular audiences could be addressed more effectively (Azzone *et al*., 1997), and how Internet-based push and pull technologies can deliver customized reports (Isenmann and Lenz, 2001; Isenmann and Marx-Gómez, 2004).

Previous content analyses of corporate environmental disclosures have focused on the environmental activities outlined (Niskala and Pretes, 1995), the amount of monetary references (Alciatore *et al*., 2004), the metrics used (Marshall and Brown, 2003), and the proportion of environmental information contained in annual reports (Belal, 2000). Esrock and Leichty's (1998; 2000) content analyses of sustainability reporting on corporate websites have highlighted the growing proportion of companies engaging in web-mediated environmental communication. Also, a number of studies have revealed

that the content of corporate environmental communication is almost exclusively positive and self-laudatory, while the amount of negative information is negligible (Wiseman, 1982; Rockness, 1985; Deegan and Gordon, 1996; Deegan and Rankin, 1996; Niskanen and Neiminen, 2001).

While previous research has focused on the content of environmental reports, it has not taken into account aspects of language and terminology and has not used web-mediated communication as a data source. Also, previous studies of corporate environmental reporting have not put their findings in perspective by comparing them with environmental information disseminated by organizations other than companies. Ultimately, samples have been relatively small. In fact, the largest sample used in the studies cited above encompasses 469 companies (cf. Stanwick and Stanwick, 2000). The present study seeks to address these shortcomings by analyzing the environmental communication efforts of a large sample of companies, news media and non-governmental organizations.

METHODOLOGY

To investigate how companies communicate their stance on solid waste management, this paper studies the websites of companies included in the 2004 edition of the *Fortune 1000.*[1] Researchers have previously used *Fortune* magazine's company rankings to investigate web content (Perry and Bodkin, 2000), online marketing strategies (Palmer and Griffith, 1998) and corporate web-usage patterns (McManis *et al*., 2001). Furthermore, a number of studies assessing companies included in the *Fortune* ranking have concentrated on aspects of corporate social responsibility (e.g. Esrock and Leichty,

[1] http://www.fortune.com/fortune1000/

1998; Weaver *et al*., 1999; Reicher *et al*., 2000; Esrock and Leichty, 2000; Kolk *et al*., 2001).

Eleven sites were excluded from the analysis, as their sites were not accessible at the time of data collection, which resulted in a sample of 989 corporate websites. In addition to these websites, the sample included websites of 39 environmental organizations[2] and 114 news media[3] for comparative purposes. The sampling procedure for the news media drew upon the *Newslink.org*, *Kidon.com* and *ABYZNewsLinks.com* directories to compile a list of 42 US news organizations and 72 organizations from four other English-speaking countries (Canada, United Kingdom, Australia and New Zealand).

In order to account for the dynamics of web content, a crawling agent (Scharl, 2000) mirrored the sample websites by following their hierarchical structure until it reached 50 megabytes of textual data for news media and 10 megabytes for commercial and NGO sites. These limits help compare sites of heterogeneous size and reduce the dilution of top-level information by content in lower hierarchical levels. The system then identified and removed redundant segments such as headlines and news summaries, whose appearance on multiple pages would otherwise distort frequency counts.

This study approaches the coverage of waste management from three different methodological angles. First, we determine the frequency of waste management terms on the sample websites to shed light on the breadth and depth of their environmental discourses. This analysis is based on the assumption that the frequency of particular concepts and ideas in a text is a measure of importance, attention or emphasis (Krippendorff, 1980). However, looking at the frequency with which certain concepts are addressed without taking into account the context in which they occur does not reveal any-

[2] http://www.ecoresearch.net/election2004/eco
[3] http://www.ecoresearch.net/election2004/media

thing about the associations among words (Weber, 1985), thus limiting the explanatory power of relative concept frequencies. As a second step, we therefore assessed the semantic orientation of each concept in all three samples by taking into account the number of positive and negative words associated with each waste management term. This approach follows the notion that the numerical balance of positive and negative attributes of a concept is a measure of attitudinal direction or bias (Krippendorff, 1980). The semantic orientation calculated thus highlights the intensity and direction of the websites' positive or negative attitudes towards waste management. Ultimately, we present a tool to automatically generate a waste-management ontology from semantic associations identified within our corpus, i.e. the texts collected from the corporate websites in our sample. This ontology captures the major concepts that corporations associate with waste management, reveals semantic relations among these concepts, and gives insights into corporations' shared worldviews.

CONCEPT FREQUENCY

To investigate the discourse on waste management on the sample websites, a list of 311 technical terms related to solid waste management was compiled from almost 100 different English resources, including books on waste management, articles in journals and magazines, as well as the websites of NGOs, governments and supranational organizations. In order to measure the extent of coverage on a particular topic, a case-insensitive pattern-matching algorithm processed regular expression queries (formalisms describing a set of strings without enumerating its elements) representing these 311 terms.

