

모바일 웹 브라우징 환경에서의 전화 번호 추출 및 자동 다이얼링*

전혁수⁰ 신동욱 김태환 최중민

한양대학교 컴퓨터공학과

hyeksu@gmail.com⁰ foremostd@gmail.com kimth@islab.hanyang.ac.kr jmchoi@hanyang.ac.kr

Phone Number Extraction and Auto Dialing in Mobile Web Browsing Environment

Hyuksu Jun⁰, Dongwook Shin, Taehwan Kim, Joongmin Choi

Department of Computer Science and Engineering, Hanyang University

1. 서 론

모바일 기기의 장점은 언제 어디서나 원하는 정보를 검색 가능하다는 점이다. 이러한 장점을 기반으로 모바일의 사용량은 기하급수적으로 증가하고 있다. 실제 모건 스탠리(Morgan Stanley)의 모바일 인터넷 보고서[1]에서 2010년 현재 노트북(notebook)과 스마트폰(smart phone)을 사용하는 사용자의 수가 데스크톱을 사용하는 사용자의 수를 능가한 상태이고, 2012년에는 스마트폰을 사용하는 사용자의 수가 노트북과 데스크톱을 포함한 모든 컴퓨터의 사용자 수를 능가할 것으로 예상된다. 이러한 스마트폰의 열풍에 힘입어, 모바일의 특징을 이용한 다양한 상업용 소프트웨어나 연구가 진행되고 있다. 실제 언제 어디서나 인터넷에 접속할 수 있는 모바일의 최대 장점을 이용한 소셜 네트워크 서비스(Social Network Service)나 모바일의 또 다른 특징인 위성위치확인시스템(Global Positioning System)을 이용한 다양한 소프트웨어가 개발되고 있다.

하지만 스마트폰은 기본적으로 전화기이기 때문에 가장 중요한 기능은 전화 기능이다. 모건 스탠리에서 아이폰 사용자를 대상으로 기능별 사용빈도를 조사한 결과와 같이 실제 전화를 걸기 위한 사용 빈도는 45%로 실제 스마트폰 이용량 중 약 절반을 차지하는 비율이다. 그럼에도 불구하고, 실제 스마트폰 기능 중 사용자가 좀 더 편리하게 전화 기능을 사용하기 위한 소프트웨어나 연구는 상대적으로 활발히 진행되지 않고 있다. 현재 보급된 스마트폰의 경우, 웹 페이지에서 전화번호로 예상되는 연속된 숫자를 사용자가 선택하면 전화 연결을 시도하지만, 계좌번호, 사업자 등록 번호 혹은 지역 번호가 생략된 전화번호 등 올바르지 않은 번호에 대하여 전화 연결 시도도 허용한다.

본 논문에서는 이러한 문제점을 해결하기 위해, 전화번호 추출을 위한 패턴을 정의한 후, 웹 페이지에서 전화번호를 추출하고, 추출된 전화번호 중 올바르지 않은 형태의 전화번호나 지역 번호가 생략된 전화번호의 정제를 통해 사용자에게 올바른 전화번호를 자동으로 제공하는 시스템을 제안하고자 한다.

2. 본 론

정보통신기술의 발달과 단말기의 획기적인 성능 향상으로 인해 스마트폰의 보급이 확산되고, 언제 어디서나 스마트폰을 이용한 인터넷이 가능하게 되었다. 하지만 현재의 웹 환경은 화면이 크고 접근이 용이한 데스크톱에 맞춰져 있기 때문에 화면 크기가 작은 스마트폰으로의 웹서핑이 용이하지 않다. 이러한 문제를 해결하기 위해 활발한 연구가 진행 되었다. 일반적인 웹 페이지를 모바일 상에서 보기 위해서는 변형이 일어나야 하는데, 변형이 일어나는 장소에 따라 서버, 프록시, 클라이언트 세 가지로 나뉜다[2]. 하지만 기존의 연구들은 모두 사용자가 모바일 환경에서 웹에 접속하는데 있어서 좀 더 편리한 환경을 구축하기 위한 부분에만 초점이 맞추어져 있다. 모바일의 주된 기능은 전화연결 기능이고, 사용자의 실제 모바일 사용량 중 대부분이 전화에 사용된다는 점은 많은 설문 조사를 통해 검증되었다. 또한, 모바일 환경에 적합하게 웹 페이지를 보여주기 위한 많은 방법들의 경우, 여전히 데스크톱 환경에 비해 많은 단점들이 존재한다.

본 논문에서 제안하는 모바일 웹 브라우징 환경에서의 전화번호 추출 및 자동 다이얼링 시스템의 구조는 모바일 환경의 낮은 컴퓨팅 파워로 인한 느린 응답속도를 높이기 위해 서버-클라이언트 구조로 시스템을 설

* 본 논문은 지식경제부 산업원천기술개발사업(10035348, 모바일 플랫폼 기반 계획 및 학습 인지 모델 프레임워크 기술 개발)의 지원으로 수행되었음

계하였다.

