

Development of technology to improve information accessibility of information vulnerable class using crawling & clipping

Seong-Bae Jeong*, Kyung-Shin Kim**

Abstract

This study started from the public interest purpose to help accessibility for the information acquisition of the vulnerable groups due to visual difficulties such as the elderly and the visually impaired. In this study, the server resources are minimized and implemented in most of the user smart phones. In addition, we implement a method to gather necessary information by collecting only pattern information by utilizing crawl & clipping without having to visit the site of the information of the various sites having the data necessary for the user, and to have it in the server. Especially, we applied the TTS(Text-To-Speech) service composed of smart phone apps and tried to develop a unified customized information collection service based on voice-based information collection method.

▶ Keyword : Information Vulnerable Layer, Crawling, Clipping, Text Summarize, Text-To-Speech, Information gathering

I. Introduction

본 연구는 노인과 시각 장애인 등과 같이 시각에 어려움으로 인해 정보취득에 취약한 계층의 정보 취득에 도움을 주고자하는 공익적 목적에서 출발하였다. 본 연구의 핵심은 스마트폰 자체 기능만을 사용해 크롤링&클리핑을 활용하고, 안드로이드 사용자 스마트폰 앱(app)으로 구성된 음성변환장치(TTS : Text-To-Speech) 서비스를 적용하여 음성 중심 정보 수집 방법을 중심으로 획일화 및 단일화 된 맞춤형 정보 수집 서비스 개발에 목표를 두고 연구가 진행되었다.

지식정보 사회를 살아가는 인간에게 정보는 사회·경제·문화 등 다양한 활동을 하는데 필수적인 자원이며, 이 정보는 정보통신기술의 발달에 힘입어 개인용 컴퓨터나 스마트폰으로 인터넷에 접속하여 간편하게 정보를 획득 할 수 있다.

개인용 컴퓨터나 스마트폰으로 인터넷을 통해 정보를 취득하려면 PC나 스마트폰과 같이 인터넷에 접속 할 수 있는 디지털 기기를 보유하고, 디지털 기기와 웹페이지를 조작하는 능력이 있어야 하고,

실제 인터넷을 이용한 경험이 있는 것이 중요하다.

그러나 모든 사용자가 동등한 수준의 정보화 환경에 있는 것이 아니다. 정보취약계층은 디지털 기기에 대한 접근성과 인터넷 활용 능력이 일반 사용자에 비해 현격하게 떨어지는 정보격차가 발생되고 있다.

우리나라의 경우 정보취득에 취약한 정책과 법률을 살펴보면, 2001년 '정보격차해소에 관한 법률'이 제정되었지만, 2008년 '정보격차해소에 관한 법률'은 '국가정보화기본법'으로 통합되면서 스마트 정보 격차에 대한 정책개발에 아쉬움이 느껴진다.

그동안 정부는 오랜 기간 동안 정보격차를 해소하기 위해 다양한 정책들을 추진해 왔지만, PC이용 상황에서 발생하는 정보격차해소에 중점을 두어 스마트폰이나 모바일 기기를 이용하는 상황에서 발생하는 이른바 '스마트 정보격차'를 해소하는데 필요한 정책은 노력하지 않는 것으로 보여진다. 인터넷 이용 환경이 급격하게 모바일 중심으로 전환 된 현실을 감안한다면 스마트 정보격차 해소

• First Author: Seong-Bae Jeong, Corresponding Author: Kyung Shin, Kim

*Seong-Bae Jeong (cellist915@Chosun.ac.kr). Dept. of Public Administration & Social Welfare, Chosun University

**Kyung Shin, Kim (rapports@daum.net), adjunct professor in the Department of preschool education at Mok-po Science University., Korea

• Received: 2018. 01. 23, Revised: 2018. 02. 01, Accepted: 2018. 02. 12.

• Following are results of a study on the "Leaders in INdustry-university Cooperation +" Project, supported by the Ministry of Education and National Research Foundation of Korea

는 더 이상 지체 할 수 없는 중요한 사회적 이슈라고 볼 수 있다[1].

2016년 과학기술정보통신부의 디지털정보격차실태조사와 같은 국내 실태조사 결과에 따르면, 유·무선 정보통신환경에서 정보취약계층(장애인, 장·노년층, 저소득층, 농어민)의 디지털정보화수준은 일반국민의 58.6%로 전년(52.4%) 대비 6.2%p 개선되었다.

조사 부문별로 살펴보면, 디지털 접근수준(84.5%)은 일반국민과의 격차가 15.5%로 상대적으로 작았으나, 디지털역량수준(45.2%)과 디지털활용수준(59%)은 격차가 각각 54.8%, 41.0%로 큰 차이를 보이고 있으며, 계층별로 살펴보면, 저소득층(77.3%)은 일반국민과의 격차가 22.7%로 상대적으로 작았으나, 장애인(65.4%), 농어민(61.1%), 장·노년층(54.0%) 격차는 각각 34.6%, 38.9%, 46.0%로 큰 차이를 보이고 있으며, <표 1>과 <표 2>와 같다[2].