In view of the large number of terms considered, we calculated concept frequencies rather than individual term frequencies for each of the three samples (companies, NGOs and news media), grouping the terms into the following nine concepts: *solid waste, waste management, waste prevention, waste collection, waste processing, resource recovery, biological waste treatment, thermal waste treatment,* and *waste disposal*.

To account for differences in corpus sizes, the aggregated term counts for each concept were divided by the total number of words contained in a corpus. Overall, only 176 of the 311 waste management terms were found at least once across the websites in the sample. As depicted in Figure 1, *resource recovery* was the most prevalent concept in two of the sub-samples, despite being the smallest concept group with only 16 terms. The use of *resource recovery* and *solid waste* on NGO sites exceeds the use of any other concept in the three samples. Also, *waste disposal* receives significantly more attention on NGO sites than on corporate or media sites. The top three concepts on both corporate and media sites are *resource recovery*, *waste collection* and *waste disposal*. Notably, relative frequencies on corporate sites tend to be greater than those on media sites. While the latter pay next to no attention to the concept of *waste prevention*, companies and NGOs pay the least attention to *biological waste treatment* and *waste processing*, respectively.

FIGURE 1 ABOUT HERE

SEMANTIC ORIENTATION

Based on the notion that there is a conceptual connection between words and their adjacent text (Giora, 1996), the semantic orientation towards a word within a sentence is calculated by measuring the distance (in words) between a term and a pre-defined list of

words known to have positive or negative connotations (Scharl *et al*., 2003). An initial set of about 3,500 sentiment words was taken from the tagged dictionary of the *General Inquirer* project (Stone, 1997). To a large extent, the validity of this approach depends on the size of the tagged dictionary. It is therefore essential that *all* instances of the sentiment terms in the corpus are included in the analysis and not just their base forms.

Lemmatization, also referred to as word stemming, is a form of linguistic processing that groups a word's base form, e.g. the infinite tense of a verb or the singular of a noun, with all its inflected forms. The list of positive and negative words from the *General Inquirer* was subjected to a reverse lemmatization procedure, which added about 4,000 inflections such as plural forms, past tense suffixes and other morphological variations to the initial list of sentiment words (e.g. manipulate → manipulates, manipulating, manipulated).

The semantic orientation values of all terms subsumed under one concept were averaged to determine the attitudes of the sample towards solid waste management concepts. The goal of this semantic analysis was to verify whether corporate environmental communication indeed contains primarily positive information (cf. Deegan *et al*., 1996; Niskanen *et al*., 2001). The results indicate that companies view waste management activities overwhelmingly positively, yielding positive results in all nine categories and the highest scores in eight of them. NGOs largely mirror the positive attitudes of companies for most categories, albeit to a smaller extent. However, they top the score of the corporate sample in the categories of *waste processing* and *resource recovery*. NGO attitudes towards *waste disposal* are almost neutral, but slightly on the negative side. The sites in the media sample express positive attitudes towards most concepts, but do so to a lesser extent than companies and NGOs. Notably, semantic orientation in the

media sample is highly negative for *thermal treatment* of waste and slightly negative for *solid waste*.

FIGURE 2 ABOUT HERE

WASTE MANAGEMENT ONTOLOGY

While conflicting definitions of "ontology" abound (Guarino, 1997), there is consensus in the information systems literature that the term refers to a designed artifact formally representing shared conceptualizations (e.g. Gahleitner *et al*., 2005; Jarrar and Meersman, 2002). In the context of this research, ontologies are explicit formal specifications of terms used in a particular domain together with a set of hierarchical relations among them (Gruber, 1993). Specifying how terms are related to each other, ontologies not only represent hierarchically organized knowledge but also provide a common vocabulary for communicating about a particular topic.

A vital characteristic of ontologies is that they capture consensual knowledge by representing an information structure accepted by a particular group (Studer *et al*., 1998; Nirenburg and Raskin, 2005). Since a lack of shared understanding among a community of interest is one of the biggest barriers to transferring and managing knowledge, the major motivations behind building ontologies are to reduce conceptual confusion, unify different viewpoints, and establish a shared terminology (Edgington *et al*., 2004). In information retrieval, ontologies can support query-term expansion and disambiguation, relevance ranking of search results, and web-resource annotation (Abdelmoty *et al*., 2005).

Automated approaches to building ontologies aim at facilitating and accelerating the time-consuming process of identifying and hierarchically positioning relevant domain concepts. Also, an automated process is not guided by decisions on inclusion or exclusion made by the researcher and is thus more objective (Gruber, 1995). Rather than using simple heuristic rules common in information retrieval, the semi-automatic extension and refinement of domain-specific ontologies necessitates the fine-grained processing of textual data (Navigli and Velardi, 2005). Therefore, the learning of taxonomic relations from unstructured textual data (cf. Cimiano *et al*., 2005) is an important step in automating the creation and validation of ontologies.

