An Inductive Classification Scheme for Green IT Initiatives

Tim S. McLaren  
*Ryerson University*, tmclaren@ryerson.ca

Priscilla R. Manatsa  
*Ryerson University*, pmanatsa@ryerson.ca

Ron Babin  
*Ryerson University*, rbabin@ryerson.ca

Follow this and additional works at: [http://aisel.aisnet.org/amcis2010](http://aisel. aisnet.org/amcis2010)
An Inductive Classification Scheme for Green IT Initiatives

Tim S. McLaren
Ryerson University
tmclaren@ryerson.ca

Priscilla R. Manatsa
Ryerson University
pmanatsa@ryerson.ca

Ron Babin
Ryerson University
rbabin@ryerson.ca

ABSTRACT
“Green IT” can be loosely described as a set of organizational initiatives undertaken to reduce the environmental impact of Information Technology. Although many trade publications are full of examples of Green IT initiatives, there is a lack of coherence around what constitutes a “Green IT” initiative, which practices could be considered Green IT, and even what the goals and motivations for a Green IT initiative should be. In this paper, we describe a centering resonance text analysis on documentation of the Green IT initiatives undertaken by seven large technology-intensive firms. The findings from the analysis are used to propose a new classification scheme for Green IT initiatives and help bring further clarity to the concepts, goals, and motivations underlying Green IT initiatives.

Keywords
Green Information Technology, Green IT, sustainability, centering resonance analysis, text mining, concept maps.

INTRODUCTION
“Green IT” can be loosely described as a set of organizational initiatives undertaken to reduce the environmental impact of Information Technology. Green IT involves the study and practice of designing, manufacturing, using, and disposing of information and communication technologies (ICT) in a manner that is resource efficient and has little or no impact to the environment. Although many trade publications are full of examples of Green IT initiatives, there is a lack of coherence around what constitutes a “Green IT” initiative, which practices could be considered Green IT, and even what the goals and motivations for a Green IT initiative should be (Murugesan, 2008). For example, some might feel that minimizing impact on the environment should be the key focus of Green IT, with cost reduction as a desirable side effect. Others might feel that conservation of resources should be the main focus of Green IT, while reducing impact on the environment is a secondary benefit.

The goal of this paper is to contribute a clearer understanding of the concepts, practices, goals, and motivations underlying Green IT initiatives. We begin by outlining the prior studies that form the conceptual foundation for this paper. Next, we describe the research methods used to analyze documentation of the Green IT initiatives undertaken by seven large technology-intensive firms. Finally, we present the findings from the text mining analysis and use the results to propose a new conceptualization of Green IT and its underlying constructs. The paper concludes with a brief discussion of how the findings can be used a basis for further research on Green IT.

BACKGROUND
Recent studies of Green IT in the IS literature would suggest that Green IT is a set of initiatives whose goal is to achieve energy efficiency, cost savings, and improved system performance in order to address one of the most important global issues: environmental sustainability (Murugesan, 2008; Olson, 2008). Elliot and Binney (2008) provide an overview of the diverse literature on the business and organizational challenges of sustainability, and propose “the development of a corporate practice for the environmental sustainability of ICT.” Similarly, Melville (2010) calls for “a research agenda for information systems and environmental sustainability [which] focuses on informing beliefs, enabling actions and transforming outcomes.” and to “galvanize IS research on environmental sustainability.” Thus, Green IT initiatives can be seen a specific application of a broader goal of achieving organizational and environmental “sustainability”. Sustainability in
this context has been loosely defined as the ability to meet current needs does not hinder the ability to meet the needs of future generations in terms economic, environmental and social challenges (Carter and Rogers, 2008). Sustainability is a widely studied, inter-disciplinary phenomenon; however, research on its application in IT is still rudimentary. Only within the last few years has the topic invoked discussions in the wider science and technology community (Kates et al. 2005; Hirsch, et al. 2006; Carter and Rogers, 2008). Sustainability is an increasingly important topic of interest in many organizations, though, like the concept of Green IT, it is still misunderstood and inconsistently defined (Molla et al., 2008; Carter and Rogers, 2008). This is mainly because both Green IT and sustainability in general are broad concepts that cover a wide range of issues (Gasparatos et al. 2009; Mebratu, 1998). This lack of clarity is not surprising since they are relatively new phenomena whose definition and usage is still evolving (Gladwin et al. 1995; Carter and Rogers, 2008).

