

즐거리에 기반한 모바일 검색 시스템

김병기*, 조경빈, 최소영, 김유섭, 김진, 고영웅
한림대학교 컴퓨터공학과
e-mail: bkkim@hallym.ac.kr

Synopsis based Mobile Search System

Byung-Ki Kim*, Gyung-Bin Cho, So-Young Choi, Yu-Seop
Kim, Jin Kim, Young-Woong Ko
Dept. of Computer Science, Hallym University

요 약

모바일 단말기의 수요가 증가함에 따라 사용자들은 다양한 형태의 검색 시스템을 요구하고 있다. 모바일 환경에서는 키 입력 방식이 일반적인 개인용 컴퓨터와 상이하기 때문에 모바일 단말기의 특성에 적합한 검색 방식이 필요하다. 본 논문에서는 PDA와 같은 모바일 단말기에서 사용자가 원하는 TV 프로그램을 검색 할 수 있는 시스템을 설계 및 구현하였다. 전체 시스템은 검색 서버와 모바일 클라이언트로 구성되며, 클라이언트는 모바일 환경에서 검색어 입력을 효율적으로 할 수 있게 데이터베이스에 기반한 검색어 입력 방법을 사용하였다. 시스템의 유용성을 확인하기 위한 실험에서 이 시스템은 검색어 입력 방식과 드라마 검색 결과를 모두 만족하는 성능을 보여주었다.

키워드 : 모바일, 드라마, 동영상, 즐거리 기반, 검색, 불리안 모델

1. 서론

인터넷의 발전과 모바일 환경의 급성장에 따라 PC 환경에서 가능한 서비스들이 모바일 단말기에서 다양한 형태의 서비스로 제공되고 있다. 본 연구에서는 즐거리에 기반한 TV 시리즈물 검색 시스템을[1] 모바일 단말기에서 서비스 가능한 효과적인 시스템으로 구현하였다. 즐거리에 기반한 TV 시리즈물 검색 시스템이란 드라마 제작사에서 제공하는 각 회당 스토리를 기본적인 불리안 모델과[2] 결합시켜 사용자가 원하는 회차를 검색하는 시스템을 말한다.

일반 사람들이 영화를 검색할 경우 영화이름, 출연 배우 등의 정보를 이용하여 검색을 하는데 단편의 영화 같은 경우 검색어의 범위가 한정되기 때문에 원하는 정보를 쉽게 얻을 수 있는 반면[3], 장편의 TV 시리즈 같은 경우 사용자가 원하는 회차를

검색하기 위해 영화 제목, 배우 이름과 같은 방법으로 검색을 한다면 원하는 결과를 얻기가 어렵다. 따라서 본 논문에서는 불리안 모델을 사용한 즐거리에 기반한 TV 시리즈물 검색시스템을 모바일 환경에 적용시켰다[4,5,6].

본 논문에서 제안하는 모바일 환경에서의 즐거리 기반 검색 시스템은 크게 검색 서버와 모바일 클라이언트로 구성된다. 검색 서버는 모바일 클라이언트에서 입력한 검색어를 수신하고 분석하여 검색 결과를 클라이언트 시스템에 전송하는 역할을 한다. 서버에서는 검색 결과에 랭킹을 부여하여 PDA의 클라이언트 시스템에 전송한다. 모바일 클라이언트에서는 검색어 입력 인터페이스를 제공하고, 서버에서 전송해준 검색 결과를 사용자에게 보여준다.

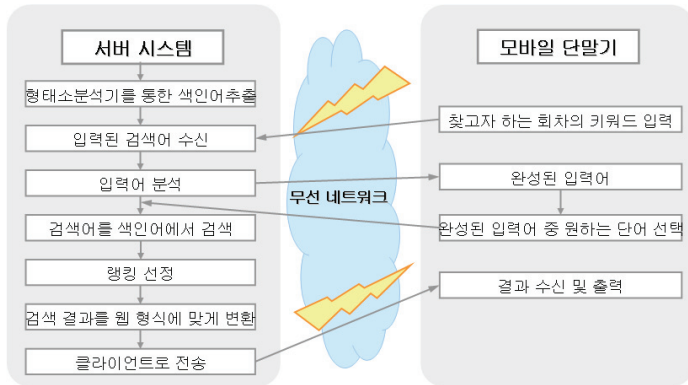
실험 결과 데이터베이스에 기반한 검색어 입력 인터페이스에 대해 대부분의 피실험자들이 만족 하였으며, 드라마 회차 검색 결과에서도 상위 2위까지 검색결과가 90%로 높게 나왔다.

