Chapter 6
I-FGM as a Real Time Information Retrieval Tool for E-Governance

Eugene Santos Jr.
Dartmouth College, USA

Eunice E. Santos
Virginia Polytechnic Institute & State University, USA

Hien Nguyen
University of Wisconsin, USA

Long Pan
Virginia Polytechnic Institute & State University, USA

John Korah
Virginia Polytechnic Institute & State University, USA

Huadong Xia
Virginia Polytechnic Institute & State University, USA

ABSTRACT

Homeland security and disaster relief are some of the critical areas of E-governance that have to deal with vast amounts of dynamic heterogeneous data. Providing rapid real-time search capabilities for such applications is a challenge. Intelligent Foraging, Gathering, and Matching (I-FGM) is an established framework developed to assist users to find information quickly and effectively by incrementally collecting, processing and matching information nuggets. This framework has been successfully used to develop a distributed, unstructured text retrieval application. In this paper, we apply the I-FGM framework to image collections by using a concept-based image retrieval method. We approach this by incrementally processing images, extracting low-level features and mapping them to higher level concepts. Our empirical evaluation shows that our approach performs competitively compared to some existing approaches in terms of retrieving relevant images while offering the speed advantages of distributed and incremental process and unified framework between text and images.
INTRODUCTION

One of the main challenges in e-governance is to effectively and efficiently find relevant information from vast amounts of dynamic heterogeneous sources quickly under the pressures and limitations of time, supporting tools, and resources. For instance, when natural disasters such as Hurricane Katrina (2005) or the Asian Tsunami of 2004 happen, we need to quickly locate the areas that are most affected and collect information in order to estimate the needed aid for items such as medicines, foods, and drinking water. Unfortunately, in such a situation, frontline communications are typically chaotic (or non-existent) and/or there are too many channels of information from different sources that make the retrieval of relevant pieces of information a lot harder. For “hot spots” such as disaster relief areas, combat zones, and so forth information is changing rapidly and as such there is only a small window of time for information to remain valid. Additionally, various types of data representation are used such as images, blogs, maps, news reports, audios, and videos. Each type of data format contains important and indispensable information for the various governmental agencies. Therefore, in order to better assist these agencies in addressing these challenges, there is a clear and urgent need to develop a system that rapidly provides real-time retrieval capabilities of heterogeneous sources of information. There are three main issues that we need to address: (1) how to gather and retrieve information quickly in a real-time setting given the limitations of resources and time; (2) how to address the problem of heterogeneous data; and, (3) how to improve retrieval success.

We address the aforementioned issues by developing a framework for intelligent foraging, gathering, and matching (I-FGM) that incrementally and distributively gathers, processes, and matches information nuggets to assist users at finding information quickly and effectively. In our previous work (Santos, Santos, Nguyen, Pan, & Korah, 2005, Santos et al., 2006), I-FGM has been empirically demonstrated to be an effective tool for text retrieval on large and dynamic search spaces. Even though unstructured text is a typical format for most databases/sources, images are also popular with significant support from commercialized search engines such as Google, Yahoo!, and MSN. In order to demonstrate that I-FGM is a general framework for information retrieval, it is necessary to study the system’s ability at effectively handling such heterogeneous data which contains at least text and images. In this article, we apply the I-FGM framework on image collections by using a concept-based image retrieval method. We approach this by incrementally processing the images, extracting low-level features, and then mapping them to higher level concepts. The novelties of our approach lie with the distributed storage and incremental processing and matching of information nuggets extracted from a region-based wavelet image retrieval scheme. We deploy a concept-based image retrieval algorithm that maps low-level features of the images to high-level concepts. In this way, we are also able to translate the visual information of images into document graphs (Santos et al., 2005), which are used in I-FGM as a common representation of information for heterogeneous data types. Thus, I-FGM provides a seamless integration of text and image through a single unifying semantic representation of content. By implementing and testing our image retrieval algorithm in I-FGM, we can validate the I-FGM framework as a method for future unified rankings of heterogeneous documents.

This article is organized as follows: In the methodology section, we first provide a brief background on current image retrieval systems. Next, we give an overview of the image retrieval algorithm that was implemented in I-FGM. We then describe the architecture of I-FGM and the implementation details of the current prototype. This section is followed by a description of our system evaluation procedure along with simulation results. We validate the performance of the