In this paper, we investigate how our sample of corporate websites conceptualizes waste management to gain insights into the extent to which their understanding of waste management corresponds to that of domain experts. A waste-management ontology was automatically generated from the corpus of corporate websites using the ontology extension module developed by Liu *et al*. (2005). This tool integrates co-occurrence analysis, subsumption analysis and other linguistic methods via a spreading activation algorithm (a technique inspired by the human brain, where neurons fire activations to adjacent neurons) to create semantic networks and hierarchically position concepts within these networks:

- Co-occurrence analysis assumes that two semantically related terms regularly co-occur in the same documents (Roussinov and Zhao, 2003).
- Assuming that general terms occur more frequently than specific terms, the subsumption approach (Sanderson and Croft, 1999) generates concept hierarchies by comparing the frequencies of co-occurring words.

Figure 3 presents a conceptual view of the system architecture. The ontology building process starts with the selection of a small set of top-level concepts as a seed ontology, which functions as an anchor for retrieving further concepts from the web corpus. The terms chosen are formulated as regular expressions, which are sequences of characters that form a template for use in pattern searching. The following regular-expression query describes the set of concepts chosen as a seed ontology. Its indented structure reflects the tree structure underlying these concepts:

```
(garbage|waste|trash) management
   resource recover(y|ies|ing)
      recycl(es?|ed|ers?|ing|ables?|ability)
      compost(ed|ing)?
   (garbage|waste|trash) disposals?
```

The terms included in the seed ontology are then fed into the *Lexical Analyzer*. Co-occurrence analysis at both the sentence and the document level limits the influence of popular terms not related to the domain (Roussinov and Zhao, 2003). Terms are selected according to a threshold value on the co-occurrence significance. The *WordNet Lexical Database* (Fellbaum, 1998)[4] was used to determine a word's intended meaning by selecting an appropriate sense from a set of choices (Navigli and Velardi, 2005).

FIGURE 3 ABOUT HERE

A further lexical analysis is conducted to search the corpus for terms connected through trigger phrases, which indicate parent-child relations (Joho *et al*., 2004). For example, in the phrase *"resource recovery activities* ***such as*** *recycling or composting",*

[4] http://wordnet.princeton.edu/

the trigger phrase SUCH AS indicates a hierarchical relationship between the concept *resource recovery* as the super-ordinate parent and *recycling* and *composting* as child nodes. This approach helps determine which of two concepts is more general within a hierarchical structure. The terms obtained are then connected with the seed ontology via directed weighted links.

Once the network is established, spreading activation identifies the terms most relevant within the domain and suggests their incorporation into the seed ontology. *WordNet*, head nouns[5] and subsumption analysis are then used to determine the semantic relationships between the terms identified and the terms included in the seed ontology. Optionally, domain experts are consulted for terms not confirmed automatically, before another iteration of spreading activation over the newly acquired terms takes place.

Figure 4 shows the extended ontology after two iterations of spreading activation. Black nodes depict the seed ontology, while the gray and white nodes were added after the first and second iterations, respectively. Arrows indicate confirmed hierarchical relations and the dotted lines link semantically related terms whose exact type of relationship could not be determined automatically. The relation values (r) indicate the strength of the relationships.

FIGURE 4 ABOUT HERE

Two iterations have identified and appropriately positioned a number of relevant terms, including *leachate*, *landfill gas*, *disposal*, *emissions*, *dioxin* and *environmental*

[5] Head nouns (e.g. *disposal* in *garbage disposal*), which often subsume noun compounds, are added to the network as potential super-ordinate terms.

matters. The following examples illustrate our approach, while at the same time indicating shortcomings of the current development stage of the system.

(i) The co-occurrence analysis has yielded semantically related terms at the sentence and document level. For example:

> *By **recycling** paper, instead of sending it to **landfills**, we avoided 9,764 tons of CO2 **emissions**.*

(ii) Trigger phrases provide additional evidence for semantic relations. In the following example, the phrase AND OTHER appropriately links the concept *landfill(s|ing)?* and *(garbage|waste|trash) disposal*:

> *In some cases, these waste sites are located at Olin manufacturing locations; in other instances, they are off-site **landfills** and other **waste disposal** sites that may have been shared with other companies.*

(iii) Due to performance limitations, co-occurrences are currently computed for up to two-word concepts only. As a result, no co-occurrence evidence for concepts represented by more than two terms is included in the semantic network. This may lead to the negligence of potentially relevant terms, as the example below illustrates.

