

질의응답시스템을 활용한 지능형 캐릭터 시스템에서 지식의 획득과 저장

Acquirement and Storage of Knowledge in Intelligent Character System Using Question Answering System

박홍원, 이기주^o, 이수진

(주)두레소프트 부설 휴먼인터페이스 연구소

walts123@hanmail.net, kjkj94@netian.com, isuuu@orgio.net

Hong-won Park, Ki-Ju Lee^o, Su-Jin Lee

DooReSoft Corp. Human Interface Laboratory

{논문요약}

질의응답시스템을 활용한 지능형 캐릭터 시스템에서는 지능형 캐릭터가 사용자(게이머)의 질의에 대해 응답할 때 해당 캐릭터에 특정 지식이 주어지지 않은 경우에 대비하여 해당 지식을 질의의 주체인 사용자 혹은 다른 외부변수로부터 획득하는 방법론과 획득한 지식을 지능형 캐릭터의 지식구조에 저장하는 방법론에 대한 연구가 병행되어 왔다.

본 논문에서는 지능형 캐릭터가 사용자가 입력한 자연어 문장으로부터 특정 지식을 획득하고 획득한 지식을 정해진 방법에 따라 지능형 캐릭터가 이해할 수 있는 지식구조로 구조화하는 방법론에 대해 구체적인 예를 통해 상세하게 설명한다.

1. 서론

1-1. 연구의 배경 및 필요성

본 연구에 앞서 온라인 게임에서 불특정 다수의 게이머와 자연어 형태의 대화를 통해 게이머에게 다양한 정보를 제공하거나 인터넷 쇼핑몰이나 안내 시스템에서 사용자가 요구한 정보에 대해 자연어 형태의 정보를 제공할 수 있는 “다국어 기반의 질의응답시스템을 활용한 지능형 캐릭터 시스템(박홍원 외, 2002)”을 발표한 바 있다.[1]

지능형 캐릭터 시스템을 실제 응용 시스템에 활용하고자 할 경우 사용자(게이머)의 질의에 대해 지능형 캐릭터¹⁾가 응답할 지식을 갖고 있지 않거나 사용자(게이머)로부터 특정한 지식을 획득하여 대화를 이끌어가야 할 경우에는 사용자가 입력한 문장으로부터 지식을 획득하여 지능형 캐릭터의 지식구조로 변환해야 할 경우가 발생한다.

이에 본 논문에서는 사용자(게이머)의 자연어 문장으로부터 지식을 획득하여 지능형 캐릭터가 이해할 수 있는 지식으로 구조화하는 방법론과 이 과정에서 필수적으로 고려되어야 할 사항에 대해 예를 통해 상세하게 살펴보고자 한다.

1-2. 연구의 범위

본 논문에서는 본 연구에 앞서 진행된 “다국어 기반의 질의응답시스템을 활용한 지능형 캐릭터 시스템(박홍원 외, 2002)”에 추가적으로 적용하기 위한 지식의 획득과 저장에 관한 방법론을 주로 논의하고자 한다.

따라서, 본 논문에서는 질의응답시스템이 주체적인 대화를 이끌어가기 위해 필요한 대화플랜이나 의도되지 않은 사용자의 질문에 대한 예외처리 같은 실제 시스템 운영에 필요한 자료나 모듈에 대해서는 언급하지 않으며, 자연어 기반의 질의응답 시스템에서 필수적인 형태소분석, 구문분석, 단어 사전 등에 대해서도 구체적인 설명은 생략한다.²⁾

1) 여기서 사용하고 있는 ‘지능형 캐릭터’라는 용어는 “다국어 기반의 질의응답시스템을 활용한 지능형 캐릭터 시스템”에서 사용한 개념과 동일한 개념임. 기존 연구에서는 ‘대화 대행자(강동혁, 1996)’라는 유사한 개념이 사용된 바 있으나 본 논문에서는 ‘지능형 캐릭터’라는 용어로 통일하여 사용함.

2) 형태소 분석, 구문 분석, 단어 사전은 (주)두레소프트의 다국어 번역 시스템(DMTS)의 방법론 및 구조와 동일함.