논문의 구성은 다음과 같다. 2장에서는 전체 시스템의 구조, 즐거리 검색 서버의 구현 및 클라이언트

This research was supported by the Program for the Training of Graduate Students in Regional Innovation which was conducted by the Ministry of Commerce Industry and Energy of the Korean Government. This work was supported by the Industry University Research Institute Consortium grant from the Small & Medium Business Administration.

시스템의 구현에 대해 설명한다. 그리고 3장에서는 구현한 시스템의 실험결과에 대해 설명하고 마지막으로 4장에서는 결론을 맺고 향후 연구 내용에 대해 설명한다.

2. 시스템의 설계 및 구현



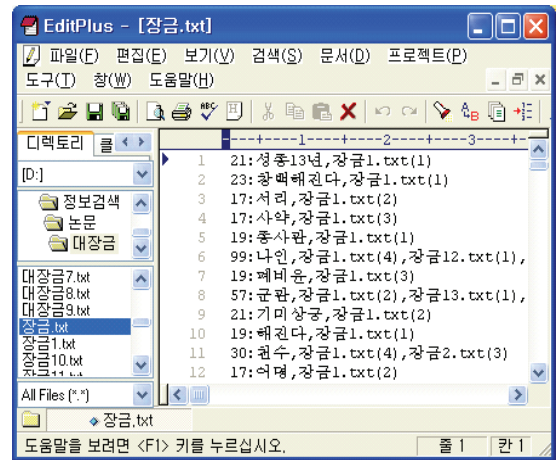
[그림 1] 시스템 구성도

그림 1과 같이 줄거리 기반 검색 서비스는 서버와 모바일 클라이언트로 구성된다. 서버는 클라이언트의 요청을 대기하고 있다가 클라이언트의 접속 요청을 받고 클라이언트로부터 입력된 검색어를 분석하여 결과를 돌려준다. 모바일 장치는 입력공간이 협소한 관계로 모바일 단말기는 단어의 초성들만 서버로 전송한다. 예를 들어, 사용자가 '대장금'을 검색하고자 하는 경우 초성인 'ㄷㅈㄱ'만 입력한다. 서버는 클라이언트에서 입력한 초성을 분석하여 완성된 검색어 목록을 클라이언트로 전송한다. 사용자는 목록에서 원하는 검색어를 선택하여 검색어를 서버에 전송하고, 서버에서는 전송받은 검색어와 추출된 색인어를 이용하여 검색을 하고 검색된 결과에 랭킹을 부여한다. 서버에서 검색된 결과는 텍스트, 이미지, 동영상 등의 형태로 클라이언트에 전송을 하도록 설계하였다.

2.1. 검색 서버

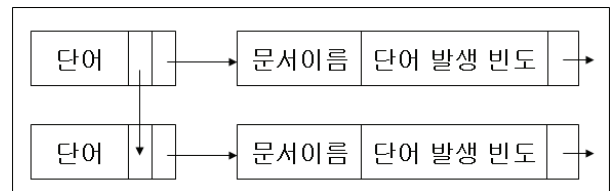
줄거리 검색 서버는 각 회에 해당하는 줄거리를 텍스트파일로 저장하고 있다. 각 텍스트 파일에 대해 형태소 분석기를 이용해 색인어를 추출한다. 추출된 색인어를 가지고 추출된 색인어의 발생 빈도를 포함하는 새로운 파일을 생성한다. 이 과정은 사용자가 입력한 검색어를 검색할 기초 데이터로 사용된다. 첫 번째 숫자는 ':'문자 이후로 한 라인의 바이트 수를 나타내고 그 음에 오는 ':'는 단순한 구분자이며 구분자 다음에 색인어가 위치한다. 색인어 다음에는

색인어가 출현한 문서의 이름과 '(')안에 색인어가 해당 문서에 출현한 횟수를 나타낸다. 색인어와 문서, 문서와 문서는 ','로 구분한다.



