

온톨로지 및 사례기반추론을 이용한 맞춤형 통합 정보 생성 프레임워크의 제안

이현정
고려대학교 경영대학
(hjlee5249@yahoo.com)

손미애
성균관대학교 공과대학 시스템경영공학과
(mysohn@skku.edu)

.....

다양한 정보자원들로부터 사용자가 요구하는 맞춤형된 정보를 추출해 내는 것은 더욱 어려워지고 있다. RSS를 비롯하여 개선된 다양한 정보 수집 방법들이 개발되었지만, 여전히 정보가공자인 사람의 도움 없이 필요한 정보들을 수집하여 정리 및 가공하는 작업이 쉽지는 않다. 따라서 본 연구에서는 정보사용자들이 사용 목적에 맞게 정보를 가공하는 부담을 줄여주기 위해 사례기반추론과 온톨로지에 기반한 맞춤형 통합정보생성 프레임워크를 제안한다. 본 프레임워크는 세 단계로 구성된다. 첫째, 수집된 웹 정보를 정보가공의 용이성을 위해 사례로 변환한다. 둘째, 동적 유사도 검색을 통해 수집된 사례들로부터 정보 사용자의 동적 요구사항에 적합한 사례를 검색한다. 셋째, 전 단계에서 추출된 사례를 정보사용자의 요구사항에 보다 적합한 지식으로 가공하기 위해 집중 유사도를 적용한다. 본 프레임워크는 여행자들의 정보수집을 위한 여행정보시스템에 적용되어 그 효과를 입증하였다.

.....

논문접수일 : 2009년 11월 18일 논문수정일 : 2009년 12월 04일 게재확정일 : 2009년 12월 15일 교신저자 : 손미애

1. 서론

다양한 정보자원의 등장과 함께 정보 사용자들이 탐색해야 할 정보 자원의 종류는 매우 다양해졌으며 이로부터 수집할 수 있는 콘텐츠의 양 또한 폭발적으로 증가했다. 그러나 정보 사용자들이 정보 검색 및 추출을 위해 현존하는 모든 정보 자원을 확인하기란 어렵고, 설령 필요한 자원들을 모두 인지하더라도 시간적 제약 등을 이유로 이를 모두 확인하여 해당 자원들로부터 필요한 정보들을 찾아 맞춤형된 지식을 생성하기란 매우 어렵다. 정보자원이 사용자들에게 풍부한 자원과 정보를

제공함에도 불구하고 여전히 사용자들이 원하는 형태의 가공된 지식을 추출하기에는 어려운 상황이 계속되고 있다. 게다가 정보자원이 가지고 있는 정보들이 대부분 비구조화된 형태를 띠고 있다는 점도 정보 가공의 큰 제약이 되고 있다.

이러한 문제점들을 해결하고자하는 노력의 일환으로 비구조화된 형식의 정보로부터 맞춤형 정보를 생성하는 방법으로서 Resource Description Framework (RDF)를 이용한 주석 또는 온톨로지를 적용하는 방법 등이 제안된 바 있다. 그러나, 이들 방법으로는 사용자가 기대하는 수준의 맞춤형 통합정보, 즉 다양한 정보원로부터 사용자의 요구

* 본 연구는 고려대학교 BK21 경영 전문서비스 인력양성단의 지원을 받았음. 본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었음 (UD080042AD).

를 의미상으로 충분히 만족시키는 정보의 조각들을 선별적으로 추출해 이를 종합적으로 통합 정보를 생성할 수 없기 때문에 사용자의 부담을 줄이는 데 한계가 있다. 이런 이유 때문에 맞춤형 통합 정보의 생성을 효과적으로 지원할 수 있는 방법의 개발이 절실히 요구된다.

본 연구에서는 동적으로 변하는 정보자원 및 그 환경에서 역시 동적으로 변화되는 사용자의 요구사항을 만족시켜줄 수 있는 맞춤형 통합정보 생성 프레임워크를 제안하고자 한다. 즉, 본 프레임워크는 수집된 정보를 가공해서 사용자의 요구사항을 만족시키는 맞춤형 통합정보를 자동적으로 생성할 수 있도록 한다. 정보의 자동 생성을 위해 웹 정보는 Really Simple Syndication (RSS)를 이용해 자동적으로 수집되는 것으로 가정한다. 맞춤형 통합정보의 효과적인 생성을 위해 본 프레임워크에서는 수집된 비구조적 정보를 구조적 정보의 한 형태인 사례로 변환하였으며, 사용자의 요구사항에 부합하는 유사사례를 찾기 위한 새로운 유사도 계산 방법인 동적 유사도 검색과 집중 유사도 검색을 제안하였다.

온톨로지 상에 존재하는 클래스(개념)들간의 유사도를 계산하는 기존 방법들은 온톨로지의 구조만을 고려하기 때문에 비교 대상인 클래스(개념)들간에 존재하는 컨텍스트에 대한 고려가 이루어지지 않았던 문제점이 있다. 따라서 본 연구에서는 온톨로지 구조와 함께 온톨로지의 클래스들이 포함된 컨텍스트에 대한 유사도 관계를 동시에 고려하는 동적유사도 계산 방법검색을 제안하고자 한다. 즉 온톨로지 상의 구조가 같더라도 포함하는 클래스들이 포함된 컨텍스트에 따라 유사도 값이 동적으로 변화할 수 있어 온톨로지 상에서의 유사도 검색의 정확성을 높일 수 있는 장점이 있다. 또한 요구사항과 다중 클래스들의 값들 간의 유사도

검색을 위해 하나의 요구사항을 중심으로 클래스 값들의 집중된 정도를 확인하는 집중유사도 계산 방법을 제안한다. 집중유사도는 직접 집중유사도와 간접 집중유사도로 구성된다. 즉 직접 집중유사도란 맞춤 사례 생성을 위해 추출된 세부-사례가 포함하는 값들 중 사용자 요구사항과 일치하는 값들을 대상으로 수행하는 유사도 계산이며, 간접 집중유사도검색이란 직접 집중유사도검색에서 사용된 값들을 제외한 나머지 값들과 사용자 요구사항과의 유사도검색을 통해 세부-사례가 포함하고 있는 값이 얼마나 사용자 요구사항에 근접하는가에 대한 유사도 계산이다.

본 연구에서 제안된 동적유사도와 집중유사도는 동적으로 변화하는 다양한 정보자원으로부터 사용자 요구사항을 만족시켜 줄 수 있는 제안되는 맞춤형 통합지식 생성을 위한 프레임워크를 지원한다.

2. 관련 연구

정보 통합은 다양한 정보 자원들로 부터의 다양한 자원들의 의미를 고려한 종합적인 정보의 가공을 의미한다(Bernstein, 2008). 즉, 이를 위해 웹 등의 정보자원들로부터 정보를 분류하고 사용하고 가공하는 것은 매우 중요하다. 그러나 이러한 자원들의 대부분이 정보 가공자의 도움 없이는 유용한 정보생성이 어렵다. 다양한 이유가 존재하겠지만 이들 중의 하나가 비구조화된 정보를 인식하고 분류하고 사용하기에는 매우 어렵다는 것이다. 이를 위한 다양한 연구가 진행되어 왔고, 의미상 같은 내용을 서로 다른 방법으로 표현된 정보를 인식하기 위해 ETL(Extraction, Transformation and Loading)이나 관련 스키마 등이 제안되었다(Bernstein, 2008; Alexiev et al., 2005). 비구조화된 문서로 부

터의 단순히 주제어 검색 및 기 정의된 정보 구조 (예를 들어, 이름, 주소, 상품명 등) 등을 이용한 문서 요약 등에 대한 연구가 있어 왔다. 또한 RDF와 같은 주석을 이용하여 정보를 요약 및 통합하고자 하는 노력도 있어 왔다. 문서의 의미적 통합을 위한 온톨로지를 활용한 방법들이 보다 진보된 문서 통합 방법으로 주목 받고 있다(Halevy et al., 2005).

