HTMLtoVoiceXML 변환기에 관한 연구

최 훈 일°, 장 영 건
청주대학교 전산정보공학
{choihi, ygjang}@chongju.ac.kr

A Study on HTMLtoVoiceXML Converter

Hoon-il Choi°, Young-gun Jang
Dept. of Computer Information Engineering, Chongju University

요 약

음성 기술의 발달과 VoiceXML 1.0의 제정으로 인해 표준화된 방식으로 이동 단말기와 전화를 통해 음성으로 웹 컨텐츠에 접근할 수 있게 되었다. 거의 모든 웹 컨텐츠들은 HTML로 작성되어 있으며, 기존의 HTML로 작성된 수많은 웹 컨텐츠에 음성으로 접근하기 위해서는 HTML 문서들을 VoiceXML 문서로 변환하여야 한다. 이를 수동으로 변환하기 위해서는 많은 시간과 비용이 필요하게 된다. 본 논문에서는 이 문제를 해결하기 위해여 HTML 문서를 VoiceXML 문서로 자동 변환하는 HTMLtoVoiceXML 변환기의 설계 방안을 제시하였다.

1. 서론

1990년 WWW(World Wide Web)이 소개된 이후 웹은 엄청난 속도로 확산되기 시작하여 오늘날 거의 모든 정보들은 웹을 통해 얻을 수가 있게 되었다. 이러한 웹 정보들은 HTML (HyperText Markup Language)이라는 마크업 언어로 작성되어 넷스케이프나 익스플로러와 같은 웹 브라우저에 의해 해석되어 컴퓨터 모니터를 통해 정보를 수집하였다.

오늘날, 무선 인터넷의 발달과 음성 기술의 발달로 인하여 웹 브라우저를 통해 컴퓨터 모니터 상에서만 얻을 수 있었던 웹 정보를 이동 단말기와 전화를 통해서도 얻을 수 있게 되었다. 이동 단말기나 전화를 통해 웹 정보를 얻기 위해서는 웹 정보가 HTML이 아닌 다른 마크업 언어로 작성되어 있어야 한다. 특히 음성으로 웹 정보를 제공하기 위해서는 AT&T, IBM, 루슨트 테크놀로지, 모토롤라 등의 기업이 만든 VoiceXML 포럼[1]에서 2000년 5월 22일 W3C에 제안한 VoiceXML(Voice eXtensible Markup Language)[2]이라는 마크업 언어를 이용하여 문서를 작성하여야 한다. 하지만 HTML로 작성된 기존의 수많은 웹 정보를 VoiceXML 문서로 변경하여 제공하기 위해서는 엄청난 비용과 시간의 낭비를 초래한다. 이런 문제를 해결하기 위해서는 기존의 HTML을 그대로 유지하면서 필요시 HTML 문서를 자동으로 VoiceXML 문서로 변환하면 될 것이다.

그러나 HTML 문서를 VoiceXML 문서로 변환하기 위해서는 많은 문제점들이 있다.

첫째, 시각적으로 정보를 제공하는 HTML 문서와 음성적으로 정보를 제공하는 VoiceXML 문서와의 정보 제공 방법의 차이와 또한 시각적인 제공되는 정보와 음성적으로 제공되는 정보에 대해 사용자가 이해할 수 있는 정보량의 차이로 인해 하나의 HTML 문서를 하나의 VoiceXML 문서로만 변환 할 수 없다. 즉, 하나의 HTML 문서에 대해 여러개의 VoiceXML 문서가 생성되어야 한다.

둘째, HTML과 VoiceXML의 정보 제공 방법의 차이로 인해 HTML 태그와 VoiceXML 태그는 서로 사용 용도가 확연히 다르기 때문에 HTML의 어떤 태그를 VoiceXML의 어떤 태그로 태그들을 대치시켜 문서를 변환할 수 없다.

