

# 시맨틱 웹 기반의 온톨로지와 연계한 사례기반 추론

고은정<sup>o</sup> 김여정 김운 강지훈  
충남대학교 컴퓨터학과  
{brain08<sup>o</sup>, innias, wkim, jhkang}@cs.cnu.ac.kr

## Case-Based Reasoning using Ontology in Semantic Web Environment

Eun-Jung Ko<sup>o</sup>, Yeo-Jung Kim, Yun Jin, Ji-Hoon Kang  
Dept. of Computer Science, Chungnam National University

### 요 약

Semantic Web은 웹상에 존재하는 정보들을 사람뿐만 아니라 컴퓨터 프로그램과 같은 기계들이 이해할 수 있도록 만들어진 차세대 웹이다. 이러한 Semantic Web을 수행하기 위해서는 Ontology가 가지고 있는 사실과 규칙들의 의미를 컴퓨터가 자동적으로 이해하기 위한 추론기술이 필요하다. 본 논문에서는 이러한 Semantic Web 환경에서 Ontology와 연계한 사례기반 추론 시스템을 제안한다. 사례기반 추론 시스템은 사례베이스로부터 현재 사례와 가장 유사한 사례를 검색하여 그 해결책을 제시하는 추론 방법으로 검색 시 빠른 해결책을 제시한다는 장점을 가지고 있다. 본 논문에서 제안하는 시맨틱 웹 기반의 온톨로지를 이용한 사례기반 추론시스템은 사용자의 요구사항을 의미적으로 정확하게 판단 할 수 있고, 검색 시 효율적인 알고리즘을 수행하여, 검색 성능 향상을 도모하였다.

### 1. 서론

시맨틱 웹[1]은 웹상에 존재하는 정보들을 사람뿐만 아니라 컴퓨터 프로그램과 같은 기계들이 이해하고 작업하기 용이하도록 표현한 차세대 웹이다. 이러한 시맨틱 웹을 지원하기 위해서는 온톨로지를 기술 할 수 있는 언어인 RDF/S[2], DAML[3], OIL[4], OWL(DAML+OIL)[5]등으로 온톨로지를 표현하고, 온톨로지가 가지고 있는 사실과 규칙들의 의미를 컴퓨터가 자동적으로 이해해야 한다. 이렇게 하기 위해서는 사실들에 규칙을 적용하여 새로운 지식을 유도할 수 있는 온톨로지와 연계한 추론 기술이 필요하다.

현재 시맨틱 웹에 적용하는 추론 기술은 논리기반(Logic-based)의 추론과 룰 기반(Rule-based) 추론이 있다. 논리기반 추론과 룰 기반 추론은 시맨틱 웹에 많이 적용이 되고 있지만, 몇 가지 단점이 있다. 사실들과 일반적 규칙에 근거하여 추론하는 논리기반의 추론론의 한 방법인 FOL추론은 단조 추론만 가능 하며, true/false 값만 대상으로 다루어져야 하고, 학습 기능이 없다. 또한 룰 기반 추론방법은 룰 시스템에서 제공하는 지식의 영역을 벗어나는 문제는 해결을 할 수 없다는 단점이 있다.

사례기반 추론(Case-Based Reasoning)[6]은 위의 추론 방법과는 달리, 과거의 사례를 바탕으로 문제를 해결하기 때문에 비록 문제가 복잡하더라도 이미 해결된 사례를 통해 해를 빨리 도출할 수 있다. 그래서 지식이 잘 파악되지 않은 대상 영역에 있어서도 추론을 가능하게 한다. 그리고 정확히 일치되는 사례를 검색하지 못했을 경우, 가장 유사한 사례를 변형하여 새로운 문제를 해결하도록 하며, 이러한 사례는 다시

새로운 사례로 저장이 된다.

본 논문에서는 사례기반 추론을 기반으로, 시맨틱 웹 환경에서 온톨로지와 연계한 Travel Agent System을 구현하도록 한다. Travel Agent System은 사용자 인터페이스를 통해, 사용자의 요구사항을 입력으로 받아, 사례기반 추론 시스템을 통해 검색을 하고, 검색 시 의미적으로 정확한 검색을 하기 위해 지역 온톨로지를 이용한다. 검색이 된 사례는 다른 예약 시스템을 통해 스케줄링을 하고, 그 결과는 다시 새로운 사례로 사례 베이스에 저장이 된다.

