Addressing Climo-Skepticism: Towards an Integrated Repository of Climate Change Research Findings

Herna L Viktor[1] *and Eric Paquet*[1,2]

[1]*School of IT and Engineering, University of Ottawa, ON, K1N 6N5, Canada*
[2]*National Research Council of Canada, Ottawa, ON, K1A 0R6, Canada*
Email: hlviktor@site.uottawa.ca, eric.paquet@nrc-cnrc.gc.ca

Increasingly, the findings of climate change research groups have come under attack by skeptics. Such attacks usually focus on inconsistencies between the findings of different studies and then aim to convince that the changes in climate have been overrated, or are non-existent. This is especially evident in the mass media, where discrepancies are often highlighted and opinions are being biased. It is therefore not surprising that a 2010 study of Ipsos Mori in the UK indicated that merely 31% of adults were convinced that climate change is definitely real, while only another 29% said that it "looks like it could be real" [1]. Furthermore, 2006 polls showed that about half of the populations of the USA (53%), France (51%) and Spain (44%) expressed little or no concern about climate change. In addition, 2009 polls in the USA and China concluded that the general public does not believe that the earth is getting warmer because of human activity, such as burning fossil fuels. This disbelief stands in stark contrast with the scientific findings, as the science academies of all major industrialized countries have now agreed that human activities are, indeed, a major cause of concern.

How, then, to address this climo-skepticism and, perhaps even more worrisome, the lack of public concern? Would the development of a condensed, readily available repository of validated climate change findings help to alleviate this problem? It follows that the creation of such a repository is not a trivial task. However, the data mining, information extraction and retrieval, data visualization and data warehousing communities have developed numerous techniques for addressing the collection, management and analysis of such very large-scale repositories. We aim to utilize this know-how to aid to convince the indifferent, or undecided, that climate change is, indeed, real. To this end, the development of an intelligent web-enabled software environment comes to mind.

Especially, we are interested in applying techniques such as cluster analysis, information retrieval, link and text mining, and content-based search, in order to produce an integrated version of the findings of diverse climate change studies. Next, we discuss some of the components, from a technological perspective, which need to be included in such a system.

1. Managing, analyzing and presenting climate change findings

The first step is to design and implement a data warehouse, to act as a repository when to **store** the climate change finding documents [2]. Such a data warehouse will provide us with an integrated snapshot of findings obtained from various sources, collected over time. To this end, we aim to design an intelligent web search engine, which would start

its search with the findings from known experts. The spider would then crawl the web to automatically find related works, using an authoritative web mining approach [3]. These relevant works are then selected, ranked and inserted into our data warehouse. For each work, we record the authors' details, dates of publication, abstracts and keywords, amongst others. During this phase, we also extract all images or other multimedia components from the text. This information is stored in our data warehouse, in order to allow us to also conduct a multimedia content-based search [4]. It follows that copyrighted materials pose a problem and a solution for this would need to be found.

The second step would consist of **classifying** or **grouping,** the content of the data warehouse into appropriate subject areas, such as e.g. carbon dioxide emissions, sea level rising, ecosystems change, impact on coral reefs, etc. We propose the use of mining techniques such as cluster analysis, classification and association analysis to "group" findings based on these predefined subject areas [4, 5]. That is, at the end of this phase, related work, in terms of subject areas and authors, are positioned into groups of related topics. Of course, this is in essence a multi-labeled learning problem, since findings may belong to more than one grouping. We propose using link-based classification and cluster analysis approaches, which also consider the *relationships* between entities (i.e. the climate change literature, in our case). Link-based methods have been used for hyperlink classification, bibliographic search and social network analysis, and are therefore highly suitable for our task [5].

Next, we are ready to **explore** the climate change data warehouse using text mining methods [5, 6]. There are numerous text mining approaches to be used, ranging from keyword-based methods, tagging approaches and information extraction techniques. The end result of this step would be a "listing" of papers within the data warehouse. These papers thus contain supportive results regarding a topic of interest, such as e.g. the impact on fresh water lakes or hotspot disease outbreaks due to climate change. We also aim to employ our existing content-based indexing and similarity search system to analyze the related images, videos and 3D objects [3], together with the text they were extracted from.

The last, and arguably the most important, component of the proposed system is the part to present of our finding to the climo-skeptics. We propose using a **recommender system** that learns a user's profile and compares it to those of others, to find the k-nearest neighbors. The system then recommends the most appropriate portion of the climate change data warehouse's literature, based on the preferences of earlier users [7]. This may be extended to include a social networking facet, in order to link a climo-skeptic to others who have previously held similar disbeliefs, but have changed their mind after using our system. Our recommender system should also include a content-based similarity search portion to locate and display multimedia objects, since we are of the opinion that the 4000 year old saying "a picture speaks a thousand words" is still very relevant today.

2. Conclusion

This paper describes our initial study regarding the development of an environment designed to provide an integrated, condensed version of the findings of diverse climate change studies. We showed how techniques such as data warehousing, text mining and content-based indexing and similarity search may be used to build such a system. Hopefully, the existence of such an environment may convince the climo-skeptics to change their viewpoints and, subsequently, their potentially harmful behaviors. Particularly, we are interested in targeting the significant portion of the populations that currently say "it is looking like climate change could be real", in order to convince them to change their opinions to "it is definitely real".

References

[1] Ipsos Mori Climate Change Omnibus Survey Results. http://www.ipsos-mori.com/researchpublications/researcharchive/poll.aspx?oItemId=2552 (Accessed 20 April 2010)

[2] R. Kimball et. al, The data warehouse lifecycle toolkit, 2nd Edition, John Wiley & Sons, 2008.

[3] E. Paquet and H.L. Viktor, Visualisation, exploration and characterization of virtual collections, the 20th ISPRS Congress, Istanbul, Turkey, July 12-23, 597-602, 2004.

[4] J Han and M Kamber, Data mining concepts and techniques, 2nd Edition, Morgan Kaufmann, 2006.

[5] Y. Sen and L. Getoor, Link-based classification, University of Maryland Technical Report CS-TR-4858, http://www.cs.umd.edu/projects/linqs/projects/lbc/index.html (Accessed 20 April 2010)

[6] M. Sahami and A. Srivastava, Text Mining: Classification, Clustering, and Applications, Taylor and Francis Group, 2009.

[7] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering **17** (6): 734–749, 2005.