사용자가 스마트폰을 사용하여 웹 페이지에서 전화 번호 추출을 요청하면 서버로 해당 웹 페이지의 URL 이 전송된다. 그 후, Phone Number Extractor가 웹 페이지에 존재하는 전화번호를 추출한다. Preprocessor는 웹 페이지의 콘텐츠에서 HTML 태그와 Script 태그 등을 모두 제거한 후, Phone Number Detector가 정의된 패턴과의 매칭을 통해 전화번호를 추출한다. Phone Number Validator는 추출된 전화번호가 올바른 전화번호인지 아닌지 검증한 후, 잘못된 전화번호를 제거한다. 이 때, 추출되는 전화번호는 지역번호를 포함하는 전화번호와 지역번호가 생략된 전화번호에 해당하는 두 가지 경우가 존재한다.

Area Code Extractor는 지역번호가 생략된 전화번호에 적합한 지역번호를 할당한다. Area Code Detector는 Text Boundary Detector에서 타겟 전화번호에 대한 텍스트 블록을 설정한다. 그 후, Region Name Extractor가 웹 페이지의 콘텐츠에서 지역명을 나타내는 명사들을 추출한다. Area Code Searcher는 Area Code Database에서 각 지역명에 해당하는 지역번호를 탐지하여 후보 지역번호를 추출한다.

그 후, Area Code Selector에서 확률을 기반으로 가장 적합한 지역번호를 전화번호에 할당한다.

Phone Number Coordinator는 사용자에게 추출된 전화번호를 적합한 형태로 구성하여 반환하는 모듈로, String Tagger는 전화번호에 대한 추가적인 정보를 제공하기 위해, 웹 페이지에서 전화번호가 나타난 위치를 기준으로 전·후 문자열을 함께 제공한다. Multiple Number Splitter는 연속 전화번호(예 : 02)991-9991~4)를 올바른 전화번호 형태로 구성한다.

Area Code Modifier는 시스템에서 판단한 지역번호가 올바르지 않을 경우, 사용자에게 음성으로 피드백을 받아 지역번호를 수정한다.

시스템의 성능 평가를 위해, 총 115개의 웹 페이지와 관련된 963개의 전화번호에 대하여 실험하였다. 실험에 사용된 전화번호의 형식과 각 형식별 전화번호의 수는 [표 1]과 같다.

[표 1] 전화번호 형식 및 형식별 전화번호의 수

번호	전화번호 형식	전화번호 수
1	지역번호를 포함하는 단일번호	376
2	지역번호가 생략된 단일번호	507
3	지역번호를 포함하는 연속번호	53
4	지역번호가 생략된 연속번호	27

[표 2] 시스템 성능 평가

형식	정확률	재현률	F-지수
SN(AC)	0.96	0.87	0.91
SN(NAC)	0.89	0.85	0.87
MN(AC)	0.85	0.83	0.84
MN(NAC)	0.6	1	0.75

실험은 정확률과 재현률, F-지수를 이용하였다. 실험 결과는 [표 2]와 같다. SN(AC)는 [표 1]의 번호 1에 해당하는 전화번호 형식으로, 지역번호를 포함하는 단일번호의 경우 매우 높은 성능을 보이는 것을 알 수 있다. SN(NAC)는 번호 2에 해당하는 형식으로, 시스템에서 지역번호를 자동으로 추출하였음에도 불구하고, 매우 높은 성능을 보인다. MN(AC)는 번호 3에 해당하는 형식으로, 전화번호가 연속번호일 경우에도 대부분 올바르게 분할하는 것을 알 수 있다. 마지막으로 MN(NAC)는 번호 4에 해당하는 형식으로, 다른 형식들에 비해 정확률이 떨어지는 것을 확인할 수 있다. 그 이유는 지역번호를 포함하지 않는 연속된 번호이기 때문에 지역번호의 올바른 추출과 연속번호의 올바른 분할, 두 가지를 모두 만족시켜야 하기 때문에 다른 형식에 비해 낮은 정확률과 F-지수를 보인다. 하지만 모든 형식을 고려한 시스템의 성능은 정확률 89.8%, 재현률 86.0%, F-지수 87.9%로 만족할만한 성능을 보이는 것을 확인할 수 있다.

3. 결 론

본 논문에서는 모바일 웹 브라우징 환경에서의 전화번호 추출 및 자동 다이얼링 시스템을 제안하였다. 웹 페이지의 콘텐츠에서 전화번호를 추출하고, 지역번호가 생략되었거나 연속번호로 구성된 전화번호의 경우에도 올바른 형태의 전화번호로 구성하여 사용자에게 반환한다. 또한 올바르지 않은 지역번호의 수정에 대한 사용자 편의를 제공하기 위해 음성인식을 통한 수정 방법도 제공한다. 제안한 시스템은 실제 웹 페이지를 기반으로 한 실험에서도 만족할만한 성능을 보였다.

앞으로의 연구를 통하여 전화번호 추출 패턴과 지역번호 선택 방법의 정제를 통해 좀 더 정확한 시스템을 구축하고, 모바일 상에서 모든 처리가 이루어지는 시스템을 구축하고자 한다.

참고문헌

[1] Morgan Stanley : The Mobile Internet Report.
http://www.morganstanley.com/institutional/techresearch/pdfs/mobile_internet_report.pdf, 2009

[2] T. Laakko, T. Hiltunen, Adapting Web Content to Mobile User Agents, IEEE Internet Computing, vol. 9, no. 2, pp. 46-53, 2005.