

# 순차적 패턴 매칭 기법을 이용한 대화형 도우미 에이전트

김 수영, 조 성배  
연세대학교 컴퓨터과학과

## Conversational Help Agent Using Sequential Pattern Matching Technique

Soo-Young Kim and Sung-Bae Cho  
Computer Science Department, Yonsei University

### 요 약

인터넷이 급속도로 성장함에 따라 웹사이트의 숫자도 늘어나고, 많은 정보들이 등록되었다. 웹사이트들은 사용자의 정보획득을 위해 다양한 하이퍼링크를 제공하고, 전문(full-text) 검색엔진을 도입하기도 하나, 웹사이트에 등록되는 정보의 양이 많아지면서 전문검색엔진의 유용성이 점점 줄어들고 있다. 따라서 본 논문에서는 사용자가 좀 더 친근하고 빠른 방법으로 웹사이트에 있는 정보를 습득할 수 있도록 하는 대화형 도우미에이전트를 제안한다. 즉, 사용자가 일상적으로 사용하는 자연어로 된 문장을 웹사이트 내의 도우미 에이전트와 주고받음으로써, 사용자가 원하는 정보를 얻을 수 있도록 한다. 도우미에이전트의 지식을 패턴-답변형태로 저장하고, 순차적 패턴매칭 기법을 이용하여 사용자가 원하는 대화를 이끌어낸다.

### 1. 서론

인터넷이 급속도로 성장함에 따라 인터넷의 웹사이트 숫자도 늘어나고, 많은 정보들이 등록되고 있다. 인터넷에 점점 많은 정보들이 등록되면서, 웹사이트는 여러 페이지에 나누어서 정보를 담고 있으며, 사용자의 정보획득을 위해 다양한 하이퍼링크를 제공하고 있다. 대개 웹사이트의 구조는 정보 제공자가 통일된 표준이 없이 구성되기 때문에 인터넷 사용자들이 낮은 웹사이트에서 원하는 정보를 쉽고 빠르게 찾는 것이 쉽지 않은 경우가 많다.

인터넷 사용자들에게 좀 더 쉬운 정보 검색방식을 위해 제안된 방식이 전문(full-text)검색이며, 사용자들은 자신이 원하는 정보를 대표하는 키워드(keyword)를 입력하면 검색엔진이 이를 포함하고 있는 정보들을 보여준다. 전문검색엔진은 하이퍼 링크에 의존하던 정보 검색 방식에 큰 도움을 주고 있으나, 홈페이지에 수록 되는 정보의 양이 많아지면서 전문검색엔진의 유용성이 점점 줄어들고 있다.

따라서 사용자가 좀더 친근하고, 빠른 방법으로 웹사이트에 있는 정보를 습득할 수 있도록 하는 대화형 도우미 에이전트를 제안한다. 즉, 사용자가 일상적으로 사용하는 자연어로 된 문장을 도우미에이전트와 주고받음으로써, 웹사이트내의 정보를 얻을 수 있도록 한다.

외국에는 영어로 대화할 수 있는 대화형 로봇들이

많이 연구되고 상용화되고 있으나, 우리나라에는 한글로 대화할 수 있는 대화형로봇에 관한 연구나, 관련 제품들이 그다지 많지 않은 실정이다.

본 논문에서는 대화형 로봇의 초기 단계로서, 순차적 패턴매칭기법을 이용하여, "패턴-답변"형태로 저장되어있는 스크립트파일로부터 적절한 대화를 이끌어낸다.

### 2. 관련연구

최초의 대화형 로봇으로는 1966년 MIT대학의 Weizenbaum 교수가 만든 ELIZA[1]가 있다. ELIZA는 사람과 기계사이의 자연어 의사소통(natural language communication)을 연구하기 위해 만들어진 프로그램으로, 사용자가 던진 문장을 분석하여, 대답에 필요한 키워드들을 추출한후, 이미 프로그램 되어 있는 문장에 키워드를 치환하여 대답한다.

ALICE (Artificial Linguistic Internet Computer Entity)[2]라는 대화형 로봇은 자연스러운 대화를 이끌어내기 위해, 패턴매칭 기법을 사용하였다. AIML (Artificial Intelligence Markup Language)이라는 언어를 사용하여, 질문/대답 패턴을 기술하였다. AIML에 기술된대로, 사용자의 질문에 해당하는 대답을 찾아 사용자와 대화하게 된다.

상용화된 대화형 로봇으로는, Neuromedia사의 "Shallow Red"[3]가 있다.

채터봇을 만들수 있는 저작도구인 NeuroStudio로 만들어진 로봇이다. (Neuromedia사는 현재 NativeMinds라는 회사로 이름을 바꾸고 업그레이드된 가상 대리자(virtual representative)를 선보이고 있다.)