셋째, HTML 태그는 정보의 시각적 표시 방법만을 나타낼 뿐 XML 태그처럼 정보에 대한 의미를 포함하고 있지 않기 때문에 컨텐츠를 분리하기가 힘들다.

본 논문에서는 위에서 열거한 문제점들을 해결하기 위해 변환 가능한 HTML 문서 형태를 게시판처럼 정보의 형태가 비슷한 컨텐츠가 나열되어 있는 문서들로 제한하고, 이런 형태의 HTML 문서의 구조를 분석하여 컨텐츠를 추출하여 이를 VoiceXML 문서에 맞게 재구성하는 자동화 시스템인 HTMLtoVoiceXML 변환기의 설계 방안을 제시하였다.

2. 관련 연구

2.1 VoiceXML

음성 인식/합성 기술이 발달함에 따라 웹에 접속하기 위한 사용자 인터페이스가 키보드 및 전화기의 키패드에서 음성으로의 전환이 거론되기 시작했다. 특히, AT&T, IBM, 루슨트 테크놀로지, 모토롤라 등 정보통신 분야의 4개 거대 기업이 모여 VoiceXML 포럼을 설립하여, 1999년 8월 웹 컨텐츠와 웹 기반의 응용 등을 대화식 음성 응답을 통해 제공받을 수 있는 음성 마크업 언어인 VoiceXML 0.9 버전을 발표하였고, 2000년 3월 0.9 버전을 보완한 1.0 버전을 W3C에 제안하여 W3C에서 2000년 5월 VoiceXML 1.0을 웹의 대화형 마크업 언어 표준으로 발표하였다. 이로 인해 많은 업체들이 VoiceXML에 대해 관심을 갖기 시작하였고, 많은 연구가 활발히 진행되고 있는 상태이다. 외국의 경우 Bevocal, Nuance, Tellme, VoiceGenie 등에서 VoiceXML 해석기를 개발하였으며, 몇몇 업체에서는 VoiceXML 시범 서비스를 제공하고 있다. IBM에서는 VoiceXML 응용을 개발할 수 있는 도구를 제공하고 있으나 표준을 완벽하게 구현하고 있지는 못하다. 국내에서는 성신여자대학교에서 VoiceXML 해석기의 구현에 관한 연구[3,4]를 수행하였으며, 미디어포드[5]와 브레인21[6,7]에서도 VoiceXML 해석기를 개발하였다. 그러나 아직까지는 거의 모든 컨텐츠가 HTML로 작성되어 있어, 그 이용율이 매우 저조하다. VoiceXML이 이용할 수 있는 단말기의 대중성과 이동성, 대화

형이라는 편의성 등의 많은 장점에도 불구하고, 대중화가 되지 않은 가장 큰 이유는 이용할 수 있는 컨텐츠와 서비스의 부족, VoiceXML로 웹 저작을 할 수 있는 전문인력의 부족 등이 있다. 따라서 기존의 HTML로 작성된 웹 컨텐츠를 자동으로 VoiceXML로 변환하여 서비스할 수 있는 기술이 요구된다.

2.2 변환기

VoiceXML의 등장 전에 그 원형 중에 하나인 VoxML이 모토로라에 의하여 발표되었으며, HTML을 VoxML로 변환하려는 연구가 Goose 등에 의하여 최초로 이루어졌다. Goose 등은 WWW의 전형적인 3층 구조에 HTML을 VoxML로 변환하는 기능을 갖는 VoxML-Agent를 추가하여 클라이언트, 에이전트, 웹 서버, 데이터베이스의 4층 구조를 갖는, 전화기를 사용하여 웹 접근이 가능한 Vox 포탈에 대한 연구를 발표하였다. VoxML-에이전트에 대해서는 Vox 포탈과의 상호 작용에 대하여 중점적으로 언급되어 있고, 핵심 부분인 변환 기능의 설계에 대한 내용은 없다[8]. VoiceXML 1.0의 등장이 2000년 5월이므로, HTML을 VoiceXML로 변환하는 연구는 발표된 것이 거의 없다.