Table 1. Trend of digital informatization

| Division | 2014 | 2015 | 2016 |
|-------------------|------|------|------|
| Access level | 72.3 | 73.7 | 84.5 |
| Competency level | 34.6 | 37.4 | 45.2 |
| Utilization level | 47.7 | 51.6 | 59.0 |
| *Average | 50.1 | 52.4 | 58.6 |

(*When the digital informatization level of the whole people is taken as 100, it means the digital informatization level)

Table 2. Level of digital informatization by information vulnerable layer

| Division | 2014 | 2015 | 2016 |
|------------------|------|------|------|
| Disabled | 60.2 | 62.5 | 65.4 |
| Low income group | 72.5 | 74.5 | 77.3 |
| Farm fishermen | 51.4 | 55.2 | 61.1 |
| Elderly people | 42.4 | 45.6 | 54.0 |
| *Average | 50.1 | 52.4 | 58.6 |

(*When the digital informatization level of the whole people is taken as 100, it means the digital informatization level)

우리사회는 정보의 홍수에 직면해 있어도 노인과 장애인 시각장애인 등은 정보에 취약한 계층이다. 이는 정보 취득이 장애를 가진 이용자보다 장애를 가지지 않는 이용자를 중심으로 발전했기 때문이라 할 수 있다. 그런데 더 문제는 이러한 정보취약계층의 정보 격차는 결국 경제적·사회적 격차로 확대 될 가능성이 매우 높다는 것이다.

우리 사회에서 가장 많은 정보를 취득하는 방법은 스마트폰과 같은 모바일 기기 사용임에 이견은 없을 것으로 생각한다. 하지만, 노인과 시각 장애인과 같은 정보취약 계층은 그 사용 기술과 방법에 한계를 가질 수밖에 없다.

그동안 정보격차해소방안에 대한 연구는 2000년대 초·중반에 집중되고 2000년대 후반부터는 연구가 미진하다는 비판이 있다. 이와 같이 정보격차에 대한 사회적·국가적 관심의 저조라고 보는 견해는 정보격차해소에 관한 기술적 연구 저조와 관련 연구자도 줄어들고 있기 때문으로 보인다[3].

실제 정보취약계층을 위한 정보격차 해소방안에 관한 연구는

손에 꼽을 정도로 많지 않다. 그 가운데 한 연구는 장애를 가진 이용자의 정보격차해소에 관한 강력한 법적강제력 필요성, 저소득층의 정보격차해소를 위해 경제적 지원과 평생교육의 활용, 노년층을 위한 정보격차 해소 방안으로 노인정보화 교육 필요성을 주장하는 정책적 제언에 관한 연구가 있었다[4].

또 다른 연구는 정보취약계층이 사용하는 스마트 폰 애플리케이션을 디자인할 때 인터페이스 디자인 전략, 커뮤니케이션 디자인 전략, 정보 구조 전략, 사용자 경험 디자인 전략, 유니버설 디자인 전략 등 5가지의 디자인 전략을 제시한 연구가 있어 주목된다[5].

이러한 정책적 제언과 디자인 제언연구는 스마트 정보격차를 해결하는데 상당한 의의가 있는 연구라 할 수 있다. 하지만, 그 기술적인 한계를 극복하기에는 다소의 아쉬움이 있는 연구로 보여진다. 이에 정보취약계층의 정보취득을 위한 현재의 기술적 한계를 극복하기 위해 더 발전적인 기술개발 연구가 필요해 보인다.

따라서 본 연구에서는 노인, 시각장애인 등 스마트 기기를 사용하면서 정보 격차를 겪고 있는 정보취약계층의 쉽고 편리한 정보취득을 위하여 스마트폰 자체 기능만을 사용하여 크롤링&클리핑을 활용한 맞춤 정보 수집 서비스 제안을 통해 현재의 기술적 한계를 뛰어넘어 정보취약계층의 정보취득을 위해 조금이라도 기여하고자 하는 기술개발이 본 연구의 궁극적 목적이다.

본 연구에서 제안된 서비스는 통합관리서버와 안드로이드 사용자 스마트폰 앱(app)으로 구성이 된다. 데이터 처리 시 저작권 문제와 서버의 부하를 줄이기 위해 통합관리서버에서는 데이터 변경 방법에 관한 정보만 보관을 하고, 데이터 다운로드 및 변경되는 부분 처리는 사용자 앱(app)에서 동작하도록 설계하고 구현하도록 노력하였다.

본 논문의 구성은 다음과 같다. 제 1장에서는 서론으로 정보취약계층의 실태, 정보취약계층의 정보 취득 필요성 등을 서술하고 본 연구의 목적을 제시하겠다. 제 2장에서는 관련연구로 정보취약계층의 정보취득을 위한 그동안의 선행연구와 그 한계점을 기술하고 본 연구와 관련된 기술을 소개하겠다. 제 3장에서는 제안 시스템으로 시스템 구성과 통합서버관리에 관해 소개하겠고, 스마트폰에서 독자적으로 RSS 피드를 수집하고 원하는 정보의 원문을 요약하여 내용을 사전에 파악 할 수 있는 기술을 소개하겠다. 제 4장에서는 결과 및 결론으로 웹 서비스와 스마트폰 앱(app)에서 노인 및 시각장애인을 위한 음성변환장치(TTS : Text-To-Speech) 서비스 적용에 있어, 획일화 및 단일화 된 포맷으로 정보를 생성하여 정보취약계층에 보다 간결하고 명확하게 정보를 제공 할 수 있는 사례와 그 가능성으로 결론을 맺고자 한다.