그러나, 기존의 정보검색이나 주제어 검색 등과 같은 방법으로 사용자의 요구사항에 적합한 문서를 생성해 내는 것은 역부족이다. 주제어 검색은 많은 정보는 나열하지만 단지 그 검색용어가 포함되는 경우의 정보도 검색하여 사용자가 필요로 하는 정보를 정제하거나 맞춤정보의 추출에는 한계가 있다. 심지어 RSS를 사용해서 웹 정보 등의 정보 변화를 인식하고 이를 정보 인식에 반영하더라도 여전히 사용자의 요구사항에 맞는 통합 정보를 종합적으로 생성하여 제공하는 데는 어려움이 있다(Philippe et al., 2000). 웹 정보 사용자들을 실제적으로 수많은 정보 자원들로부터 맞춤 정보를 생성하기 위해 많은 시간과 노력을 낭비하고 있다(Mani et al., 1998; Teufel and Moens, 1997). 따라서, 정보의 맞춤화를 위한 정보 추출, 분류, 통합 등에 대한 연구 요구는 매우 현실적이고도 필요성에 대한 인식이 증대되고 있다고 할 수 있다. 이를 위해 본 연구에서는 온톨로지 기반 사례추론을 이용한 맞춤형 통합지식 생성 프레임워크를 제안한다.

3. 맞춤형 통합정보 생성 프레임워크

정보자원들로부터 수집된 다양한 정보로부터 통합정보를 자동적으로 생성하기 위한 조건은 다음과 같다. 첫째, 맞춤 정보의 자원이 될 수 있는 웹 정보 등이 정보사용자인 사람의 도움 없이 시

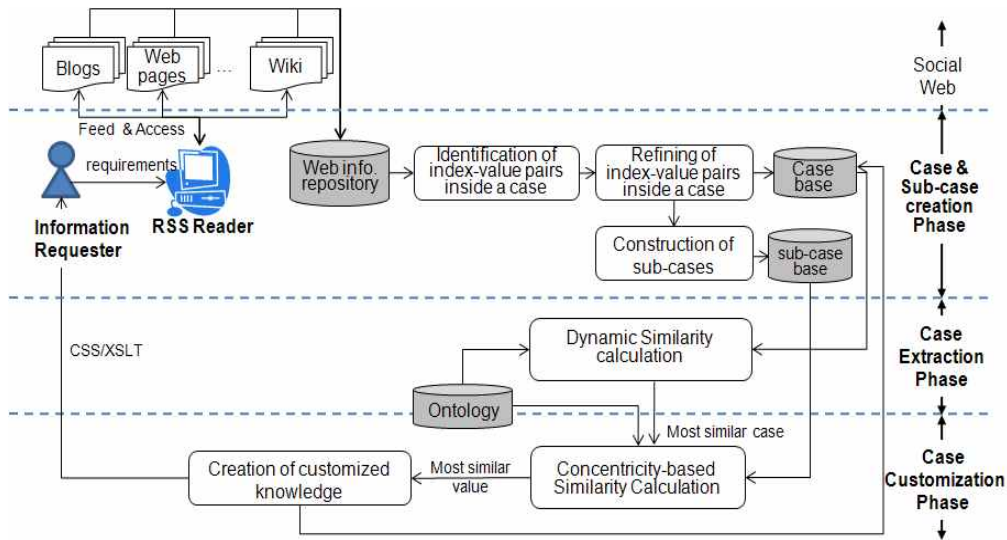
스템에 의해 인식 및 처리가 가능해야 한다. 둘째, 시스템이 사용자가 필요로 하는 정보만을 자동적으로 추출할 수 있어야 한다. 셋째, 추출된 정보로부터 사용자의 요구를 의미상 및 종합적으로 만족하는 정보들로만 구성된 통합 지식을 생성할 수 있어야 한다. 이를 위해 본 프레임워크는 다음과 같이 세단계로 구성한다.

3.1 사례(Case) 및 세부-사례(Sub-case) 생성 단계

사용자 요구에 맞는 맞춤형 통합 지식을 생성하기 위한 첫 단계는, 수집된 비구조적 정보들을 지식 표현 방법 중의 하나인 사례로 변환하는 것이다. 이때 사례의 구조는 인덱스와 값의 쌍들로 구성된다. 인덱스는 사용자 요구사항, 요구사항들의 조합 및 연관 요구사항으로 구성될 수 있다. 인덱스에 대응되는 값은 해당 인덱스를 포함하고 있는 문장으로 구성된다. 사용자 요구사항에 따라 인덱스와 값의 쌍들로 구성된 사례는 사례베이스에 저장된다. 세부-사례는 하나의 인덱스를 포함하고 있는 서로 다른 값들이 여러 사례들로부터 모아져 인덱스-값들로 생성된다.

3.2 유사 사례 추출 단계

사례베이스로부터 사용자의 요구에 가장 인접한 사례를 추출하는 단계이다. 유사사례를 추출하기 위해 사용자 요구사항과 과거 사례의 인덱스들 간의 유사도를 계산한다. 이를 위해 본 연구에서는 동적 유사도를 제안한다. 이 방법은 'Edge-based similarity calculation' 방법(Slimani et al., 2006; Song et al., 2007)을 개선한 것으로서, 비교 대상이 되는 두 클래스들이 공유하는 클래스들은 어떤 것인지 그리고 공통의 클래스가 몇 개나 있는지에 따라



<그림 1> 웹 정보를 이용한 맞춤형 통합지식 생성 프레임워크

유사도가 달라져야 한다는 것에 착안한 것이다. 기존 'Edge-based similarity calculation' 방법의 경우 비교 대상이 되는 두 클래스들의 유사도는 온톨로지의 구조가 변하지 않는 한 항상 동일한 값을 가지며, 이러한 유사도 측정 방법을 사용해 생성된 맞춤형 통합정보의 사용자 요구 충족도는 낮을 수밖에 없다. 본 논문에서 제안된 프레임워크에서는 이를 보완하기 위해, 비교 대상이 되는 두 클래스들이 어떤 클래스를 공유하는지, 그리고 그 수에 따라 유사도가 다르게 계산된다.

3.3 사례를 이용한 맞춤정보 생성 단계

사용자 요구사항과 가장 근접한 사례가 추출되었다 하더라도, 이것이 사용자의 요구사항을 모두 만족시키지 못할 수 있다. 이 경우 사례에 대한 맞춤화가 수행된다. 첫 단계는, 사용자의 요구사항을 만족시키지 못하는 사례가 포함된 인덱스들을 확

인하고 이에 해당하는 세부-사례를 찾는 것이다. 세부-사례는 해당 한 인덱스가 가질 수 있는 다양한 값들로 쌍을 이루므로, 해당 인덱스의 세부-사례가 포함하는 값들 중에서 사용자의 요구사항을 제일 잘 만족하는 값을 찾아 먼저 단계에서 생성된 사례를 맞춤 한다. 사용자의 요구사항과 세부-사례의 값들 간의 유사도는 집중 유사도를 이용해 계산한다. 맞춤화 과정을 통해 수정된 사례는 정보 사용자에게 전달됨과 동시에 추후 재사용을 위해 사례베이스에 저장된다.