1-3. 연구의 방법론

본 연구는 사용자(게이머)가 입력한 자연어 문장으로부터 지식을 획득하고 저장하는 방법론³⁾에 대한 것으로 주로 표를 통해 예시하고 그림을 통해 진행 과정을 설명하는 방법으로 논의를 전개해 나가고자 한다.

먼저, 지식의 획득에 대해서는 획득할 수 있는 문장의 유형에 대해 설명하고 이미 제안한 바 있는 지능형 캐릭터의 지식구조에 대비하여 자연어 문장에서 문장의 어떤 요소가 지식구조에서 어떤 항목을 구성하는지에 대해 예를 통해 상세하게 설명하고 마지막으로 지식의 획득 절차를 그림을 통해 설명한다.

다음으로 지식의 저장에 대해서는 지식의 주체와 객체를 어떻게 구분하는지에 대해 살펴보고 지식의 신뢰도 개념이 무엇인지를 다양한 신뢰도가 적용된 예문을 통해 그 차이점을 설명하고 마지막으로 지식의 저장 절차를 그림을 통해 설명한다.

2. 지식의 획득과 저장

2-1. 지식의 획득

2-1-1. 지식의 획득이 가능한 문장의 유형

본 연구가 선행 연구인 “다국어 기반의 질의응답시스템을 활용한 지능형 캐릭터 시스템(박홍원 외, 2002)”의 후속 연구이므로 지식의 획득이 가능한 문장의 유형은 지능형 캐릭터의 지식구조에 기반하여 설명되어야 할 것이다.

선행 연구에서 지능형 캐릭터의 지식구조는 보어 구문, 타동사 구문, 자동사 구문을 구조화된 지식으로 변환하는 과정을 보였다. 또한, 부사구(부사절)를 통해 장소, 시간, 방법, 이유를 부가적으로 기술(description)하는 방법도 제시하였다.

<표1>에서는 보어 구문과 형용사 술어구문을 비교하여 설명하는 것으로 예문을 보더라도 보어 구문과 형용사 술어구문은 통사적으로 다른 형태를 가지고 있지만 선행 연구에서 이 두 구문을 공통된 방법론으로 처리하였으므로 지식을 획득하는 과정에서도 이들을 동일한 지식종류로 간주한다.

3) 문장형태의 지식을 개념형태의 지식으로 바꾸는 과정을 '의미구조'에서 '개념구조'로의 전환으로 해석하는 시각(김명철, 1987)도 있으나 본 논문에서는 '지식의 구조화'라는 용어로 통일해서 사용하고 내부적인 처리순서에 따라 '지식의 획득'과 '지식의 저장'이라는 용어로 구분해서 사용하기로 함.

	지능형 캐릭터의 지식구조	한국어 문장
(a)	Jane.ATR.JOB;a student	제인은 학생이다.
(b)	Jane.ATR.APR;pretty	제인은 예쁘다.

<표1> 보어 구문과 형용사 술어 구문의 비교

(a)와 (b)는 지능형 캐릭터의 지식구조에서는 'ATR'이라는 동일한 지식종류⁴⁾로 분류하고 있지만 문장형태의 지식일 때에는 (a)는 '주어 + 조사 + 보어 + 서술격 조사' 형태의 보어 구문이 되고 (b)는 '주어 + 조사 + 서술어(형용사)' 형태의 형용사 술어 구문으로 각각 다르게 나타난다.

<표1>에서 알 수 있듯이 지식의 획득이 가능한 문장의 유형은 지능형 캐릭터의 지식구조가 구조화할 수 있는 문장이라면 어떤 것이든 가능하다. 지능형 캐릭터 시스템에서 현재 처리하고 있지 않은 내포문, 이중주격문(접주어문) 등은 지식의 획득 대상에서 제외한다.

2-1-2. 문장의 구성요소와 지식구조

선행 연구에서는 지능형 캐릭터의 지식구조를 설명함에 있어 자연어 문장의 구성요소를 중심으로 설명하지 않고 통일화된 데이터베이스 필드에 지식을 구조화하는 방법론을 중심으로 논의를 진행하였다.