3. HTMLtoVoiceXML 변환기

3.1 변환 가능한 HTML 문서 유형

오늘날 웹 저작자는 다양한 시각적 효과를 사용하여 가능한 많은 정보를 한 페이지에 표현하고자 한다. 이 정보는 시각적으로 단체화하여 조각나 있다. 즉, 메뉴, 광고, 내용 등의 조각이 하나의 페이지에 표현된다. 이 조각 중에서 어느 조각이 이 페이지에서 제공하고자 하는 주요 정보인지를 판단하고, 해당 정보를 내용 단위로 분리하여, 그 내용 단위를 기준으로 VoiceXML 문서로 변환하여야 한다. 그러나 HTML 태그는 정보의 의미를 포함하고 있지 않게 때문에 조각 단위로 분리하는 것은 거의 불가능하다.

따라서 본 논문에서는 HTML 문서의 구조 분석을 통해 컨텐츠의 조각을 분리할 수 있는 HTML 문서의 유형을 제안하고, 이런 문서 유형을 변환 대상으로 한다. 본 논문에서 제안한 변환 가능한 HTML 문서 유형은 비슷한 컨텐츠가 나열되어 있는 형태로써 그림 1과 같이 게시판 유형, 리스트 유형, 검색결과 유형으로 나눌 수 있다. 이런 유형의 HTML 문서는 트리 구조로 표현할 때 동일한 형태의 자식 노드를 많이 갖고 있고, 그 내용 속에 많은 문자를 포함하고 있는 점을 이용하여 컨텐츠의 위치를 구할 수 있다

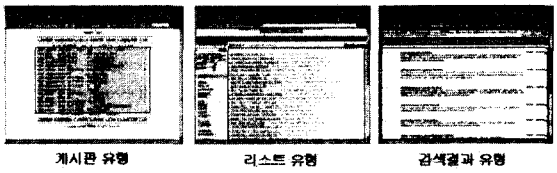

그림 1. 변환 가능한 HTML 문서 유형

3.2 HTMLtoVoiceXML 변환기 구성도

HTMLtoVoiceXML 변환기는 해당 URL의 웹 페이지를 읽어 웹 페이지에서 HTML 페이지의 구조 분석을 통해 적절한 컨텐츠를 추출하여 이 추출된 컨텐츠를 대상으로 하여 음성 시나리오를 생성하고 이를 VoiceXML 문서로 변환한다. HTMLtoVoiceXML 변환기의 구성도는 그림 2와 같다.

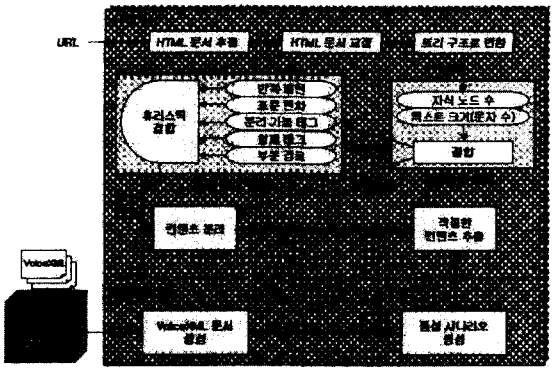

그림 2. HTMLtoVoiceXML 변환기 구성도

3.2.1 HTML 문서 트리 구조로 구성

읽어들인 HTML 문서는 트리 구조로 구성한다. HTML 문서를 트리 구조로 구성하는 이유는 트리 구조가 HTML 문서 구조를 분석하기가 좀더 용이하기 때문이다. HTML 문서를 트리 구조로 구성할 때 DOM(Document Object Model)을 사용하면 쉽게 트리화 할 수 있다.