본 논문의 구성은 다음과 같다. 2절은 관련연구를 알아보고, 3절에서는 시맨틱 웹 환경의 온톨로지와 연계한 Travel Agent 사례기반 추론 시스템을 소개하고, Travel Agent System의 각 세부 시스템에 대해 알아보도록 한다. 4절에서는 결론을 내린다.

### 2. 관련연구

사례기반 추론 기법은 주어진 새로운 문제를 과거의 유사한 사례를 바탕으로 주어진 문제의 상황에 맞게 응용하여 문제를 해결하는 추론 방법이다. 사례기반 추론기법은 과거의 전문가 시스템에서 사용하던 지식의 추론을 통해서 해를 얻는 방법보다는 단순하면서도, 문제 영역이 잘 정형화되지 않는 분야에서도 적용이 가능한 방법이다.

사례기반 추론기법의 장점은 다음과 같다. 첫째, 추론자가 완전히 이해하지 못한 분야에 대해서도 해답을 제시하며, 둘째, 해답을 평가할 수 있는 적절한 알고리즘이 존재하지 않을 때에도 평가의 수단을 제시해 준다. 그리고 마지막으로

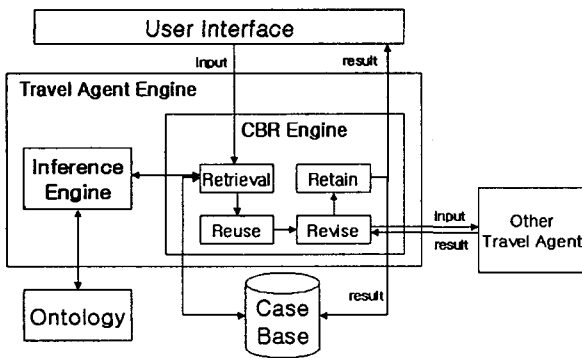
과거의 경험은 과거에 발생했던 문제에 대해 같은 실수를 반복하지 않도록 경고하는 기능을 한다.

사례기반 추론 방법은 도메인의 지식이 부족 할 때 유용하게 쓰인다. 또한 해결책의 정당성을 판단하거나 설명 할 때도 유용한 장점이 있다. 하지만 사례기반 추론기법은 충분한 유사 사례가 존재하지 않으면 해결책으로 제시되는 해결책이 부정확할 수도 있으며, 기존 사례와 다른 사례가 입력으로 들어오면 입력으로 들어온 사례를 인식하지 못할 수도 있는 단점이 있다.

이러한 사례기반 추론시스템을 이용하여, E-Commerce에 적합한 사례로 Analog Device[7], WEBSSELL[8], READEE[9] 등이 있다. 이와 같은 사례기반 추론 시스템들이 E-Commerce에 사용되는 이유는 쇼핑몰 등에서 좀더 고객의 기호에 맞는 지능적인 assistant를 원하고 있고, 더 빠른 검색 속도를 위한 인덱싱 기법이 사례기반 추론시스템에 적용이 되어 있기 때문이다.

### 3. 시맨틱 웹 환경의 온톨로지와 연계한 Travel Agent 사례기반 추론 시스템

시맨틱 웹 환경에서 온톨로지와 연계한 Travel Agent 사례기반 추론 시스템은 [그림1]과 같다. 이 시스템은 Inference Engine, CBR Engine, CaseBase, Ontology로 크게 네 부분으로 구성되어 있다.



[그림 1] 시맨틱 웹 기반의 온톨로지와 연계한 사례기반 추론 시스템

사용자는 사용자 인터페이스를 통해 사용자의 요구사항을 입력하고, CBR Engine에서는 그 요구사항에 맞는 사례를 CaseBase로부터 검색을 한다. CaseBase에 유사한 사례가 검색이 되면, 그 해결책을 재사용하여, 다른 Travel Agent에 예약이 가능한 정보를 조회하여 사용자에게 검색 결과를 보여준다.

CaseBase에서 검색시, 유사한 사례가 검색이 되지 않으면, Inference Engine을 통해, 지역 온톨로지 정보를 이용한다. 사용자가 입력한 출발지역이나 도착지역을 보고, 이와 가까운 거리에 위치한 지역 정보 온톨로지를 Inference Engine을 통해 이끌어 낸다. 그래서 이 새로운 정보를 CaseBase에서 재검색을 하여 해결책을 사용자에게 보여준다. 이 사례는 새로운 사례로 CaseBase에 저장된다.