> *Dow's focus is on reducing emissions of **persistent, bioaccumulative, and toxic substances (PBTs)** such as **dioxin**, rather than on phasing out products that contribute positively to society.*

Automated approaches to language processing are, for the most part, unable to distinguish between common nouns and proper nouns. Therefore, the tool has also placed company names (*McDonald's, Monsanto, Renessen, Sunoco, Pharmacia*), chemical

substances (*DEHP, RoundupR*), and environmental projects (*Redirectory, GRI, Balanced Lifestyles*) in the ontology. They are all relevant to the domain represented, thus strengthening the validity of the approach.

Clearly, an automated process cannot generate a completely accurate positioning of all related terms in view of the complex semantic networks spanning super-ordinate and subordinate terms. For example, business terms such *borrower*, *asset*, *issue series* and *revenue bonds* may not seem relevant to domain experts. Still, the algorithm has placed them in the ontology, as they frequently co-occur with their super-ordinate concepts within the same document yet not in the same sentence. Only *asset* co-occurs with *disposal* also at the sentence level, e.g. as in

> *IBM's Global **Asset** Recovery Services unit oversees materials sale, reuse, **disposal**, refurbishing, repair, remanufacturing, dismantling and warehousing associated with all IBM returning and excess inventory worldwide.*

These complexities notwithstanding, the ontology obtained illustrates that companies concern themselves with waste management and participate in environmental initiatives. Quite plausibly, they conceptualize it as a business processes rather than framing it from an ecological perspective, as the inclusion of business terms in the ontology suggests.

CONCLUSION AND OUTLOOK

The goal of this paper has been to assess the content of environmental online communication using an automated approach. A comparison of three sub-samples has helped to corroborate the findings and validate the approach adopted, as the findings reflect what is to be expected from the three sub-samples. The analysis of term frequen-

cies has highlighted that corporate attention to solid waste management is small compared to that of environmental NGOs, but compares favorably to that of news media. In particular, the little attention both companies and news media pay to waste prevention and biological waste treatment may indicate that these organizations embrace sustainability as a response to stakeholder demands, rather than being truly committed to the issue (Cormier *et al*., 2004). Also, considering that the terms included in the frequency analysis were largely technical terms, the results suggest that corporate environmental reporting either shies away from specialized language or lacks scientific rigor altogether.

Further, corporate attitudes towards waste management have turned out to be highly positive, confirming that companies do in fact avoid negative information in environmental communication, as previous research has shown. The results of the semantic analysis suggest that companies do not frame waste management as a problem. News media, meanwhile, have expressed negative attitudes towards thermal waste treatment and solid wastes, while NGOs have viewed waste disposal slightly negatively. Those concepts viewed negatively seem to reflect the most pressing problems in waste management. These findings fit in well with what previous research has reported on positive and negative information in environmental reports (e.g. Deegan and Gordon, 1996). The dominance of words with positive connotations may be explained by corporate policy documents that use proactive terminology, outlining for the most part only goals and visions rather than actions (Preuss, 2005).

The ontology building process has proved to be another valuable source for gaining insights into corporate attitudes toward waste management. By applying various natural-language processing technologies we were able to identify important concepts, key-

words, and linguistic associations characteristic of corporate environmental reporting. The analysis has also uncovered relationships linking the concepts of waste management and garbage disposal with companies addressing these issues (e.g. *McDonald's*, *Monsanto*, *Renessen*, *Sunoco* and *Pharamcia*) and with environmental projects such as *GRI*, *Redirectory* and *Balanced Lifestyle*. Further, the fact that potentially harmful chemical substances (e.g. *DEHP*, *RoundupR*, *dioxin*) were identified as relevant concepts in the ontology building process suggests that companies also take up delicate issues in their environmental reporting initiatives.

The lexical, semantic and terminological analyses have proved useful in investigating environmental online communication, adding to the existing body of research on corporate environmental reporting in several respects. First, the approach adopted uses a large and varied sample including companies, news media and environmental NGOs. Second, the paper presents a specific, in-depth coverage of environmental reporting practices in one particular domain. Third, this paper captures information disseminated via the World Wide Web, which includes both environmental reports made available for download and information disseminated via regular web pages. Ultimately, this paper pays particular attention to terminological aspects, which previous research has largely ignored.

The findings suggest that further refinements of the approach for future studies in environmental online communication are worthwhile. Incorporating qualitative insights from discourse analysis or interviews will help grasp the underlying attitudes and conceptual understandings. In particular, capturing more subtle connotations in between the two poles of positive and negative attitudes would increase the utility of the semantic orientation component. Similarly, capturing the distribution of semantic orientation val-

ues across sites would help discern commonly agreed knowledge from the coverage of controversial issues.