Organizations are motivated to formulate and adopt sustainability initiatives because of the perceived and real market and operational benefits associated with doing so (Bowen et al., 2001). Those incorporating sustainability goals into their overall goals can benefit from the resulting efficiencies. It also encourages a lifecycle, holistic approach in managerial decision making (Bowen et al., 2001). Through reduction of waste, energy conservation, reusing and recycling material, cost savings are realized (Shrivastava, 1995). Other motivations for greening organizational processes and procedures include managing reputational risks (Drumwright 1994; Bowen et. al. 1998), managing liability for environmental damage and avoiding potential increase in cost of waste or waste disposal (Lamming and Hampson 1996; Min and Galle 1997). It can lower the delivery of present and future legislative compliance (Green et al. 1996, Lamming and Hampson 1996; Min and Galle 1997). Further it can improve product or service quality (Cramer 1996; Noci 1997) and allow organizations to meet market expectations (Hutchinson 1996). A study Mathur and Mathur (2000) analyzed stock price reactions after corporate announcements of green marketing activities in 73 firms. Even though investors had reservations about green marketing activities, they were more comfortable and believed in the credibility of the activities of big firms with a strong financial performance and whose operations were designed with environmentally sensitive issues in mind. This means companies have to adopt and practise sustainability initiatives on an ongoing basis, while simultaneously achieving financial success.

Implementing sustainability initiatives can lead to improvements of the triple bottom line i.e. cost, environment and corporate social responsibility (Seuring and Müller, 2008). Figure 1 below shows a representation of the triple bottom line and its components. The concept ‘triple bottom line’ was developed by Elkington (1998) and outlines that firms should focus not only on adding economic value but on adding environmental and social value as well. Hence, the triple bottom line implies that organizations can simultaneously undertake activities affecting the environment and society, and still meet economic goals (Carter and Rogers, 2008). This suggests that when measuring the performance of an organization, environmental and social actions should be measured beyond economic actions, to ensure that the organization is fulfilling its obligations to stakeholders in the long run (Norman and MacDonald, 2004). Examples of stakeholders are shareholders, employees, customers, suppliers, investors, the government and the community at large (Clarkson, 1995). Many organizations are adopting sustainability initiatives such as using environmentally friendly technologies and taking an aggressive stance towards carbon dioxide emissions to remain competitive and demonstrate their level of corporate social responsibility to stakeholders (McWilliams et al. 2006; Molla et al. 2008).
RESEARCH METHODOLOGY

In order to uncover how the concept of Green IT is used in practice, we performed a comparative text mining analysis of Green IT documentation from the corporate websites of seven large IT-intensive firms. We used both computerized and manual searches of the websites to locate documents relating to Green IT. We then used Crawdad text mining software to analyze and compare the Green IT concepts described for each of the seven firms.

The seven firms (Dell, HP, Microsoft, Google, Yahoo, NVidia, and Symantec) were selected for the study using a convenience sampling approach (Miles and Huberman 1994) as they had readily available documentation on the Green IT initiatives they have undertaken, perhaps because their business models highly rely on Information and Communication Technologies (ICT). ICT-intensive organizations may be particularly interested in spearheading corporate sustainability initiatives as they are believed to account for almost 2% of global carbon dioxide emissions, the same amount emitted by the aviation industry (Molla et al. 2008). An initial search for documentation on corporate Green IT initiatives across several other industries failed to produce sufficient documentation for a comparative analysis of that industry.

