Frank Lambert
University of Western Ontario

The *mycommunityinfo.ca* Approach to Online Networked Community Information Provision

Abstract: *Mycommunityinfo.ca* offers online community information (CI) by rejecting the traditional CI directory model of excessive metadata. This dynamic and sustainable approach to providing online CI through single window access to local community and three levels of government information sources also offers an unprecedented glimpse into CI needs in Southwestern Ontario.

Résumé : *Mycommunityinfo.ca* offre de l'information communautaire (IC) virtuelle en rejetant le modèle traditionnel des répertoires de l'IC utilisant de nombreuses métadonnées. Cette approche dynamique et renouvelable offrant de l'IC virtuelle, grâce à une fenêtre unique pour l'accès à la communauté locale et à trois niveaux de sources d'information gouvernementales, propose également une perspective inédite sur les besoins en IC du sud-ouest de l'Ontario.

1. Introduction

Mycommunityinfo.ca (MCI) delivers a simple, innovative, cost-effective and sustainable approach to providing online community information to residents of Middlesex County and the City of London during a time when community information (CI) providers are faced with the extreme difficulty of remaining financially viable. MCI was first conceived in 1999 when representatives of various ministries, departments and agencies from the municipal, provincial and federal governments met to address the question of how these groups could provide "…a cost-effective integration of information service offerings from three levels of government" (Cummings, 2004, 1) in the Middlesex-London region of southwestern Ontario. Virtually from the outset, the proposed solution for providing access to community information (CI) in an *electronic* format was to eschew the traditional, considerably more expensive and difficult to maintain networked directory model of organized, static links to government and non-profit organization information resources that have to be monitored on an ongoing basis by a complement of paid staff. Instead, a search engine technology approach for indexing and retrieval purposes that specifically targets local public sector Web sites through its own index was adopted. This offers many benefits to the information seeker as the online directory CI model assumes a search method where he or she does not have a formal expression of information needs, as he or she is able to navigate such Web sites through the chain of links found in the sites' pages. "However, when some specific information is searched, this point-and-click access paradigm is unpractical, and the effectiveness of the results strongly depends on the starting page." (Herrera-Viedma & Pasi, 2006, 511) While search engines have their own challenges, MCI's approach allows the information seeker to formally express their needs to the best of their abilities without having to attempt to match the Web site's metadata to retrieve relevant information. This is particularly useful when one considers that "databases [directories] are always built in the language, syntax, and semantics of the database builders and have targeted user bases. As a result, it is difficult to search them using natural language." (Cummings, 2005) Additionally, queries launched from MCI's search engine portal may be aimed at provincial and federal government Web sites with the click of a button, offering true

single window access to information sources from all three levels of government and local community public organizations. The technical launch of MCI occurred in October 2002 with the official media launch following in May 2003 (Cummings, 2004).

MCI operates out of the City of London's Technology Services Division (TSD), an arrangement that works well since MCI's technology relies very much on the skills of TSD's technicians. Ongoing operational governance of MCI is by consensus through a Senior Advisory or Steering Committee that is comprised of regional Provincial Directors from several ministries, Chief Administrative Officers and other senior municipal staff (Cummings, 2004). Delegates from the Steering Committee form the Working Group, who advises when the senior group should meet. The Project Manager serves as the Secretary to both groups. MCI is an unincorporated not for profit organization funded by tax dollars. There is no Board of Governors that separates it from the founding partners. All "…strategic and operational decisions are based on [the] goodwill and consensus of Federal, Provincial and Municipal partners because all agree that citizen-centric service is a primary government objective." (Cummings, 2004, 1).

From October 2003 to October 2004, 381,809 query submissions were sent through MCI's Google Search Appliance from the main MCI search engine, as were queries submitted through the London Police Services, London Public Library, City of London, County of Middlesex and Region of Waterloo Web sites (*Municipal Starter Kit*, 2005). The latter three Web sites also include an option of launching a search of the whole MCI collection of Web sites if the user did not find the desired information results from their in-site search. As of early 2005, nearly 170 local municipal governments, their associated agencies and local non-profit community organizations in London and Middlesex County (~150) and in the Region of Waterloo (~20) have had their Web sites indexed and made accessible through MCI's search engine. This paper examines only the first four month's queries launched from the Web sites mentioned above where, from August to November 2005, nearly 150,000 unique queries were recorded in a Microsoft Access database. This is but a small sample of what will be a one-year case study for a LIS dissertation.

The notion of applying a Web-based technology, in particular a search engine, to form the essential component in assisting the local population find community information instead of adopting a directory approach with a large number of organized fixed links is a novel approach. It deserves further investigation and analysis since: 1) this model deviates from the traditional LIS model of the public library being one of, if not the only, primary source of a community's information (the topic of a large body of research in the literature) without wholly displacing it; 2) its online community information service is a much more dynamic information structure from all three levels of government with the added contribution of local community and non-profit groups; and, 3) community information services are also offered through human intermediaries, provided in partnership with MCI through other community and governmental organizations without imposing the financial burden of staff training, salary, benefits and other compensatory expenses.