[그림 2] 기초 데이터 예시

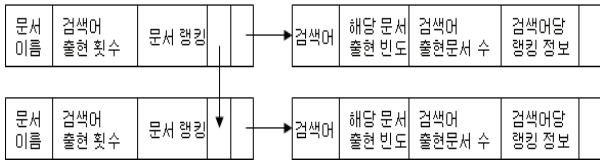
기초 데이터가 생성되면 서버에서는 사용자가 입력한 검색어를 이용하여 해당 줄거리를 검색하고 랭킹을 하는 과정을 처리한다. 검색 서비스의 요청이 발생하면 서버는 기초 데이터 파일의 색인어를 기준으로 서버에 연결 리스트로 로딩한다.



[그림 3] 색인어 데이터 구조

다음 그림 4의 자료구조는 검색된 단어를 포함하는 문서를 기준으로 생성된 것으로 문서이름, 검색어 출현 횟수, 문서 랭킹 정보, 해당 검색어 문서 출현 횟수, 검색어 출현 문서 수, 검색어 랭킹 정보를 포함한다. 문서 이름은 검색어를 포함하고 있는 줄거리 파일의 이름이며, 검색어 출현 횟수는 해당 문서에 사용자가 입력한 검색어가 몇 개나 출현했는지 그 수를 의미한다. 그리고 문서의 랭킹 정보는 해당 문서에 출현한 각각의 검색어에 대한 랭킹 정보의 값을 더한 값이며, 검색어는 사용자가 입력한 검색어 중에서 해당 문서에 출현한 검색어를 의미한다. 해당 문서 출현 횟수는 문서 파일을 기준으로 해당 검색어의 출현 횟수를 나타내고, 검색어 출현 문서 수는 해당 검색어가 전체 스토리 파일 중에서 몇 개의 파일에 출현했는지를 나타낸다. 마지막으로 검색어 랭킹 정보는 해당 문서에 검색어가 출현한 횟수

나누기 검색어가 출현한 문서의 수이다.



[그림 4] 검색 결과 데이터 구조

검색 결과 데이터에 순위를 지정하기 위해 다음의 식을 이용하여 R 값이 가장 큰 파일에 대해 높은 순위를 부여한다. 순위가 결정되면 결과를 클라이언트에 전송한다.

$$R = \sum_{i=1}^n \frac{TF_i}{N_i}$$

- n = 해당 문서에서 검색된 단어의 수
- TF = 해당 검색어가 해당 문서에 출현한 횟수
- N = 해당 검색어가 출현한 문서의 수
- R = 문서 랭킹 정보

2.2. 모바일 클라이언트

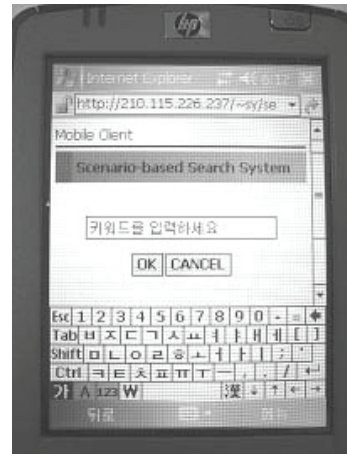
모바일 단말기에서 검색 서버로 웹 접속을 하면 단말기의 웹브라우저에 검색을 위한 입력창이 나타난다. 모바일 단말기의 입력 공간이 협소한 관계로 단말기의 터치 패드를 이용해 검색하고자 하는 단어의 초성들만을 입력한다. 예를 들면 ‘대장금’이라는 검색어를 키워드로 하고 싶다면 ‘ㄷㅈㄱ’과 같이 입력을 하여 서버로 전송을 한다.