To assess corporate environmental reporting from a broader angle, future research should also extend the approach to other environmental domains. These domains need to be broad enough to ensure a certain level of depth in the frequency analysis but at the same time should be narrow enough to build a meaningful seed ontology. Issues in environmental management meeting these criteria include, for example, *energy management* or *clean technologies*.

REFERENCES

Abdelmoty AI et al. 2005. A critical evaluation of ontology languages for geographic information retrieval on the Internet. *Journal of Visual Languages and Computing*, **16**(4): 331-358.

Alciatore M, Callaway C, Dee PE. 2004. Changes in environmental regulation and reporting: The case of the petroleum industry from 1989 to 1998. *Journal of Accounting and Public Policy* **23**: 295-304.

Annandale D, Morrison-Saunders A, Bouma G. 2004. The impact of voluntary environmental protection instruments on company environmental performance. *Business Strategy and the Environment* **13**(1): 1-12.

Azzone G, Brophy M, Noci G, Welford R, Young W. 1997. A stakeholders' view of environmental reporting. *Long Range Planning* **30**(5): 699- 709.

Belal AR. 2000. Environmental reporting in developing countries: Empirical evidence from Bangladesh. *Eco-Management and Auditing* **7**(3): 114-121.

Bell D. 1973. *The Coming of the Post-Industrial Society: A Venture in Social Forecasting*. Basic Books: New York.

Campbell D. 2003. Intra- and intersectoral effects in environmental disclosures: Evidence for legitimacy theory? *Business Strategy and the Environment* **12**(6): 357-371.

Cerin P. 2002. Communication in corporate environmental reports. *Corporate Social Responsibility and Environmental Management* **9**: 46-66.

BioCycle. 2004. Composting and sustainable waste management help solve climate change problems. *BioCycle*, 45(12): 11.

Cimiano P, Pivk A, Schmidt-Thieme L, Staab S. 2005. Learning taxonomic relations from heterogeneous evidence. *Ontology Learning from Text: Methods, Applications and Evaluation*. Eds. P. Buitelaar et al. IOS Press.

Connelly J, Limpaphayom P. 2004. Environmental reporting and firm performance: Evidence from Thailand. *The Journal of Corporate Citizenship* **13**: 137-149.

Cormier D, Gordon IM, Magnan M. 2004. Corporate environmental disclosure: Contrasting management's perceptions with reality. *Journal of Business Ethics* **49**(2): 143-165.

Cormier D, Magnan M. 2003. Environmental reporting management: A continental European perspective. *Journal of Accounting and Public Policy* **22**: 43-62.

Cormier D, Magnan M. 2004. The impact of the Web on information and communication models: The case of corporate environmental disclosure. *International Journal of Technology Management* **27**(4): 393-416.

Corporate storytelling. *The Economist*, 11 June 2004: 13-14.

Dando N, Swift T. 2003. Transparency and assurance: Minding the credibility gap. *Journal of Business Ethics* **44**(2/3): 195-200.

Deegan C, Gordon B. 1996. A study of the environmental disclosure practices of Australian corporations. *Accounting and Business Research* **26**(3): 187-199.

Deegan C, Rankin M. 1996. Do Australian companies report environmental news objectively? An analysis of environmental disclosures by firms prosecuted successfully by the Environmental Protection Authority. *Accounting, Auditing & Accountability Journal* **9**(2): 50-67.

Di Vita G. 2004. Renewable resources and waste recycling. *Environmental Modeling and Assessment* **9**: 159-167.

Edgington T et al. 2004. Adopting ontology to facilitate knowledge sharing. *Communications of the ACM* **47**(11): 85-90.

Elkington J. 1994. Towards the sustainable corporation: Win-win-win business strategies for sustainable development. *California Management Review* **36**(2): 90-100.

Esrock SL, Leichty GB. 1998. Social responsibility and corporate Web pages: Self-presentation or agenda-setting? *Public Relations Review* **24**(3): 305-319.

Esrock SL, Leichty GB. 2000. Organization of corporate Web pages: Publics and functions. *Public Relations Review* **26**(3): 327-344.

Fellbaum, C. 1998. *WordNet – An Electronic Lexical Database.* MIT Press, Boston.

Gahleitner E, Behrendt W, Palkoska J, Weippl E. 2005. Knowledge sharing and reuse: On cooperatively creating dynamic ontologies. *16th ACM Conference on Hypertext and Hypermedia (Hypertext-2005).* Salzburg, Austria: ACM Press.

Giora R. 1996. Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory. *Journal of Pragmatics* **27**: 17-34.

Gray RH Javad M, Power DM, Sinclair CD. 2001. Social and environmental disclosure and corporate characteristics: A research note and extension. *Journal of Business Finance and Accounting* **28**: 327-357.

Gruber TR. 1993. A translation approach to portable ontology specification. *Knowledge Acquisition* **5**: 199-220.

Gruber, TR. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies* 43(5-6): 907-928.