II. Related Research

1. Web Accessibility

월드 와이드 웹(World Wide Web)을 창시한 팀 버너스 리(Tim Berners-Lee)는 웹이란 '장애에 구애 없이 모든 사람들이 손쉽게

정보를 공유할 수 있는 공간'이라고 주장하였으며, 웹과 관련된 소프트웨어와 콘텐츠를 제작할 때에도 이용자의 장애에 관련 없이 누구나 쉽게 이용하고 접근할 수 있도록 제작되어야 한다고 주장하였다.

이렇듯 웹의 기본적 철학은 웹 접근성에 장애가 없어 누구나 쉽게 접근할 수 있어야 한다. 그럼에도 불구하고 실제 웹 접근성과 그 웹을 이용하는 입장에서 취약한 계층은 존재할 수밖에 없다.

웹의 접근성이란 노인과 시각 장애인과 같은 장애가 있는 사람들을 포함한 모든 사람이 정보통신 기기나 웹 서비스를 어렵지 않게 이용할 수 있도록 제작되어야 한다는 전제를 갖는다. 그동안의 접근성 향상 노력은 많은 장애를 가진 이용자들에게 도움이 되었지만, 어떻게 보면 일반인들에게 많은 혜택이 돌아간 것도 사실이다. 예를 들면, 리모콘, 전화, 오토메틱 자동차, 자동문 등과 같은 제품들은 장애를 가진 사람들이 편리하게 사용하기도 하지만, 이러한 편리성 제품들이 제작되고 보급되면서 궁극적으로는 많은 사람들이 편리한 생활이 이루어진 것이다.

이러한 접근성 개념은 정보통신 분야에서 다양한 제품 및 서비스 간의 호환성 문제, 이동 정보통신 기기 등으로 그 필요성 및 중요성이 더욱 증대되고 있다. 인터넷 전자도서관인 Wikipedia는 웹 접근성의 개념을 웹 콘텐츠 뿐 만 아니라 인터넷을 통하여 전달될 수 있는 모든 콘텐츠로 확대하고 있다[6][7].

2. Existing Technology limitations

2.1 Web accessibility failure

웹 접근성을 준수한 사이트라도 장애인이 정보에 접근하기 위해서는 많은 어려움이 있다. 하나의 예로 시각장애인의 경우 네이버 뉴스를 접근하기 위해서는 브라우저 또는 앱(app)을 실행 시키고 뉴스 메뉴를 찾아서 눌러야 하며, 뉴스 페이지에 들어가서 원하는 기사를 선택하여야 한다.

이처럼 정보 접근을 위해서 현재 서비스 되어 지고 있는 많은 사이트 또는 앱(app) 들의 인터페이스는 일반인들을 위한 것으로 시각적 기능을 적극 활용한 정보취득 개념이다. 하지만, 노인이나 시각장애인과 같은 시각적 기능이 부족한 이용자들에겐 현실에 맞지 않는 인터페이스가 대부분이다.

2.2 Information Filtering Failure

인터넷과 정보통신이 발달하면서 필요한 지식을 모으기는 쉬워졌다. 하지만, 컴퓨터는 지식을 수집할 뿐 내가 필요한 모양새로 추려주지는 못 한다. 또한 골라낸 정보가 길고 어렵다면 정보취득과 거리가 멀다.

노인과 시각 장애인과 같은 정보취약계층 뿐만 아니라 일반 사용자들에게도 해당이 되는 문제다. 즉, 수많은 신문기사 내용에서 제목만을 보고 원하는 정보인지 아닌지 판단을 할 수 없으며 긴 본문을 어느 정도 읽어야 어떤 내용인지를 확인 할 수 있다면, 일반 사용자도 불편함을 느끼는 문제이지만 노인과 시각장애인과 같은 정보취약계층은 정보의 선택에 있어 더 차별을 받을 수밖에 없는 부분이다.

그러면 이러한 정보취득 장애를 조금이라도 극복한다면 리모콘, 전화 오토메틱 자동차처럼 취약계층은 물론이고 일반들에게도 더 편리한 정보취득의 기회가 열릴 수 있을 것이라는 가능성도 적지 않다.

3. Crawling

크롤링(crawling)은 스크래핑(scraping)이라고 부르며, 웹 페이지를 그대로 가져와서 거기서 데이터를 추출해 내는 기술을 말한다. 웹 상에서 활동하기 때문에 웹 크롤러가 하는 작업을 웹 크롤링(web crawling) 혹은 스파이더링(spidering)이라고도 부르는 경우가 있다.

검색 엔진과 같은 여러 사이트에서는 데이터의 최신 상태를 위해 웹 크롤링한다. 웹 크롤러는 대체로 방문한 사이트의 모든 페이지의 복사본을 생성하는 데 사용되며, 검색 엔진은 이렇게 생성된 페이지를 보다 빠른 검색을 위해 데이터 구조를 단편화시키기 위해 인덱싱(Indexing) 한다. 또한 크롤러는 링크 체크나 HTML(Hyper text markup language)코드 검증과 같은 웹 사이트의 자동 유지 관리 작업을 위해 사용되기도 하며, 자동 이메일 수집과 같은 웹 페이지의 특정 형태의 정보를 수집하는 데도 사용된다[8][9].

