

정규표현식을 이용한 안드로이드 단말기의 단문 메시지 분류

○ 최재혁*, 정성수*, 이운호*, 김용석*
*강원대학교 컴퓨터정보통신공학과
e-mail: idnamehyuk@naver.com

Classification of Short Messages on Android Phones Using Regular Expresson

Jae-Hyuk Choi*, Sung-Soo Jung*, Yoon-Ho Lee*, Yong-Suk Kim*
*Dept of Computer Engineering, Kangwon University

요 약

기존 안드로이드 단말기에서 제공되는 문자기능은 단순히 수신만 가능한 것에 그쳐 사용자가 원하는 정보만 얻음에 있어 다소 불편함이 있었다. 기존 불편함을 보완하고자 정규표현식을 이용한 단문 메시지 파싱 기능을 구현하여 원하는 정보를 분류해서 볼 수 있게 구현하였다.

1. 서론

인간의 정보 습득에 대한 욕망은 정보통신기술의 빠른 발전을 이룩하게 하였고 현재에 이르러서는 인터넷 없이는 문화생활을 누릴 수 없을 정도로 우리 생활에 깊숙이 자리 잡고 있다. 하지만 더 빠르게 정보를 얻고자 하는 유저들의 욕구를 충족시키기에는 PC에기반한 인터넷은 부족함이 많았고 결국 그들은 모바일기술에 눈을 돌리기 시작하였다. 이러한 추세는 모바일이 전성기를 맞는 계기가 되었고 모바일 유저는 폭발적으로 증가 하였다. 시대에 발 맞추어 해당 기업들은 다양한 모바일 전용 플랫폼을 개발하여 출시하였다. 지금은 블랙베리OS와 심비안OS, 애플의 아이폰OS가 시장을 크게 양분하고 있다. 그중 후발 주자로서 오픈소스로 호평을 받고 있는 안드로이드가 새롭게 각광을 받고 있다.

안드로이드 플랫폼은 설계자들이 포괄적인 방식으로 개발하여 안전한 운영체제를 기반으로 풍부한 어플리케이션을 개발할 수 있는 프레임워크를 구축 하는 완전성, 오픈소스 라이선스를 통해 제공되어 개발자들의 자유로운 접근이 가능한 개방성, 그리고 개발 부대비용이 무려라는 특징을 가지고 있다.

본 연구는 안드로이드 플랫폼의 신기술과 자바의 문자 파싱기술을 응용하여 SMS(Short Message Service) 메시지의 파싱구현에 초점을 두었다. 기존의 메시지 서비스는 단지 검색 기능만을 제공하기 때문에 정보화시대에 넘치는 불필요한 정보들을 사용자가 일일이 분류하기에는 많은 시간과 자원이 소모가 된다. 이점을 보완하고자 사용자가 필요한 정보만을 얻을 수 있도록 정규표현식을 이용하여 필요한 문장을 찾고 메시지를 구분 할 수 있도록 구현

하였다. 예를 들어 카드 사용 내역 통지 서비스, 통장의 입출금 통지 내역 서비스 문자에 이 기술을 적용하여 가계부 프로그램을 만든다면 문자의 사용처, 입출금 금액, 날짜 등을 추출하여 데이터베이스에 자동 저장하고 이 데이터베이스 내용을 참조하여 사용자가 가계부에 문자의 입출금 내역을 입력하는 불편을 줄일 수 있을 것이다.

2. 본론

2.1 종래기술

기존의 메시지 서비스는 새로운 문자메시지를 수신하는 경우, 수신한 문자메시지를 무조건 메모리에 저장하고 있다. 이로 인하여, 사용자에게 불필요한 정보와 필요한 정보 모두 한 곳에 저장되며 필요한 정보를 보고 싶을 때마다 검색하는 방식을 쓰고 있다. 또한 불필요한 정보를 매번 삭제해 줘야하는 번거로움이 있다. 이와 같은 문제점을 해결하기 위하여 다음과 같은 기술을 창안하였다.

2.2 개요

안드로이드에서 문자메시지를 수신하여 원하는 정보를 얻으려면 문자메시지 수신을 담당하는 통신 수신자, 수신된 문자에서 원하는 정보를 추출하는 파싱함수, 추출한 정보를 저장할 수 있는 데이터베이스, 이 모든 것을 처리 하는 모바일 디바이스에 최적화된 가상머신이 필요하다. 그림 1에서 보는 것처럼 메시지를 수신하면 SMS리스너에서 수신된 문자를 SMS파싱함수에 전달하여주고 SMS파싱함수는 파싱된 문자를 데이터 베이스에 저장한다.



(그림 1) 동작

2.3 가상기계

DVM(Dalvik Virtual Machine)이란 모바일 디바이스를 위해 최적화 되어 있는 가상머신을 말한다. 기본적으로 안드로이드는 Java를 이용하지만 안드로이드 SDK로는 Java소스를 Dalvik 바이트코드로 직접 컴파일 할 수 없다. 처음에는 Java컴파일러를 사용하여 Java 바이트코드를 생성한 뒤에, 안드로이드 SDK에 dx라는 툴을 이용하여 Dalvik 바이트코드로 변환한다. 변환된 코드는 DVM위에서 작동하며 Linux Kernel을 사용하기 때문에 프로세스 독립성을 가지고, 메모리 관리와 스레딩을 상용 할 수 있다. 안드로이드 어플리케이션은 기본적으로 사용자눈에는 보이지 않지만 DVM 위에서 작동하게 된다.