Guarino N. 1997. Understanding, building and using ontologies. *International Journal of Human-Computer Studies* **46**(2-3): 293-310.

Haddad WD, Draxler A. 2002. The dynamics of technologies for education. In *Technologies for Education – Potential, Parameters and Prospects*. Haddad WD, Draxler A. (eds). United Nations Educational, Scientific and Cultural Organization: Paris; 1-17.

Hooghiemstra R. 2000. Corporate communication and impression management - New perspectives why companies engage in corporate social reporting. *Journal of Business Ethics* **27**(1/2): 55-68.

Isenmann R, Lenz C. 2001. Customized corporate environmental reporting by internet-based push and pull technologies. *Eco-Management and Auditing* **8**(2): 100-110.

Isenmann R, Marx-Gómez J. 2004. How to provide customized environmental reports properly. In *Environmental Online Communication*, Scharl A (ed). Springer, London; 173-182.

Jarrar M, Meersman R. 2002. Formal Ontology Engineering in the DOGMA Approach, *International Conference on Ontologies, Databases and Applications of Semantics (Lecture Notes in Computer Science, Vol 2519)*. Berlin: Springer. 1238-1254.

Joho H, Sanderson M, Beaulieu M. 2004. A study of user interaction with a concept-based interactive query expansion support tool. In *Advances in Information Retrieval, 26th European Conference on Information Retrieval*, McDonald S (ed). Springer, Berlin; 42-56.

Kolk A. 2005. Environmental reporting by multinationals from the triad: Convergence or Divergence? *Management International Review* **45**: 145-166.

Kolk A, Walhain S, van de Wateringen S. 2001. Environmental reporting by the Fortune Global 250: Exploring the influence of nationality and sector. *Business Strategy and the Environment* **10**(1): 15-28.

Krippendorff K. 1980. *Content Analysis. An Introduction to Its Methodology*. Sage, Beverly Hills.

Kunst M, Witlox N. 1993. Communication and the environment. *Communication Research Trends* **13**(1): 1-31.

Lev B. 1992: Information disclosure strategy. *California Management Review* **34**(4): 9-32.

Liu W, Weichselbraun A, Scharl A, Chang E. 2005. Semi-automatic ontology extension using spreading activation. *Journal of Universal Knowledge Management* **0**(1): 50-58.

Marshall RS, Brown D. 2003. Corporate environmental reporting: What's in a metric? *Business Strategy and the Environment* **12**(2): 87-106.

McManis BL, Ryker R, Cox C. 2001. An examination of Web usage in a global context. *Industrial Management and Data Systems* **101**(8/9): 470-478.

Navigli R, Velardi P. 2005. Structural semantic interconnections: A knowledge-based approach to word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(7): 1075-1086.

Nirenburg, S, Raskin, V. 2004. *Ontological Semantics*. MIT Press, Boston.

Niskala M, Pretes M. 1995. Environmental reporting in Finland: A note on the use of annual reports. *Accounting, Organization and Society* **20**(6): 457-466.

Niskanen J, Neiminen T. 2001. The objectivity of corporate environmental reporting, a study of Finnish listed firms' environmental disclosures. *Business Strategy and the Environment* **10**: 29-37.

Noci G. 2000. Environmental reporting in Italy: current practice and future developments. *Business Strategy and the Environment* **9**(4): 211-223.

Norman W, MacDonald C. 2004. Getting to the bottom of "Triple Bottom Line". *Business Ethics Quarterly* **14**(2): 243-262.

OECD. 2004. *Towards Waste Prevention Performance Indicators*. http://www.oecd.org/dataoecd/12/1/1954291.pdf [20 February 2006]

Palmer JW, Griffith DA. 1998. An emerging model of Web site design for marketing. *Communications of the ACM* **41**(3): 44-51.

Perry M, Bodkin C. 2000. Content analysis of Fortune 100 company Web sites. *Corporate Communications: An International Journal* **5**(2): 87-96.

Pick T, Menger M, Jensen S, Lethen J. 2000. Access to environmental information – Towards a digital global knowledge marketplace. In *14th Symposium Computer Science for Environmental Protection*. Metropolis: Bonn. http://enviroinfo.isep.at/UI%20200/ui2000-etc-cds-paper2.el.hsp.pdf [30 June 2005]

Preuss, L. 2005. Rhetoric and reality of corporate greening: A view from the supply chain management function. *Business Strategy and the Environment* **14**(2): 123-139.

Reicher AK, Webb MS, Thomas EG. 2000. Corporate support for ethical and environmental policies: A financial management perspective. *Journal of Business Ethics* **25**(1): 53-64.

Rockness JW. 1985. An Assessment of the relationship between US corporate environmental performance and disclosure. *Journal of Business Finance & Accounting* **12**(3): 339-354.