웹 크롤러는 봇이나 소프트웨어 에이전트의 한 형태이다. 웹 크롤러는 대개 시드(seed)라고 불리는 URL(Uniform Resource Locator) 리스트에서 부터 시작하는데, 페이지의 모든 하이퍼링크를 인식하여 URL 리스트를 갱신한다. 갱신된 URL 리스트는 재귀적으로 다시 방문한다. 기존에 개발된 웹 크롤러는 단일 크롤러와 분산 크롤러로 분류 할 수 있다. 일반적인 크롤러는 수집할 웹 문서 URL 목록인 시드(Seed)가 거의 변화지 않는 고정으로, 수집 방법 및 필터를 정해 놓으므로 어느 정도 정확한 정보를 수집할 수 있는 장점이 있다. 하지만, 웹 크롤러 기술을 이용하여 웹 문서를 분석하는 필터 기술에 자동화 알고리즘이 있음에도 불구하고 크롤링(crawling)기술만을 가지고 정확한 정보를 추출하는 데는 아직까지 완벽하지 않다는 견해가 많아 웹 크롤러(Web crawler)나 클리핑(Clipping)이 적절히 조화를 이루어야 한다[10].

4. Clipping

가독성 높은 웹 문서를 만들기 위해서는 CSS(Cascading Style Sheets) 언어가 무엇보다 중요하다. 웹 크롤러(Web crawler)로 웹 문서에서 정확한 정보를 추출하여 표현하면 가독성이 현저하게 떨어진 웹 문서를 가지게 된다. 또한 더 문제는 해당 정보를 재가공 하기 위해 웹 디자이너와 웹 개발자 필요에 따라서 기획자도 필요하다는 불편함이 있다. 이런 이유로 유지비용 문제가 크다는 단점이 있어 좋은 서비스가 나와도 많은 이용자들에게 알려지기가 쉽지가 않다. 이러한 문제점을 해결하기 위해서 본 연구에서는 클리핑(Clipping)을 제안한다.

클리핑(Clipping)의 사전적 의미는 '오려낸다'는 뜻을 가진다. 웹(Web) 상에서 나타내는 활동으로 웹 클리퍼(Web

Clipper)는 웹 크롤러(Web crawler)로 추출한 웹 문서의 뼈대에 글자 크기, 색, 굵기 그리고 이미지 등 꼭 필요한 요소만을 서버에서 정의하고 클라이언트에서 실행하여 문서 가독성을 향상 시켜주는 개념을 뜻한다.

웹 클리퍼(Web Clipper)를 이용해 신문이나 잡지를 읽다가 좋은 내용이 있으면 가위로 오려서 파일에 모으기, 시험지를 오려 공책에 붙여서 오답 노트 만들기와 같은 개념이다. 웹 상에서 인터넷 기사나 블로그 포스팅을 스크랩하는 경우인데, 바로 웹 클리퍼(Web Clipper)를 통해 텍스트, 이미지, 링크 등을 포함해 웹페이지와 같은 원하는 정보를 손쉽게 스크랩할 수 있는 것이다. 에버노트 웹 클리퍼 기능을 이용하면 클릭 한 번으로 웹페이지에서 원하는 부분만 스크랩할 수 있다[8].

웹 클리퍼(Web Clipper)는 기본적으로 사용자 스마트 폰 환경에서 CSS(Cascading Style Sheets) 언어를 동적으로 옮겨주고, 자바스크립트(JavaScript)언어를 생성하여 스마트 폰 상황에서 사용자가 즉시 실행하는 방법으로 이루어져 해당 결과로 나온 화면이 스마트 폰 기계의 시각 틀에 딱 맞게 해주는 기능이 있다.

5. Text Summarize

문장요약(Text Summarize)은 텍스트랭크(TextRank) 알고리즘을 이용한다. TextRank는 Mihalcea(2004)이 제안한 알고리즘으로 텍스트에 관한 graph-based ranking model 로써 Google의 Page Rank를 활용한 알고리즘이다[11].

PageRank는 Brin and Page(1998)이 제안한 알고리즘으로 하이퍼링크를 가지는 웹 문서에 상대적 중요도에 따라 가중치를 부여하는 방법이다. 서로 간의 인용과 참조로 연결된 임의의 묶음에 적용할 수 있다. 이러한 PageRank 알고리즘을 활용한 것이 바로 TextRank이다. TextRank는 PageRank의 중요도가 높은 웹 사이트는 다른 많은 사이트로부터 링크를 받는다는 점에 착안하여 문서 내의 문장(or 단어)를 이용하여 문장의 Ranking을 계산하는 알고리즘이다[12][13]. (그림 1)은 PageRank 알고리즘을 TextRank 알고리즘에 적용하는 그림이다.