그러나, 본 연구는 지식구조의 설계가 논의의 핵심이 아니라 지식의 획득 방법론이 논의의 핵심이므로 자연어 문장의 구성요소가 지능형 캐릭터의 지식구조에서 어떤 항목에 어떤 인식자로 기술되어야 하는지에 대해 설명한다.

<표2>에서 보는 바와 같이 문장의 주요한 구성요소는 지능형 캐릭터의 지식구조에서도 역시 중요한 역할을 담당한다. 자연어 문장에서 주어는 지식구조에서도 해당 지식(1개의 레코드)의 주어 역할을 수행한다. 또한, 목적어와 보어는 지능형 캐릭터의 지식구조에서 '속성대상' 항목에 기술된다.

한편, 서술어의 경우는 그 형태가 보다 다양하게 나타나는데 보어를 동반하는 보어 구문이나 형용사 술어문은 'ATR'이라는 지식종류로 분류된다. 'ATR' 지식은 동작을 수반하지 않는 지식으로 주어의 상태나 주어의 속성을 설명하는 지식이다.

4) '지식종류'는 주로 문형을 결정짓는데, 지능형 캐릭터의 지식구조에서는 보어구문(ATR), 자동사 구문(VI), 타동사 구문(VT)이 있음.

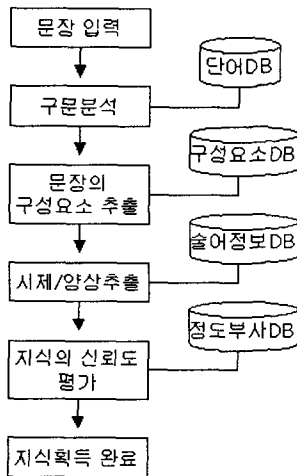
'VT'는 목적어를 수반하는 동작을 설명하는 지식이고, 'VI'는 목적어를 수반하지 않는 동작을 설명하는 지식이다.

(a)	철수는 학생이다.
	(주어+보어+서술어[서술격 조사]) 철수.ATR.JOB;학생;PRESENT
(b)	철수는 똥똥하다.
	(주어+서술어[형용사]) 철수.ATR.APR;똥똥하다;PRESENT
(c)	영희는 사과를 좋아한다.
	(주어+목적어+서술어[타동사]) 영희.VT.좋아하다;사과;PRESENT
(d)	인호는 죽었다.
	(주어+서술어[자동사]) 인호.VI.죽다;NULL;PAST

<표2> 문장의 구성요소와 지식구조의 관계

지능형 캐릭터의 지식구조에서 보어 구문과 형용사 술어문은 공히 'ATR' 지식종류로 분류되지만 '구체속성'의 인식자에 의해 사실상 다른 지식으로 재분류되어 구조화된다. (a)는 주어의 직업을 설명하는 'ATR.JOB'으로 구조화되고, (b)는 주어의 외모를 설명하는 'ATR.APR'로 구조화된다.

2-1-3. 지식의 획득 절차



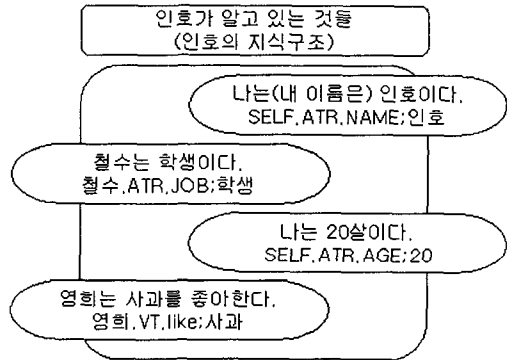
<그림1> 지식의 획득 절차

2-2. 지식의 저장

2-2-1. 지식의 주체와 객체

지능형 캐릭터 시스템을 실제 응용 프로그램에 적용하였을 경우 지능형 캐릭터의 이름, 사용자(게이머)의 이름, 자연어 문장의 주어 등의 3가지 요소가 어떤 관계에 있는가에 따라 해당 지식을 주체로 취급할지 객체로 취급할지가 결정된다.