The selection of individual documents followed a “snowball” approach where additional documents were located as more details were required until the researchers determined that “theoretical saturation” was reached – i.e., when additional documents would add relatively little additional information (Miles and Huberman 1994). The selection criteria for the documents included: (1) the documents were publicly available in order to allow the statements to be scrutinized in public and (2) the documents focused on the firms’ own internal Green IT initiatives that they have already undertaken to help ensure the documents focused on actual rather than hypothetical practices. An initiative was identified as Green IT-related if it appeared to be undertaken to reduce the environmental impact of the firm’s ICT. As in other qualitative analysis approaches, the documents were selected for relevance to the model rather than to be statistically representative of a population (Eisenhardt 1989).

Crawdad text analysis software was used to perform a Centering Resonance Analysis (CRA) on the documents (Corman et al. 2002). The CRA generates a network or matrix of the most influential terms (more specifically, noun phrases) in the documents for each software package individually, as well in aggregate. The CRA networks form a concept map of the most influential terms in the documents. Prior studies have validated the concept maps produced using Crawdad and have shown

Figure 1 - Sustainability: The Triple Bottom Line (Adapted from Carter and Rogers, 2008)
the concept maps produced are very similar to what researchers would produce using more intensive hand coding techniques (Corman et al. 2002; Lee and James 2007). Unlike hand coding, CRA describes the concepts in statistical terms, which are suitable for cluster analysis.

Most prior text analysis software determines word importance from the frequency of the term in the text or across texts. In contrast, CRA uses linguistic centering theory to infer the importance of a term from the amount of influence it has on creating coherence in the text (Corman et al. 2002). In centering theory, important terms are ones that are not necessarily the most frequently used, but ones that form links between other important terms. Thus, even if a word does not appear frequently in the text, if it connects together other influential concepts, it is deemed an influential word. As a result, the list of most influential terms provided by CRA and the concept maps describing their linkages are very similar to what a researcher would produce if they were interpreting the text. In contrast to frequency analysis, CRA more closely mimics the human linguistic process of determining word value from the structural position of the words and their coherence in communicating ideas (Corman et al. 2002; Lee and James 2007).

In essence, the Crawdad text mining software was used to create CRA networks that were similar to a qualitative researcher's concept maps, without requiring researcher interpretation or the lengthy time required to manually code documents. Furthermore, unlike traditional concept maps, the CRA networks are described in quantitative terms (the resonance, distance, and betweenness of terms described in Corman et al., 2002) that are well suited to further statistical techniques such as hierarchical cluster analyses.

Prior to generating the CRA matrices, the documents were converted to plain text files and aggregated into one file for firm. The texts were then pre-processed to remove tables, lists, headers or footers, and any text not related to the firm’s Green IT initiatives, such as descriptions of the firm or its customers. Once the CRA matrices were generated, Crawdad was used to identify the most influential terms in each of the seven files as well as in an aggregate file containing all text documents. Crawdad was also used to generate concept maps for each file, which graphically indicated the most influential terms and their relationships. The following section describes the findings from the CRA text analysis of the key concepts embedded in the documentation of the Green IT initiatives.

RESULTS

In this section, we present the most significant terms found in the CRA text mining analysis of the Green IT documentation of the seven firms. We first describe the individual terms and concept maps that were generated for each firm and then we describe the terms that were most influential across all the documents.

Following Lee (2007), Crawdad CRA software was used to determine the significance of each term from its statistical influence score. A word or word pair received a higher influence score the greater number of times it co-occurs with another influential word. Therefore, the terms with the highest influence scores are those that have the greatest significance or meaning within the document. Using this technique, the Crawdad CRA analysis is able to identify significant words and relationships the words, in a manner that has been found to be similar to what an experienced researcher would do if they were to code the documents and create concept maps manually (Corman et al. 2002; Lee and James 2007). A major benefit of using CRA over hand coding is it enables large volumes of documentation to be coded and analyzed in a much shorter amount of time. The statistical CRA techniques may also highlight word influences and linkages between terms that might not have been as noticeable to a human reading and coding the text.

The primary output of the CRA analysis are the CRA matrices that are computed for each firm’s documentation and contain a list of all the significant terms found in the document as well as the calculated linguistic properties of each term (the resonance, distance, and betweenness of the words). While the individual CRA matrices themselves are hard to interpret at first glance, they can be used to generate concept maps and statistically analyzed further to uncover commonalities and differences between the document matrices.