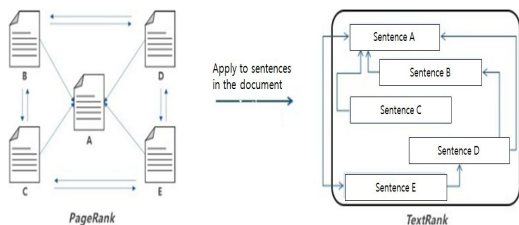


Fig. 1. PageRank > TextRank Application situation

TextRank 식은 아래와 같다.

$$TR(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} TR(V_j)$$

- TR(V_i): 문장 또는 단어(V_i)에 대한 TextRank 값
- w_{ij}: 문장 또는 단어 i 와 j 사이의 가중치

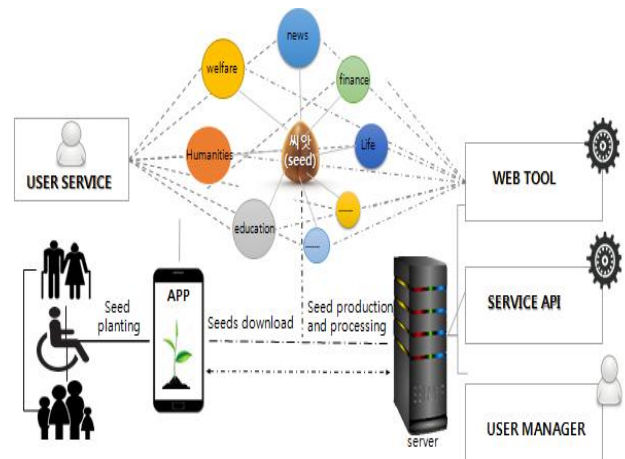
- d:dampingfactor, PageRank에서 웹 서핑을 하는 사람이 해당 페이지를 만족하지 못하고 다른 페이지로 이동하는 확률로써, TextRank에서도 그 값을 그대로 사용
- TextRank TR(V_i)를 계산 한 뒤 높은 순으로 정렬

III. Research System

1. System Configuration

본 제안 시스템은 크게 통합관리서버와 사용자 안드로이드 스마트폰 앱(app)으로 구성된다. 본 서비스를 이용할 경우 먼저 씨앗(seed)이라는 정보수집 기능을 활용한다. 씨앗은 정보 수집을 위한 URL(Uniform Resource Locator) 및 가독성과 편의성을 위한 데이터 가공 패턴 정보가 저장되어 있는 파일 또는 프로토폴로이다. 씨앗이라는 기능은 사용자가 원하는 정보와 관련한 단어, 뉴스, 분야 등을 선택해 선택된 단어와 관련한 정보를 찾을 수 있는 기능이다. 즉, 씨앗이란 정보수집 기능은 사용자가 찾고자하는 정보와 관련된 일종의 키 워드(key word) 나 핵심 검색어를 정리한 용어로 뉴스 주제, 단어, 제목, 분야 등 관련정보를 요청하는 가장 처음 단계라 본 연구에서는 씨앗이라 명하였다. 예를 들면, 정치, 사회, 경제, 음식, 여행 등과 관련한 단어가 해당된다.

통합관리서버는 씨앗이라는 정보수집 기능을 프로토폴 생성을 위한 WEB TOOL과 SERVICE API 및 사용자 관리(USER MANAGER) 기능으로 구성한다. 사용자 스마트폰 앱(app)은 씨앗이라는 정보 관련 검색어를 다운로드 받아 분석하여 실행할 있도록 하였다. (그림 2)는 서비스 구성도이다.



This task is implemented as an RSS news gathering service.

Fig. 2. Service configuration diagram

1.1 Server Management

WEB TOOL은 웹 크롤링 & 클리핑 패턴을 설정, 생성하고 저장한다. 앱(app)을 이용해 다양한 정보를 가져오기 위해서는 인터넷을 통해 다양한 매체에 접근하여 정보를 가져오는데, 구

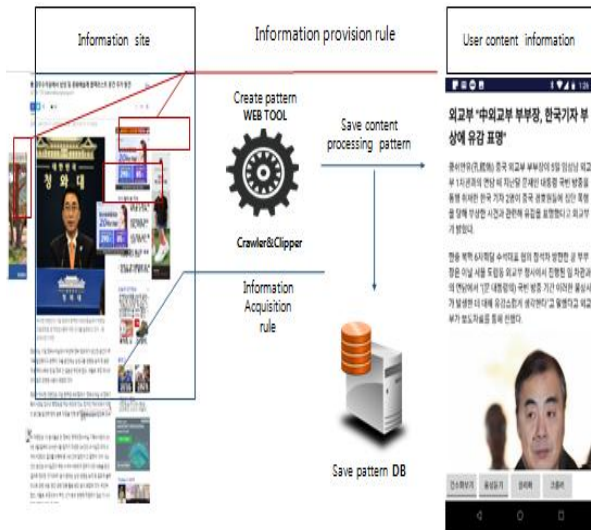


Fig. 5. Construction diagram of Crawler & Clipper pattern using WEB TOOL

뉴스 정보는 RSS를 기본으로 사용하므로 주기적으로 최신 정보를 갱신하여 사용자들에게 제공한다. 결론적으로 필터링 기술과 스마트폰이라는 특수한 기기를 이용하여 지속적인 정보 취득을 할 수 있는 서비스를 제공한다.