#### 3.1 User Interface

사용자는 사용자 인터페이스를 통해 출발날짜, 도착날짜,

출발지역, 도착지역, 동반인원, 예상경비, 여행의 목적 등을 사용자의 요구사항에 맞게 입력을 한다.

검색의 결과로, 교통정보는 기본적으로 보여주고, 숙박정도, 여행일정등의 정보는 사용자가 선택하여 원하는 정보를 볼 수 있도록 한다.

[그림 2]는 사용자 인터페이스의 모습을 나타내고 있다.

[그림 2] 사용자 인터페이스

#### 3.2 CBR Engine

사용자 인터페이스를 통해 입력 받은 정보를 바탕으로, Retrieval, Reuse, Revise, Retain과정을 통해, 사례기반 추론이 진행된다.

Retrieval은 가장 유사한 사례를 CaseBase로부터 검색을 하는 과정으로, CaseBase에 있는 사례를 인덱싱하여 해당 사례와 기존의 CaseBase에 있는 사례가 얼마나 유사한지 측정하여, 가장 특성이 유사한 사례들을 검색한다. Retrieval과정은 사례기반 추론방법의 검색 속도를 결정하는 과정으로 효율적인 인덱싱과 검색 알고리즘 선정이 가장 중요하다.

Reuse은, 검색된 사례의 해결책을 재사용하는 방법으로, 검색된 해결책을 복사하여, 현재 입력으로 들어온 사례에 해결책을 적용하는 과정이다.

Revise는 제안된 해결책을 수정하는 과정으로, Reuse과정에서 적용된 해결책을 평가하여, 사례의 해결책이 실패일 경우 도메인 지식을 이용하여 해결책을 제시한다.

Retain은 현재 해결된 사례를 미래에 재사용하기 위해 CaseBase에 저장하는 과정으로, 사례기반 추론의 학습에 해당한다.

#### 3.3 Inference Engine과 Ontology

Inference Engine은 온톨로지에 있는 정보를 기반으로, 질의가 들어왔을 시, 추론 과정을 통해 사용자에게 추론 결과를 보여주는 Engine이다.

현재 사용 가능한 Inference Engine으로는 FaCT[10]와 Racer[11]가 있으며, 본 논문에서는 포함 관계 등을 추론하기 위해서, FaCT Inference Engine을 사용한다. 온톨로지는 지역 온톨로지를 구축하여, 국내, 국외 지역을 중심으로 구축한다.

### 3. 4 CaseBase

CaseBase는 사례기반 추론을 통해 학습된 사례를 저장하는 곳으로, 사례의 효율적인 저장을 위해 설계하였다. 데이터 교환의 표준을 고려하여 XML[12]기반의 XML Schema[13]를 기본으로, Case Schema를 정의하였으며, Case Schema에 맞게 표현된 Case XML문서의 예는 [그림2]와 같다.

```

<?xml version="1.0" encoding="EUC-KR"?>
<case>
  <description>
    <departDate>
      <year>2003</year>
      <month>9</month>
      <day>20</day>
    </departDate>
    <departRegion>대전</departRegion>
    <arriveRegion>제주도</arriveRegion>
    <people>1</people>
    <expenditure/>
    <round>편도</round>
    <purpose>개인여행</purpose>
  </description>
  <solution>
    <path>
      <departure>대전</departure>
      <destination>청주</destination>
      <transport>버스</transport>
    </path>
    <path>
      <departure>청주</departure>
      <destination>제주도</destination>
      <transport>비행기</transport>
    </path>
  </solution>
</case>
    
```

[그림 3] XML로 표현된 Case 예제

[그림 2]에서 나타난 것과 같이 Case는 Description과 Solution으로 구성이 된다. Description은 Case의 특성을 나타내며, CBR Engine에서 Retrieval시 검색 대상이 된다. Solution은 해당 사례의 해결책을 기술하는 부분으로, CBR Engine을 통해 검색이 된 사례의 해결책을 Reuse, Revise하기 위해 필요하다.