2.3 통신 수신자

안드로이드는 시스템 이벤트를 각 어플리케이션에 전달하기 위해 통신 수신자(Broadcast Receiver)클래스를 이용한다. 어플리케이션은 브로드캐스트 수신자를 구현하기 위해서 Broadcast Receiver클래스를 상속받은 후 onReceive 이벤트 핸들러를 오버라이드 해야한다. 브로드캐스트 수신자에서는 간단한 처리만 수행해야 하는데, onReceive이벤트 핸들러에서는 10초라는 제한된 시간 동안에만 작업을 해야 하기 때문이다. 따라서 문자메시지를 수신에 대한 이벤트를 처리하기 위해서 Broadcast Receiver클래스를 상속받아 문자메시지를 수신하여 구문 처리하는 Parsing함수로 내용을 넘겨주어야 한다. 다음 그림 2와 같이 통신 수신자 클래스를 이용하면 된다.

```

public class SMSReceiver
extends BroadcastReceiver {
private static final String ACTION
="android.provider.Telephony.SMS_RECEIVED";
public void onReceive(Context context,
Intent intent) {
    if(intent.getAction().equals(ACTION) {
    }
}
}
    
```

(그림 2) 통신 수신자 함수

2.4 정규표현식

문자를 파싱하기 위해서 정규표현식을 사용하였다. 정규표현식이란 텍스트를 찾고 조작하는데 쓰는 문자열이다. 정규 표현식은 정규 표현 언어를 사용해 만들지만 완전한 프로그래밍 언어가 아니며 다른 프로그래밍 언어나 제품에 포함된 작은 언어일 때가 많다.

<표 1> 정규표현식

정규식	설명
^	문자열의 시작을 의미
\d	숫자를 의미
\$	문자열의 끝을 의미
\s	문자를 의미

^와 \$를 통해 문자의 시작과 끝을 알리고 \d나 \s를 이용하면 메시지 내용중 숫자나 문자를 추출해낼 수 있다. 따라서 위에 정규표현식을 가지고 원하는 정보를 얻어내는데 도움을 얻을 수 있다. 예를 들면, "are\$"로 "who are you?"를 파싱하였을 때, "who are"가 검색되며 "^are"로 파싱하면 "are you?", "\d"로 검색하면 "who are you", "\s"로 검색하면 문자열에 숫자가 없기 때문에 결과가 나오지 않는다.

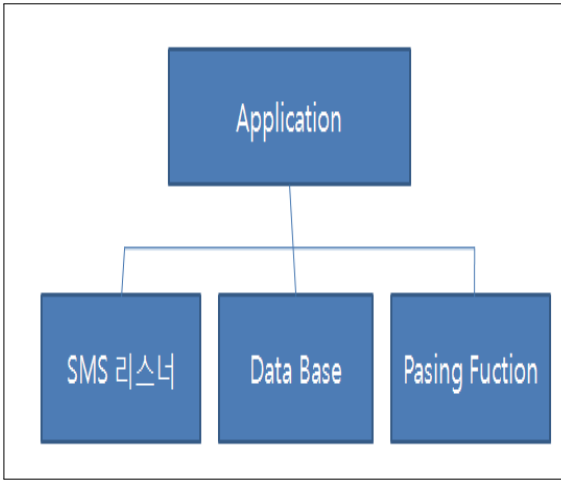
2.5 데이터베이스

파싱된 문자를 저장하기 위해서는 데이터베이스를 써야하는데 SQLite는 Oracle이나 MS-SQL, MySQL처럼 무겁지 않으며, 아주 가볍고 빠르고 간결한 DB엔진이다. 위에 언급한 MySQL같은 데이터베이스들은 큰 규모의 데이터베이스 프로그래밍을 지원하기에 적합하며 간단하고 빠른 용도의 데이터베이스로는 적합하지 않다. SQLite는 해당 장치에서만 사용가능하며 원격 데이터베이스로의 연결이 발생하지 않기 때문에 데이터로의 접근이 매우 빠르며, 표준 SQL구문을 모두 지원하기 때문에 데이터의 관리도 용이하다. 따라서 모바일 데이터베이스용으로 SQLite가 적합하며 실제로 모바일용으로 널리 쓰이고 있다.

2.6 구현방법

본 프로그램은 Java를 이용하여 문자내용을 파싱함수를 구현하고, 안드로이드 SDK를 이용하여 SMS Manage 기능을 얻어 원하는 문자내용을 데이터베이스에 저장하는 것이다. 프로그램은 크게 다음 그림 3과 같이 나누어진다.