정보필터링 문제는 정보 선택을 개선하기 위해서 크롤러(Crawler) 기술로 필터링한 데이터를 사용하여 뉴스 정보 사이트에서 문자 데이터만 추출하여 문자 요약(Text Summary) 모듈로 기사 요약 문장을 만들어 정보 선택을 보다 효율적으로 할 수 있도록 하였다. (그림 6)은 신문기사 요약 모듈 흐름도이다.

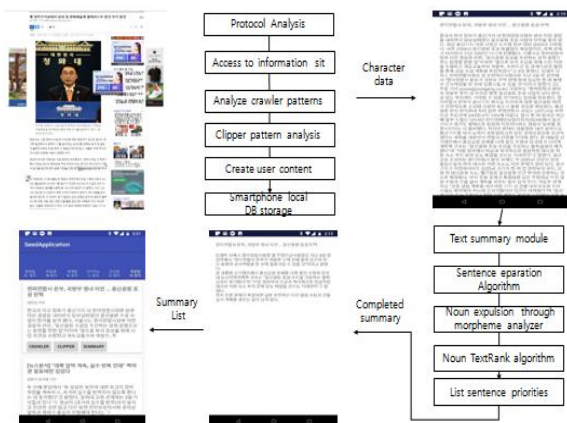


Fig. 6. Newspaper article summary module flow chart

1.1 Server Management Web Tool

웹 페이지에서 제목영역, 본문영역, 제외대상 영역을 지정하고 각 웹페이지별 패턴을 저장한다. (그림 7)은 웹 크롤링 툴 동작 화면이다.

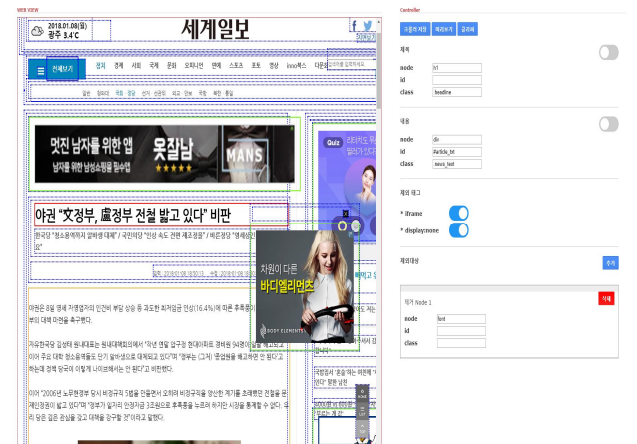


Fig. 7. Web Crawl Tool Action Screen

뉴스정보 페이지를 해당 페이지 크롤러 패턴을 적용하여 확인할 수 있다. (그림 8)은 웹 크롤링 패턴 적용 결과 확인을 나타내었다.



Fig. 8. Check the results of applying a Web crawl pattern

뉴스정보 페이지를 해당 페이지 크롤러 패턴을 적용하여 확인 할 수 있다. 크롤러 패턴이 적용 된 뉴스정보에 사용자 가독성을 위해 정렬방법, 폰트 사이즈, 색깔 등 표현 스타일을 조절 할 수 있다. (그림 9)는 웹 클리퍼 동작화면을 표현한 것이다.



Fig. 9. Web Clipper Action Screen

1.2 User smartphone app

본 서비스를 이용할 경우 씨앗이라는 정보수집 기능을 이용해 이용자가 찾고자하는 정보와 관련된 일종의 키 워드(key word) 나 핵심 검색어를 제시하면 그와 관련된 정보가 생성된다. (그림 10)은 씨앗 정보 요청 받아오기 화면이다.



Fig. 10. Get seed information using app

스마트폰 앱(app) 로딩시 통합관리서버의 씨앗이라는 정보 수집 기능을 통해 관련 정보를 받아와 설치 여부를 확인 할 수 있다. (그림 11)과 (그림 12)는 새로운 씨앗 추가 화면이다.