<그림2>에서 보는 바와 같이 이름이 '인호'인 캐릭터가 이름이 '철수'인 사용자와 대화를 나누고 있다고 한다면, '철수'가 "나는 학생이다."라고 입력하였을 경우, 이름이 '인호'인 지능형 캐릭터는 자신의 지식구조 데이터베이스에 "나.ATR.JOB;학생"이라고 기술하는 것이 아니라 "철수.ATR.JOB;학생"이라고 기술해야 한다. '인호'의 입장에서 보면 '나(=철수)'는 객체이기 때문이다.



<그림2> 지능형 캐릭터('인호')의 지식구조

반면, 이름이 '인호'인 지능형 캐릭터가 이름이 '철수'인 사용자와 대화를 나눈다고 가정하면, '철수'가 "너는 똥똥하다."라는 문장을 입력했을 경우, 지능형 캐릭터는 자신의 지식구조 데이터베이스에 "SELF.ATR.APR;똥똥하다"라고 기술해야 한다.⁵⁾

5) 지능형 캐릭터 입장에서 보면, 제3자에 대한 지식이나 대화 상대방에 대한 지식은 자연어 문장의 내용대로 여과없이 지식구조로 변경할 수 있지만 상대방이 자신(지능형 캐릭터)에 대한 욕설을 하거나 비상식적인 말을 할 경우에는 그 지식을 그대로 믿고 자신의 지식구조 데이터베이스에 저장하는데 어려움이 있음. 이러한 내용에 대해 2-2-2에서 상세하게 다루고 있음.

한편, 문장의 주어가 지능형 캐릭터나 대화 참가자 둘중에 그 어느 쪽도 아닐 때에는 그냥 객체로 취급하여 자연어 문장의 주어에 해당 지식의 주어로 기술하면 된다.

2-2-2. 지식의 신뢰도

지능형 캐릭터의 지식구조에서 '지식의 신뢰도' 개념을 사용하는 이유는 크게 2가지로 나누어 볼 수 있다. 첫째는 사용자(게이머)가 입력한 자연어 문장 내에 강조, 추측, 불확실성, 의심 등의 양상이 포함되어 있어 해당 지식을 완전히(100%) 신뢰할 수 없다는 것이고, 둘째는 사용자(게이머)가 입력한 자연어 문장 자체가 지능형 캐릭터에 대한 욕설이거나 비상식적인 내용의 지식이어서 여과없이 받아들이기 어렵다는 것이다.

특히, 후자의 경우 지능형 캐릭터의 지식이 인간의 그것과 동일하다고 가정한다면 지능형 캐릭터가 어떤 자연어 문장 형태의 지식을 구조화된 지식으로 변환하여 자신(지능형 캐릭터)의 지식구조 데이터베이스에 저장하는 것은 인간으로 말하면 사실 여부에 관계없이 그 사실을 인정하는 것이므로 결코 바람직한 일이라고 볼 수 없다.

이에 본 연구에서는 전자의 경우에 대해서는 추측이나 확신을 나타내는 부사나 양상에 따라 지식의 신뢰도를 설정하고 후자의 경우에 대해서는 허위지식 판별 데이터베이스를 별도로 고안하여 지능형 캐릭터 시스템에 적용하였다.⁶⁾

(a)	철수는 틀림없이 영화를 좋아한다.
	철수.VT.like;영화;TRUST=100%
(b)	철수는 영화를 좋아한다.
	철수.VT.like;영화;TRUST=80%
(c)	철수는 아마도 영화를 좋아할 것이다.
	철수.VT.like;영화;TRUST=50%
(d)	철수는 어쩌면 영화를 좋아할지도 모른다.
	철수.VT.like;영화;TRUST=30%

