Chapter 10
Multimodal Search on Mobile Devices:
Exploring Innovative Query Modalities for Mobile Search

Xin Fan
University of Sheffield, UK

Mark Sanderson
University of Sheffield, UK

Xing Xie
Microsoft Research Asia, China

ABSTRACT

The increasingly popularity of powerful mobile devices, such as smart phones and PDAs, enables users to search for information on the move. However, text is still the main input modality in most current mobile search services although some providers are attempting to provide voice-based mobile search solutions. In this chapter, we explore the innovative query modalities to enable mobile devices to support queries such as text, voice, image, location, and their combinations. We propose a solution to support mobile users to perform visual queries. The queries by captured pictures and text information are studied in depth. For example, the user can simply take a photo of an unfamiliar flower or surrounding buildings to find related information from the Web. A set of indexing schemes are designed to achieve accurate results and fast search through large volumes of data. Experimental results show that our prototype system achieved satisfactory performance. Also, we briefly introduce a prospective mobile search solution based on our ongoing research, which supports multimodal queries including location information, captured pictures and text information.

INTRODUCTION

As search technology has proliferated on the desktop, search service providers are competing to develop innovative technologies and offer new services in order to attract more users and generate more revenue from advertisements. Nowadays, the competition is moving to mobile platforms since mobile phone, smart phone and PDA users are a

larger community than computer users. Major Internet search providers, such as Google, Yahoo!, Microsoft or Ask.com are actively promoting new mobile services, including traditional information search and more mobile specific services such as product search and local search. Many such services use text-based interfaces, which, due to the constrained input/output modalities make the services inconvenient to use. However, mobile phones can support rich multimodal queries, composed of both text combined with image, or audio, or video, or location information.

Research into technology and user interaction in modalities other than text is becoming increasingly important. Since the basic function of mobile phones is to enable voice communications, voice is the most natural way on a mobile to input search queries. The ready availability of the microphone makes non-speech audio queries a natural option; embedded cameras offer the possibility of visual input; also the increasing prevalence of built-in GPS provides location information as well. Not only can GPS in a mobile offer the user a chance to navigate around, it can provide the search engine useful geographical information based on the user's current location. Therefore, with such technologies, the mobile phone can be turned into a powerful tool to carry out multimodal search and acquire information on the go.

In this chapter, we will examine mobile search based on multimodal queries composed of text, local information, image and audio. In the next section, we will investigate existing work on location based mobile search, visual mobile search and audio mobile search from both an industrial and academic perspective, followed by typical system design of a multimodal mobile search service.

A mobile search system by visual query is introduced in detail as an example system. As audio queries based on audio matching and speech recognition are relatively mature in the research perspective, we will mainly emphasise on visual and location modalities when introducing the implementation details. Finally, we propose a

prospective mobile search solution to process multimodal queries including location information, captured pictures and text information, followed by a conclusion.

STATE OF ART

Initially, mobile search services were simple counterpart versions of web search ones and the content was mainly limited to news, weather, sports, etc. In recent years, remarkable efforts have been made on both interface and content of mobile search services in the research community. The search providers are also shifting their focus on multimodal interfaces and richer content types. A location-aware system (e.g. GPS), embedded camera and micro-phone can be successfully exploited to collect multimodal information and generate multimodal queries for better search experiences on mobile devices. In this section, we will specifically study the present research work and commercial search services on mobile audio search, mobile visual search and location based mobile search.

Mobile Audio Search

A considerable number of mobile audio search services have been described by the major search engine companies in recent years; the use scenarios include but are not limited to the following aspects (Xie, et al., 2008):

- To search for local information, such as restaurants, travel destinations etc, through a spoken keyword or dialogue;
- To search for text information on the Web that is relevant to the corresponding text of an audio query;
- To search for similar audio materials, such as songs or ringtones, to the audio query;

Related Content

Network Layer Mobility Management Schemes for IP-Based Mobile Networks: A Survey
Paramesh C. Upadhyayand Sudarshan Tiwari (2010). *International Journal of Mobile Computing and Multimedia Communications (pp. 47-60).*
www.irma-international.org/article/network-layer-mobility-management-schemes/46123

Context-Aware and Adaptive Units of Learning in mLearning
Sergio Gómezand Ramón Fabregat (2012). *International Journal of Handheld Computing Research (pp. 1-25).*
www.irma-international.org/article/context-aware-adaptive-units-learning/67094

Mobile Electronic Commerce
Samuel Pierre (2009). *Mobile Computing: Concepts, Methodologies, Tools, and Applications (pp. 18-26).*
www.irma-international.org/chapter/mobile-electronic-commerce/26485

Relay Selection Scheme for Cooperative Communication Systems in Fixed Decode-and-Forward Mode
Jyh-Horng Wen, Jheng-Sian Li, Hsiang-Shan Houand Cheng-Ying Yang (2014). *International Journal of Mobile Computing and Multimedia Communications (pp. 68-77).*
www.irma-international.org/article/relay-selection-scheme-for-cooperative-communication-systems-in-fixed-decode-and-forward-mode/129001

Virtual EZ Grid: A Volunteer Computing Infrastructure for Scientific Medical Applications
Mohamed Ben Belgacem, Nabil Abdennadherand Marko Niinimaki (2012). *International Journal of Handheld Computing Research (pp. 74-85).*
www.irma-international.org/article/virtual-grid-volunteer-computing-infrastructure/64366