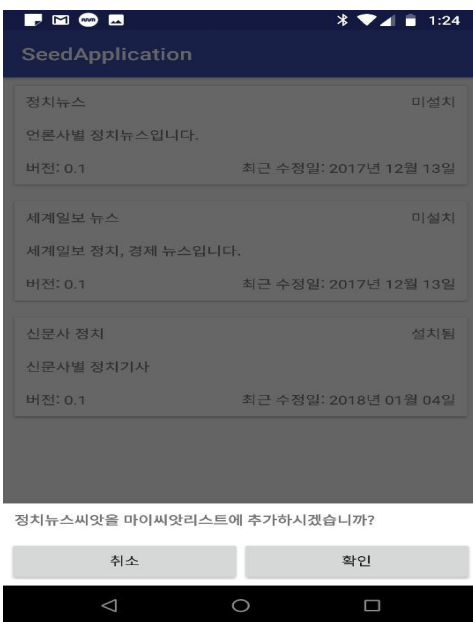


Fig. 11. Add New Seed Screen 1

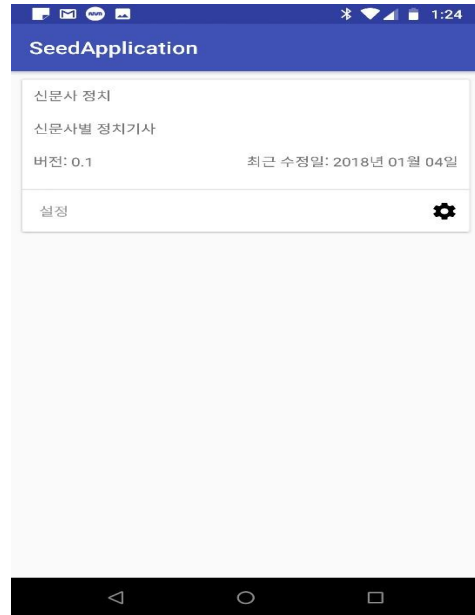


Fig. 12. Add New Seed Screen 2

씨앗이라는 정보수집 기능을 이용해 정보 내용을 보면서 정보를 요약해 화면으로 표출할 수 있으며, (그림 13)은 씨앗을 이용해 요약정보를 화면으로 나타낸 그림이다.



Fig. 13. Summary information screen

씨앗정보를 이용해 관련기사나 내용 또는 그림을 선택할 수 있고, 관련 정보의 원본을 볼 수 있다. (그림 14)는 씨앗정보를 이용한 요약정보를 화면으로 표현한 것이고, (그림 15)는 관련 기사나 내용의 원본을 볼 수 있는 그림을 표시하였다.



Fig. 14. Selected information screen

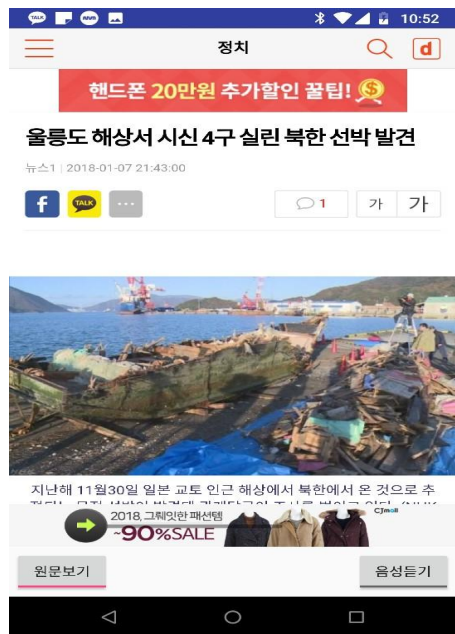


Fig. 15. Show original web page

1.3 Development of interface considering accessibility

정보취약 계층의 또 다른 문제는 스마트폰에서 접근하는 모바일 앱(app)의 사용자 인터페이스가 접근성을 고려하지 않고 있다는 점이다. 그러한 점을 개선하기 위해 본 연구에서는 스마트폰의 인터페이스에서 변화를 주지 않고 외부 인터페이스를 활용하여 정보수집 서비스를 이용할 수 있도록 노력하였다. (그림16)은 접근성을 고려한 사용자 경험디자인 UX(User Experience Design)을 나타내었다.

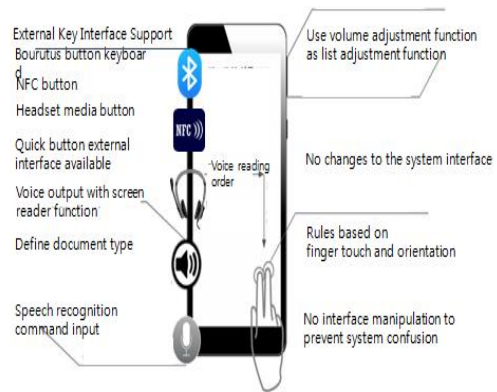


Fig. 16. UX Design considering accessibility

2. conclusion

본 연구는 노인과 장애인 시각 장애인 등과 같이 정보취득에 취약한 계층의 정보 취득을 돕기 위해 크롤링&클리핑을 활용한 안드로이드 사용자 스마트폰 앱(app)으로 구성된 음성 변환장치(TTS : Text-To-Speech) 서비스 적용에 있어, 획일화 및 단일화된 포맷으로 정보를 생성하고, 정보 수집 서비스 개발하여 정보취약 계층의 정보취득에 도움을 주고자하는 공익적 목적에 연구의 의의를 두고자 하며, 다음과 같은 연구 결과를 제시 할 수 있겠다.

본 연구에서는 웹 접근성 문제를 해결하기 위해 웹(web)이 아닌 앱(app)으로 정보를 제공하는 방법에 중점을 두고자 했다. 그 이유는 스마트폰이 대중화되고, 활성화 되어 있는 시점에서 정보 접근을 위한 인터페이스가 웹(web) 보다는 앱(app)이 편리하다고 판단하였기 때문이다. 또한, 지속적인 서비스가 되기 위해서는 서비스 관리 비용이 낮아야 하므로, 서버 자원을 최소한으로 사용하고 사용자 스마트폰에서 대부분 처리되는 서비스로 구현하였다.

