

모바일 기기를 위한 정보검색 시스템

김재훈† · 김형철*

(원고접수일 : 2009년 2월 12일, 원고수정일 : 2009년 3월 3일, 심사완료일 : 2009년 4월 20일)

Information Retrieval System for Mobile Devices

Jae-Hoon Kim† and Hyung-Chul Kim*

Abstract : Mobile information retrieval is an evolving branch of information retrieval that is centered on mobile and ubiquitous environments. In general, mobile devices are characterized by lightweight, low power, small memory, small display, limited input/output, low bandwidth, and so on. Some of these characteristics make it impossible to apply general information retrieval to mobile environments without any modification. In order to relieve this problem, we design and implement an information retrieval system for mobile devices like wireless phones, PDA and handheld devices. We use document summarization techniques to alleviate the limitation of small display and user profiles to retrieve the most proper documents for each individual user for personalized search. Futhermore we use meta-search to lighten some burdens visiting several portal sites. In this paper, we have implemented and demonstrated the proposed mobile information retrieval system on the domain of travel and received good evaluation from users subjectively.

Key words : Mobile information retrieval(모바일 정보검색), Meta-search(메타검색), Document summarization(문서요약), User profile(사용자 프로파일)

1. 서 론

최근 많은 사람들은 적어도 하나의 모바일 기기를 가지고 다니면서 언제 어디서나 자신이 원하는 서비스를 받고자 한다. 이와 같은 환경이 U-환경 또는 모바일 환경이다. 또한 많은 기업들은 모바일 환경에 적합한 상품이나 서비스를 제공하고 있는데 이를 총칭해서 U-Biz라고 한다^[1]. 모바일 환경은 일반적인 유선 환경에 비해 아래와 같은 특징을 가

지고 있다^[2]. 첫째, 모바일 환경은 시공간적인 제약이 없다. 사용자들은 모바일 기기를 휴대하고 다니면서 언제 어디서나 원하는 정보를 찾고자 한다. 둘째, 모바일 환경은 매우 개인적이다. 대부분의 모바일 기기는 개인이 소유하고 있으며 개인의 취미나 관심분야에 따라 원하는 정보가 서로 다르다. 따라서 개인의 특성에 따라 표면적으로 같은 내용 이더라도 실질적으로 질의 내용이 다를 수 있다. 셋째, 모바일 기기는 소형으로 입출력에 많은 제약

† 교신저자(한국해양대학교 컴퓨터공학과, E-mail:jhoon@hhu.ac.kr, Tel: 051-410-4574)

* 한국해양대학교 컴퓨터공학과

을 가지고 있다. 모바일 환경에서 작은 화면을 통해 여러 페이지를 장황하게 보여주는 것보다는 사용자가 원하는 내용만을 추출하여 보여주는 것이 훨씬 효과적이다. 넷째, 일반적으로 모바일 환경은 자료의 전송률은 유선 환경에 비해 매우 낮으며 가격은 더 비싸다. 다섯째, 일반적으로 모바일 기기는 메모리와 처리 능력의 한계를 가지고 있다. 모바일 기기 자체에서 많은 양의 정보를 처리할 수 없다.

본 논문에서는 이와 같은 모바일 환경의 특성을 고려하여 모바일 환경에 적합한 정보검색 시스템을 제안한다. 제안된 시스템은 문서요약^[3]과 사용자 프로파일^[4]을 메타검색 모델에 적용한 모바일 정보검색 모델이다. 문서요약 기술은 작은 화면과 낮은 전송률 문제를 완화시킬 수 있고 사용자 프로파일은 모바일 환경에서 개인화된 정보검색 모델을 사용할 수 있도록 한다.

제안된 시스템은 여행 영역을 대상으로 PDA 환경에서 구현되었으며 검색시간이 다소 지연되는 문제가 있었으나 학생들을 대상으로 주관적 평가를 수행했을 때 좋은 반응을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 모바일 정보검색과 문서요약 그리고 사용자 프로파일에 대해서 간단히 살펴본다. 3장에서 모바일 환경에 적합한 정보검색 시스템을 설계하고, 4장에서 모바일 정보검색 시스템의 구현 방법과 분석 및 평가에 대해서 기술한다. 5장에서 제안된 시스템의 문제점과 앞으로의 연구 방향을 살펴보고 결론을 맺는다.