<그림 1>은 본 논문에서 제안하는 맞춤형 통합지식 생성 프레임워크를 여행정보 생성에 적용한 워크플로우이다. 여행자의 경우 여행지와 목적에 맞는 정보를 생성하기 위해 다양한 웹 정보로부터 필요한 정보를 추출하고 통합하는 작업을 수행한다. 여행자는 먼저 자신의 RSS 리더를 통해 여행 관련 정보를 모을 수 있는 환경을 구축한다. 모아진 정보 중 관심정보 만을 선별해 자신의 웹 정보

저장소에 저장한 후, 앞서 서술한 맞춤화 정보 생성의 세단계를 거쳐 통합 지식을 생성한다.

4. 사례 및 세부-사례 생성

동적으로 변하는 사용자의 요구를 만족시킬 수 있는 지식을 자동적으로 생성하기 위한 첫 단계는 사람만이 이해하고 처리할 수 있는 비구조적인 정보를 시스템이 이해하고 처리할 수 있는 형식으로 변환하는 것이다. 시스템을 위한 지식 표현 방법에는 규칙(rules), 프레임(frames), 및 사례(cases) 등이 있으며, 사례는 다른 두 방법에 비해 비 구조화된 형식에서 구조화된 형식으로 변환 시 정보의 손실이 적고 규칙이나 프레임 등으로는 표현하기 어려운 비구조적 정보들을 표현할 수 있다는 장점을 가지고 있다. 그래서 우리는 정보자원으로부터 수집된 HTML 형식의 비구조적인 정보를 사례 및 세부-사례로 변환하는 방식을 채택하였다. 사례 및 세부-사례로의 변환과정은 다음과 같다.

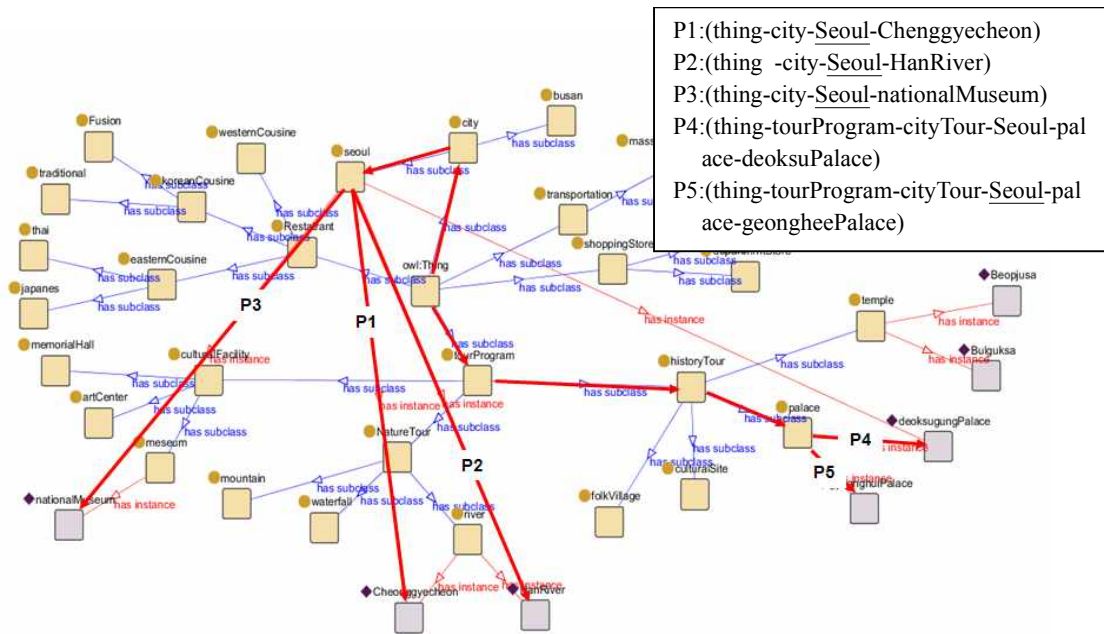
4.1 사례(Case) 생성

정보자원으로부터 수집된 문서를 사례로 변환하는 과정은 HTML 형식의 웹 정보를 문장 단위로 파싱하는 것으로부터 시작된다. 이후 파싱된 문장들을, 생성하고자 하는 사례의 잠재적인 값들로 취한 후, 사용자의 요구사항과 잠재적인 값들 간의 연관관계를 확인한다. 그리고 모든 잠재적인 값에 대해, 한 값이 사용자 요구사항 또는 요구사항 조합을 포함하고 있는지를 확인한다. 만약 한 값이 요구사항 또는 요구사항 조합을 포함하고 있다면 그 값의 인덱스로 요구사항 또는 요구사항 조합을 채택한다. 예를 들어, 한 여행자의 요구사항이 ‘Seoul’, ‘Korea’ 및 ‘Cuisine’이고, 웹 정보로부터

다음과 같은 문장을 식별했다면 이 문장은 사용자 요구사항인 ‘Seoul’, 및 ‘Korea’를 포함하고 있으며 이들이 다음 예시 문장의 인덱스가 된다.

‘In the downtown area, there are several great walking courses that will take you around Seoul’s ancient palaces and royal shrines and reveal Korea’s rich historical legacy.’

해당 값이 사용자의 요구사항 또는 요구사항의 조합을 포함하고 있을 수도 있지만, 그렇지 않은 경우도 고려해야 한다. 즉, 직접적인 요구사항에 해당하는 용어는 아니지만 요구사항과 연관 있는 용어들 (연관요구사항)을 포함하고 있는 경우이다. 연관요구사항은 사용자 요구사항과 의미적 연관성이 있는 용어들로 구성되며 이를 도출하는 과정은 다음과 같다. 사용자 요구사항과 온톨로지의 클래스들로 존재하고 있는 지를 확인한다. 존재한다면, 해당 클래스를 중심으로 생성될 수 있는 모든 경로들을 찾아 낸다. 이때 이 경로의 최상위 노드는 ‘owl : thing’ 이며 최하위 노드는 인스턴스로 구성된다. 예를 들어, 사용자 요구사항이 ‘Seoul’인 경우 <그림 2>의 온톨로지로부터 추출할 수 있는 경로는 모두 5개이다. 사용자의 요구사항 또는 요구사항의 조합을 포함하지 않는 값들이 경로들을 구성하는 클래스들이나 인스턴스를 포함하고 있는지 확인한다. 만약 어떤 값이 경로 상의 한 클래스를 포함하고 있다면, 해당 클래스에서 시작해서 사용자 요구사항에 이르는 경로의 일부분을 연관 요구사항으로 취한다. 예를 들어, 한 값이 ‘Seoul’은 포함하지 않지만 ‘Chenggyecheon’을 포함하고 있다면, 경로 P1으로부터 (Seoul, Chenggyecheon) 이 연관요구사항으로 도출된다. 만약 한 값이 ‘Seoul’은 포함하지 않지만 ‘geongbokPalace’와 ‘geong-



< 그림 2> 연관요구사항 인덱스 추출을 위한 온톨로지 상에서의 경로 예

heePalace’를 포함하고 있다면 <그림 2>에서 보는 것처럼 연관요구사항은 (Seoul, palace, geongbok-Palace, geongheePalace)가 될 것이다.

