ABSTRACT
This paper presents the Faceted Retrieval of E-Government Documents (FRED) system and describes the ongoing user-centred design process in which we are engaged, to develop and refine the system. FRED is an experimental vertical search engine for Canadian federal government content. The paper will highlight the interface design process as well as summarizing the guiding principles that emerged from a multi-year study of e-government search behavior.

Author Keywords
Electronic government, faceted search, user-centred design

ACM Classification Keywords
H.3.4 Information Storage and Retrieval: Systems and Software
H.5.2 Information Interfaces and Presentation: User Interfaces: User-centered Design

INTRODUCTION
A major motivation for the move to electronic government (e-government) is the opportunity to deliver government information and services to the public directly and efficiently. In Canada, as in many countries, e-government initiatives have resulted in the creation of thousands of websites, repositories, portals, and more recently, social media sites, containing government content for use by the public. Unfortunately, this wealth of content is not truly accessible to the public unless they have the tools and skills needed to find it and make use of it, which is not always the case. In Canada, poor findability of federal e-government content has been identified as a major public concern [4].

The design of effective search systems for e-government is challenging, and largely due to the convergence of several factors. First, access to government information is the right of all citizens regardless of their position in society. As a result, systems must be designed for the widest possible spectrum of users. Second, the majority of the user community has low domain knowledge and limited search experience with e-government content. Typically, members of the general public have little knowledge of the structures of government, interact with government information only occasionally, and search for it much less frequently than for other kinds of online information, such as news, or health information [2]. Third, government information is often complex, reflecting the nature of government itself: websites are large and layered with different kinds of content written for diverse audiences and purposes, and documents are often lengthy and jargon-laden. At the same time, there are also advantages to designing for the e-government domain. Although heterogeneous in nature, government information is more consistent in form and function than general web content; further, the application of metadata standards and templates ensures that certain conventions are generally followed in its description. In addition, the work tasks and information tasks that prompt searching within this domain follow relatively clear and consistent patterns.

To date, e-government search tools have done little to respond to the specific challenges and opportunities of e-government search. In the Canadian context, e-government search tools are generic and lack sophistication with most employing only simple site search. By way of comparison, other national governments have developed more sophisticated search systems, such as Search.usa.gov, powered by Microsoft’s Bing, and Australia.au.gov, powered by the FunnelBack search engine. Nevertheless, numerous studies have indicated that general commercial search engines, particularly Google, are the first choice for searching for e-government content [11, 2], even though these do not provide dedicated access to government content.

In this paper, we present FRED, an experimental domain-specific search engine for Canadian federal government content. We describe the ongoing user-centred design process in which we are engaged to develop and refine the system. The paper highlights the design goals that emerged from a multi-year study of e-government search behavior and provides an overview of the iterative design process.

PRIOR RESEARCH
The design of e-government search systems has not received substantial attention from the research community [9]. A small number of studies focus on particular application areas, such as statistical [8] or election information [9] or on particular features of e-government search, such as the display of metasearch results [13] or faceted search [10]; however, general design frameworks are lacking.
Studies of user needs and behaviours indicate a steady, if somewhat slow, uptake of e-government by the public. In 2009, 82% of American internet users searched for information or used services online [11]. Common e-government activities are to search for information on agencies and services, download forms, and search for employment; the levels of direct communication with or participation in e-government remain low [11].

Since 2008, our research group has conducted a series of studies focused on public access to e-government information through the ‘E-Informing the Public’ Project1. Completed studies include: two analyses of metadata use in the Government of Canada (GOC) domain; a study of e-government content production through interviews and document analysis; a survey of librarians’ perceptions of the impact of e-government on public access to information; and three user studies examining selection criteria for e-government content, e-government web search behaviour and an experimental study of faceted searching using a version of our FRED prototype. The results of these studies have informed the initial design and ongoing development of the FRED system, and have allowed us to identify a set of design goals to guide the development of system tools and features. More information on the E-Informing the Public project and a full list of publications are available from the project website.

**DESIGN GOALS**

Our five design goals are grounded in characteristics of the e-government domain, although they may be relevant to other domains as well.

1. **Capitalize on what defines e-government as a distinct information use environment (IUE).**

   Taylor [12] considers an IUE to consist of the people, problems, acceptable solutions, and characteristics of the setting that influence the flow and use of information in that domain. We study these elements of the e-government domain in order to exploit patterns and consistencies and provide specialized features and tools.

2. **Design for the whole public**

   E-government information carries with it the requirement that it be accessible by all members of the public, despite widely varying levels of skill, knowledge, and motivation. Our approach is to provide users with an intuitive interface that requires little expertise to use effectively, along with tools that allow for easy customization of the interface according to a user’s level of skill or knowledge.

3. **Move beyond search to information interaction and use**

   Searching for government information is often part of a broader task or project, which may include activities such as reading, making sense of numerous documents, filling out forms, visiting physical service points, and sharing content with others. By providing tools and workspaces to help accomplish these tasks, we can increase the value of the system to the user.

4. **Provide an unobtrusive learning environment for the user**

   Rather than functioning as a black box, an e-government search system should provide the user with opportunities to learn about the structure of government and the nature of government information through interaction with the system. The system should expose and feed information to the user and provide tools to support exploration and engagement, thereby contributing to an informed public.