2. 관련 연구

2.1 모바일 정보검색

휴대폰이나 PDA와 같은 모바일 기기가 대중화되면서 포털 업체나 이동통신사들의 모바일 정보검색에 대한 관심이 날로 증가되고 있다. 해외에서는

2006년도에 야후, 구글, 노키아, 마이크로소프트 등이 모바일 검색 엔진을 출시했고, 국내에서도 SKT, KTF, LGT와 같은 이동통신사들이 포털 업체와 손잡고 모바일 검색 시장에 뛰어들었다¹⁾. 모바일 검색의 시장규모는 2013년에 약 48억 달러에 달할 것으로 전망하고 있다^[5]. 모바일 검색과 연계된 다양한 멀티미디어 서비스로 인한 수익이 현재 모바일 시장에서 가장 주도적인 위치를 차지하고 있는 광고 수익을 크게 상회할 것으로 예측하고 있다. 앞에서 언급한 모바일 기기의 특성으로 인해 모바일 정보검색은 일반 검색에 비해 여러 가지의 어려움이 있다²⁾. 가장 큰 어려움은 작은 화면이며 이로 인해 한꺼번에 많은 내용을 볼 수 없는 것이다. 또한 검색어의 길어도 점점 짧아지고 있는데 검색어의 길이가 짧아질수록 원하는 문서의 검색이 어렵다. 이러한 어려움에도 불구하고 모바일 검색의 수요는 지속적으로 증가하고 있다. 모바일 정보검색은 정보의 제공 방법에 따라 사용자 주도 검색과 제공자 주도 검색으로 나눌 수 있다^[2]. 사용자 주도 검색은 기본 질의어에 사용자의 주변 상황 정보를 더하여 새로운 질의를 생성하는 반면에 제공자 주도 검색은 기본 질의어를 그대로 사용한다.

2.2 자동 요약

자동 요약은 원문서의 의미를 유지하면서 원문서의 길이나 정보의 복잡도를 줄이는 작업이다^[3]. 일상적인 생활에서 뉴스의 머리기사, 각종 회의의 의사록 등이 문서요약의 대표적인 예이다. 최근 자동 요약은 단순한 하나의 문서의 내용을 요약할 뿐 아니라 여러 문서의 내용을 하나로 요약하기도 하고, 심지어는 문서가 아닌 이미지, 오디오, 비디오와 같은 멀티미디어 정보를 요약하기도 한다^[6]. 기술적인 측면에서 자동 요약 기술은 크게 문장추출과 문서요약 기술로 나눌 수 있다^[3]. 전자는 문서 내에 속하는 가장 중요한 문장을 추출하는 방법이

1) <http://blog.diquest.com/entry/다이퀘스트-검색트렌드-태동기의-모바일-검색>
<http://www.kisti.re.kr/yesKISTI/Briefing/Trends/View.jsp?cn=GTB2008030820>
 2) <http://searchenginewatch.com/showPage.html?page=3624288>

고, 후자는 원문서의 내용을 새로운 문장으로 재작성하는 방법이다. 일반적으로 문서요약 방법이 이상적인 방법이지만, 아직 기술적으로 여러 가지 문제를 가지고 있을 뿐 아니라 구현도 매우 어렵다. 또한 자동 요약은 단순한 문서의 내용을 요약하는 방법과 주어진 질의에 관련된 내용을 요약하는 방법으로 분류할 수 있으며 후자를 질의 기반 요약이라고 한다^[7].

2.3 사용자 프로파일

정보검색에서 사용자 프로파일은 사용자에게 관련된 정보를 저장하고 이를 이용하여 사용자의 질의를 확장하는 방법이다^{[4][8]}. 사용자 프로파일에는 사용자들의 관심 사항이 저장되며, 정적 혹은 동적인 방법으로 구축된다. 정적 구축 방법은 사용자가 직접 자신의 관심 사항을 편집하고 수정하는 방법이다. 반면에 동적 구축 방법은 사용자의 이전 질의나 연관 피드백 등을 통해서 사용자의 정보요구에 관련된 정보를 수집하는 방법이다. 동적 프로파일은 처음에는 문서 순위화에 크게 기여하지 못하지만 시간이 경과함에 따라서 사용자의 관심 사항이 사용자 프로파일에 정확하게 기술되어 사용자에게 적합한 문서를 검색할 수 있을 뿐 아니라 정확률도 크게 개선할 수 있다^[9].