이외에도 Artificial Life사의 SmartBot[4], Virtual Personalities사의 Verbot(Verbal Software Robot)[5]등이 있다.

### 3. 대화형 도우미 에이전트의 구현

#### 3.1 구조

대화형 도우미 에이전트는 웹페이지를 통해 사용자와 대화를 주고받는 인터페이스로 되어 있으며, 구조는 다음과 같다.

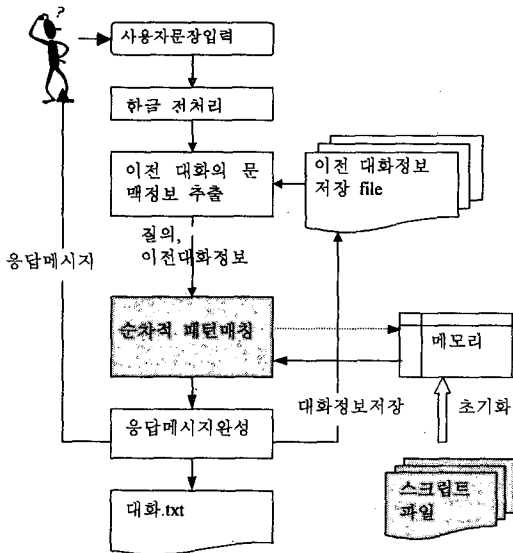


그림 1. 대화형 도우미 에이전트의 구조

사용자가 도우미 에이전트에게 원하는 것을 문장으로 입력하면, 도우미 에이전트는 한글 전처리를 수행한 후, 이전 대화의 문맥정보를 추출한다.

사용자가 입력한 문장과 문맥정보를 가지고 순차적 패턴매칭을 실행한 후, 응답메시지를 완성하여 사용자에게 보내주고, 대화내용은 텍스트파일에 기록하여 로그를 남긴다. 도우미 에이전트 관리자는 주기적으로 로그파일을 분석하여, 지식을 지속적으로 축적해 나가게 된다.

#### 3.2 한글 전처리

간단한 한글 전처리를 수행한다. 한글은 영어와 달라서 조사의 변화가 다양하므로, 도우미 에이전트가 인식할 수 있는 형태로 조사를 통일시킨다.

#### 3.3 순차적 패턴 매칭

도우미 에이전트가 대화를 나눌 수 있는 지식은 XML 형식을 가진 스크립트 파일에 저장된다. 스크립트 파일은 패턴-대답의 기본적인 패턴을 가진다.

다음은 스크립트 파일의 예이다.

```
<qna>
<pattern>계 이름은 * 입니다.</pattern>
<previous>당신의 이름은 무엇입니까?</previous>
<response>
<random>
<li>안녕하세요? <star/>님! 무엇을 도와드릴까요?</li>
<li>반갑습니다. <star/>님! 무엇든 물어보세요.</li>
</random>
</response>
</qna>
```

<qna>...</qna>는 하나의 입력패턴에 대한 패턴-답쌍을 기술해 놓은 것이다.

<pattern>...</pattern>은 사용자가 입력한 문장 청시키는 패턴을 기술해 놓은 것이다. 하나의 패턴에는 하나의 '\*'만을 허용한다.

<previous>...</previous>은 도우미 에이전트가 전에 대답했던 문장을 뜻한다. 생략 가능하다.

<response>...</response>은 입력한 문장에 대해 기술해 놓은 것이다.

<response>...</response>사이의<random>...</random>은 답변에 변화를 주기 위해 후보 답변들을 여러개 기술한 것. 도우미 에이전트는 후보 답변들중 랜덤하게 하나만 선택하여 사용자에게 답변을 한다.

<star/>는 입력패턴중 \*에 해당하는 단어를 뜻함.

도우미 에이전트는 특정 사이트에 관한 정보가 기술되어 있는 많은 <qna>..</qna>를 가지게 되는데, 사용자가 입력한 문장과 에이전트가 가지고 있는 패턴-대답쌍들중에서 가장 잘 매칭이 되는 패턴을 선택하게 된다. 이때, 도우미 에이전트는 자신이 가지고 있는 패턴-답변쌍들을 정렬해놓고 사용자가 입력한 문장과 순차적으로 매칭하다가, 일치하는 패턴이 나타나면, 그것을 선택하여 답변을 내보내게 된다.

매칭시에 사용자가 입력한 문장뿐만 아니라, 도우미 에이전트와 사용자와의 직전 대화에서의 도우미 에이전트의 대답까지 고려하여(<previous>...</previous>) 적절한 패턴을 찾아낸다.