| 초성 | 0 | 1 | 2 | 3 | .. | .. | .. | n |
|-----|----|----|----|----|----|----|----|----|
| ㄱ | 가 | 갸 | 거 | 겨 | .. | .. | .. | 강 |
| ㄴ | 나 | 냐 | 너 | 녀 | .. | .. | .. | 녕 |
| ㄷ | 다 | 댜 | 더 | 더 | .. | .. | .. | 딩 |
| ㄹ | 라 | 랴 | 러 | 려 | .. | .. | .. | 링 |
| ... | .. | .. | .. | .. | .. | .. | .. | .. |
| ㅎ | 하 | 햐 | 허 | 혀 | .. | .. | .. | 형 |

[표 1] 문자 테이블

서버는 이미 색인어를 추출하여 저장을 하고 있기 때문에 글자 ‘대’라는 글자가 있으면 문자 테이블을 참조하여 이 문자의 초성 ‘ㄷ’을 찾아서 2차 색인어로 저장한다. 예를 들면 ‘대장금’이라는 단어가 있으면 ‘ㄷㅈㄱ’을 2차 색인어로 저장을 한다. 클라이언트가 ‘ㄷㅈㄱ’을 입력하면 서버는 2차 색인어를 이용해 ‘ㄷㅈㄱ’이 초성으로 이루어진 모든 단어를 찾아오게 된다. 서버에서 똑같은 단어가 검색이 된 경우 검색된 모든 단어들을 클라이언트로 전송을 하고, 클라이언트는 원하는 검색어를 선택하고 다시 서버로 전송을 한다. 중성과 종성의 입력을 제거하고 필

요시 검색된 단어 중 원하는 단어를 선택하기 위한 한 번의 클릭만 추가 되므로 입력시간을 단축한다.



[그림 5] 클라이언트 실행 화면

모바일 단말기의 인터페이스는 그림 5와 같이 키워드를 입력할 수 있는 공간과 터치패드 형태의 키보드를 보여주고 있다. 사용자는 펜을 이용하여 키워드에 해당하는 단어의 초성만 입력하면 된다. 검색어를 입력하고 전송을 하면 출력되는 검색 결과는 다음 리스트형태로 보여준다.

1. 검색된 결과에 해당하는 회차
2. 검색된 회차의 줄거리 요약
3. 동영상 재생 링크



[그림 6] 검색 결과 화면

검색된 결과화면에서 텍스트 형태의 줄거리를 간단히 요약한 내용을 보여주고 동영상 파일의 링크를 클릭하면 해당 회차의 동영상을 감상할 수 있다.

3. 실험 결과

본 실험에서 제안한 모델의 유용성을 검사하기 위해 MBC 드라마 시리즈 ‘대장금’의 줄거리를 실험 데이터로 사용하여 테스트 하였으며 얼마나 정확한 결과가 상위에 랭크가 되는지에 초점을 맞추었다. 실험은 20명의 피실험자를 2인 1조로 10개의 팀으로 나

누어 대장금 32회를 시청하도록 한 후 각각 5단어로 구성된 검색어 집합을 입력하도록 하여 검색 시스템을 실험해 보았다.

| 실험자 | 랭킹순위 | | | | |
|-----|------|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 |
| A | 31 | 32 | 27 | 33 | 11 |
| B | 32 | 24 | 31 | 30 | 11 |
| C | 32 | 24 | 31 | 29 | 7 |
| D | 32 | 31 | 7 | 33 | 49 |
| E | 32 | 33 | 11 | 31 | 49 |
| F | 28 | 31 | 29 | 32 | 8 |
| G | 32 | 27 | 29 | 31 | 29 |
| H | 32 | 33 | 7 | 31 | 11 |
| I | 33 | 32 | 11 | 31 | 30 |
| J | 32 | 27 | 11 | 31 | 30 |

[표 2] 실험 결과

[표 2]에 보이는 실험 결과는 가장 상위에 랭크된 결과 5가지를 선정 하였다. 표2와 같이 실험 결과 피실험자의 특성상 각자 다른 의미에서 검색어를 입력한 경우를 제외하면 32회가 동일하게 상위 랭크된 결과를 볼 수 있다. 검색에 실패한 피 실험자들은 ‘한 상공의 죽음’이라는 검색어를 사용하여 검색에 실패하였다. 또한 검색어 입력은 키보드와 동일한 위치의 키 배열을 사용한 터치패드 형태의 입력도구를 사용해 사용자들이 익숙한 환경에서 문자입력이 가능했다.

