Designing A Better Web Portal for Digital Government: A Web-mining Based Approach

Xiao Fang
Department of Operations, Information and Technology Management
The University of Toledo
+1-419-530-7724
xiao.fang@utoledo.edu

Olivia R. Liu Sheng
School of Accounting and Information Systems
The University of Utah
+1-801-585-9071
olivia.sheng@business.utah.edu

ABSTRACT
This paper addresses an important research area in digital government development: intelligent design of digital government Web portals, a primary means of realizing the core function of digital government – online government service delivery. In particular, we propose and study the e-service selection problem and introduce a web-mining based approach to e-service selection. The proposed approach is evaluated using real world data collected from the Utah State Government Web portal.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Data Mining; H5.2 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms
Digital government web portal, e-service selection

Keywords
ServiceFinder, web mining

1. Introduction

Digital government (hereafter called *e-government*) can be defined as the use of adequate information and telecommunication technologies to support political conduct of government and enhance its services to citizens and various constituent communities [1]. The core function of current e-government practices is to provide a wide range of government services through Web portals [1], namely *e-services*. Critical to e-government development is how to deliver an array of e-services to citizens and constituent communities through Web portals effectively. As quoted from former Utah State Governor Mike Leavitt, "our goal is to use technology (i.e., Web portal) to make state government services more efficient and accessible" [2].

A common way of fining e-services in an e-government portal is to use site-specific search engines. As pointed out in [3], using search engines suffers from the disadvantage of returning too much irrelevant information. A wealth of research has attempted to address this disadvantage [4]. An alternative popular way of seeking e-services is to surf through hyperlinks. This research focuses on enhancing the effectiveness of seeking e-services through hyperlink surfing. As users click through a set of hyperlinks to find their desired e-services, placing appropriate hyperlinks in Web pages is critical in improving their searching effectiveness. In particular, this research studies how to select a small number of service-links (i.e., hyperlink pointing to a e-services) for the homepage of an e-government portal.

The homepage of an e-government portal is the starting point for seeking e-services. Usually, a small number of service-links are selected and featured in the homepage. E-government portals with a good selection of featured service-links guide users to locate e-services they seek effectively and attract more users; while e-government portals with a poor selection of featured service-links make e-service searching difficult and might lose users who are time pressed or impatient [5]. As such, the time and costs required of providing or obtaining government information or services increase due to in-effective selection of e-government portal links. Hence, an important problem in e-government portal development is how to select a small number of service-links to be featured in the homepage of an e-government portal so that users can be directed to find e-services they seek effectively. We name this problem *e-service selection*.

A well-designed e-government portal normally features 5 to10 service-links in its homepage. Currently, a typical e-government portal provides 100-200 e-services. And the number of e-services provided in an e-government portal grows over time as more and more government agencies provide e-services. It is computationally too expensive to exhaust all combinations of several service-links from a pool of several hundred service-links and find the set that is the most effective in guiding users to locate e-services they seek. For the case of Utah.gov, the number of combinations of selecting 6 service-links out of 145 service-links is 1.16E+10 (i.e., C_{145}^{6}). Current practice of e-service selection relies on domain experts' (e.g., webmasters) experiences, namely *expert selection*. Obviously, expert selection is subjective. In addition, it reflects domain experts' perspectives on what service-links should be selected, not users' perspectives. The latter should be emphasized because the main purpose of e-service selection is to reduce users' e-service searching efforts, not domain experts'. In this research, we propose metrics to measure the quality of e-service selection and formally define the e-service selection problem. ServiceFinder, a heuristic solution to the e-service selection problem, is then presented and evaluated using real world data collected from Utah.gov.

2. The E-Service Selection Problem

We propose three metrics to measure the quality of e-service selection: effectiveness, efficiency and utilization. All three metrics are calculated using web logs. A web log can be broken down into sessions with each session representing a sequence of consecutive web accesses by the same visitor. Effectiveness is measured as the *degree of easiness* to locate *user-desired e-services*. We denote S as a set of service-links provided in an e-government portal and S_C as a set of chosen service-links, where $S_C \subset S$. Obviously e-services pointed to by service-links in S_C, which are featured in the homepage, can be easily located by Web surfers. For a service-link s, where $s \in S$, we denote $X(s)$ as a set of service-links placed

in the web page pointed to by s. We define service set S_X as $S_X = \bigcup_{\forall s \in S_C} X(s)$. Usually, a Web surfer explores one level from the homepage of an e-government portal, if a user's desired e-service cannot be located from S_C, by clicking on a related service-link in S_c. Therefore, e-services pointed to by service-links S_U, where $S_U = S_C \cup S_X$, are considered to be easily located.