위의 예제는 사용자가 대전을 출발해서 제주도를 여행 갔을 때의 요구사항과 해결책을 보여주고 있는 예제이다. 사용자는 2003년 9월 12일에 출발하여, 대전을 출발해서 제주도로 도착하는 여행으로, 또한 동반인원은 1명이며, 예상경비는 30만원이고, 편도여행을 원하고, 여행의 목적은 개인여행이라고 명시한 것이다. 이러한 사용자의 요구사항, 즉 사례를 description부분을 기술한 뒤, 이에 해당하는 해결책을 기술하게 된다.

해결책으로는 교통정보로, 대전에서 청주까지는 버스를 타고 이동하여, 청주에서 제주도까지는 비행기를 이용하라는 해결책이 명시되어 있는 사례이다. 교통 예약 및 숙박 예약 등은 다른 Travel Agent를 이용하여 Revise과정에서 사용자에게 실시간으로 정보를 제공하게 된다.

이렇게 XML로 기술된 사례들을 더 효율적으로 관리하기 위해, 본 연구에서는 데이터베이스를 이용하여 사례를 저장 및 관리한다.

### 4. 결론

시맨틱 웹은 컴퓨터에 고도의 인공지능 기능을 부여하여 웹상에 존재하는 각종 정보를 사람 대신 컴퓨터 프로그램이 인식하고 가공하여 새로운 정보를 만들어 내는 기술이다.

이러한 시맨틱 기술을 활용하면, 웹상에 존재하는 문서들이 단순히 사람만이 인식할 수 있도록 표현되는 것이 아니라 컴퓨터가 쉽게 읽을 수 있도록 표현되어, 컴퓨터프로그램이 이러한 문서를 원래의 의미(semantic)에 적합하게 처리함으로써 전자상거래 등 다양한 응용영역에서 데이터의 공유와 가공이 가능하게 된다.

이러한 시맨틱 웹 기술을 이용하기 위해서는 온톨로지를 표현하고, 이 온톨로지가 가지고 있는 사실과 규칙들의 의미를 컴퓨터가 자동적으로 이해하기 위한 추론기법이 필요하다. 현재 추론 기법들은 비단조론, 퍼지이론 불확실성등 많지만, 시맨틱 웹 환경에 적합한 추론 기법으로, 본 논문에서는 사례기반 추론기법을 이용하였다.

사례기반 추론기법은 검색속도가 빠르며, E-Commerce에서는 고객의 기호를 더 정확히 알 수 있다는 장점을 가지고 있다. 본 논문에서는 사용자가 입력한 요구사항의 정보를 바탕으로, 정확히 일치하는 사례가 검색이 되지 않을 때에는 지역 온톨로지 정보를 Inference Engine을 이용하여, 사용자에게 보다 폭 넓은 정보를 제공한다. 또한 사례기반 추론기법을 사용하기 때문에 검색이 빠르며, XML을 기본으로 효율적인 사례 관리를 하기 때문에, 분산환경에서도 사례의 교환이 가능하다.

### 5. 참고문헌

- [1] <http://wemanticweb.org>
- [2] W3C, RDF Vocabulary Description Language 1.0:RDF Schema, Working Draft, 23, January 2003.
- [3] <http://daml.org>
- [4] <http://www.ontoknowledge.org/oil>
- [5] <http://www.daml.org/2001/03/daml+oil-index.html>
- [6] A.Adamodt, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and system Approaches", AI Communications ISO Press, Vol.7:1, p.39~59(1994)
- [7] <http://analog.com>
- [8] P. Cunningham, "WEBSSELL: Intelligent Sales Assistant for the World Wide Web", Kunstliche Intelligenz (KI), January, 28~32. Arendtap, Bremen, Germany
- [9] P.Oehler, "READEE - Decision Support fir IP Selection using a Knowledge-Based Approach", IP.98
- [10] L. Horrocks, "FaCT: Fast Classification of Terminologies", <http://www.cs.man.ac.uk/~horrocks/FaCT/>
- [11] V. Haarslev "RACER: Renamed ABOX and Concept Expression Reasoner" <http://www.cs.concordia.ca/~faculty/haarslev/racer/index.html>
- [12] W3C "Extensible Markup Language(XML)" Recommendation, February 1998
- [13] W3C "XML Schema Part0: Primer", Recommendation, May 2001