| | 설문 내용 | 만족 | 보통 | 불만 |
|---|-------------|----|----|----|
| 1 | 문자 입력 방법 | 75 | 20 | 5 |
| 2 | 문자 입력 시간 | 70 | 15 | 15 |
| 3 | 문자 입력 인터페이스 | 60 | 25 | 15 |
| 4 | 사용 방법 | 80 | 20 | 0 |
| 5 | 결과 출력 화면 | 80 | 10 | 10 |

[표 3] 설문 조사

[표 3]은 20명의 피실험자를 대상으로 입력 방법에 대한 만족도를 조사한 결과이다. 문자입력 방법에 대해서 5%의 사용자가 불만을 표현 하였으며, 문자 입력 시간에 대한 불만은 15%의 사용자가 불만을 표현하였다. 문자 입력 방법과 문자 입력 시간에 대한 불만을 표현한 사용자들은 핸드폰 키패드 조작 능력이 높은 사용자였고, 핸드폰에 비해 문자를 입력하는데 불편함을 표현하였다. 문자 입력 인터페이스에 대해 60% 사용자가 만족을 하였다. 보통이나 불만을 표현했던 사용자들은 가상 키보드의 그래픽에 대해 지적을 하였으며, 입체적인 가상키보드 수준을 원했다. 하지만 대부분의 사용자들이 모바일

단말기와 친숙하기 때문 처음에 입력하는 방법만 설명하면 검색어를 입력하는데 있어 별다른 문제가 발견되지 않았으며 대체로 만족하는 결과를 얻을 수 있었다.

4. 결론 및 향후 과제

본 논문에서는 모바일 환경에서 드라마 동영상의 검색을 쉽게 하기 위해 즐겨리 기반 TV 시리즈물 검색 시스템에 데이터베이스에 기반한 검색어 입력 방법을 사용한 시스템을 제안하였다. 제안한 시스템은 초성, 중성, 종성을 모두 입력하는 방법에 비해 초성만 입력하여 자동으로 검색할 단어를 완성하기 때문에 버튼의 클릭횟수가 줄어들어 검색어 입력시간이 크게 줄어들었다. 본 시스템은 모바일 환경에서 멀티미디어 콘텐츠 서비스를 이용하는 사용자에게 보다 편리한 검색 서비스를 제공 할 것이다. 향후 연구로는 드라마의 회차 뿐만 아니라 원하는 장면을 검색하는 기능을 추가 하고자 한다.

원하는 장면을 검색하기 위해서는 동영상 프레임의 메타 데이터에 장면에 대한 설명이 있어야 하고 동영상을 프레임 단위로 검색 전송할 수 있는 스트리밍 서버와 스트링 동영상 플레이어에 대한 연구가 필요하다.

참고문헌

- [1] 조진표, 천영우, 김유섭, 고영웅, “즐거리에 기반한 TV 시리즈물 검색 시스템의 설계 및 구현” HCI2006 , pp.1256-1261.
- [2] Gerard Salton, Edward A. Fox, and Harry Wu, “Extended Boolean information retrieval”, Communications the ACM, 26(11):1022-1036, 1983. 11.
- [3] <http://cinexox.com/>
- [4] 이용일, 허혜선, 홍윤식, “PDA를 사용한 H.263 영상데이터 전송 시스템” 한국 정보처리학회, 13회 1-2 pp 181~186 2004.2
- [5] 허혜선, 이용일, 홍윤식, “유무선 통합 인프라스트럭처 네트워크에서 PDA를 사용한 데이터 전송 성능분석”, 한국 인터넷 정보학회 논문지 제5권 제 3호, pp.99-108, 2004. 6.
- [6] 이계숙, 용환승, “데이터베이스에 근거한 자동 키보드의 입력방법”, 멀티미디어 학회 논문지 Vol. 8, No. 1, 2005.1