After computing the CRA matrices for each firm’s collection of Green IT documents, concept maps were generated in Crawdad to show the terms that had influence values greater than 0.015 as well as the structural connections between terms in the texts. The concept maps can reasonably depict only the top 10-20 most influential terms, but they are useful for graphically showing some of the larger similarities and contrasts in highly influential terms, such as the contrast between Google and Nvidia’s Green IT documents shown in Figures 2 and 3).
Figure 2. Crawdad Concept Map for Google

Figure 3. Crawdad Concept Map for Nvidia
The terms highlighted with black boxes had very high influence scores (> 0.2), those in grey boxes had high influence scores (> 0.1) and the remaining terms had moderate influence scores (> 0.015) within the document. Similarly, terms that had very high betweenness scores (a statistical measure of their co-occurrence) were joined with heavy black lines, terms that had high betweenness were joined with medium grey lines, while terms that had moderate betweenness were joined with light grey lines. As can be seen in the examples in Figures 2 and 3, the Green IT documents from Google (primarily a web-based computing services provider) had many similarly influential terms as those from Nvidia (primarily a manufacturer of computer peripherals such as video cards). However, Google’s concept map contains much fewer concepts suggesting their Green IT documentation is highly focused around a few key terms (mostly energy and water related). In contrast, Nvidia’s concept map is much more diverse with many more terms that were influential throughout the documents.

These comparisons and contrasts can be described more precisely by using Crawdad to perform a theme analysis to compare the influence values for each of the top 100 terms in each of the seven document sets. Table 1 excerpts the 25 most influential terms and their influence values for each of the seven firms. The shaded cells highlight influence values greater than 0.2 (words considered highly influential). As can be seen in Table 1, there were not many significant differences between the firm’s usage of the five most influential terms (energy, data, center, server, and power). However, the firm’s clearly placed different emphasis on the next 20 most influential terms. For example, water was a highly influential term in the Green IT documentation from Google and Nvidia, but had negligible influence in the documentation from Dell and Yahoo. Explaining these differences in Green IT documentation is beyond the scope of this paper, we merely call attention to the fact that each of the firm’s appeared to emphasize energy usage initiatives in their Green IT documents, but differed in their emphasis on some of the other components of Green IT, such as paper recycling or water usage.

To explore the commonalities further, a list of the 50 most significant terms in the aggregate documents was generated using Crawdad. Table 2 presents these terms in descending order of influence. As page size limits the concept map to presenting only the 10-20 most influential terms, it is not presented here. As seen in the shaded cells in Table 2, the five most significant or influential terms used in the documents appear to relate to energy usage (Terms 1 and 5) and data center facilities (Terms 2, 3, 4). In addition to generic terms such as “green” or “environmental”, the remaining significant terms in the documents include water, computer, recycling, solar, cost, cooling, carbon, and waste.
Table 1. Word Influence Values by Firm