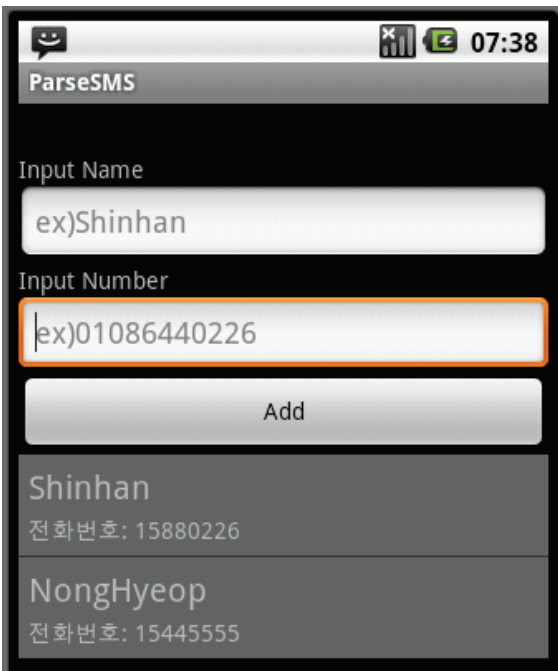
먼저 안드로이드 프로젝트 생성시 BroadCast Receiver 클래스를 상속받는다. 상속받은 클래스에서 onReceive 이



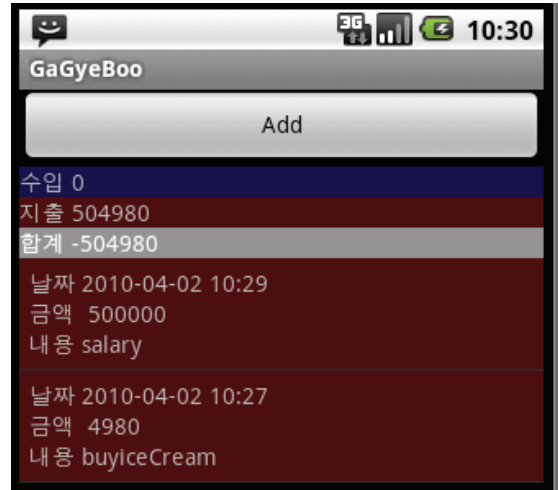
(그림 3) Application 구성

벤트 핸들러를 오버라이드 한다. onReceive 이벤트 핸들러에서는 단문 메시지가 수신됐을 시에 단문 메시지의 내용과 번호를 얻을 수 있다. 이 얻어진 정보를 Java의 정규표현식을 이용한 문자과싱함수에 전달하게 된다. 문자과싱함수는 전달받은 정보를 가지고 원하는 정보만 추출해서 얻을 수 있다.

예를 들면, 원하는 문자가 있는 단문 메시지만 분류할 수도 있으며 원하는 번호가 있는 단문 메시지만 따로 분류하여 저장 할 수 있다. 단문 메시지를 분류할 때는 사용자가 원하는 문자나 번호를 입력받아 그 정보만 추출할 수 있도록 옵션 데이터 베이스에 저장한 뒤 문자과싱함수에서 옵션 데이터 베이스에 저장된 정보를 이용하여 정보를 분류한다. 문자과싱함수에서 분류된 정보들은 분류된 내용을 저장하는 데이터베이스에 전달되어 저장한다. 사용자는 이 데이터베이스에 접근하여 원하는 정보를 볼 수 있게 된다.



(그림 4) 응용예 - 전화번호 등록 액티비티



(그림 5) 응용예 - 가계부

3. 결론

본 연구에서는 개방형 모바일 플랫폼인 안드로이드에서 정규표현식을 이용한 문자 파서를 구현하였다. 이를 통해 문자 내용에서 사용자가 원하는 정보만을 추출하여 데이터베이스 필드에 자동 저장되도록 하였다. 또한 단문 메시지의 발신자 전화번호를 분류하는 기능을 통해 등록된 번호에서 온 문자만 문자과싱 처리를 받도록 하여 모든 문자가 문자과싱 함수 처리를 받는 오버헤드를 피하도록 하였다. 위의 두 기능을 통해서 특정발신자의 특정 정보만을 담은 데이터베이스를 만들 수 있고 이를 이용하는 응용프로그램을 개발할 수 있다.

이번 연구를 통해 문자 파싱을 이용한 사용자 편의 프로그램의 좋은 표본이 될 것이며, 문자를 이용한 응용프로그램 개발에 도움이 될 것이다.

참고문헌

- [1] 리토 마이어, "프로페셔널 안드로이드 애플리케이션 개발", 제이펍, 2009
- [2] Shane Conder/류광, "[시작하세요!] 안드로이드 프로그래밍 : 모바일 소프트웨어 개발", 위키북스, 2009
- [3] 케이 S. 호스트만 저 | 오세만 역 "Big JAVA(제2판)", 생능출판사, 2006
- [4] <http://java.sun.com/j2se/1.5.0/docs/api/java/util/regex/Pattern.html>
- [5] <http://java.sun.com/j2se/1.5.0/docs/api/java/util/regex/Matcher.html>