After examining one year's worth of query log data, meeting and discussing various important issues with staff members including MCI's Project Manager and TSD technicians and seeing how quickly traffic to its Web site and services were growing on a monthly basis, a study was designed and subsequently approved in mid 2005 to form the basis of the author's dissertation. In this paper, preliminary findings based on the analysis of a small sample of queries collected through MCI's search engine, three

municipal Web sites and the Web site of a municipal agency, all of which are indexed by MCI, are presented with the focus of the analysis covering August to November 2005. The goal of this paper, and one of the four goals in the dissertation, is to answer the following research questions based on using Web log analysis as a research method: What types of CI are most being sought in an online environment through MCI's CI service approach? Are there differences in the types of CI being sought between MCI and other local online services affiliated with MCI? Will the collection and analysis of this data show different results than the findings of other researchers who rely primarily on other approaches (i.e. Durrance & Pettigrew, 2002)?

2- Literature Review

Durrance and Pettigrew (2002) note that there is a lack of comprehensive data regarding citizen's use of networked CI services. In-depth examinations are also lacking regarding citizen's information behaviour in networked CI environments. While they make these observations within the context of "the types of situations that prompt citizens to seek CI online, how the information helps, and the role of intermediaries such as reference librarians" (Durrance and Pettigrew, 2002, 144), they do not express any specific concern over CI users' online navigational or querying behaviour. The research questions are designed to help fill the voids identified by Durrance and Pettigrew primarily through the analysis of Web query logs. This will generally follow the approach that Savolainen (1998) defines as a non-work study of use that focuses on finding information to solve everyday problems (340-341).

Until 2002, studies that examined what types of CI people are searching for have reported limited areas of topicality. Using household interviews, focus groups and survey instruments, Bishop, Tidline, Shoemaker and Salela (1999) examined community information use and computer access of a low-income locale in the United States populated primarily by African-Americans. Their sample group reported seven subject areas of information that they would like to have access to online: community services and activities, resources for children, healthcare, education, employment, crime and safety and general reference tools such as dictionaries (372).

Through the use of online surveys posted on three community networks (CNs) across the United States (NorthStarNet in Northeastern Illinois, Three Rivers Free-Net (TRFN) in Pittsburgh, PA and CascadeLink in Portland, OR), Durrance and Pettigrew (2002) found that electronic CI needs are far more diverse than those reported by Bishop, Tidline, Shoemaker and Salela, discerning twenty categories of interest from their respondents (n = 197) (2). All three of these CNs follow the model of a directory of organized fixed links to Web sites. Notable differences between Durrance and Pettigrew's findings and CI studies reported prior to the Internet lie in searching for employment opportunities, volunteerism, the availability of social services, items for sale, local history and genealogy, local news, computer and technical information and other people (Durrance and Pettigrew, 2002, 23). Pettigrew, Durrance and Unruh (2002) were able to supplement this survey data with follow-up in-depth interviews with users to determine: "(1) how the public is using networked CI systems for daily problem solving, (2) the types of barriers users encounter, and (3) how individuals and physical communities are befitting [*sic – benefiting?*] as a result of public library-community networking initiatives and the emergence of information communities." (Pettigrew, Durrance and Unruh, 2002, 896) However, the data collection methods used by

Durrance and Pettigrew (2002) or Pettigrew, Durrance and Unruh (2002) do not reveal any ranking of their results as to the urgency of what people of the communities under investigation are most concerned with finding to meet their information needs. Also, no studies published thus far give any indication of how successful the participants were in finding the information they needed.

What is particularly interesting in the context of the current study is Durrance and Pettigrew's report of various "information-related" barriers that arose from the follow-up interviews with 27 of the original survey respondents. One of these barriers was information overload "[d]ue to poor search engines and site indexing," resulting in the retrieval of so many "hits" that the user would be sidetracked from his/her original task. Some of the other relevant barriers were poor organization (classification), out-of-date and inaccurate information, missing information and dead links (Durrance and Pettigrew, 2002, 9).

In a separate but related article, Pettigrew and Durrance (2001) suggest, based on their findings, fourteen ways in which online CI systems may be improved (Pettigrew and Durrance, 2001, 141-142). What is most obvious about these suggestions is how task- or work-oriented they are, requiring many hours of labour from those responsible for maintaining the CI system (for example, indicate when the CI on a page was last updated and the source and credentials of the CI source, use Sales' (1994) taxonomy to organize and index CI records and make it available online). Other suggestions are potentially beyond the available skill sets of those who work with the CI system on a daily basis (for example, improve the capability of search engines and search fields, incorporate anticipatory search fields, incorporate features that allow the systems to be used by users with slower computers) (Pettigrew and Durrance, 2001, 142). While the suggested improvements are worthy and with foundation, the cost associated with implementing all or even some of these upgrades would itself be a significant barrier for most CI organizations.