특정 값이 요구사항, 요구사항 조합 및 연관요구사항 중 어느 것도 포함하고 있지 않다면 이 값은 제거된다. 이 변환 과정을 거치고 나면, 웹으로부터 수집한 비구조적인 정보는 사용자 요구사항들과 연관된 인덱스 및 값들의 쌍들로 구성된 하나의 사례로 변환된다.

4.2 세부-사례(Sub-case) 생성

생성된 사례는 인덱스-값의 쌍들로 구성되며 사례를 구성하는 각각의 인덱스를 기준으로 인덱스-값의 쌍으로 분할한 것이 세부-사례다. 즉 세부-사례 역시 사례의 한 유형으로서 인덱스-값의 형

식으로 표현된다. 그러나 세부-사례는 하나의 인덱스만 가질 수 있고, 그 인덱스는 여러 개의 값을 가질 수 있다. 새롭게 사례가 구성될 때 사례를 구성하는 인덱스-값의 쌍들 중 어느 인덱스가 세부-사례 베이스에 저장되어 있는 특정 세부-사례의 인덱스와 동일하다면, 생성된 사례의 해당 인덱스의 값은 그 인덱스를 포함하는 세부-사례의 값으로 추가될 수 있다. 만약 신규 사례를 구성하는 인덱스-값의 쌍들 중에서 특정 인덱스가 세부-사례 베이스에 저장된 어떤 세부-사례에 포함된 인덱스와도 같은 값을 갖지 않으면 해당 인덱스-값을 쌍으로 하는 새로운 세부-사례를 생성한다. 이상의 과정을 거쳐 생성된 세부-사례들은 사용자 요구와 직간접으로 관련된 모든 정보의 조각들을 값들로 갖는다. 이 값들은 사용자 요구사항과 유사한 맞춤형 정보 생성에 활용된다.

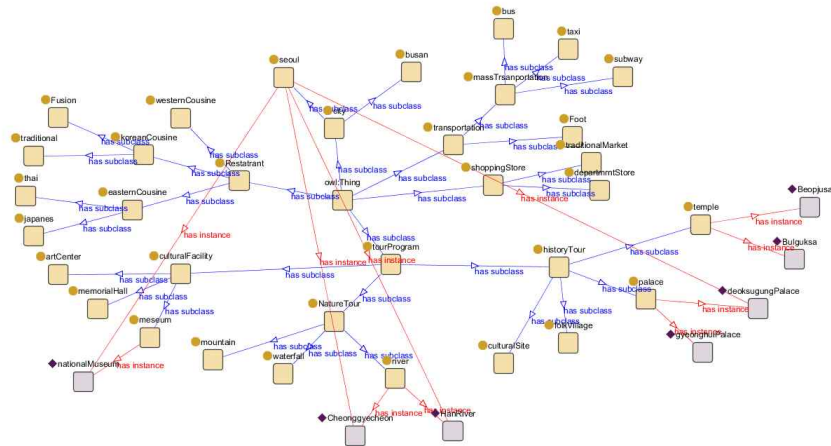
5. 온톨로지 기반 동적 유사도를 이용한 사례 추출

사용자의 요구사항들을 만족하는 사례들을 추출하기 위해서는 본 논문에서 생성한 사례 구조에 대한 이해가 선행되어야 한다. 앞서 서술한 바와 같이, 본 논문의 사례는 사용자 요구사항에 따라 수집된 웹 정보를 기반으로 생성된다. 이 사례는 웹 정보에 포함되어 있는 사용자 요구사항, 요구사항 조합 및 연관요구사항 들을 인덱스로 하며, 웹 정보를 구성하고 해당 인덱스를 포함하는 문장들을 그 인덱스의 값으로서 취하기 때문에, 사용자 요구사항들과 웹 정보의 내용에 따라 다양한 인덱스-값의 쌍들이 생성된다. 사례의 인덱스로는 요구사항, 요구사항 조합 및 연관요구사항 등을 사용하기 때문에, 사용자 요구사항과 비교 대상이 되는 사례들의 인덱스들 간에는 상이한 값이 존재할 수 있다. 따라서 온톨로지를 이용해 상이한 개념들 간의 유사도의 차이를 계산해 내는 것이 필요하다. 온톨로지에 포함된 클래스들간의 유사도 계산 방법으로 ‘Feature-theoretic approach’, ‘Edge-based approach’, ‘Information-theoretic model and fuzzy method’ 등이 제안된 바 있다(Slimani T., 2006; Song et al., 2007). 이들 방법은 비교 대상이 속한 컨텍스트들을 고려하지 않은 채, 단순히 두 클래스들의 관계만을 근거로 유사도를 계산하기 때문에 온톨로지의 구조에 변화가 없는 한 비교 대상인 두 클래스들이 항상 같은 값의 유사도를 갖는다. 그러나 비교 대상이 되는 클래스들은 그 클래스들이 포함되어 있는 컨텍스트에 따라 상이한 관계로 해석될 수 있으며, 결과적으로 클래스들 간의 유사도 역시 관계에 따라 달라질 수 있다. 예를 들어, 비교해야 할 클래스들(hanRiver, Beopjusa)이고, 이

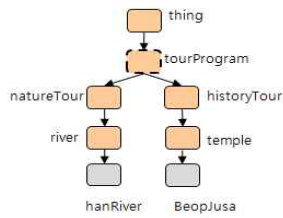
들이 공통적으로 포함하는 클래스가 ‘historyTour’와 ‘Seoul’인 경우를 살펴보자. <그림 3(b)>는 온톨로지 <그림 3(a)>에 존재하는 두 클래스들간의, 유사도를 계산하기 위해 ‘Edge-based method’가 사용하는 비대칭 그래프(acyclic graph)로서 항상 최상위 클래스를 상위 노드로 하고 비교 대상이 되는 두 클래스들을 하위 노드로 갖는다. 그러나 비교 대상인 두 클래스들이 공유하는 클래스를 기준으로 비대칭 그래프를 추출하면 <그림 3(c)>와 같이 상이한 구조의 비대칭 그래프가 도출됨을 알 수 있다. <그림 3(c)>로부터 알 수 있는 것은, 비교 대상인 두 클래스들이 어떤 클래스를 몇 개나 공유하느냐에 따라 연관되어 있는 클래스들도 다르고 클래스들의 개수도 달라질 수 있다는 것이다. 비교 대상인 두 클래스들이 공유하는 클래스들 역시 이들이 포함되어 있는 사용자 요구사항과 사례 인덱스들의 값에 따라 달라질 수 있으므로, 비교 대상인 두 클래스들이 공유하는 클래스들 정보를 컨텍스트로 이용하는 것은 충분한 가치가 있다.

이에 본 연구에서는 비교 대상 클래스들의 컨텍스트-즉 클래스가 포함되어 있는 문서의 내용이나 다른 클래스들과의 관계-에 따라 동적으로 변화되는 유사도 계산 방법인 동적 유사도를 제안하였다. 동적 유사도에서는 사용자 요구사항들과 사례들의 인덱스들을 비교한다. 정적 및 동적 유사도 과정은 다음과 같다.