3. 모바일 정보검색 시스템의 설계

일반적인 정보검색에 비해 모바일 정보검색의 두드러진 특징은 개인화와 입출력의 제약이다. 본 논문에서는 개인화를 위해서 사용자 프로파일을 사용하여 개인에 가장 적합한 문서를 검색할 수 있도록 한다. 이렇게 함으로써 불필요한 통신 요금을 줄일 뿐 아니라 짧은 시간 내에 원하는 정보를 찾을 수 있다. 입출력 제약을 위해서는 문서요약 기술을 이용해서 문서 전체의 내용을 확인하지 않고 일부의 내용만 확인하여 원하는 정보를 조기에 찾을 수 있

도록 한다. Fig. 1은 본 논문에서 제안된 모바일 기기에 적합한 정보검색 시스템의 구성도이며, 크게 모바일 기기, 포털 사이트, 모바일 정보검색 시스템으로 구성되고, 모바일 정보검색 시스템은 메타검색, 사용자 프로파일 관리, 문서요약으로 구성된다. 각각은 이하에서 좀 더 자세히 설명할 것이다.

3.1 모바일 기기

모바일 기기는 휴대폰이나 PDA와 같은 기기로서 사용자 인터페이스를 제공한다(자세한 구성은 Fig. 3과 4를 참조). 사용자 인터페이스는 크게 질의 인터페이스와 표제 인터페이스 그리고 본문 인터페이스로 분류된다. 질의 인터페이스는 사용자가 원하는 내용을 질의하는 인터페이스이고, 표제 인터페이스는 검색결과에 해당하는 문서의 제목을 보여주고 해당 제목의 문서를 선택할 수 있는 인터페이스이다. 본문 인터페이스는 선택된 표제에 해당하는 문서의 전체 내용이나 요약물 보여주는 인터페이스이다.

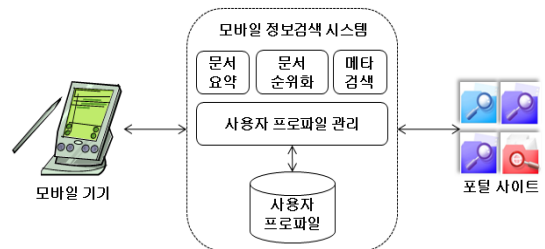


Fig. 1 System configuration of the proposed system

3.2 포털 사이트

포털 사이트는 인터넷의 입구에 해당하는 거대한 웹사이트이다. 검색엔진 등을 중심 서비스로 제공하며, 부가적으로 뉴스, 주가 전자계시판 등의 정보 제공 서비스를 무료로 제공한다. 한국에는 대표적인 포털사이트로는 네이버³⁾, 다음⁴⁾, 구글⁵⁾, 야후⁶⁾ 등이 있다. 본 논문에서는 메타검색(3.3.1절

3) <http://www.naver.com/>
 4) <http://www.daum.net/>
 5) <http://www.google.co.kr/>

참조)을 위해서 네이버, 다음, 야후를 이용한다.

3.3 모바일 정보검색 시스템

3.3.1 메타 검색

메타 검색은 자체 검색엔진을 가지지 않고, 기존의 검색엔진을 이용하는 방식이다^[10]. 즉 사용자 요구(질의)가 입력되면 그 요구를 여러 검색엔진에 동시에 질의하여 각 검색엔진으로부터 검색 결과를 받아 이들을 결합하는 방식이다. 같은 질의를 사용하더라도 색인 방법이나 검색엔진에 따라 서로 다른 문서들이 검색될 수 있다. 본 논문에서는 3.3.2 절에서 설명할 사용자 프로파일을 이용해서 문서의 순위를 재결정하기 때문에 검색된 문서들을 하나의 검색결과로 단순히 결합한다.

3.3.2 사용자 프로파일 관리

사용자 프로파일 관리 시스템은 사용자 프로파일을 수정하고 검색하는 일을 담당한다. 본 논문에서 사용자 프로파일은 크게 영역 프로파일과 이력 프로파일로 구성되며, 모든 프로파일은 용어 벡터⁷⁾로 표현된다.

영역 프로파일은 사용자가 주로 관심을 가지는 영역에 속하는 용어 벡터이며 정적으로 생성된다. 예를 들면 질의 영역이 여행이라면 그 벡터는 “($w_{여행}$, $w_{관광}$, $w_{호텔}$...)”이 될 수 있으며 모든 가중치는 $1/N$ 이다. 여기서 N 은 영역 프로파일에 속하는 용어의 수이다. 영역 프로파일에 속하는 용어는 수작업으로 선정하거나 다음 디렉토리⁸⁾와 같은 용어의 범주 정보로부터 자동으로 구할 수 있다. 본 논문에서는 후자의 방법을 사용한다.

이력 프로파일은 질의 이력과 표제 이력이 있으며 동적으로 생성된다. 질의 이력은 사용자가 원하는 질의에 관련되고 표제 이력은 사용자에게 의

서 선택된 문서의 표제에 관련된다. 이력 프로파일도 영역 프로파일과 마찬가지로 용어 벡터로 표현되고 이력이 새로 발생할 때마다 용어 빈도가 더해진다. 본 논문에서 사용자 프로파일의 가장 중요한 역할은 메타 검색을 통해서 검색된 문서를 사용자 개인에 맞게 재순위화하는 것이며 3.3.4절에서 자세히 기술할 것이다.

3.3.3 문서요약

문서요약은 사용자에게 의해서 요구될 때 질의의 내용을 토대로 문서를 요약한다. 본 논문의 문서요약은 추출 요약이며 질의 기반 요약이다. 문서요약의 입력은 질의 Q 와 문서 D 이다. Q 는 질의 벡터 (qq_1, qq_2, \dots, qq_T)이고, qq_i 의 값은 i 번째 용어가 Q 에 속하면 1이고 그렇지 않으면 0이다. D 는 여러 개의 문장 (S_1, S_2, \dots)으로 나누고, i 번째 문장 S_i 는 문장 벡터 ($s_{i,1}, s_{i,2}, \dots, s_{i,T}$)이다. $s_{i,j}$ 의 값은 S_i 에서 j 번째 용어의 정규화된 용어 빈도이다. 본 논문에서 문서 요약은 질의 Q 에 가장 유사한 문장 순으로 문서를 재구성하여 가장 유사한 몇 개의 문장을 추출하는 것이다. 즉 식 (1)을 이용해서 Q 와 S_i 사이의 코사인 유사도 $sim(Q, S_i)$ 를 구하고, 각 문장을 유사도에 따라 내림차순으로 정렬한다. 이와 같은 방법으로 재구성된 문서로부터 처음 세 문장을 추출하고 추출된 문장을 D 의 요약으로 간주한다.

$$sim(Q, S_i) = \frac{Q \cdot S_i}{\|Q\| \times \|S_i\|} = \frac{\sum_{t=1}^T (qq_t \times s_{i,t})}{\sqrt{\sum_{t=1}^T (qq_t)^2} \times \sqrt{\sum_{t=1}^T (s_{i,t})^2}} \quad (1)$$

3.3.4 문서 순위화

모바일 정보검색 시스템은 사용자에게 적합하도록

6) <http://kr.yahoo.com/>

7) 용어 벡터는 (w_1, w_2, \dots, w_T)로 표현된다. 여기서 w_i 는 i 번째 용어의 가중치이며 T 는 정보검색에서 사용되는 용어 수이다. 일반적으로 w_i 는 용어 빈도 tf_i 와 역문서 빈도 idf_i 의 곱이다^[10]. tf_i 는 어떤 문서에 나타난 i 번째 용어의 빈도수이고 idf_i

는 i 번째 용어가 나타난 문서의 수의 역이다. 본 논문에서 tf_i 는 정규화된 용어 빈도 $tf_i / \sum_{j=1}^T tf_j$ 를 사용하고 idf_i 는 모두 1

로 간주한다. 즉 정규화된 용어 빈도만 사용한다.

8) <http://directory.search.daum.net/>

문서의 순위를 재조정하고 전체 시스템을 제어한다. 문서 재순위화를 위해서 질의 Q 와 문서 D 의 유사도는 코사인 유사도를 사용하며 식 (2)와 같다.

$$sim(Q, D) = \frac{Q \cdot D}{\|Q\| \times \|D\|} = \frac{\sum_{i=1}^T (q_i \times d_i)}{\sqrt{\sum_{i=1}^T (q_i)^2} \times \sqrt{\sum_{i=1}^T (d_i)^2}} \quad (2)$$

Q 는 (q_1, q_2, \dots, q_T) 이지만, D 는 문장으로 분리되지 않고 하나의 문서 벡터 (d_1, d_2, \dots, d_T) 이다. d_i 의 값은 D 에서 i 번째 용어의 정규화된 용어 빈도이고, q_i 는 사용자 질의를 사용자 프로파일로 확장한 질의이며, 식 (3)과 같이 구한다.

$$q_i = qq_i + qd_i + qh_i \quad (3)$$

여기서 qq_i, qd_i, qh_i 는 각각 사용자 질의, 영역 프로파일, 이력 프로파일에서 i 번째 용어의 가중치이다. qq_i 의 값은 3.3.3절에서와 같이 i 번째 용어가 사용자 질의에 속할 경우 1이고 그렇지 않으면 0이다. 예를 들면 영역 프로파일에 속하는 단어가 100개 있다면 각 단어의 가중치는 $1/100$ 이다. 또한 이력 프로파일에서 각 용어의 빈도가 “{(해운대, 1), (호텔, 5)}”이라면 용어 “해운대”의 가중치는 $1/6$ 이고, “호텔”의 가중치는 $5/6$ 가 된다. 만약 “해운대”가 질의와 이력 프로파일에 속한다면 $q_{\text{해운대}}$ 는 $1 + 1/6 = 1.167$ 이 되고 “호텔”이 영역 프로파일과 이력 프로파일에 속한다면 $q_{\text{호텔}}$ 을 $1/100 + 5/6 = 0.84$ 가 된다.

3.4 시스템 제어 및 자료의 흐름

모바일 정보검색 시스템(서버)에서 시스템 제어 및 전체 자료의 흐름은 Fig. 2와 같다.

사용자는 모바일 기기를 통해서 원하는 정보를 질의한다. 이 정보에 사용자 정보를 추가하여 서버로 전송한다.

서버는 사용자 질의를 그대로 메타 검색 시스템으로 보낸다.

메타 검색 시스템은 사용자 질의를 이용해서 여러 검색엔진으로부터 문서를 수집하고 중복된 문서를

제거한다. 이렇게 수집된 문서를 서버로 전송한다.

서버는 사용자와 질의를 사용자 프로파일 관리 시스템으로 전송한다.

사용자 프로파일 관리 시스템은 사용자와 질의를 이용해서 질의 이력 프로파일(3.3.2절을 참조)을 수정하고 사용자 프로파일을 서버로 전송한다.

서버는 사용자 프로파일(이력과 영역 프로파일)과 질의를 이용해서 질의 벡터를 확장하여 문서의 순위를 재조정한다(3.3.4절 참조). 상위 10개 문서의 제목을 모바일 기기에 전달한다.

사용자는 표제 인터페이스를 통해서 원하는 문서의 요약 정보를 요구한다.

서버는 선택된 문서와 질의를 문서 요약 시스템으로 보낸다.

문서 요약 시스템은 요약 문서를 서버에 전달한다.

서버는 요약 문서를 모바일 기기에 전달한다.

필요에 따라서 사용자는 문서의 전체 내용을 요구한다.

서버는 웹 사이트로부터 전체 내용을 검색하여 모바일 기기에 전달한다.

서버는 전체 내용을 요구한 문서에 대해서 사용자와 제목을 사용자 프로파일 관리 시스템에 보낸다. 사용자 프로파일 관리 시스템은 사용자와 제목을 이용해서 표제 이력 프로파일을 수정한다.



Fig. 2 Information flow in the proposed system

4. 모바일 정보검색 시스템의 구현 및 분석

4.1 구현 환경

이 장에서 모바일 정보검색 시스템의 구현 방법을 기술한다. 전체 프로그램의 통합 및 관리를 위하여 C#을 이용하였으며, 데이터의 검색 및 가공 작업이 진행되는 서버는 Windows XP를 기반으로 모든 통신은 TCP/IP를 사용하여 구현하였다. 모바일 기기로는 PDA를 이용하였고, 메타 검색을 위해서 검색엔진으로 네이버, 다음, 야후를 사용하였다. PDA는 Windows Mobile 5.0이 탑재된 HP사의 hx2790 모델을 사용하였으며, 개발은 Visual Studio 2005의 스마트클라이언트 개발환경에서 Windows Mobile 5.0 SDK를 이용하였으며 언어는 C#을 사용하였다.

또한 한국어 색인을 위해서는 기존 시스템^[11]을 이용하였다. 대상 영역은 국내여행 영역으로 한정하였으며 문서의 경우도 모든 웹 문서가 아니라 블로그 문서로 한정하였다. 영역 프로파일은 다음 디렉토리의 “여행, 세계정보>국내여행” 아래에 연결된 모든 문서에서 빈도수가 높은 100개의 단어를 선정해서 구축되었다.

각 프로파일 및 개인 사용자 정보 등은 DB를 이용하여 관리하는데, 데이터베이스는 SQL-Server 2005를 이용하였다. 질의 및 표제 이력 프로파일은 동적으로 확장된다. 각각의 개인별로 사용했던 질의어 및 표제어에 대하여 색인어와 값의 쌍으로 표현된 자료구조를 이용하여 저장되며 색인어에는 질의어 및 표제어의 내용이 저장되고 값에는 빈도수가 저장된다. 3.5절에서 설명한 예를 이용해서 용어 빈도를 관리하는 데이터베이스 스키마를 설명하면 Table 1과 같고, Table 2는 정규화 빈도를

Table 1 Database for term frequency

사용자 (user)	용어 (term)	빈도 (freq)	종류 (kind)
사용자A	호텔	5	질의
사용자A	해운대	1	질의
사용자A	호텔	1	영역
...			

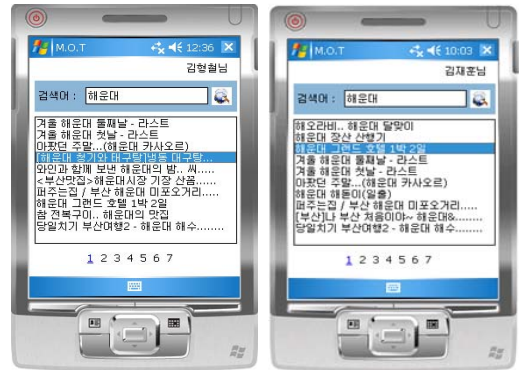
Table 2 Database for total term frequency

사용자 (user)	빈도총합 (tot_freq)	종류 (kind)
사용자A	6	질의
사용자A	100	영역
...		

구하기 위한 빈도 총합을 저장하는 데이터베이스 스키마이다.

4.2 사용자 프로파일을 이용한 개인화 실험

Fig. 3는 사용자 A(김형철)와 B(김재훈)가 ‘해운대’를 질의한 결과이며 각 개인의 관심사에 따라서 서로 다른 결과를 얻을 수 있었다. 이 실험에서 사실 영역 프로파일은 같고 이력 프로파일만 사용자에게 따라서 다르며 이 실험 전에 수차례의 검색을 통해서 사용자 프로파일이 축적되었다. Fig. 3에서 화면의 상단부는 입력을 위한 질의 인터페이스이고 하단부는 문서목록을 위한 표제 인터페이스이다. 본문 인터페이스는 Fig. 4에서 보여줄 것이다.



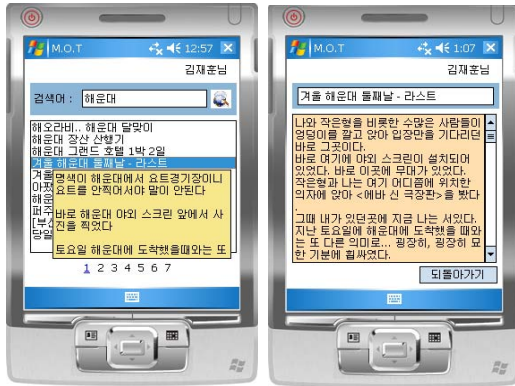
(a) User A (b) User B

Fig. 3 Search results according to user profiles

4.3 문서요약 실험

Fig. 4a는 사용자 B에 의해서 검색된 문서 “겨울 해운대 둘째날 - 라스트”의 요약 결과이며 질의한 내용과도 매우 유사함을 알 수 있었다. 즉 추출된 모든 문장에 ‘해운대’와 관련된 내용들이며 이 문장들은 Fig. 4b에 출현된 원문서로부터 추출된

것이다. Fig. 4b는 원문서로서 이는 사용자의 요구에 의해서만 보여진다. 전체 내용에서는 사용자가 찾고자하는 정보가 문서 가운데 있다면 가운데로 이동하여 원하는 내용을 찾을 수 있다.



(a) Summarized text (b) Original text

Fig. 4 The result of automatic text summarization

4.4 효용성 평가

본 논문에서 제안된 시스템의 효용성(편의성과 신속성)을 평가하기 위해서 학부생 30명을 무작위로 추출하여 본 제안된 시스템과 A 검색엔진을 1시간 동안 사용하게 하였다. 학생들에게는 검색 영역을 여행 영역으로 한정하여 사용한 후에 아래와 같은 설문 평가를 실시하였다.

질문1(편의성): 여행 정보를 찾고자 할 때 A 검색엔진과 비교하면 어느 시스템이 원하는 정보를 더 쉽게 얻을 수 있었습니까?

질문2(신속성): 여행 정보를 찾고자 할 때 A 검색엔진과 비교하면 어느 시스템이 원하는 정보를 더 빨리 얻을 수 있었습니까?

Table 3 Questionnaire analysis for Question 1 and 2

	질문1	질문2
본 논문 > A 엔진	19(63%)	17(57%)
본 논문 < A 엔진	3(10%)	6(20%)
판단할 수 없음	8(27%)	7(23%)
합계	30(100%)	30(100%)

Table 3에서 제한된 영역에서는 A 검색엔진에 비해서 효용성이 우수한 것으로 평가되었다. 그러나 Table 3의 평가 결과는 평가자들의 주관적인 평가이며 충분한 평가라고는 생각되지 않는다. 객관적인 평가를 위해서는 평가집합이 개발되어야 할 것이다.

4.5 제안된 시스템 특성 및 활용 방안

이 절에서는 본 논문에서 제안된 시스템의 장점 및 단점을 분석하고 이 시스템의 활용 방안에 대해서 기술한다. 먼저 제안된 시스템의 장점은 다음과 같다. 첫째, 메타검색을 통해서 여러 검색엔진으로부터 정보를 찾아야 하는 번거로움을 해결할 수 있었다. 이것은 모바일 환경에서 전송률과 요금문제의 해결에 큰 도움이 될 것이다. 둘째, 사용자 프로파일을 통해서 사용자 개인의 관심사를 먼저 검색할 수 있게 되었다. 모바일 환경에서 개인화 기술의 발전에 기여할 수 있을 것으로 생각된다. 셋째, 질의 기반 문서 요약을 통해서 문서 전체를 읽지 않아도 원하는 정보를 보다 쉽게 파악할 수 있었다. 이것도 모바일 환경에서 전송률과 요금문제의 해결에 큰 도움이 될 것으로 생각된다. 넷째, 원문서의 내용을 원하지 않을 경우에는 불필요한 정보를 제공하지 않으므로 개인화와 입출력 제약을 해결하는데 도움이 되었다.

제안된 시스템은 일반적인 정보검색 시스템과 비교하면 여러 가지 단점을 가지고 있다. 첫째, 검색 대상이 블로그로 그 대상이 매우 제한된다. 블로그는 문서의 내용이 다소 주관적이므로 경우에 따라서는 사용자의 의도와 전혀 다른 내용들이 검색될 수도 있다. 따라서 검색 영역을 좀더 다양화할 필요가 있다. 둘째, PDA에서 검색어를 서버로 전송 후 서버에서 자료를 처리하는 동안에도 PDA가 서버와 통신을 유지하고 있으므로 서버에서 처리시간 동안은 불필요한 통신요금이 발생할 수 있다. 이 문제는 앞으로 개선해야할 큰 연구과제이다. 셋째, 오랜 사용을 통해 개인 프로파일의 양이 커지기 전에는 오히려 사용자의 의도와 빗나간 검색결과를 도출할 수도 있다. 이 문제 또한 앞으로의 연구과제 중 하나이다.

제안된 시스템은 선박과 같이 필수적으로 무선통신 환경에서 인터넷을 활용해야 할 경우 매우 유용할 것이다. 또한 PDA뿐 아니라 휴대폰이나 다른 모바일 기기에도 쉽게 적용할 수 있으며 또한 검색 영역을 더욱 제한하여 통신량을 최소화할 수 있을 것이다.

4.6 상용 시스템과의 비교 분석

사실 모바일 검색 서비스를 제공하는 상용 시스템들은 검색 기능 이외에 많은 부가적인 기능이 있기 때문에 제안된 논문과 전체적인 시스템을 비교하는 것은 의미가 없다. 따라서 본 논문에서는 모바일 검색 서비스를 제공하고 있는 구글⁹⁾, 다음¹⁰⁾, 네이버¹¹⁾을 중심으로 모바일 정보검색에 대해서만 그 특징을 살펴보고 제안된 논문과 비교 분석한다. 세 모바일 검색 서비스 모두가 휴대전화와 휴대기기용 웹사이트 검색을 위해 별도로 제작되었다. 검색 결과는 일반 검색과 동일하다. 다만 모바일 기기의 작은 화면 때문에 문서내용을 줄여서 보여준다. 그러나 문서내용을 줄이는 방법은 문서요약과 같은 기술을 사용하지 않고 단순히 제목과 첫 번째 문장 그리고 URL을 보여준다. 또한 사용자 프로파일을 사용하지 않기 때문에 모든 사용자의 검색결과는 같을 뿐 아니라 메타검색 서비스는 모두 제공하지 않는다. 다음 모바일 검색 결과는 일반 검색 결과와 달리 사전, 카페, 블로그 등의 순으로 재정렬해서 보여주는 특징이 있으며 제안된 시스템과 같이 초기 검색 결과에는 제목만 보여준다.