Two research streams have emerged in Web log analysis. One is the examination of users' navigation behaviour within a Web site as they move from one page to another. The other is scrutinizing search engine log files where data about what the user is actually interested in is recorded (Thelwall, Vaughan & Björneborn, 2005, 95). Space restrictions for this paper only allow for consideration of the latter.

Research published on the analysis of Web query logs has focused on querying behaviours through commercial search engines. It is this literature that is most relevant to the current study. Silverstein, Henzinger, Marais and Moricz (1999) examined a data set recorded over 43 days from the AltaVista search engine comprising over 900 million total requests. Their interest lay in determining which queries were most common, the average length of query strings, how many queries were submitted during an individual session and, especially, correlations between query terms and other field values (Silverstein, Henzinger, Marais & Moricz, 1999, 6). Among other things, they found that only the first screen result is viewed for 85% of the submitted queries and that 77% of the sessions contained only one unmodified query. The authors concluded that the "average Web user differs significantly from the user model assumed by the information retrieval community." (Silverstein, Henzinger, Marais & Moricz, 1999, 12).

Spink, Wolfram, Jansen, & Saracevic (2001) conducted a similar study involving over one million query logs (531,416 of which were unique) sent to the Excite search engine on one day by 211,063 users. They examined the number of queries submitted per

identified user (48.4% submitted a single query, 20.8% two queries and 31% three or more queries – the mean was 2.16), measured the change in unique queries submitted by each identified user, the number of results pages viewed (the median number was 8) and whether multi-term queries used advanced search features such as Boolean operators (less than 5%).

There are further, relevant studies related to this topic, but they will be considered in context with the data collection and analysis methods for the study.

3 – MCI's Technology and the Web Query Logs

MCI's Google search appliance[i], the primary tool it uses to provide its service, is located on the 6th Floor of London City Centre, which houses the City of London's Technology Services Division (TSD). MCI purchased its latest model, a GB-1001, in October 2004, on a two-year service contract for $32,000 (U.S). It connects to the Internet via LARG*net, whose Internet switch is located at the University of Western Ontario (S. Cummings, personal communication, February 8, 2005). This primary search appliance is operated and maintained by TSD who is also responsible for a second search appliance. This second appliance's primary function is to index information on the City of London Intranet, serving only this one domain behind a firewall and inaccessible by the greater population. It also works as a backup unit to the primary MCI appliance should it suffer a malfunction, greatly reducing any "down" time from days to hours (Cummings, 2004). MCI's search appliance uses the Google search appliance (gsa)-crawler to retrieve a new copy of each accessible page of the targeted or specified Web sites. (Mycommunityinfo.ca). The crawl frequency for the latest search appliance is different than MCI's first model; the original version initiated a complete crawl of the Web sites in its index daily whereas the crawl frequency of the 2004 model determines the crawl frequency using a timing algorithm. Regardless of the frequency, MCI's index is always extremely up to date. Each of the search appliances were originally licensed with the capacity to crawl up to 150,000 URLs or Web pages, with the MCI search appliance currently crawling about 110,000 URLs when it updates its index. In 2006, Google announced that when MCI renews its license for its search appliance, the cost would be reduced to $30,000 for two years and the crawling capacity was immediately increased to 500,000 Web pages (Personal communication, March 3, 2006). The crawl takes anywhere from six to eight hours to complete using a two-megabyte Internet connection. The current model has the potential to crawl up to 1.5 million documents. To gain even more capacity, MCI would simply pay Google a higher licensing fee. Google, in turn, would send an alphanumeric key that unlocks the extra capacity (Cummings, 2004). This scalability means that not only can MCI easily expand its service offerings from within the immediate geographical area, but additional regions outside of London and Middlesex County may also be added with little effort as was demonstrated with the addition of the Region of Waterloo Web sites in 2004.

The search appliance captures query data from all queries submitted through MCI's search engine and through the in-site search capability of five municipal Web sites (the City of London, London Police Services, the Region of Waterloo, the County of Middlesex and London Public Library, although this last Web site has not been using the MCI search bar since a Web site redesign went live in December 2005). Only the London Police Services Web site does not offer the option of widening the search to the whole collection of Web sites indexed by MCI through the MCI search bar. This is due

to political, rather than technological, reasons, as MCI is still trying to convince all participating municipalities, boards & commissions that they should also incorporate the MCI search bar on their Web sites (S. Cummings, personal communication, February 4, 2005). When conducting an *in-site* search of one of the five municipal Web sites, once the information seeker has entered his or her query and presses the "GO" or "SEARCH" button, the search is automatically routed to MCI's search appliance, asking it to find a server on the Web that, for instance, contains a collection of Web pages called "City of London" or "Region of Waterloo". This is accomplished through the use of various Java and C++ scripts in the Web site's in-site search bars that have been written by TSD technicians. Tied in to this, the technicians have written additional scripts that capture this query data from the search appliance that then "dumps" the query and other relevant information into an Access database. So, rather than using an off-the-shelf Web log analysis solution such as WebTrends, MCI, with the cooperation of TSD, has adopted a customized solution to provide a valuable added value service to its partners and clients.