<table>
<thead>
<tr>
<th></th>
<th>Dell</th>
<th>Google</th>
<th>HP</th>
<th>Microsoft</th>
<th>Nvidia</th>
<th>Symantec</th>
<th>Yahoo</th>
</tr>
</thead>
<tbody>
<tr>
<td>energy</td>
<td>0.17546</td>
<td>0.16184</td>
<td>0.25311</td>
<td>0.02197</td>
<td>0.09858</td>
<td>0.22596</td>
<td>0.20871</td>
</tr>
<tr>
<td>data</td>
<td>0.00108</td>
<td>0.15038</td>
<td>0.18238</td>
<td>0.14118</td>
<td>0.12554</td>
<td>0.19723</td>
<td>0.21806</td>
</tr>
<tr>
<td>center</td>
<td>0.00062</td>
<td>0.06148</td>
<td>0.06054</td>
<td>0.03475</td>
<td>0.00409</td>
<td>0.17607</td>
<td>0.13631</td>
</tr>
<tr>
<td>server</td>
<td>0.00192</td>
<td>0.18461</td>
<td>0.15868</td>
<td>0.02315</td>
<td>0.04167</td>
<td>0.0235</td>
<td></td>
</tr>
<tr>
<td>power</td>
<td>0.03495</td>
<td>0.07821</td>
<td>0.04763</td>
<td>0.08343</td>
<td>0.00005</td>
<td>0.05261</td>
<td>0.06005</td>
</tr>
<tr>
<td>facility</td>
<td>0.00857</td>
<td>0.02214</td>
<td>0.08315</td>
<td>0.00283</td>
<td>0.00854</td>
<td>0.06289</td>
<td>0.01885</td>
</tr>
<tr>
<td>company</td>
<td>0.03061</td>
<td>0.00024</td>
<td>0.03168</td>
<td>0.0673</td>
<td>0.07401</td>
<td>0.01303</td>
<td></td>
</tr>
<tr>
<td>system</td>
<td>0.05789</td>
<td>0.03124</td>
<td>0.1844</td>
<td>0.00103</td>
<td>0.02515</td>
<td>0.01723</td>
<td></td>
</tr>
<tr>
<td>new</td>
<td>0.00722</td>
<td>0.03382</td>
<td>0.00224</td>
<td>0.02973</td>
<td>0.02244</td>
<td>0.04633</td>
<td></td>
</tr>
<tr>
<td>environmental</td>
<td>0.01781</td>
<td>0.02175</td>
<td>0.00128</td>
<td>0.00281</td>
<td>0.24461</td>
<td>0.05023</td>
<td></td>
</tr>
<tr>
<td>program</td>
<td>0.14221</td>
<td>0.0073</td>
<td>0.03306</td>
<td>0.00184</td>
<td>0.03554</td>
<td>0.00877</td>
<td>0.06943</td>
</tr>
<tr>
<td>green</td>
<td>0.02689</td>
<td>0.01294</td>
<td>0.00435</td>
<td>0.01557</td>
<td>0.0202</td>
<td>0.04336</td>
<td>0.01262</td>
</tr>
<tr>
<td>percent</td>
<td>0.00707</td>
<td>0.13597</td>
<td>0.00015</td>
<td>0.00197</td>
<td>0.16504</td>
<td>0.00855</td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>0.00364</td>
<td>0.01782</td>
<td>0.2601</td>
<td>0.0014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>office</td>
<td>0.13971</td>
<td>0.01978</td>
<td>0.00466</td>
<td>0.0994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>product</td>
<td>0.07444</td>
<td>0.00593</td>
<td>0.00107</td>
<td>0.14133</td>
<td>0.02061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>0.05817</td>
<td>0.01388</td>
<td>0.02891</td>
<td>0.00014</td>
<td>0.07593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>space</td>
<td>0.00352</td>
<td>0.12117</td>
<td>0.01012</td>
<td>0.02483</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>operation</td>
<td>0.05153</td>
<td>0.03394</td>
<td>0.03699</td>
<td>0.00931</td>
<td>0.00663</td>
<td></td>
<td></td>
</tr>
<tr>
<td>technology</td>
<td>0.02928</td>
<td>0.00020</td>
<td>0.00004</td>
<td>0.15467</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temperature</td>
<td>0.06392</td>
<td>0.07601</td>
<td>0.01664</td>
<td>0.03282</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Top 50 Most Influential Terms across All Documents

1. Energy 2. Data 3. center 4. server 5. power
11. program 12. green 13. percent 14. water 15. office
16. computer 17. product 18. IT 19. space 20. operation
21. technology 22. temperature 23. employee 24. equipment 25. electricity
26. recycling 27. use 28. management 29. solar 30. cost
31. community 32. practice 33. group 34. PUE1 35. efficiency
36. team 37. fan 38. cooling 39. global 40. library
41. network 42. kWh 43. carbon 44. lab 45. business
46. own 47. part 48. hardware 49. computing 50. waste

---

1 PUE is a commonly used acronym for Power Usage Effectiveness (a measure of computing data center energy efficiency calculated by dividing total data center energy consumption including lighting and cooling by the energy consumption of the IT computing equipment).
CONCLUSIONS AND DISCUSSION