<표3> 다양한 표현에 따른 지식의 신뢰도 설정

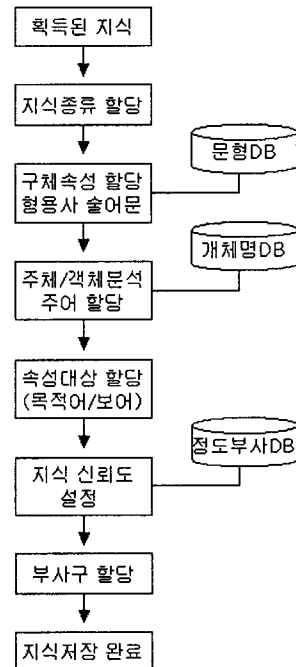
6) 허위지식 판별 데이터베이스는 욕설이나 비상식적인 논리를 구조화된 지식으로 구축한 것으로 본 논문에서 구체적인 내용을 다루지는 않음.

<표3>에서는 지식의 신뢰도가 높은 순서대로 (a)(b)(c)(d)의 순으로 해당되는 문장의 예를 들고 있다. (a)는 (b)보다는 높은 신뢰도를 가지고, (b)는 (c)보다는 높은 신뢰도를 가지고, (c)는 (d)보다는 높은 신뢰도를 가지지만, TRUST라는 변수의 값이 100%, 80%, 50%, 30%라는 수치는 큰 의미는 없다. 단지, 높고 낮음을 설정하기 위한 방법론일 뿐이다.

만든지 100%, 80%, 50%, 30%와 같이 수치가 아니더라도 HIGH, MEDIUM, LOW, LOWEST 등과 같이 서로 높고 낮음을 비교할 수 있는 표현을 대신해서 사용해도 무방할 것이다.

이때, 사용자(게이머)는 지능형 캐릭터를 속이거나 잘못된 지식을 주기 위해 악의적으로 거짓말을 하지 않는다는 것과 사용자(게이머)가 입력하는 문장에서 부사어를 표현할 때 상황에 맞는 적절한 부사어를 사용하여야 한다는 전제 조건이 필수적이다.

2-2-3. 지식의 저장 절차



<그림2> 지식의 저장 절차

3. 결론

본 논문은 질의응답시스템을 활용한 지능형 캐릭터 시스템에서 지식의 획득과 저장에 관한 방법론을 다룬 것으로 주로 자연어 문장으로부터 지식을 획득하는 방법론과 획득한 지식으로부터 지능형 캐릭터의 지식구조에 저장하는 방법론에 대해 살펴보았다.

지식의 획득에 대해서는 지식의 획득이 가능한 문장의 유형과 문장의 구성요소와 지식구조의 관계를 보이고 마지막으로 지식의 획득 절차를 그림을 통해 설명하였다.

지식의 저장에 대해서는 지식의 주체와 객체의 관계를 설명하고 지식의 신뢰도의 개념과 각각의 신뢰도가 설정된 예를 보이고 마지막으로 지식의 저장 절차를 그림을 통해 설명하였다.

계속되는 향후 연구에서는 지식의 신뢰도를 평가하고 설정하는데 보다 정교한 시스템이 필요할 것으로 생각되며, 허위사실이나 비상식적인 지식을 판단하는 방법론에 대해서도 단순히 데이터베이스에 의존하는 방법을 탈피하여 보다 논리적이고 체계적인 접근이 필요하다고 판단된다.

또한, 주체와 객체의 관계 설정에 있어서도 지능형 캐릭터에게 주어진 방식에 의해 일률적으로 적용되는 것이 아니라 사용자(게이머)나 다른 지능형 캐릭터가 한 말(지식)에 대하여 해당 객체에 대해서만 그 지식을 유효화하거나 인정하게 함으로써 객체마다 각각 다른 지식을 적용할 수 있도록 하는 방법론도 검토되어야 할 것이다.

참고문헌

- [1] 박홍원 외, “한영 기계번역 시스템을 위한 효율적인 한국어 용언 처리”, 제14회 한글 및 한국어 정보처리 학회 발표예정 논문, 2002.
- [2] 강동혁, “질의 응답을 위한 대화 대행자 모델 구현”, 연세대 석사학위 논문, 1996.
- [3] 김명철, “자연스러운 표층문 생성을 위한 한국어 표현특성에 관한연구”, 과학기술원 석사학위논문, 1897.