The relevant query log data captured and then recorded in the Access database is comprised of eleven attributes, all of which are contained in one table. These attributes include: 1) the *time and date* that the query was received; 2) the *actual text of the query* as submitted by the user; 3) the *site* field, which identifies the Web site from which the search was launched; 4) the *type of search* attribute which has two values: "0" for MCI's search appliance or "1" for Google Web API. When a search is recorded as the former, it indicates that the search is targeting the roughly 170 Web sites indexed by MCI through their search appliance as well as for the municipal in-site searches since these Web sites are already included in MCI's index. The second value is recorded *only* when a searcher expands their search from MCI's local collection to Canadian federal or Ontario provincial Web sites, with the automatic domain restriction added on to limit the search to ".gc.ca" or ".gov.on.ca" domains; 5) the *estimated number of results* returned which is the number of "hits" that result from the information seeker's query; 6) the *referrer* attribute which records the referring URL accompanying the incoming query to MCI's search appliance from the request header of the site issuing the request; 7) the *start page* which is a pagination of the number of pages of returned hits that the user has examined; 8) an artificial *session identification number* that is assigned every time the MCI search appliance is "touched" by an incoming query request, be it from an in-site search or a search from the MCI search engine; 9) the number of submitted queries or *searches* per session; 10) the *site search* attribute which has two values: "community" for queries targeting Web sites in MCI's local index (using MCI's search engine); and "site search" that indicates the search was launched on a municipal Web site from its own search bar, to search only its collection of Web pages; and 11) the incoming query's *IP address*, which is not provided to the researcher due to privacy concerns.

It should also be noted that cookies are not used to identify a unique user session (Thelwall, Vaughan & Björneborn, 2005, 95; Rubin, 2001). This is a policy decision made primarily by London's TSD. They view placing a cookie on a user's hard drive as something akin to government tracking the information seeker's every move while using online resources regardless of what they are viewing (Rubin, 2001, 208).

4 - Data Collection and Analysis Methods

MCI has been sending the Web query log data to the author on a monthly basis via email in support of the dissertation. This allows for easier cleaning of smaller query data sets

on an ongoing basis. At the end of the case study period (July 2006), the full year's worth of collected data will be analyzed to provide an overall elucidation of the CI topics that are searched for by the community.

The MCI query logs, like Ross and Wolfram's study (2000), include many examples of identical queries being submitted by the same session identifier in succession. While they assumed that "[t]hese submissions most likely represent the request to view the next page of hits for the query" (951), this is definitely the case with the MCI data. These duplicates are created by additional page views of returned hits, as identified by the session number, and thus must be deleted to create a new Microsoft Access table of "clean" queries for frequency analysis. To work with only those raw queries that are duplicates, two 'find duplicates' Access queries were created: the first pulled out all of the duplicate query session numbers; then the second Access query pulled out all of the duplicate query text grouped within these sessions. Following this, the duplicate queries based on the text were manually deleted. While this was very time consuming (30-40 hours per month of data), it afforded a level of accuracy that would not have been achieved through other, programmable means. Additionally, as much other data recorded in the respective attributes in the Access table wanted to be preserved so that, through the creation of carefully designed Access queries, it would allow for easier and fuller analysis of the query logs.

Following the parsing of multi-term queries, Ross and Wolfram (2000) used the frequency of binary term co-occurrence to determine facets of multi-term queries without having to look at every query. This was a bottom-up grounded theory method with no *a priori* categorization beforehand (Ross and Wolfram, 2000, 951). Pu, Chuang and Yang (2002) used this same approach to develop their subject taxonomy for the automatic classification of Web query terms into broad subject categories (620, 617). An empirical approach of gathering the words and terms that actually occur in the data collected by MCI, based on their frequencies, and then grouping them together in their fundamental categories or facets is being used in this study to determine *what sorts of CI people are searching for online*. It should be stressed that, while Lancaster advocates such a method to create a formal thesaurus of controlled vocabulary terms (1986, 23-24), that is *not* at all the intent of this procedure. Ross and Wolfram (2000) used the binary co-occurrence of multiple term (word) queries for their analysis with the reasoning that "…the most frequently used single search terms cannot take into account the context of the search terms... The study of cooccurring terms, therefore, provides a better understanding of query topics." (Ross and Wolfram, 2000, 950) It should be noted that Ross and Wolfram's data came from the Excite search engine, which offers substantially different options for searching the WWW. Since quotation marks may be used in query formulation, Google allows the searcher to construct more complex queries that could combine terms with single or multiple words; for example, '"real estate" listings' or '"real estate" listings [AND] agents' (with an automatic AND Boolean operator being implied). In the latter example, a searcher is still looking for housing that he or she may want to purchase. They have just happened to qualify this query by including a "property" facet word ('listings') and an "agent" facet word ('agents') that are qualified by the term 'real estate'. Thus parsing the queries for co-occurrence analysis with the MCI data becomes more problematic, with the potential of affecting the analysis of the original intent or meaning of the user's query. As a result, the study relies on using a simple frequency occurrence analysis for single and multiple word and term queries to group these types of queries into their predominant categories. Besides, nearly 77% of the top ten most frequently occurring queries captured by MCI are single word queries