기존 정보제공 서비스들의 대부분은 이벤트 요청 방식으로 서버에 데이터를 저장하거나, 서버에서 다시 다른 서버로 데이터를 요청하여 전달되는 방식으로 정보를 제공 받는 사용자들은 해당 서버에서 제공되는 데이터에서 필요로 하는 정보를 골라 볼 수밖에 없었다. 이는 검색을 잘 하는 사용자이거나, 정보 활용 능력이 뛰어난 사용자는 가능한 일이나, 그렇지 못 한 사용자는 자신을 위한 정보를 취득하기에는 어려움이 있다. 또한 사용자들에게 필요한 정보는 한 사이트에만 있는 것이 아니라 여러 사이트에 나뉘어져 있어 정보취득의 획일화와 단일화가 쉽지 않다는 기존의 정보취득 단점을 보완하려 노력하였다.

기존의 정보제공 사이트는 광고, 복잡한 메뉴, 인터페이스를 가지고 있어 노인이나 시각장애인의 경우 수많은 광고나 메뉴를 헤치고 필요한 정보를 취득하기에는 쉽지 않았다. 이를 개선하기 위해 크롤러(Crawler)와 클리퍼(Clipper) 기술을 사용하여 해당 사이트를 가공한 정보를 프로토콜로 만들어 저장하여 필요 없는 부분은 버리고 필요한 정보만 취득하기 위해 노력하였다.

또한, 기존 기술은 최종 결과물을 위한 중간 도구로써 기술을 사용하지만, 본 연구의 크롤링&클리핑 활용한 정보수집 서비스는 정보 수집 및 분석을 위한 프로토콜 생성 및 가독성과 편의성을 높이기 위한 기반 기술로 활용할 수 있게 하였다. 즉, 사용자에게

필요한 데이터를 가지고 있는 여러 사이트의 정보와 해당 사이트를 일일이 방문하지 않고서도 크롤링&클리핑 기술을 활용하여 패턴 정보만을 수집하여 가지고 있는 파일 또는 프로토콜을 서버에서 가지게 함으로써 필요한 스마트 폰에서 정보를 수집해 가져오는 방법을 구현하였다.

결론적으로 사용자는 자신한테 필요한 정보가 있는 구성정보 파일 또는 프로토콜을 스마트 폰 앱(app)으로 다운로드 받아 필요정보를 최대한 간단하고 편리하게 직접 취득할 수 있도록 함으로써 정보 취득 방법을 좀 더 간단하고 쉬운 방법으로 취득 할 수 있도록 노력하였다.

무엇보다 노인 및 시각장애인과 같은 시각적 기능에 자유롭지 못한 이용자들에게 취득하고 하는 정보를 음성변화시스템인 TTS 서비스를 적용해 정보취득에 있어 청각기능을 활용해 획일화 및 단일화된 포맷으로 정보를 생성하여 정보취약계층에 보다 간결하고 명확하게 정보를 제공 할 수 있을 것으로 기대한다.

하지만, 본 연구는 노인이나 시각장애인의 정보취득을 위한 사용에 아직은 대중화된 많은 실험을 거치지 않았기 때문에 집적적인 사용자의 불편함에 대한 많은 정보는 취합하지 못했다는 한계를 가진다. 추후 연구에서는 이용자 입장에서 충분한 실험적 과정을 거친 연구가 요구된다.

REFERENCES

- [1] H. K. Kim, "Mobile divide in smartphone era," KT Economic Research Institute, 20. Jul 2011.
- [2] National Information Society Agency(NIA), "2016 Digital Information Gap Survey" Ministry of Science, ICT and Future Planning, 20. Feb 2017.
- [3] H. J. Seo, S. W. Meong. "Domestic Digital Divide Research Trend". Journal of Korean Society of Information Science and Technology, Vol. 19, No 4, pp. 151-187. Dec. 2016
- [4] Y. M. Kim, "A Study on the Solution of the Digital Divide for Information Disadvantaged Groups", Hanyang University Master Thesis, Feb 2007
- [5] Y. B. Jang, "A Study on Mobile Application Design Strategy for Information Alienated Groups" Korea Society of Design Culture, Vol. 19, No. 1. pp.410-417. Mar. 2013
- [6] <http://www.wah.or.kr/Accessibility/define.asp>
- [7] Rada Mihalcea, Paul Tarau, "TextRank: Bringing Order into Texts"
- [8] https://en.wikipedia.org/wiki/Web_scraping
- [9] https://en.wikipedia.org/wiki/Web_crawler
- [10] <http://it.donga.com/16627/>
- [11] <http://excelsior-cjh.tistory.com/entry/TextRank>
- [12] <https://en.wikipedia.org/wiki/PageRank>
- [13] <https://web.stanford.edu/class/cs54n/handouts/24-GooglePageRankAlgorithm.pdf>

Authors



Seong-Bae Jeong received his Ph.D. from the University of Daejeon in 2006. In 1998, he was a professor at Seong Hwa University. He is currently an assistant professor of administrative and welfare at Chosun University. The elderly, social welfare,

and addiction rehabilitation are areas of interest.



Kyung-Shin Kim received the Ph.D degree in Special Education from Chosun University, Korea in 2008. She is a special teacher at Mok-po In-Sung Special School. Dr. Kim had worked as an adjunct professor D for 3 years at Sae Han University, Seoul, Korea

in 2012-2014. She is currently an adjunct Professor in the Department of preschool education at Mok-po Science University. She is interested in intellectual disability of transition education and life long education.