User-desired e-services are e-services visited in sessions. We denote n as the number of sessions in a web log, and e_i as a session, for $i = 1,2,\cdots n$. We denote $D(e_i)$ as a set of service-links pointing to e-services visited in e_i.

Definition 1: For a session e_i, the session level effectiveness $effectiveness(e_i)$, is

$$effectiveness(e_i) = \frac{|D(e_i) \cap S_U|}{|D(e_i)|}. \quad (1)$$

And the log level effectiveness, $effectiveness(\log)$, is

$$effectiveness(\log) = \frac{\sum_{i=1}^{n} effectiveness(e_i)}{n}. \quad (2)$$

Definition 2: For a session e_i, the session level efficiency, $efficiency(e_i)$, is

$$efficiency(e_i) = \frac{|D(e_i) \cap S_U|}{|S_C|}. \quad (3)$$

And the log level efficiency, $effeciency(\log)$, is

$$efficiency(\log) = \frac{\sum_{i=1}^{n} efficiency(e_i)}{n}. \quad (4)$$

Definition 3: The log level utilization, $utilization(\log)$, is

$$utilization(\log) = \sum_{i=1}^{n} |D(e_i) \cap S_U|. \quad (5)$$

Based on the three metrics presented above, we formally define the e-service selection problem as below.

Given: (1) a set of service-links provided in an e-government portal, S;
(2) the number of service-links to be featured in the homepage of an e-government portal, N,
where $N < |S|$.
Select N service-links from S.
Objective: maximize effectiveness, efficiency and utilization.

4. ServiceFinder

In this section, we present a heuristic method, namely ServiceFinder for the e-service selection problem defined in Section 3. ServiceFinder is based on patterns extracted from the structure of an e- government portal and those mined from a Web log, which records Web surfers' behaviors.

Input: S, a set of service-links provided in an e-government portal;
N, the number of service-links to be featured in the homepage of an e-government portal
Output: S_C, a set of selected service-links
Discover structure relationships;
Discover access relationships;
Calculate preferences of individual service-links;
Calculate preferences of service-link sets;
Select N service-links.

Figure 1: Sketch of ServiceFinder

5. Performance Evaluation

To evaluate the performance of ServiceFinder, we compare it with expert selection. One-month Web log collected from Utah.gov in November 2003 was used for the experiment. The Web log contained 49 million records. The Web log was divided into 313,780 sessions using the widely applied rule of thumb, in which the maximal session length cannot exceed 30 minutes. 232,842 sessions of Web log, which spanned the first 23 days of November 2003, were used as the training data, and the remaining 80,938 sessions in the last 7 days of the November Web log were used as the testing data.

	ServiceFinder	Expert Selection		
		Maximum	Minimum	Mean
Effectiveness	0.853983	0.677307	0.02754	0.19664
Efficiency	0.166129	0.133071	0.005394	0.03868
Utilization	16076	12877	522	3743

Table 1: Performance Comparison Between ServiceFinder and Expert Selection

Acknowledgements

We gratefully thank the Utah State Government and Utah Interactive, for kindly providing us the data and information used in this study.

References:

1. A.K. Elmagarmid, and W.J., Jr. McIver, **"**The ongoing march toward digital government," *IEEE Computer*, vol. 34, no. 2, Feb. 2001, pp.32 – 38.
2. Center for Digital Government "Utah State Portal Ranks No. 1," http://www.centerdigitalgov.com/center/highlightstory.phtml?docid=69811, 2003.
3. B. A. Huberman et al., "Strong regularities in World Wide Web Surfing," *Science*, vol. 280, no. 3, pp 95-97, 1998.
4. J. Chakrabarti, "Data mining for hypertext: a tutorial survey," *ACM SIGKDD Explorations*, vol. 1, no. 2, pp1-11, 2000.
5. J. Nielson, and A. Wagner, "User interface design for the WWW, " In *Proceedings of ACM CHI 96*, British Columbia, Canada, Apr. 1996, pp.330-331.