Roussinov D, Zhao JL. 2003. Automatic discovery of similarity relationships through Web mining. *Decision Support Systems*, **35**: 149-166.

Runkel M. 2003. Product durability and extended producer responsibility in solid waste management. *Environmental and Resource Economics* **24**: 161-182.

Sanderson M, Croft, WB. 1999. Deriving concept hierarchies from text, *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Berkeley, USA, 206-213.

Scharl A. 2000. *Evolutionary Web Development*. Springer, London.

Scharl A. 2004. Web coverage of renewable energy. In *Environmental Online Communication*, Scharl A (ed). Springer, London; 25-34.

Scharl A, Pollach I, Bauer C. 2003. Determining the semantic orientation of Web-based corpora. In *Intelligent Data Engineering and Automated Learning,* Lecture Notes in Computer Science, Vol. 2690, Liu J, Cheung Y, Yin, H (eds). Springer: Berlin; 840-849.

Schlegelmilch BB, Pollach I. 2005. The perils and opportunities of communicating corporate ethics. *Journal of Marketing Management* **21**: 267-290.

Stanwick SD, Stanwick PA. 2000. The relationship between environmental disclosures and financial performance: an empirical study of US firms. *Eco-Management and Auditing* **7**(4): 155-164.

Stone PJ. 1997. Thematic text analysis: New agendas for analyzing text content. In *Text Analysis for the Social Sciences*, Roberts C (ed). Lawrence Erlbaum, Mahwah; 35-54.

Stray S, Ballantine J. 2000. A sectoral comparison of corporate environmental reporting and disclosure. *Eco-Management and Auditing* **7**(4): 165-177.

Studer R, Benjamins VR, Fensel D. 1998. Knowledge engineering: Principles and methods. *Data and Knowledge Engineering* **25**: 161-197.

Tilt CA. 1997. Environmental policies of major companies: Australian evidence. *British Accounting Review* **29**: 367-394.

UNEP. 1972. *Report of the United Nations Conference on the Human Environment*. http://www.unep.org/Documents/Default.asp?DocumentID=97 [20 February 2006]

Watson M, MacKay J. 2003. Auditing for the environment. *Managerial Auditing Journal* **18**(8): 625-630.

Weaver GR, Treviño LK, Cochran PL. 1999. Corporate ethics practices in the mid-1990s: An empirical study of the Fortune 1000. *Journal of Business Ethics* **18**(3): 283-294.

Weber RP. 1985. *Basic Content Analysis*. Sage, Newbury Park.

Weingart P. 2002. The moment of truth for science. *EMBO Reports* **3**(8): 703-706

Wiseman J. 1982. An evaluation of environmental disclosures made in corporate annual reports. *Accounting, Organizations and Society* **7**(1): 53-63.