(Appendix A). Thus, their frequencies are also examined in context with the higher frequency multiple word query occurrences.

The ten most frequently occurring queries were coded with categories based on the conceptual grouping of the query terms. To test the reliability of the coding, an associate of the author engaged in the same exercise. Two rounds of coding, with changes to the categories following the first round, both resulted in inter-coding reliability in the lower eighth decile. Following a discussion, it was determined that this relatively low reliability was due to different knowledge of the local community as well as the ambiguity of some of the queries. The discussion resulted in the final table of nineteen categories as presented in Appendix B. One of the other difficulties in coding such data, especially single word queries, is that some of the queries could potentially belong to more than one category. However, the Web site through which the query was submitted did provide some clue to the intent of the information seeker and offered some assistance with the categorization.

5 – Results – MCI's Web Query Log Data

The results of the categorization effort are displayed in Figure 1. Underneath each of the italicized categories are the relevant query word(s) or term(s) in bold as submitted exactly by information seekers with, enclosed in parentheses, the Web site that the query was submitted through and the rank of the query frequency for that Web site out of ten. The actual frequencies for the queries are displayed in Appendix A.

Not surprisingly, information about *Job seeking/opportunities* is most sought after with queries related to this subject being submitted through all of the Web sites mentioned with the exception of MCI Middlesex-London. This demonstrates what may be a lack of knowledge on behalf of the user as to the extent and scope of coverage that this search engine indexes compared to the other Web sites if one is searching for job openings in general. If one is searching the London Police Services Web site for employment, perhaps they are indeed interested in a career in policing and this would obviously be the Web site of choice. However, in-site searches of the other Web sites are limited *only* to those Web sites and the MCI Region of Waterloo search engine is limited to only twenty Web sites. In the case of the latter example, the municipal Web sites of Kitchener, Waterloo and Cambridge are part of this collection, thus potentially increasing the likelihood of returning relevant information for the topic of job opportunities. What is astonishing is that the search engine with the greatest potential for returning the most information for job openings, MCI Middlesex-London, did not record any relevant queries in the top ten for this subject; after all, it indexes 150 Web sites thus making it a far larger collection. The only other conclusion that may be made is that those who are looking for information related to finding employment must really want to work for their respective municipality.

Interesting patterns begin to emerge when the types of queries submitted through the municipal Web sites that serve primarily rural populations, County of Middlesex and Region of Waterloo, are compared. Not including the major urban areas contained within each of these areas (London for County of Middlesex and Kitchener, Waterloo and Cambridge for Region of Waterloo), both geographic areas have roughly the same population (roughly 55-60,000 residents).

Figure 1. Categorization of queries from MCI search engine, 3 municipal Web sites and 1 municipal agency Web site

Legend: "CoL" = City of London; "RoW" = Region of Waterloo; "CoM" = County of Middlesex; "LPS" = London Police Services; "MCI-ML" = *mycommunityinfo.ca* search engine covering Middlesex-London; "MCI-RW" = *mycommunityinfo.ca* search engine covering Region of Waterloo

Animal Care & Control	*Community Activities & Events*	*Crime*	*Education/Training*
Pet adoption (MCI-ML 10)	**Spectrum** (CoL 1) **Santa Claus parade** (CoL 7) **Kiwanis** (MCI-RW 3) **Events** (MCI-RW 4) **Swimming** (MCI-RW 9)	**Graffiti** (LPS 5) **Fraud** (LPS 6) **Drugs** (LPS 8) **Wanted** (LPS 9)	**Education** (CoM 4) **Schools** (CoM 5) **School** (CoM 9)
Health Information	*Housing*	*Information & Reference Tools*	*Job Seeking/ Opportunities*
Flu shot (RoW 1; MCI-RW 10) **Flu shots** (RoW 3; MCI-RW 5) **Flu clinics** (RoW 4)	**Real estate** (MCI-ML 9)	**Maps** (RoW 8)	**Employment** (CoL 2; CoM 2; RoW 7; LPS 4; MCI-RW 8) **Jobs** (CoL 3; CoM 1; LPS 7) **Employment opportunities** (CoM 7)
Leisure/ Entertainment	*Local Attractions*	*Policies, Bylaws & Regulations*	*Neighbourhood Services*
Restaurants (CoL 4)	**Jail** (CoM 3) **County jail** (CoM 8)	**Parking** (CoL 10) **Zoning** (CoM 6) **Noise** (LPS 3) **Water** (RoW 6)	**Library** (CoL 8)
Place Names	*Private Sector*	*Statistical Information*	*Taxation*
Strathroy (CoM 10) **Toronto** (RoW 2) **Cambridge** (MCI-RW 1) **Kitchener** (MCI-RW 7)	**Greyhound** (RoW 5) **Funeral homes** (MCI-RW 6)	**Population** (CoL 5; MCI-RW 2) **Statistics** (LPS 10)	**Middlesex farm taxes** (MCI-ML 1) **Farm taxes** (MCI-ML 2) **Tax** (MCI-ML 6) **Taxes** (MCI-ML 7)
Transactions	*Waste & Waste Reduction*	*Other*	
Auction (LPS 1) **Auctions** (LPS 2)	**Garbage** (CoL 6) **Recycling** (RoW 9) **Yard waste** (RoW 10)	**Licence** (MCI-ML 3) **License** (MCI-ML 5) **Program** (MCI-ML 4) **Programme** (MCI-ML 8) **Safety** (CoL 9)	

However, their most urgent information needs diverge quite significantly with residents of County of Middlesex being primarily concerned about taxation and Region of Waterloo citizens more anxious about health concerns, primarily influenza. Considering the problems that farmers have recently been encountering with falling incomes, it seems odd that information needs that have some focus on *Taxation* would not be a priority for most rural dwellers that rely mainly on a farm income. Either the latest round of municipal tax increases, if any, were not a concern in Region of Waterloo or there was genuine alarm through the population over the prospect of an influenza epidemic in this

area. Another interesting note with the Region of Waterloo is the number of queries that imply some need to find travel information. This is demonstrated by the *Private sector* subject query 'Greyhound' and the *Place names* subject query 'Toronto', which was the second most submitted query through the Region's Web site. Finding information about how to reach this particular destination seems to be of significant value for the information seekers on this Web site.

CI seekers using the City of London's Web site seem to have much different information priorities than their rural cousins. *Community activities and events* take precedence, exemplified by their search for an online version of the *Spectrum* magazine that lists and describes a large and diverse number of programs and recreational activities sponsored by the city for little or sometimes no cost to its residents that is published biannually. *Leisure/entertainment* based queries take precedence over more "serious" queries such as those that are topically focused in *Education/training*, *Health information* and *Housing*. Even when it comes to searching for CI on the London Police Services Web site, Londoner's priorities lie with what may be more broadly described as recreational activities such as police auctions of goods seized from criminals even though, for the purposes of this study, such activities are better defined as *transactions* between local government and its citizens. In fact, crime hardly seems to be a major concern in London with the most searched for criminal activity, graffiti, ranking fifth and much more serious crimes not ranking in the top ten at all.

Inquiries over regular municipal services and activities also prevail through querying activity. *Waste and waste reduction* is a relatively important concern for the City of London and Region of Waterloo, although this may be in relation to pick-up schedules or what materials may or may not be disposed of through the respective waste management agencies. *Policies, bylaws and regulations* also rate some importance. While queries about 'water' through the Region of Waterloo Web site might seem ambiguous and lead to questions as to why this query is placed in this category, cleaning the raw query data gives some strong clues. Almost all instances of this query were sent in August when watering restrictions are still in place. Indeed, 'water restrictions' and other variations occurred rather frequently. Since watering restrictions are a municipal regulation, 'water' may be placed here with some degree of confidence.

What results from this initial analysis is not only a diversity of CI needs from three relatively close communities in southwestern Ontario (County of Middlesex, London and Region of Waterloo with the latter being one hour's drive east of the former two) but even distinct differences may be seen between two communities that share similar demographic characteristics. Previous studies either focused on only one community (Bishop, Tidline, Shoemaker & Salela, 1999) or, for those that examined multiple communities (Durrance & Pettigrew, 2002), no distinction was made of the CI needs for each municipality. Also, no previous studies have been able to conclude what types of CI are of the highest priority for the communities under examination. By and large, Bishop, Tidline, Shoemaker and Salela's findings (1999) are confirmed by this study by demonstrating that online community networks may indeed address the needs of the community they studied. Durrance and Pettigrew (2002) found more specific examples of CI needs sought for on an online environment but many of the categories are similar to this study's. Their methods allowed for deeper probing into their research participants' CI needs. However, considering the sample size of those interviewed (n = 27) from three rather large communities, it may be difficult to argue that these CI needs are transferable to the three respective populations. Using an unobtrusive method such

as Web log analysis also reduces the potential for any reactivity since those using MCI are not aware that they are under observation and that, should they inquire whether they were under observation, that their identities are completely anonymous.