5. **Create a social information space in support of civic engagement**

   An important goal of e-government is to facilitate civic engagement. The search system should support this by facilitating contact and interaction between the government and users, and within and between the community of system users through social features.

The next section describes the design process for FRED, highlighting the relationship between the Design Goals and current and future versions of the system.

**SYSTEM DESCRIPTION**

FRED was designed as a testing environment for e-government search features and functionality. The baseline system was constructed using Apache Solr (3.5.0), which has built-in support for faceted search, and the Apache Nutch (1.4) Web crawler. The Nutch index-metatags plugin is used to fetch metadata, which is then filtered and mapped to the facet schema displayed on the interface using the Nutch-Solr field mapping component.

![Figure 1: Version 1.0 of the FRED Interface](image)

**Version 1.0**

The initial version of FRED (Figure 1) was created in response to design goals 1, 2 and 4. We chose a faceted

---

1 http://diigubc.ca/research/egovernment/
search approach because facets expose the particular characteristics of the domain to the searcher, making use of the available metadata, and enable exploration and learning [14]. Facets offer support to the novice searcher and greater control to experts [10].

The main design challenge for FRED 1.0 was to determine how to make use of the metadata to create the search facets. The GOC has a Metadata Standard that requires certain elements, but offers considerable latitude to departments in the use of other elements. We analyzed the metadata available in a limited crawl of approximately 500 thousand pages from 24 federal department and agency websites and identified three elements with potential value as search facets: Subject, Audience and Type (i.e. document type). More detail on this analysis is available in [3].

Two additional facets are drawn from other sources of information: Department is extracted from the URL and Location is extracted from the page content itself and mapped to the GOC Regions of Canada Schema.

One of the challenges we faced in using author-assigned metadata to provide search facets was the great diversity of values in the dataset, which extends beyond what a typical searcher might make sense of. We found approximately 40 different values for Audience, ranging from broad categories such as “Parents” to highly specific ones such as “Pension Agents”; and 229 values for Type, such as Budget, Case Study and Media Release. We conducted a card-sorting focus group in which the researchers worked with two participants to identify groupings and assign labels to values within the Department, Audience and Type Facets. The result is 4 top level categories for 24 Departments; 4 top level categories for 20 aggregated Audience values, and 14 top level categories for 107 aggregated Type values.

The Subject facet proved too unwieldy for the card-sorting exercise, as the corpus yielded 6,000 unique values. We decided that this facet would introduce complexity to the interface without adding much value over the keyword search function, and removed it from the facet schema.

Version 1.1

This next version of FRED did not introduce new features, but instead refined the existing interface in preparation for user testing. In order to ensure a basic level of usability we conducted an expert assessment in the form of a group Cognitive Walkthrough [7]. Two assessors worked together through a series of five tasks simulating different types of searches, some of which required use of the facets. As a result, a number of changes were made to the interface, including: more use of colour and a new logo, a clear indication of the content collection being searched, the addition of a query prompt in the searchbox, a help file, a wider pane to display search results, greater consistency in snippet length, and setting the default to display the top level of facets. The revised interface is presented in Figure 2.

User testing, based on this version of the system, is now underway, and is focusing on the relationship between different e-government search task types and the use of facets. The study will also collect data on usability of the system and the facets in general.

Version 2.0

We are still in the planning stages of Version 2.0 of FRED, which will address our design goals more fully and improve the underlying functionality. An important goal in Version 2.0 is to incorporate an automatic classification component that will use the meta-tagged content as training data to assign Audience and Type metadata to the entire corpus. We have also identified a number of interface features that will contribute to our design goals (Table 1). Some of these features are presented in Figure 3, a design prototype for Version 2.0.
Proposed Interface Features | Goals
--- | ---
Reading Tools, These include annotation and highlighting tools, as well as look-up tools linked to other GOC information tools, such as staff directories, dictionaries, etc. These types of tools have been developed and implemented in other information use contexts [6]. | 2, 3, 4
Information Feeds: query dependent display of GOC press releases, open data updates, social media feeds. | 4, 5
Tabbed search results option showing results by Department or by Type. This type of display has been found to be effective for e-government [13]. | 1, 2, 4
Tools to contact government, including a GIS feature to provide maps showing local government offices. | 3, 5
The ability for users to control interface complexity and a reading level filter. Prior work [1, 5] indicates that older and low-literate users can benefit from such tools. | 2
Task-based query suggestion feature, to support common e-government information tasks | 1
Social information tools: tagging, rating, sharing, forum for Q&A | 5
An information management workspace for taking notes and saving and viewing files. | 3, 4

Table 1: Interface features mapped to design goals

CONCLUSION

FRED is a work in progress and serves as a convenient testing environment for e-government search features and tools. The main contribution of this research to date is the implementation of facets based on existing metadata. An evaluation of faceted and non-faceted versions of the interface under different task conditions is now underway. The future potential of FRED lies in the implementation and testing of some of the more novel features we have envisioned, based on our design goals: the reading tools and learning and social components. Given the central role that access to information plays in our society, the value of designing effective, inclusive, engaging and dedicated search systems for government information cannot be overestimated.

ACKNOWLEDGEMENTS

This work is supported by the Social Sciences and Humanities Research Council of Canada and by the GRAND Network of Centres of Excellence.

REFERENCES