Liam Whalen
The University of Western Ontario

Diane Rasmussen Neal
The University of Western Ontario

Nadine Desrochers
The University of Western Ontario

Kayley Viteo
The University of Western Ontario

Trends, Icons, and Feelings: Notions of Affect within Canadian User-Generated Content

Abstract: This paper will examine how emotion- or affect-based discussions of Canadian content occur within Internet discourse. A random sample of YouTube videos, Blogger.com entries, and Twitter posts will be qualitatively and quantitatively analyzed for connections between searches, tags, comments, and other metadata with the intention of improving information exploration.

Résumé: Cette communication examine comment se manifestent les discussions sur les émotions ou les affects relativement au contenu canadien sur Internet. Un échantillon aléatoire de vidéos YouTube, d'entrées sur Blogger.com et de microbillets sur Twitter a été analysé de façon qualitative et quantitative pour déceler des connections entre les recherches, les étiquettes, les commentaires et autres métadonnées avec l'intention d'améliorer l'exploration de l'information.

1. Introduction

Social media provides a wealth of information relevant to current popular culture. However, the concept of trending, which establishes the popularity of online topics through numbers, views, searches, or access, does not paint an accurate landscape of iconic Canadian content. For example, Justin Bieber is a *trending* person online, while Leonard Cohen is an *iconic* Canadian figure (Desrochers et al 2012).

This project first seeks to determine how the online world discusses Canadian cultural and user-generated content from an emotion- or affect-based perspective. Then, by harnessing emotional responses to content, subject-based methods of search and retrieval can be enhanced, improving "information exploration" (Neal 2010). This concept seeks to take information retrieval beyond the standard "search box" approach and make connections among content in new ways. The first phase of our pilot study, based on a sample of Flickr content, demonstrates that while people frequently post and discuss trending online Canadian content, they may not provide many affective-oriented comments that may be of use for the purpose of enhancing information

exploration. Given these results, this second phase will continue investigating how Canadian trending topics are discussed online through a broader dataset consisting of a variety of content.

2. Literature Review and Background

Identifying Canadian content is a difficult process, complicated by the ambiguity present in attempts to define a Canadian identity. Issues emerging around the notion of a dual settler nation (Mackey 1999; O'Neal 1995), the multiplicity of points of view surrounding the history and tensions between the English and the French populations (Laforest 2004), the concept of multiculturalism and the inherent creation of a sense of "otherness" (Bannerji 2000, 8-10), along with the constant comparison to the United States, especially post 9/11 (Bociurkiw 2011), create an ambiguity that is only exacerbated when trying to apply the label "Canadian" to user-generated content.

In terms of cultural production, the Canadian Radio-Television and Telecommunications Commissions offers guidelines for use in defining Canadian content (CRTC 2008). However, popular culture often seems to trump the guidelines, creating affiliations that fall out of the MAPL system established by this official institution. Likewise, it is difficult to label social media content as Canadian solely on the basis that the person (or group of people) responsible for it is a Canadian citizen – if that information is even available. Could we, then, define an example of social media as Canadian because it represents something that is recognized as inherently Canadian? In other words, is there a democratic process that can be referenced whereby an image, video, or piece of text becomes Canadian, so to speak, because a majority of Canadians value it as representative of our country? These are avenues we will need to consider in the global online context in which we are conducting our research.

3. Methodology

Our research will examine "Canadian" YouTube videos, Blogger.com entries, and Twitter posts in order to understand how Canadian content is discussed online from an emotional/affective perspective. These three sites were chosen because they were among the highest ranked user-generated content sites within Alexa.com's top 25 sites visited by Canadian IP addresses. Alexa.com ranks websites against each other relative to the Internet traffic visiting every site in its index (Alexa 2012). Other user-generated sites were ranked higher than the three we chose, but they either did not provide easy access to their content (e.g. Facebook), or they were unlikely to provide affective comments (e.g. Wikipedia). In order to produce a list of Canadian topics, we examined Google Zeitgeist's 2011 annual summary of the most frequently searched Google terms and generated a list of 8 popular topics associated with Canada: www.census2011.gc.ca, Canada Post Strike, Jack Layton, Vancouver Canucks, Justin Bieber, Blue Jays, Stephen Harper, and *Final Destination 5* (Google 2011).

We will access the videos, blogs, and tweets via Application Programming Interfaces (APIs). APIs provide access to collections of data and methods used to manipulate those data. By providing a list of parameters that can be defined, APIs allow programmers to retrieve and manipulate data that is relevant to their needs. Parameters can specify the keywords to be used in

a search, the number of results to return, a manipulation that modifies the data, and a plethora of other elements limited only by the amount of access to data with which the API's creators want to provide the API's user-base. Our research leverages APIs to collect a random sample of social media content.