6 - Concluding Remarks

This study presented a small sample of preliminary results of the Web log analysis of CI needs expressed as natural text queries submitted by users through a CI provider that not only operates its own search engine, but also is responsible for the indexing and retrieval capabilities of three municipal Web sites and two municipal agencies' Web sites. MCI is always looking for ways to expand its CI service offerings and may do so in a way that is not only inexpensive for itself but for other communities as well. For instance, Region of Waterloo, MCI's only subscriber, pays merely $15,000 a year to have up to 40,000 public sector-based Web pages in its geographic region indexed and made available. When this is compared to the cost for the Region of Niagara to participate in the 211Niagara initiative ($135,000 for 2006 to contribute to the overall annual budget of $556,000 (211 Niagara Steering Committee, March 2004, 5-6)), MCI must seem like a bargain. Additionally, MCI is not a bottomless money pit consuming tax dollars for it to maintain its operations. In 2001, MCI received start-up funding of $206,760 from the federal government through Human Resources and Development Canada, $155,000 from the Province of Ontario through the Management Board Secretariat and "in-kind" funding from the City of London and County of Middlesex that, in total, matched the province's amount. As of early 2006, MCI is still operating on the start up funding issued five years ago.

MCI easily has the capacity and the capability to expand its services to any upper-tier municipality. With its current configuration and client agreements, MCI could sign up ten additional client communities resulting in roughly $165,000 in annual revenues, an amount that would likely result in rebates back to the client communities so that MCI could maintain its non-profit status. In other words, MCI has the potential to be a completely self-sustainable CI provider that could devote all of its time to servicing its clients and, consequently, its client's citizens rather than having to devote significant amounts of time chasing after various government grants in order to keep its "doors" open.

When the cost of subscribing to MCI's service is taken into consideration along with the added value of being able to know what sort of CI that one's citizens are actually searching for, MCI's potential becomes practically priceless. However, Web log analysis is something that should not be taken lightly. Web site administrators are not likely to have the time or the expertise to fully evaluate all of the log data that various log analysis software, whether off the shelf or custom made, may produce (Nicholas, Huntington, Williams, Lievesley, Dobrowolski & Withey, 1999). However, MCI operates virtually on its own, automatically updating its index with regular crawls of its partner and client's Web pages. This allows the Project Manager to collect and examine these log statistics and inform his client and partner communities about the types of CI that their citizens are most concerned with finding. Indeed, the MCI approach, or other models similar to its approach, may well be the solution to preserving online CI organizations and avoiding the wrecks of past failed projects that are littering the online CI highway.

References

211 Niagara Steering Committee. March 2004. *211 service for the Niagara Region: Business plan overview*. http://www.informationniagara.com/Business_ Plan_Overview.pdf (accessed February 5, 2006).

Bishop, A.P., T.T. Tidline, S. Shoemaker & P. Salela. 1999. Public libraries and networked information services in low-income communities. *Library and Information Science Research*, 21, no. 3: 361-390.

Cummings, Stephen. 2004. *Mycommunityinfo.ca.* London, Ont.: mycommunityinfo.ca.

Cummings, Stephen. Winter 2005. Powered by Google. In *Library Journal NetConnect Supplement,* http://www.libraryjournal.com/article/CA490055 (accessed February. 23, 2005).

Durrance, Joan C. & Karen E. Pettigrew. 2002. *Online community information: Creating a nexus at your library.* Chicago: American Library Association.

Herrera-Viedma, E. & G. Pasi. 2006. Soft approaches to information retrieval and information access on the Web: An introduction to the special topic section. *Journal of the American Society for Information Science and Technology*, 57, no. 4: 511-514.

Lancaster, F.W. 1986. *Vocabulary control for information retrieval*, 2nd ed. Arlington, Virginia: Information Resources Press.

Municipal Starter Kit. 2005. London, Ont.: mycommunityinfo.ca.

Mycommunityinfo.ca. n.d. *FAQs for Web authors and Web site administrators.* http://www.mycommunityinfo.ca/signup/notes.htm (accessed February 10, 2005).

Nicholas, D., P. Huntington, P. Williams, N. Lievesley, T. Dobrowolski & R. Withey. 1999. Developing and testing methods to determine the use of web sites: Case study newspapers. *Aslib Proceedings.* 51, no. 5: 144-154.

Pettigrew, K.E. & J.C. Durrance. 2001. Public use of digital community information systems: Findings from a recent study with implications for system design. In *International Conference on Digital Libraries: Proceedings of the 1st ACM/IEEE Joint Conference on Digital Libraries*, *June 24-28, 2001*, 136-143. Roanoke, Virginia. New York: ACM Press.

Pettigrew, K.E., J.C. Durrance & K.T. Unruh. 2002. Facilitating community information seeking using the Internet: Findings from three public library-community network systems. *Journal of the American Society for Information Science and Technology.* 53, no. 11: 894-903.