3.2.2 컨텐츠를 포함하는 최소 서브 트리 추출

컨텐츠를 추출하기 전에 컨텐츠가 포함되어 있는 최소 서브 트리를 먼저 추출하는데 이 최소 서브 트리를 추출하기 위해 각 노드의 자식 노드 수를 구하여 가장 많은 자식 노드를 갖는 노드를 루트로 하는 트리를 최소 서브 트리로 한다. 왜냐하면 컨텐츠가 나열되어 있는 형태의 웹 페이지에서는 자식 노드의 수가 가장 많은 노드를 루트로 하는 서브 트리에 컨텐츠가 포함되어 있을 확률이 가장 많기 때문이다[9,10].

그러나 최소 서브 트리를 추출할 때 자식 노드의 수만을 고려할 경우 메뉴가 많은 웹 페이지에서는 가장 많은 자식 노드를 갖는 노드를 루트로 하는 최소 서브 트리는 메뉴를 포함하는 서브 트리가 되기 때문에 최소 서브 트리를 구할 때 자식 노드의 개수만을 고려하는 것은 그리 좋은 방법이 아니다.

본 논문에서는 자식 노드를 많이 가지고 있는 노드들을 구하여 이 노드 내에 있는 텍스트의 크기(문자열의 수)를 구하여 텍스트의 크기가 가장 큰 노드를 루트로 하여 최소 서브 트리를 구한다. 이 방법은 메뉴를 나타내는 텍스트의 크기는 작기 때문에 메뉴를 포함하는 서브 트리는 자식 노드의 개수는 많지만 텍스트의 크기는 크지 않다는 점을 이용한 것이다.

최소 서브 트리를 추출하는 단계는 다음과 같다.

- 웹 페이지 트리 구조에서 최대 자식 노드 수를 구한다.
- 최대 자식 노드 수의 50%이상의 자식 노드 수를 갖는 노드들을 구한다.
- 각 노드를 루트로 하는 서브 트리를 구하여 이 서브 트리내의 텍스트의 크기를 구한다.
- 텍스트의 크기가 가장 큰 서브 트리를 최소 서브 트리로 한다. 만약 텍스트의 크기가 같은 경우에는 자식 노

드의 수가 많은 서브 트리를 최소 서브 트리로 한다.

3.2.3 분리 태그 추출 및 컨텐트 분리

컨텐트 분리 태그를 추출하기 전에 먼저 컨텐트 분리 후보 태그를 결정하여야 하는데 최소 서브 트리의 루트 노드의 자식(Child) 노드 태그를 컨텐트 분리 후보 태그로 한다.

분리 후보 태그가 결정되면 이 태그들을 대상으로 컨텐트를 분리하기 위해 경계가 되는 분리 태그를 추출하기 위하여 글자수에 대한 표준편차, 태그 쌍의 반복 패턴, 부분 경로의 발생 횟수, 인접 형제 태그의 발생회수 등을 이용한 휴리스틱(Heuristic)을 사용한다.

적용한 각 휴리스틱들은 특정 형태의 웹 문서에서 최적화되어 있고 또한 다른 휴리스틱들과 독립적이다. 그러므로 웹 문서의 올바른 컨텐트 분리 태그를 추출할 가능성을 향상시키기 위해 각각의 독립적인 휴리스틱을 결합시키는 것을 고려한다.

다수의 휴리스틱에 대한 최상의 결합 상태를 결정하기 위해, 스탠포드 확신도 이론(Stanford certainty theory)[11]을 이용한다. 이 이론을 사용하기 위해서는 각각의 개별적인 휴리스틱에 대한 확신도(certainty factor)를 가지고 있어야 한다.

분리 태그를 추출하면 분리 태그를 기준으로 컨텐츠를 분리하고 분리된 컨텐츠 중에서 적절한 컨텐츠를 추출한다.

3.3 시나리오 생성 및 VoiceXML 문서 작성

3.3.1 시나리오 생성

컨텐츠 추출과정을 통해 추출된 컨텐츠는 음성 인터페이스를 통해 사용자에게 정보를 제공하게 된다. 음성적 정보 제공 방법은 시각적 정보 제공 방법보다 사용자가 한번에 이해할 수 있는 정보의 양이 제한적이기 때문에 추출된 컨텐츠의 양에 따라 N 개의 음성 시나리오가 생성된다.