#### 3.4 문맥

사용자와의 대화를 원활하게 하기 위해서는 대화의 문맥을 고려해야 하는데, 본 논문에서는 초보적인 문맥 처리로서, "그것", "그곳" 등의 대명사를 처리할 수 있도록 하였다.

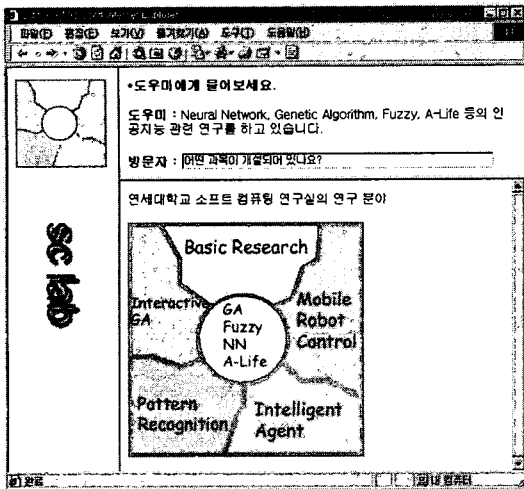
### 3.5 지식 축적

도우미 에이전트가 대화를 나누는데 필요한 특정 사이트의 지식은 패턴-답변 생성물을 사용하여 입력한다. 특정 사이트에서 나올 수 있는 질문/대답유형을 모두 정리하여, 사용자가 원하는 정보를 얻을 수 있도록 해야 한다. 패턴-답변생성물은 도우미 에이전트 관리자가 입력한 패턴을 도우미 에이전트가 인식할 수 있도록 바꾸어 스크립트 파일에 저장하는 역할을 한다.

## 4. 도우미 에이전트와의 대화

### 4.1 User Interface

특정 사이트의 한쪽 프레임에 위치하여, 사용자가 문장을 입력할 수 있는 텍스트 박스와, 도우미 에이전트가 대답을 할 수 있는 공간으로 이루어져 있다.



### 4.2 대화의 예

도우미 : 안녕하세요? 무엇을 도와드릴까요?  
 방문자 : 이 사이트는 무슨 사이트인가요?

도우미 : 이 사이트는 연세대학교 컴퓨터 과학과 소프트웨어 컴퓨터 연구실 홈페이지입니다.  
 방문자 : 어떤 분야를 연구하나요?

도우미 : Neural Network, Genetic Algorithm, Fuzzy, A-Life 등의 인공지능 관련 연구를 하고 있습니다.

방문자 : 어떤 과목이 개설되어 있나요?

## 5. 결론 및 향후 연구방향

본 논문에서는 대화형 로봇의 초기 단계로서 XML 스크립트 파일에 기초한 순차적 패턴매칭 기법을 중심으로 대화형 로봇을 제안하였다.

그러나 대화형 로봇과 사용자와의 대화가 사람끼리 대화하는 것처럼 매끄럽고, 사용자가 원하는 정보를 주기 위해서는 앞으로 많은 연구가 있어야 하겠다.

현재는 자연어 처리 기법을 거의 사용하지 않았으나, 다양한 형태의 질문을 알아듣고, 다양한 형태의 답변을 할 수 있으려면, 자연어 처리 기법이 필요하다.

사용자와 대화를 할때 재료가 되는 로봇의 지식의 표현방법이 더 연구되어야 하겠다.

사용자와의 대화를 매끄럽게 하기 위해서는 스크립트 파일 형태로 되어 있는 대화형 로봇의 지식의 양을 늘려야 한다. 로봇 관리자가 일일이 지식(문/답 패턴)을 입력하지 않고도, 지식을 축적할 수 있다면 로봇의 지식이 더욱 풍부해 질 것이다. 따라서 향후에는 검색 엔진이나, 데이터베이스를 활용할 수 있는 방안이 필요하다. 더불어, 로봇이 사용자와의 대화를 통해 스스로 학습할 수 있도록 하는 기능등이 있어야 한다.

사용자가 대화를 이끌어 나가면서 로봇이 일방적으로 묻는 말에 대답을 하는 것이 아니라, 로봇과 사용자가 대화의 주도권을 번갈아 가질 수 있다면, 실생활의 대화와 같은 형태의 대화가 이루어질 수 있다.

## 6. 참고문헌

- [1] Weizenbaum, J. "ELIZA-a computer program for the study of natural language communication between man and machine," Communications of the Association for Computing Machinery, 9(1):36-45, 1965.
- [2] <http://www.alicebot.org>
- [3] <http://www.neuromedia.com>
- [4] <http://www.artificial-life.com>
- [5] <http://http://www.vperson.com>