5. 결론 및 향후 연구

논문에서는 모바일 환경에 적합한 정보검색 시스템을 설계하고 구현하였다. 모바일 환경에서의 가장 큰 제약은 작은 화면으로 인해 많은 내용을 화면에 보일 수 없다는 것이다. 본 논문에서는 이 문

제를 완화하기 위해서 질의 기반의 문서요약 기술을 사용했는데 이 방법은 사용자 정보 요구에 가장 유사한 문장을 추출하여 보여줌으로써 문서 전체 내용을 확인하지 않고도 원하는 정보를 찾을 수 있었다. 그 결과로 좀 더 빠른 검색이 가능하게 되었고 통신비용을 절감할 수 있을 뿐 아니라 정확률과 만족도를 높일 수 있었다. 모바일 환경에서 또 다른 특징으로는 개인화이며 이를 위해서 사용자 프로파일을 사용하였다. 사용자 프로파일은 크게 영역 프로파일과 이력 프로파일로 나뉘어서 관리되었으며 사용자에게 적합한 문서가 검색될 수 있었다. 이는 같은 질의를 사용하더라도 사용자 개인의 관심사에 더 적합한 문서를 검색할 수 있게 되었다. 본 논문에서 제안된 시스템이 실용화되기 위해서는 실시간 검색이 제공되어야 할 것이며 GPS를 이용한 사용자 주변 상황을 인식할 수 있는 기능도 제공되어야 할 것으로 생각된다. 본 논문에서 실시간 처리를 위해서 문서 재순위화와 문서요약에서 간단한 유사도 계산 방법을 사용하였으나 정확률을 높이기 위해서 좀 더 정교한 계산 방법이 요구된다. 이 경우에는 분산처리를 이용한다면 실시간 처리가 가능할 것으로 생각된다. 질의응답이나 정보추출과 같은 기능이 모바일 환경에 적용된다면 매우 유용할 수 있으나 실시간 처리가 요구되는 모바일 환경에서는 좀 더 많은 연구가 필요할 것으로 생각된다.

참고문헌

- [1] 박성득 외, 유비쿼터스 백서 2005, 전자신문사, 2005.
- [2] 김학수, 장명길, "유비쿼터스 환경에서 모바일 정보검색 기술의 동향", 정보과학회지, 제24권, 제1호, pp. 48-55, 2006년.
- [3] Mani, I. and Maybury, M. T., Advances in Automatic Text Summarization, The MIT Press, 1999.
- [4] Korfhage, R., Information Storage and

9) <http://www.google.co.kr/m>

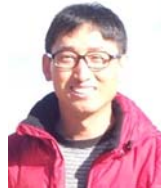
10) <http://mobile.daum.net/>

11) <http://mobile.naver.com/>

Retrieval, Wiley Computer Publishing, 1997.

- [5] Juniper Research, Mobile Search & Discovery Opportunities & Markets 2008-2013, Juniper Research Limited, 2008.
- [6] Yun, W-Q. and Kankanhalli, M. S., "Multimedia simplification for optimized MMS synthesis", ACM Transactions on Multimedia Computing, Communications, and Applications, Vol. 3, No. 1, 2007.
- [7] Bosma, W., "Query-based summarization using rhetorical structure theory", Proceedings of CLIN, pp. 29-44, 2003.
- [8] 소영준, 박영택, "사용자 프로파일 기반 개인 웹 에이전트", 정보과학회논문지: 소프트웨어 및 응용, 제27권, 제3호, pp. 248-256, 2000.
- [9] 이재원, 김응모, "서비스 제공자 목록에 의한 사용자 프로파일 기반 웹 서비스 검색의 정확도 향상", 2006년도 한국정보과학회 가을 학술발표논문집, 제33권, 제2B호, pp. 66-70, 2006.
- [10] Baeza-Yates, R. and Ribeiro-Neto, B., Modern Information Retrieval, Addison-Wesley, 1999.
- [11] 김재훈, 이공주, "사례기반 학습을 이용한 음절기반 한국어 단어 분리 및 범주 분류", 정보처리학회논문지, 제10-B권, 제1호, pp. 47-56, 2003.

저 자 소 개



김재훈(金載熏)

1964년생, 1986년 계명대학교 전자계산학과(학사), 1988년 한국과학기술원 전산학과(공학석사), 1996년 한국과학기술원 전산학과(공학박사), 1988년~1997년 한국전자통신연구원 선임연구원, 2001년~2002년 Information Sciences Institute in University of Southern California 방문연구원, 2007년~2008년 Beckman Institute in University of Illinois at Urbana-Champaign 방문연구원, 현재 한국해양대학교 컴퓨터공학과 교수, 한국마린엔지니어링학회 편집이사, 한국정보과학회 편집위원, 관심분야: 자연언어처리, 한국어정보처리, 정보검색, 정보추출.



김형철(金亨澈)

1982년생, 2009년 한국해양대학교 컴퓨터공학과 졸업, 현재 한국해양대학교 대학원 석사과정, 삼성전자 소프트웨어 멤버십 소프트웨어분과장. 관심분야: 자연언어처리, 정보검색, 텍스트 마이닝, 웹 마이닝.