Because the types of content on YouTube, Blogger.com, and Twitter are varied (videos, blogs, and tweets), the methods used to gather data from each site will also vary. However, the 8 Canadian terms mentioned previously will form 8 consistent sampling frames for our random samples. A total of 200 items will be selected for every search term. From each of these populations, a sample of 10 items will be selected. This will provide a total of 240 samples from the three social media sites. Within each item retrieved, multiple comments may exist that will need to be analyzed for the presence of affect-related content.

A computer application will be built that will retrieve these samples. To retrieve YouTube content, our 8 search terms will be used as queries to retrieve videos and their associated comments. Our application will search Blogger.com Categories (which are labels assigned by a blogger to a blog post) for the 8 Canadian keywords, and we will then store the contents of each post and the comments about those posts. Twitter content will be selected by using Twitter's Search API to retrieve tweets that contain our chosen keywords. The entirety of each tweet will be stored.

Our collected data will be analyzed via qualitative and quantitative content analysis for information such as the frequency of term occurrences, the relevance of the term to the content, the types and expressions of affect present in the data, and the content's relevance to Canadian culture. By determining the relevance of the collected content to Canadian culture as well as establishing patterns in the emotion-based content, our research will provide insight into the connections between searches, tags, comments, and other metadata for Canadian content on YouTube, Blogger.com, and Twitter. If connections are found, then it could be possible to leverage this content in order to create new methods of information exploration.

4. Limitations
Preliminary studies have resulted in certain terms with large numbers of false positives (content that contains one of the relevant search terms but has nothing to do with Canada). These false positives represent the difficulties of working with Internet data. There are so many people, from various cultures, contributing to social media that some Canadian keyword searches will have high recall but low precision. By identifying these occurrences, our data may provide insight into identifying this situation within search result sets.

5. Conclusion
It is quite clear from our previous findings that a large amount of content related or pertaining to Canada is present in social media; however, using trending, or popularity, as the key factor in determining the sampling frame created a clear bias for the trendy (rather than the iconic). Previous findings also revealed that while using trending Canadian topics did lead to social media content, these topics did not yield many affect-based comments, particularly where connections to Canada were concerned. What is unknown at this point is whether this is specific

to the Flickr culture, to Canadians, or even perhaps to the online context, or simply due to the trending angle pursued in the sampling frame. This second phase of the study, which will harness not only visual but also textual and hybrid content, should reveal whether trending suffices as a parameter to determine perception, or whether other avenues of affiliations should be explored.

References

Alexa. 2012. Frequently asked questions: how can my rank get worse when my site traffic is getting better? Accessed January 15, 2012. http://www.alexa.com/faqs/?p=7991

Alexa. 2012. Top sites in Canada. Accessed January 15, 2012. http://www.alexa.com/topsites/countries/CA

Bannerji, Himani. 2000. *Dark side of the nation*. Toronto: Canadian Scholars' Press.

Bociurkiw, Marusya. 2011. *Feeling Canadian: television, nationalism, and affect*. Waterloo: Wilfrid Laurier University Press.

Canadian Ratio-Television and Telecommunications Commission. 2008. The MAPL System – Defining a Canadian Song. Accessed November 21, 2011. http://www.crtc.gc.ca/eng/info_sht/r1.htm.

Desrochers, Nadine, Diane Rasmussen Neal, Erica Lenton, and Pamela Saliba. 2012. Using Affect-Based Labels in Whole Collection Retrieval. Poster presented at the 2012 ALISE Annual Conference in Dallas, Texas.

Google. 2011. Google Zeitgeist: Global. Accessed January 15, 2012. http://www.googlezeitgeistd.com/en/top-lists/global

Google. 2011. Google Zeitgeist: Canada. Accessed January 15, 2012. http://www.googlezeitgeist.com/en/top-lists/ca

Laforest, Guy. 2004. *Pour la liberté d'une société distincte*. Sainte-Foy: Les Presses de l'Université Laval.

Mackey, Eva. 1999. *The house of difference: cultural politics and national identity in Canada*. London, New York: Routledge.

Neal, Diane. 2010. Breaking in and out of the silos: what makes for a happy photograph cluster? Paper presented at the 2010 Document Academy Conference (DOCAM '10), University of North Texas, Denton, Texas, USA.

O'Neal, Brian. 1995. *Distinct society: origins, interpretations, implications*. [Ottawa]: Library of Parliament, Research Branch.