In summary, we have argued that there is currently a lack of coherence around what constitutes a “Green IT” initiative, which practices could be considered Green IT, and even what the goals and motivations for a Green IT initiative should be. This paper contributes a clearer understanding of the concepts underlying Green IT initiatives at seven large ICT-intensive firms. As with most case studies, the seven firms were chosen as they could provide sufficient documentation to conduct the analysis – they were not chosen to statistically represent a larger population. Rather, based on the lack of documented evidence on many firms’ Green IT initiatives, we assume that the seven firms analyzed are likely among the leaders in Green IT initiatives. Regardless, a centering resonance analysis of the Green IT documentation from these firms reveals several interesting phenomena.

The analysis of Green IT documentation revealed a high degree of similarity among the five most significant concepts, which collectively related to energy usage and data center facilities. As shown in the text mining analysis, these two concepts figured prominently in the collected documentation from each of the seven firms (see Table 1). Other significant terms in the documentation that were less prominent included water, computer, recycling, solar, cost, cooling, carbon, and waste, as well as generic terms such as green or sustainability. However, the usage and significance of these additional terms varied from firm to firm, especially when one looks at the usage of the significant terms in the source documents. This suggests that all seven firms primarily conceptualized Green IT initiatives as a focus on energy and data center issues, while differing in their emphasis on related concepts of usage and recycling of computers, water, and paper; carbon emissions, and cost reduction.

Reducing energy usage appeared to be the major focus of the Green IT initiatives in the firm’s studied, while other concepts such as reducing paper usage had much less significant in the documents. This finding was surprising given that the term “green” has a much broader meaning than just energy-related. However, this emphasis on energy issues is consistent with the focus of emerging regulations for IT data centers, such as the European Commission Institute for Energy’s “Code of Conduct on Data Centers Energy Efficiency” (European Commission, 2008). This focus is likely in response to the finding that electricity usage in IT data center servers has doubled over the period 2000 to 2005 (Koomey 2008). Energy usage has not only a large economic impact on a firm's IT operations, but also a large environmental impact since most electricity generation produces greenhouse gases, which have been linked to global warming. The combined economic and environmental impacts of energy usage are substantial contributors to a firm’s triple bottom line performance.

The comparison and clustering of Green IT concepts from the documentation from the seven firms makes an important contribution to knowledge by providing an account of the apparent similarities and latent differences in the conceptualization of the multi-faceted, yet loosely defined phenomena collectively known as “Green IT”. This paper also demonstrates a novel method for comparing and clustering organizational documentation using intelligent text mining tools. The Centering Resonance Analysis generated interesting findings both in terms of the amount of similarity between the documents from each firm, and also in the less obvious differences.

In this paper, we analyzed the firm’s publicly available documentation of their Green IT initiatives as an indicator of the way Green IT is conceptualized and practiced at the firm. We recognize that such documents may not accurately reflect the implemented initiatives in all cases, but the use of publicly available documents from the websites of publicly-traded firms is meant to mitigate this. Furthermore, we believe there is also value in analyzing how a firm describes or markets their Green IT initiatives as this can highlight the relative importance of the underlying concepts in Green IT. Once there is more comprehensive documentation available on more firms’ Green IT initiatives, future research could employ our techniques to analyze the conceptualization and practice of Green IT further. Furthermore, the CRA comparison and clustering techniques could be useful for quickly scanning and evaluating the Green IT practices evident in a firm’s publicly-available documentation.

Finally, we note this paper is meant to serve as a proof-of-concept of a novel and potentially useful technique for analyzing the usage of the Green IT concept in practice. Further research is clearly indicated in order to eventually produce a more definitive conceptualization of Green IT. However, the results of the CRA text mining analysis appear to be congruent with traditional manual coding techniques. Furthermore, the text mining approach facilitates comparisons between the Green IT documentation from several different firms, something that would be extremely time-consuming using manual coding, thematic interpretation, and concept mapping techniques.
REFERENCES