본 논문에서는 추출된 컨텐츠의 평균 문자 수가 100개 미만이면 4개의 컨텐츠로 하나의 음성 시나리오를 구성하고 100개 이상이면 2개의 컨텐츠로 하나의 음성 시나리오를 구성하도록 하여 N개의 시나리오를 생성하도록 하였다.

생성되는 음성 시나리오의 기본 구조는 다음과 같다.

컴퓨터 : 컨텐츠에 대한 음성 제공
컴퓨터 : 선택 가능한 메뉴에 대한 안내 음성 제공
사용자 : 메뉴 선택
컴퓨터 : 해당 VoiceXML 문서로 이동

3.3.2 VoiceXML 문서 작성

각 음성 시나리오는 VoiceXML 1.0 형식에 맞춰 VoiceXML 문서를 생성한다. 생성되는 VoiceXML 문서의 기본 구조는 그림 3과 같다.

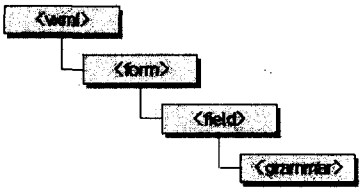

그림 3. 생성되는 VoiceXML 문서의 기본 구조

4. 결론 및 향후 과제

본 논문에서는 HTML로 작성된 기존의 컨텐츠를 이동 단말기 및 전화를 통해 음성으로 제공하기 위해 VoiceXML 문서로 재작성하는데 드는 비용과 시간을 줄이기 위해 HTML 문서를 자동으로 VoiceXML 문서로 변환하는 HTMLtoVoiceXML 변환기의 설계 방안을 제시하였다. 하지만 HTML 문서들이 갖는 많은 문제점에 의해 변환 가능한 대상 HTML 문서의 형태를 제한하였는데, 이를 해결하여 좀더 범용적으로 변환이 가능하도록 계속적인 연구가 필요할 것으로 생각된다.

5. 참고문헌

[1] VoiceXML Forum, www.voicexml.org
[2] W3C, "Voice eXtensible Markup Language (VoiceXML) version 1.0", http://www.w3.org/TR/voicexml, W3C Note 05 May 2000
[3] 김경란, 홍기영, VXML 편집기와 음성 브라우저의 설계 및 구현, 2000년 한국정보과학회 춘계학술대회 논문집, 27권 1호(B), pp414-416, 2000. 4
[4] 김경란, "VoiceXML 기반 음성 브라우저의 설계 및 구현", 성신여자대학교 석사학위 논문, 2001. 2
[5] http://www.mediaford.co.kr/
[6] 윤현주, 하준, 은성배 김병호, 강상민, 서원균, VXML 인터프리터의 설계 및 구현, 제9회 한국음성과학회 학술발표대회 논문집, 2000
[7] http://www.brain21.com/
[8] Stuart Goose, Mike Newman, Claus Schmidt, Laurent Hue, "Enhancing Web accessibility via the Vox Portal and a Web-hosted dynamic HTMLVoxML converter", WWW9, Volume 33, Numbers 1-6, pp583-592, June 2000
[9] D.W. Embley, Y.S. Jiang, and Y.-K. Ng. "Record-boundary discovery in Web documents", In Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data (SIGMOD'99), pp467-478, Philadelphia, Pennsylvania, 31 May - 3 June 1999.
[10] David Buttler, Ling Liu, Calton Pu. "A Fully Automated Extraction System for the World Wide Web", IEEE ICDCS-21, Phoenix, Arizona, April 16-19, 2001.
[11] G.F. Luger, W.A. Stubblefield, "Artificial Intelligence: Structures and Strategies for Complex Problem Solving", Third Edition. Addison Wesley Longman, Inc., 1997