Figure 1. Frequency of Waste Management Terms by Sector

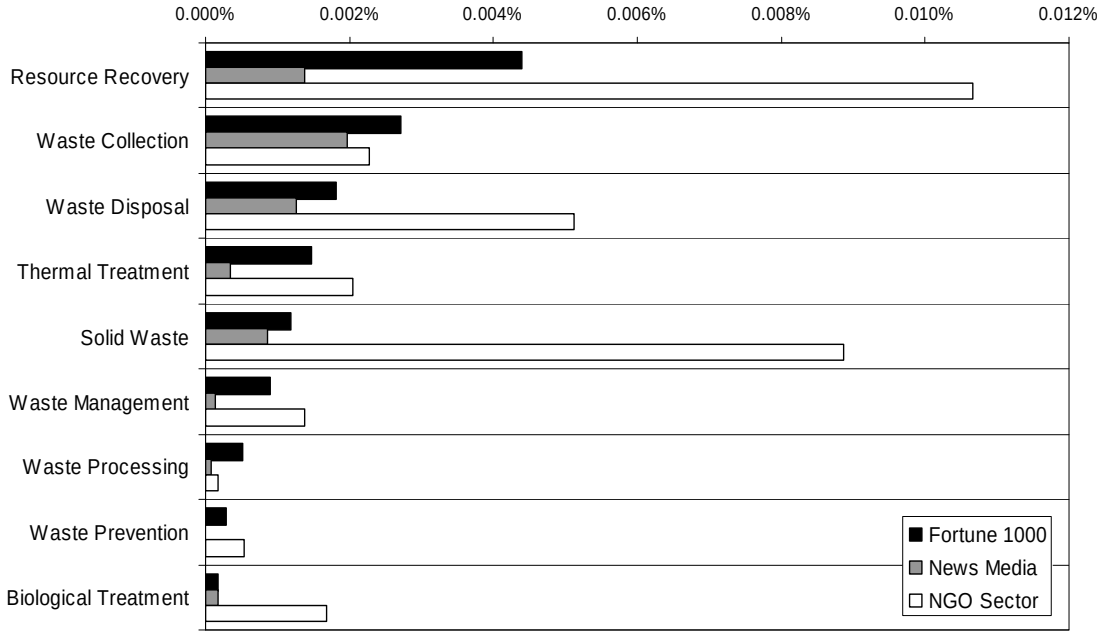

Figure 2. Semantic Orientation toward Waste Management Terms by Sector

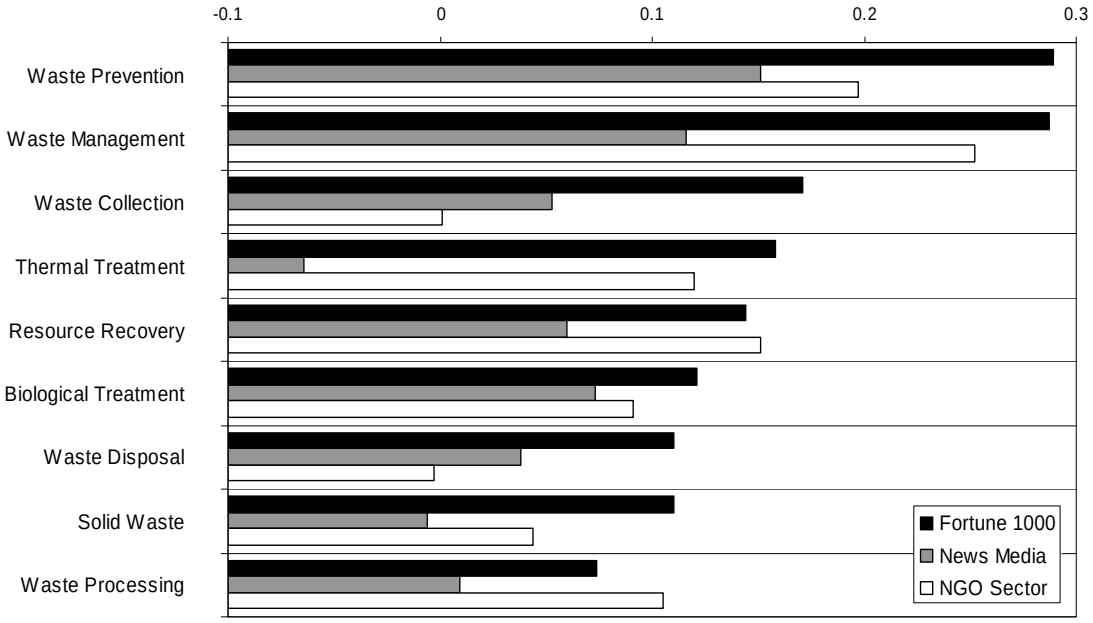

Figure 3. System Architecture of the Ontology Extension Tool (Liu et al., 2005)

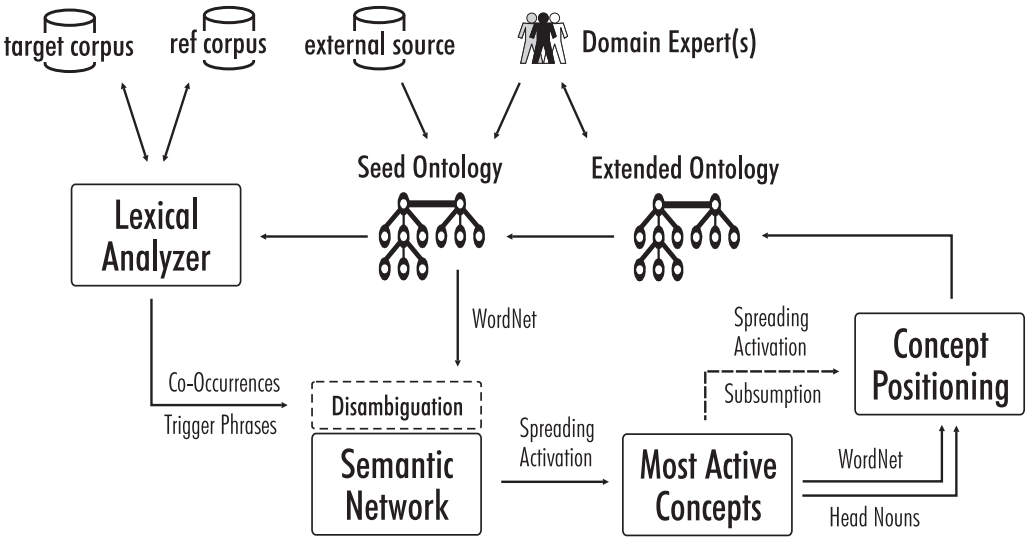

Figure 4. Concept Hierarchy after Two Rounds of Spreading Activation