Pu, H., S. Chuang & C. Yang. 2002. Subject categorization of query terms for exploring Web users' search interests. *Journal of the American Society for Information Science and Technology*. 53, no. 8: 617-630.

Ross, N.C.M. & D. Wolfram. 2000. End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine. *Journal of the American Society for Information Science.* 51, no. 10: 949-958.

Rubin, J.H. 2001. Introduction to log analysis techniques: Methods for evaluating networked services. In *Evaluating networked information services: Techniques, policy, and issues*, edited by C. R. McClure & J. C. Bertot, 197-212. Medford, New Jersey: Information Today.

Sales, G. 1994. *A taxonomy of human services: A conceptual framework with standardized terminology and definitions for the field*, 3rd ed. El Monte, California: Information and Referral Federation of Los Angeles County, Inc.

Savolainen, R. 1998. Use studies of electronic networks: A review of empirical research approaches and challenges for their development. *Journal of Documentation*. 54, no. 3: 332-351.

Silverstein, C., M. Henzinger, H. Marais & M. Moricz. 1999. Analysis of a very large Web search engine query log. *SIGIR Forum*. 33, no. 1: 6-12.

Spink, A., D. Wolfram, M.B.J. Jansen & T. Saracevic. 2001. Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology*. 52, no. 3: 226-234.

Thelwall, M., L. Vaughan & L. Björneborn. 2005. Webometrics. In *Annual Review of Information Science and Technology, 39*, edited by Blaise Cronin, 81-135. Medford, NJ: Information Today.

Appendix A

Duplicate "Cleaned" Queries from City of London Web Site

QueryText Field	NumberOfDups
spectrum	1108
employment	704
jobs	644
restaurants	578
population	560
garbage	467
santa claus parade	361
library	319
safety	266
parking	255

Duplicate "Cleaned" Queries from London Police Services Web Site

QueryText Field	NumberOfDups
auction	160
auctions	53
noise	27
employment	25
graffiti	25
fraud	23
jobs	17
drugs	16
wanted	16
statistics	15

Duplicate "Cleaned" Queries from County of Middlesex Web Site

QueryText Field	NumberOfDups
jobs	72
employment	69
jail	27
education	18
schools	17
zoning	16
employment opportunities	15
County Jail	12
school	11
strathroy	10

Duplicate "Cleaned" Queries from MCI Search Engine, London & Middlesex Collection

QueryText Field	NumberOfDups
middlesex farm taxes	556
farm taxes	553
licence	538
program	531
license	495
tax	493
taxes	491
programme	482
real estate	452
pet adoption	404

Duplicate "Cleaned" Queries from Region of Waterloo Web Site

QueryText Field	NumberOfDups
flu shot	156
toronto	127
flu shots	126
flu clinics	125
Greyhound	124
WATER	120
employment	119
maps	113
recycling	113
yard waste	107

Duplicate "Cleaned" Queries from MCI Search Engine, Region of Waterloo Collection

QueryText Field	NumberOfDups
cambridge	61
dog (N/A-TEST QUERY)	39
population	7
kiwanis	7
events	7
flu shots	6
funeral homes	6
kitchener	6
employment	6
swimming	5
flu shot	5

Appendix B – Community Information Categories

Categories	Notes
Animal Care & Control	Anything to do with animals; pets, adoption, pest control, pet licences
Community Activities & Events	Activities and events organized and held in the local community involving community organizations or government agencies that residents may participate in; involves either registration on an ongoing basis or they occur sporadically or on a periodic basis
Crime	Anything related to criminal activity(ies) including investigations, crime scene information and media produced in relation to criminal activity(ies)
Education/Training	Anything related to education, job training, etc.
Health Information	
Housing	
Information & Reference Tools	Tools to find information about the local community that are searched for through MCI, municipal Web sites
Job Seeking/Opportunities	
Leisure/Entertainment	Activities performed by individuals in their leisure time other than community activities & events
Local Attractions	Tourist attractions including historical sites
Policies, Bylaws & Regulations	Policies &/or bylaws &/or regulations created by a government conducive to the effective governance of a political entity
Neighbourhood Services	Services provided by any level of government through its annual budget and funded by tax dollars
Place Names	The names of places, communities, towns, municipalities
Private Sector	Any information request about the private sector or organizations that exist in the private sector including names of businesses
Statistical Information	Quantitative information about a politically defined entity and its associated agencies and departments
Taxation	Raising/collecting of revenues by a level of government
Transactions	Occurrences where residents engage in some form of transactional activity with any level of government; usually, but not always, involves the exchange of money. Does not include instances of taxation be it property taxation or otherwise
Waste & Waste Reduction	Anything related to waste produced by people and how the government, local or otherwise, deals with it
Other	

[i] More information about Google search appliances may be found at http://www.google.ca/enterprise/gsa/index.html