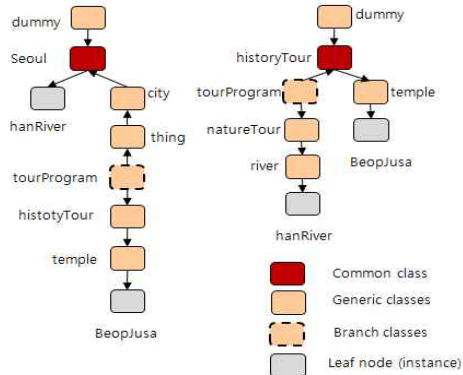
첫째, 사용자 요구사항과 사례베이스로부터 추출된 한 사례의 모든 인덱스 $f_j(j=1, 2, \dots, n)$ 를 비교한다. 사례의 인덱스와 정확히 일치하는 요구사항 만을 모아 구성된 집합을 구축한 후 이 집합이 가진 원소의 개수를 계산해 정적 유사도인 $sSim_{(R, C)}$ 값으로 취한다. R은 사용자 요구사항 및 C는 사례를 의미한다.



(a) illustrative ontology



(b) acyclic graphs for classical similarity calculation



(c) acyclic graphs for dynamic similarity calculation

<그림 3> 비대칭 그래프로 표현된 예

둘째, 사용자 요구사항과 정확히 일치하지 않는 사례의 인덱스, 즉 정적 유사도 계산에서 제외된 나머지 인덱스 $f_j (0 < j' < n)$ 을 대상으로 동적 유사도를 계산한다. 동적 유사도 계산을 위한 컨텍스트로는 전 단계에서 구축한 사례의 인덱스가 사용자 요구사항과 정확히 일치하는 요구사항 집합을

사용한다. 이를 유사도 검색에 반영하기 위해, 온톨로지로부터 정확히 일치하는 요구사항 집합의 원소를 임의의 더미 최상위 노드 (dummy top node)의 하위 노드로 하고 이 하위 노드에서 시작해 비교 대상인 두 클래스들에 이르는 노드들과 아크들로 구성된 비대칭 그래프를 도출하였다. <그림 3>

에서 보는 바와 같이, 비대칭 그래프는 정확히 일치하는 요구사항 집합에 포함된 원소의 수만큼 도출된다. 도출된 모든 비대칭 그래프를 대상으로 요구사항 r_i 와 이에 해당하는 인덱스 f_j 간의 유사도를 계산한 후, 이들의 산술 평균을 구해 요구사항 r_i 와 인덱스 f_j 의 유사도로 취한다. 요구사항 r_i 와 사용자 요구사항과 정확히 일치되지 않는 인덱스 집합에 포함된 인덱스들에 대해 위 계산을 수행한 후, 유사도가 가장 큰 값을 요구사항 r_i 와 인덱스 f_j 의 최종 유사도로 결정한다. 사용자 요구사항에 정확히 일치되는 인덱스 집합에 포함되지 않는 모든 요구사항에 대해 위 과정을 반복한 후 이를 합산하면 정확히 일치하는 요구사항 집합 (exactly matched set)의 원소가 아닌 사용자 요구사항들과 사례의 인덱스들을 대상으로 동적 유사도가 도출된다. 이를 수학적 모델로 표현하면 다음 식 (1)과 같다.

$$sim_{(r_{ij}^{EM}, n_{ik}^{AR})} = \frac{\sum_j (sim(r_{ij}^{EM}, n_{ik}^{AR}) | given r_{ij}^{EM})}{j} \quad (1)$$

where, $(sim_{(r_{ij}^{EM}, n_{ik}^{AR})} | given common node r_{ij}^{EM}) = \frac{N_3}{N1 + N2 + N3}$

식 (1)에서 N1과 N2는 최상위 클래스로부터 두 개의 비교 클래스들에 이르는 edge의 수이며, N3는 가지 클래스들로부터 공동 클래스에 이르는 edge의 수이다. r_{ij}^{EM} 는 j번째 사용자 요구사항이 인덱스와 정확히 일치하는 않을 경우를 의미하며, n_{ik}^{AR} 는 연관요구사항들로 구성된 인덱스에 연결된 가지노드의 수를 의미한다.

식 (1)을 온톨로지에 존재하는 모든 r_{ij}^{EM} 과 n_{ik}^{AR} 의 조합에 대해 수행한 후 합산해 $dsim_{(r, c)}$ 를 계산한다. 즉, $dsim_{(r, c)}$ 는 다음 식 (2)와 같다.

$$dsim_{(r, c)} = \sum_k sim_{(r_{ij}^{EM}, n_{ik}^{AR})} \quad (2)$$

식 (2)에서 r 은 사용자 요구사항, c_i 는 i 번째 사례가 포함하는 인덱스들을 나타낸다.

사용자 요구사항들과 사례가 포함하는 인덱스들과의 유사도는 정적 유사도인 $sSim_{(r, c)}$ 과 동적 유사도인 $dSim_{(r, c)}$ 를 합산함으로써 얻어진다. 즉 비교 대상인 사례들로부터 유사도 값이 가장 큰 사례를 선택해 맞춤화 작업을 한다.

6. 온톨로지 기반 집중 유사도를 이용한 사례 맞춤화

동적 유사도 계산을 통해 추출된 유사사례는 사용자 요구사항들과 정확히 일치되는 인덱스들과 일치하지 않는 인덱스들이 혼재되어 있다. 따라서 선택된 사례가 포함하지는 않지만 요구사항들의 용어의 값을 내재한 경우를 찾는 것이 집중 유사도 계산의 목적이다. 집중 유사도 계산 과정은 다음과 같다.

Step 1 : 사용자 요구사항들을 만족하는 해당 값을 찾기 위해 세부-사례 베이스에 저장되어 있는 세부-사례들 중 사용자 요구사항과 일치하는 값을 인덱스로 갖는 세부-사례와 사용자 요구사항을 포함하는 세부-사례를 모두 탐색한다.

Step 2 : 탐색된 세부-사례는 한 인덱스가 여러 개의 값들로 구성된 다중 값들로 구성된다. 이 중 사용자 요구사항에 가장 적합한 값을 찾는다. 이를 위해, 탐색된 세부-사례의 값들의 각각이 포함하고

있는 클래스들을 온톨로지로부터 찾은 후, 사용자 요구사항에서 클래스들에 이르는 edge의 수를 계산해 클래스들을 한 요구사항을 중심으로 N개의 층으로 구성된 동심원에 배치시킨다. 그림 4는 하나의 요구사항과 연관된 3개의 세부-사례들이 도출된 경우, 관련 클래스들을 하나의 요구사항을 중심으로 하는 동심원에 배치한 예이다. 이때 세부-사례는 하나의 값만을 가진 것으로 가정하였다.

Step 3 : 세부-사례의 값들이 포함하고 있는 클래스들과 해당 요구사항과의 유사도를 계산한다. 이때 세부-사례의 인덱스가 가진 값들 중 사용자 요구사항에 의미상 가장 인접한 값을 찾는 것이 중요하다. 즉, 하나의 요구사항과 다수 클래스들간 유사도 계산을 위해 집중도의 개념을 도입하였다. 집중도는 하나의 요구사항 값으로부터 추출된 클래스들이 클래스에 얼마나 집중 되어 있는지를

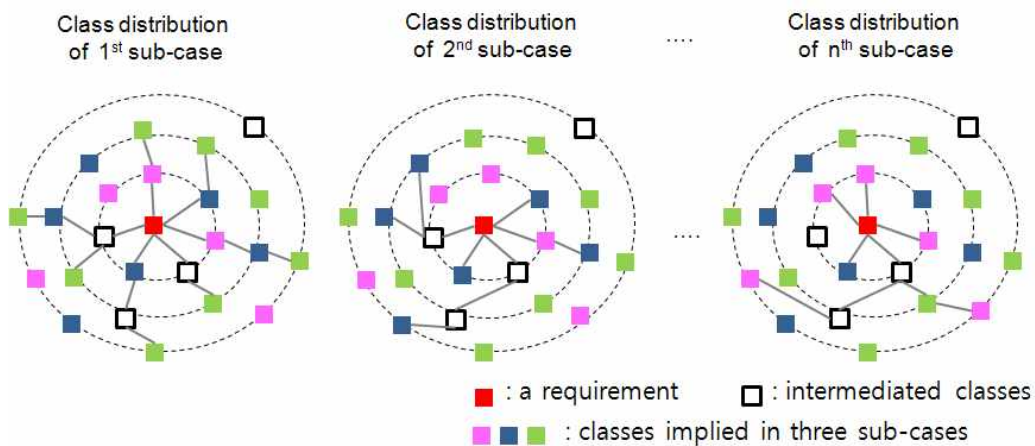
측정하는 것이다. 즉, 해당 값에 포함되어 있는 클래스들의 대다수가 매칭되는 요구사항에 집중되어 있다면 그렇지 않은 값보다는 그 요구사항을 만족할 가능성이 높다는 것이다. 이에 본 논문에서는 그 요구사항과 다중 클래스들의 값들과의 유사도를 계산하기 위해 하나의 요구사항을 중심으로 클래스 값들의 집중된 정도를 측정하는 직접 집중도 (Direct Intensity)를 사용하였다. 직접 집중도는 식 (3)과 같이 계산된다.

$$n(n_i) = n_i/N, N = \max(n_i), \text{ where } 0 \leq n(n_i) \leq 1, n_i \geq 1$$

$$con^d_{(r,v_i)} = \frac{\sum_i w_i \cdot n(n_i)}{\sum_i w_i} \quad (3)$$

식 (3)에서 i 는 웨도의 인덱스, w_i 는 i 번째 웨도의 가중치이며, n_i 는 i 번째 웨도 값, N 은 n_i 중 최대 값, v_i 는 세부-사례가 포함하는 i 번째 값이다.

또한 직접 집중도를 전체 요구사항 관점에서 보



<그림 4> 연관인덱스 정의를 위해 온톨로지로부터 도출된 경로

정하는 것이 필요하다. 이를 위해 세부-사례의 값이 요구사항 중 완전히 일치하는 요구사항을 포함한 경우를 제외한 세부-사례의 값들과 다른 요구사항들과의 간접 집중도 (Indirect intensity) $con^i_{(r', v_i)}$ 를 식 (4)와 같이 계산했다.

$$con^i_{(r', v_i)} = \sum_{k, r \neq r_k} con^i_{(r_k, v_i)} = \sum_{k, r \neq r_k} \left(\frac{\sum_j w_{jk} \cdot n(n_{jk})}{\sum_j w_{jk}} \right) / \sum_{k=1}^m k \quad (4)$$

식 (4) 에서 r 은 사용자 요구사항, v_i 는 해당 인덱스가 포함하는 i 번째 값, m 은 사용자 요구사항의 개수이며, n_j 는 k 번째 요구사항이 속한 계도의 인덱스, w_j 는 이에 대한 가중치이다.

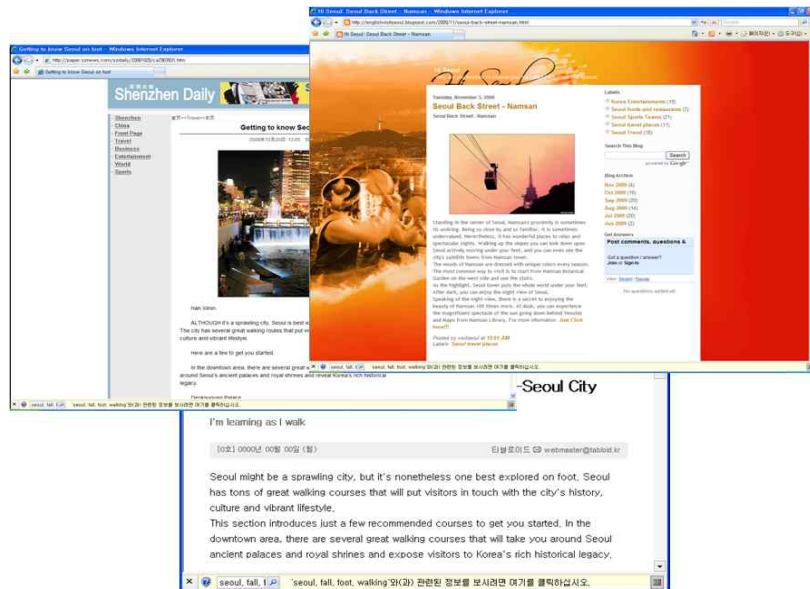
직접 집중도를 간접집중도로 보정한 집중 유사도 'intSim'은 식 (5)와 같다.

$$intSim = (w^d \cdot con^d_{(r, v_i)} + w^i \cdot con^i_{(r, v_i)}) / (con^d_{(r, v_i)} + con^i_{(r, v_i)}) \quad (5)$$

식 (5)에서 w^d 는 직접 집중도, w^i 는 간접 집중도의 전체 보정을 위한 가중치이다. 각각의 가중치 w^d 와 w^i 의 합은 1이다.

Step 4 : 탐색된 모든 세부-사례들에 대해 step 2와 step 3의 과정을 만족한 후, 집중 유사도의 값이 가장 큰 세부-사례의 값을 요구사항을 만족하는 값으로 취한다.

이상의 과정을 거치면 사용자 요구사항들과 동일한 인덱스들과 그에 해당하는 값들로 구성된 맞춤형 사례가 생성된다. 맞춤형 사례는 사례 베이스에 저장됨과 동시에 사용자에게 전송된다. 통합지식은 CSS나 XSLT를 적용해 사용자에게 전달된다.



<그림 5> 여행자의 요구사항(Seoul, Fall, Foot, Walking)에 의해 탐색된 웹 정보

7. 예제

다음은 앞서 서술한 바와 같이 여행자들의 요구 사항에 따라 본 프레임워크를 통해 맞춤형정보를 생성하는 과정을 예시하였다.

7.1 Step 1 : 웹 정보 수집

RSS 리더를 통해 수집된 웹 정보 중 여행자의 요구사항 (Seoul, Fall, Foot, Walking)과 관련된 정보가 <그림 5>와 같이 3개가 검색되었고, 수집된 HTML 문서들을 웹 정보 저장소에 저장한다.

7.2 Step 2 : 사례 생성

웹 정보 저장소에 저장된 각각의 HTML 형식의 웹 문서들을 사례로 변환한다. 이때 생성되는 사례의 인덱스 집합은 각각의 문장이 포함하는 사용자 요구사항인 'Seoul', 'Fall', 'Walking' 또는 'Foot', 이들의 조합으로 생성된 요구사항과 이들과 의미적으로 연관된 연관요구사항들로 구성된다. 인덱스와 쌍을 이루는 값은 웹 정보가 내포하고 있는 문장들이 해당 인덱스를 포함하고 있는 경우에 해당된다. 수집된 3개의 HTML 문서와 여행자 요구사항들 'Seoul', 'Fall', 'Walking' 및 'Foot'을

```
// Get to know Seoul on foot
{
  ((is-a)(case))
  ((Case-id)(1))
  ((url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm))
  ((Seoul)((Call it Seoul by sole) (In the downtown area, there are several great walking courses that will take you around Seoul's ancient palaces and royal shrines and reveal Korea's rich historical legacy)))
  ((fall)((It was historically dry through the fall and winter but would run high during the summer rainy season)))
  ((Seoul, foot) (ALTHOUGH it's a sprawling city, Seoul is best explored by foot))
  ((Seoul, palace, Doeksoogung)(This course begins at Deoksugung-Deoksugung Palace, a walled compound that contains some of the most historically significant palaces in Seoul) (Each day in front of Deoksugung Palace, the ancient changing-of-the-guard ceremony is reenacted for tourists-At 10:30 a.m., 2 p.m. and 3 p.m., except on Mondays))
}

// Seoul's back streets:
{
  ((is-a)(case))
  ((Case-id)( 2))
  ((url)(http://www.koreaherald.co.kr/NEWKHSITE/data/html_dir/2009/10/30/200910300018.asp))
  ((Seoul)(Standing in the center of Seoul, Namsan's proximity is sometimes its undoing))
  ((Seoul, walking)(Walking up the slopes you can look down upon Seoul actively moving under your feet, and you can even see the city's satellite towns from Namsan tower))
  ((Seoul, Foot) (As the highlight, Seoul tower puts the whole world under your feet))
  ((Seoul, Mountain, Namsan)(The route proceeds past an authentic old section of the Namsan Fortress Wall and Korea's prominent National Theater and then on up the mountain through a lovely dense patch of forest.)) }

// Seoul City Walking Tour
{
  ((is-a)(case))
  ((Case-id)(3))
  ((url)(http://www.tabloid.kr/news/quickViewArticleView.html?idxno = 140))
  ((Seoul)(Seoul might be a sprawling city, but it's nonetheless one best explored on foot))
  ((Seoul, Walking)(Seoul has tons of great walking courses that will put visitors in touch with the city's history, culture and vibrant lifestyle))
  ((walking)(In the downtown area, there are several great walking courses that will take you around Seoul ancient palaces and royal shrines and expose visitors to Korea's rich historical legacy))
  ((Seoul)(Seoul might be a sprawling city, but it's nonetheless one best explored on foot)) }
```

<그림 6> 인덱스와 해당 값들로 구성된 사례들

```
//Sub-Case
{
((is-a)(sub-case))
((Case-id)(Seoul))
(((url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm))(Call it Seoul by sole))
(((url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm))(In the downtown area, there are several great walking courses that will take you around Seoul's ancient palaces and royal shrines and reveal Korea's rich historical legacy))
(((url)(http://www.koreaherald.co.kr/NEWKHSITE/data/html_dir/2009/10/30/200910300018.asp))(Standing in the center of Seoul, Namsan's proximity is sometimes its undoing))) }

{
((is-a)(sub-case))
((Case-id)(fall))
(((url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm))(It was historically dry through the fall and winter but would run high during the summer rainy season)) }

{
((is-a)(sub-case))
((Case-id)( walking))
(((url)(http://www.tabloid.kr/news/quickViewArticleView.html?idxno=140))(In the downtown area, there are several great walking courses that will take you around Seoul ancient palaces and royal shrines and expose visitors to Korea's rich historical legacy)) }

{
((is-a)(sub-case))
((Case-id)(Seoul, Foot))
(((url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm))(ALTHOUGH it's a sprawling city, Seoul is best explored by foot) (As the highlight, Seoul tower puts the whole world under your feet))
(((url)(http://www.koreaherald.co.kr/NEWKHSITE/data/html_dir/2009/10/30/200910300018.asp))(Seoul might be a sprawling city, but it's nonetheless one best explored on foot)) }

{
((is-a)(sub-case))
((Case-id)(Seoul, Walking))
(((url)(http://www.tabloid.kr/news/quickViewArticleView.html?idxno = 140))(Seoul has tons of great walking courses that will put visitors in touch with the city's history, culture and vibrant lifestyle))
(((url)(http://www.koreaherald.co.kr/NEWKHSITE/data/html_dir/2009/10/30/200910300018.asp))(Walking up the slopes you can look down upon Seoul actively moving under your feet, and you can even see the city's satellite towns from Namsan tower)) }

{
((is-a)(sub-case))
((Case-id)(Seoul-palace-Deoksugung_Palace))
(((url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm))(This course begins at Deoksugung-Deoksugung Palace, a walled compound that contains some of the most historically significant palaces in Seoul) (Each day in front of Deoksugung Palace, the ancient changing-of-the-guard ceremony is reenacted for tourists-At 10:30 a.m., 2 p.m. and 3 p.m., except on Mondays))
(((url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm))(Close to Deoksugung Palace is Jungmyeongjeon, the site where the Eulsa Treaty between Korea and Japan was signed)) }

{
((is-a)(sub-case))
((Case-id)(Seoul, Mountain, Namsan))
(((url)(http://www.koreaherald.co.kr/NEWKHSITE/data/html_dir/2009/10/30/200910300018.asp))(The route proceeds past an authentic old section of the Namsan Fortress Wall and Korea's prominent National Theater and then on up the mountain through a lovely dense patch of forest.)) }

```

<그림 7> 사례들로부터 생성된 세부-사례들

이용해 생성한 사례들을 <그림 6>에 도시하였다. <그림 6>의 ‘사례 1’을 구성하는 인덱스들 중 ‘Seoul’과 ‘Fall’은 요구사항으로부터 생성된 인덱스들이고 (Seoul, Foot)는 요구사항조합이며 (Seoul, palace, Deoksugung)은 온톨로지를 이용해 식별한 연관요구사항들로부터 생성된 인덱스들이다. 생성된 사례를 사례베이스에 저장한다.

7.3 Step 4 : 세부-사례 생성

모든 사례들은 인덱스-값 쌍을 기준으로 분할될 수 있다. 인덱스를 기준으로 사례들로부터 동일 인덱스를 갖는 값들을 모아 인덱스-값들로 구성된 하나의 세부-사례를 생성한다. <그림 6>에 도시된 3개의 사례들로부터 생성된 세부-사례를 다음 그림 7에서와 같이 예시하였다. 생성된 세부-사례들은 세부-사례 베이스에 저장된다.

7.4 Step 5 : 정적 유사도

웹 정보들로부터 사례를 생성하였으며 맞춤정보생성을 세부-사례의 구성이 완료되었다. 지금부터는 사용자 요구사항을 보다 만족하는 사례의 추출을 위해 사례가 포함하는 인덱스와 요구사항 간의 동적 유사도를 계산한다. ‘사례 1’, ‘사례 2’, 및 ‘사례 3’이 포함하고 있는 인덱스들과 여행자 요구사항들을 비교함으로써 세 개의 정적유사도 $sSim_{(r,c_i)}$ 값을 구한다.

사례 1 : $sSim_{(r,c_1)} = 3(\text{seoul, fall, foot})$

사례 2 : $sSim_{(r,c_2)} = 3(\text{Seoul, foot, walking})$

사례 3 : $sSim_{(r,c_3)} = 2(\text{Seoul, walking})$

즉, 정적유사도 $sSim_{(r,c_i)}$ 은 각 사례가 포함하는

인덱스와 사용자 요구사항이 정확히 일치하는 지를 확인하는 유사도 검색이다. ‘사례 1(C₁)’과 ‘사례 2(C₂)’의 경우 3개의 인덱스와 사용자 요구사항이 일치하였으며, ‘사례 3(C₃)’ 경우는 2개의 인덱스가 사용자 요구사항과 일치하였다.

7.5 Step 6 : 동적 유사도

사례가 포함하는 인덱스들 중 연관 요구사항들에 대한 인덱스로 갖는 ‘사례 1’과 ‘사례 2’를 대상으로 동적 유사도 $dSim_{(r,c_i)}$ 를 계산한다.

$$dSim_{(r,c_1)} = \left(\frac{(\text{sim}(a,b)|\text{common node Seoul}) + \text{sim}(a,b)}{|\text{common node fall}| + (\text{sim}(a,b)|\text{common node foot})} \right) = 0.22$$

위 식은 ‘사례 1’에 대한 동적유사도이다. ‘사례 1’에서 ‘a’는 ‘Deoksugung’, ‘b’는 ‘walking’을 나타낸다. 공통 노드 ‘Seoul’, ‘fall’, ‘foot’으로부터 연관 요구사항들과의 유사도를 계산한 값이다.

$$dSim_{(r,c_2)} = \left(\frac{(\text{sim}(c,d)|\text{common node Seoul}) + \text{sim}(c,d)}{|\text{common node walking}| + (\text{sim}(c,d)|\text{common node foot})} \right) = 0.17$$

위 식은 ‘사례 2’에 대한 동적유사도 값이다. ‘사례 2’에서 ‘c’는 ‘Namsan’ 및 ‘d’는 ‘Fall’을 나타낸다. 공통노드 ‘Seoul’, ‘walking’ 및 ‘Namsan’과의 연관요구사항들 간의 유사도를 계산한 값이다.

$sSim_{(r,c_i)}$ 의 3과 $dSim_{(r,c_i)}$ 의 0.22를 합산하면 ‘사례 1’은 3.22이고 ‘사례 2’는 3.17, ‘사례 3’은 2값을 가지므로 ‘사례 1’이 가장 유사한 사례로 선택된다.

7.6 Step 7 : 맞춤 통합 정보 생성을 위한 집중 유사도

위 결과로부터 '사례 1'이 여행자 요구사항에 따른 가장 유사한 사례로 '사례 1'의 맞춤화를 통해 여행자 요구사항들 'Seoul', 'Fall', 'Walking', 및 'Foot'을 만족하는 통합 지식을 생성한다. 여행자 요구사항 관점에서 '사례 1'을 살펴 보았을 때, '사례 1'이 여행자 요구사항이 필요로 하는 'Walking,' 및 'Foot'과 정확히 일치하는 정보를 포함하고 있지 않음을 알 수 있다. '사례 1'이 포함하지 않는 요구사항인 'Walking'을 포함한 정보를 찾기 위해 세부-사례 베이스를 검색한다. 세부-사례 베이스로부터 'Walking'을 인덱스로 갖는 2개의 세부-사례를 추출한 후, 각 세부-사례가 가지고 있는 값들 중에서 여행자의 요구사항과 가장 유사한 값을 찾아내기 위해 집중 유사도를 계산한다.

'사례 1'에서 'Walking'을 인덱스로 갖는 세부-사례의 값에 포함되어 있는 온톨로지의 클래스는 'Seoul'이고, 'Seoul'은 'Walking'을 중심으로 그린

동심원에서 4번째 궤도에 위치한다. 식 (3)으로부터 $con^d_{(walking, v_{walking})} = 1$ 이고, 'Seoul'과 다른 요구사항인 'Seoul', 'Fall', 및 'Foot'들과의 간접 집중도는 0.9다. 직접 집중도와 간접 집중도에 대한 가중치에 따라 최종적인 집중유사도 'intSim'을 계산한 후, 'intSim'이 가장 큰 값을 세부-사례로부터 추출해 인덱스 'Walking'의 값으로 선택한다. 본 예제에서는 세부-사례의 값이 하나이므로 해당 값을 밖에 'Walking'의 값으로 선택하였다. 위 과정을 거쳐 생성된 여행자의 요구사항에 따른 맞춤 정보는 <그림 8>과 같다. 아래 사례는 재사용을 위해 사례 베이스에 저장된다.

8. 결론

인지 능력의 한계와 시간적인 제약을 가진 사용자가 정보자원으로부터 관련된 모든 정보를 수집한 후 자신이 필요로 하는 정보의 조각들을 이용하여 사용자 요구사항에 적합한 맞춤 지식을 생성하는 것이 쉬운 일은 아니다. 이러한 문제점을 해결하기 위한 도구로 온톨로지가 널리 활용되고는

```
// Customized knowledge for traveler
{
  ((is-a)(case)
  ((Case-id)(4))
  ((url)(http://www.korea.pe.kr))
  ((Seoul)(Standing in the center of Seoul, Namsan's proximity is sometimes its undoing))
  ((Seoul, walking)(Walking up the slopes you can look down upon Seoul actively moving under your feet, and you can even see the city's satellite towns from Namsan tower))
  ((Seoul, Foot)(As the highlight, Seoul tower puts the whole world under your feet))
  ((Seoul, Mountain, Namsan)(The route proceeds past an authentic old section of the Namsan Fortress Wall and Korea's prominent National Theater and then on up the mountain through a lovely dense patch of forest.))
  ((walking)(In the downtown area, there are several great walking courses that will take you around Seoul ancient palaces and royal shrines and expose visitors to Korea's rich historical legacy)) }

```

<그림 8> 여행자 요구사항에 따라 생성된 맞춤 정보

있으나, 해당 용어들 간의 관계만을 모델링해주는 온톨로지 만으로는 사용자들이 기대하는 수준의 맞춤 정보의 생성이 어렵다. 이에 본 연구에서는 사례와 온톨로지를 활용한 맞춤형 통합정보 생성 프레임워크를 제안하였다. 웹 정보를 효과적으로 통합 맞춤화하여 사용자 요구 충족도가 높은 정보를 생성하기 위해 동적 및 집중 유사도를 적용했으며 결과적으로 사용자들이 원하는 수준의 통합 맞춤 정보의 생성에 성공하였다.

참고문헌

- Bernstein, A. P. and L. Haas, "Information Integration in the Enterprise", *Communications of the ACM*, vol.51, No.9(2008), 72~79.
- Alexiev, V., M. Breu, J. de Bruijn, D. Fensel, R. Lara, and H. Lausen, "*Information Integration with Ontologies : Experiences from an Industrial Showcase*", Wiley, 2005.
- Halevy, Y. A. D. Bittonz, M. Carey, D. Draper, J. Pollock, A. Rosenthal, and V. Sikka, "Enterprise information integration : successes, challenges and controversies", *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, 2005, 778~787.
- Philippe, M. and P. W. Eklund, "Knowledge Retrieval and the World Wide Web", *IEEE Intelligent Systems*, 2000(May/June), 18~25.
- Mani, I. and E. Bloedorn, "Machine learning of Generic and User-Focused Summarization", *In proceedings of the National Conference on Artificial Intelligence(AAAI)*, 1998, 820~826.
- Teufel, S. and K. Moens, "Sentence extraction as a classification task", *Workshop of Intelligent and scalable text summarization, ACL/EACL*, 1997, 58~65.
- Slimani T., B. B. Yaghlane, and K. Mellouli, "A New Similarity Measure based on Edge Counting", *World Academy of Science, Engineering and Technology* