

Digital Government Portal: A Tool for Digital Government Research

Hsinchun Chen Yilu Zhou Jialun Qin Catherine Larson
Department of Management Information Systems, University of Arizona, Tucson, Arizona

The project has been supported by a grant from NSF: H. Chen, PI, DGPort: Mapping Knowledge in the Digital Government Domain, NSF 0302353, January 2003-January 2004.

Introduction

DGPort is an intelligent Web portal for digital government researchers developed in the Artificial Intelligence (AI) Lab at the University of Arizona. We will highlight the project accomplishments and goals in this paper.

While the Web has provided a convenient means for sharing research ideas, digital government researchers often find themselves facing the problem of information overload (Blair et al., 1985), in which a search on a general-purpose search engine such as Google results in thousands of hits. In addition, the diversity of research perspectives and applications areas has resulted in terminology and vocabulary differences. Researchers in different disciplines use different terminologies to describe their findings, and user search terms may be different from the indexing terminologies used in databases. Information sources and quality on the web are equally diverse (Lawrence & Giles, 1999), forcing researchers to go to multiple information sources (Web sites, search engines, online academic databases, etc.) and conduct significant manual analysis to identify quality, time-critical information. The lack of an integrated, “one-stop shopping” service is an urgent problem to be resolved.

To address the above problem, our research project aims to explore various issues in developing a digital government Web portal and human interaction with such automated systems. We report our experience in implementing and evaluating a prototype of a digital government Web portal—DGPort.

DGPort System Architecture

DGPort integrates various information sources as well as search and analysis capabilities for researchers to access government content with more personalized and comprehensive service.

Intelligent Searching Component. DGPort supports searching of a customized digital government research database. Based on extensive research, we identified then collected 3,912 quality pages, based on URL seeds suggested by experts. This initial collection building includes pages from NSF, NHPRC, DGRC, COS, Grants.gov. SpidersRUs, a digital library building toolkit developed in the AI Lab, was used to automatically download 310,389 Web pages using the seeds.

In addition to searching its own database, DGPort supports meta-searching of 11 digital government information sources suggested by domain experts. These sources include online research databases (Scirus, CiteSeer), online journals (FirstGov, GovTech, ACM, FCW, CIO), online news providers (CNN, BBC), and major search engines (Yahoo, AltaVista). The use of meta-search ensures comprehensiveness and recency in covering the domain and reduces information biases.

Keyword Suggestion. To facilitate searching with different keywords, DGPort provides a keyword suggestion function, developed using the concept space approach (Chen & Lynch, 1992). This function identifies pairs of keywords co-occurring on the same page and extracts them for use as thesaurus terms in our database. Each query to the DGPort concept space thesaurus elicits a ranked list of keywords highly related to the input keyword.

Document Summarization. The DGPort summarizer, a modified version of a text summarizer called TXTRACTOR (McDonald & Chen, 2002), can flexibly summarize Web pages using three or five sentences to allow users to get an idea about a webpage without having to read through it.

Document Categorization. The DGPort categorizer organizes the search results into various folders labeled by the key phrases appearing in the page summaries or titles, thereby facilitating understanding of

different groups of Web pages. We used the Arizona Noun Phraser (AZNP) (Tolle & Chen, 2000) to extract meaningful phrases from the titles and summaries of the search results.

Document Visualization. The DGPort visualizer was developed to reduce information overload when a large number of search results are obtained, which is a typical situation in many search scenarios. It has two versions: the Jigsaw and Geographic Information Systems (GIS) versions. Web pages are clustered onto a map, generated using the Kohonen (Kohonen, 1995) self-organizing map (SOM) algorithm, which is a two-layered neural network that automatically learns from the input Web pages and clusters them into different naturally occurring groups. Using the DGPort visualizer, digital government researchers can obtain a meaningful and comprehensive picture of a large number of search results.

DGPort Evaluation

In addition to the system perspective, we investigated the digital government portal from a human perspective. We are interested in studying what information people want from a digital government portal and how they access and process information. We also want to learn about how various segments of society (e.g. expert and novice) operate differently in the information seeking process.

We conducted two preliminary user studies of DGPort prototype's search, collection and analysis capabilities. The first user study was a field-based expert experiment to study the impact of a prototype system on digital government researchers. Protocol analysis approaches were adopted to study how DGPort impacts the learning and information searching behaviors. The positive feedback and suggestions from the experts provided valuable input to our planned technology for the new information analysis and visualization technologies to be developed. The second user study used a task-based experimental design. We tested the search and analysis capabilities of DGPort. Thirty-two subjects were recruited from the University of Arizona; each completed assigned search and browse tasks using DGPort and two benchmark systems, Google and FirstGov (www.firstgov.gov). Experimental results indicated that DGPort could significantly augment Google and FirstGov with its browse support. These preliminary results indicated directions for future system improvements, such as expanding the collection size, integrating more information resources and enhancing the system interface.

Research Impact

The intent of this research project is to investigate potential means of allowing digital government researchers to retrieve relevant information easily and effectively, thus reducing information overload. Preliminary results indicate that the pre- and post-search tools we experimented with in this project, (such as the integration of diverse information resources, and post-search document summarization, document categorization and document visualization) have the potential to effectively facilitate information retrieval in the digital government domain.

References

- Blair, D. C. and Maron, M. E.: An Evaluation of Retrieval Effectiveness for a Full-text Document-retrieval System. *Communications of the ACM*, 28(3), (1985) 289–299.
- Chen, H., Lynch, K. J.: Automatic Construction of Networks of Concepts Characterizing Document Databases. *IEEE Transactions on Systems, Man, and Cybernetics* 22 (5) (1992) 885-902
- Lawrence, S., Giles, C. L.: Accessibility of Information on the Web. *Nature* 400 (1999) 107-109
- McDonald, D., Chen, H.: Using Sentence-selection Heuristics to Rank Text Segments in TXTRACTOR. *Proceedings of Second ACM/IEEE-CS joint conference on Digital libraries, Portland, Oregon, USA (2002)*
- Tolle, K. M., Chen, H.: Comparing Noun Phrasing Techniques for Use with Medical Digital Library Tools. *Journal of the American Society for Information Science (Special Issue on Digital Libraries)* 51 (4) (2000) 352-370
- Kohonen, T.: *Self-organizing Maps*. Springer-Verlag, Berlin (1995)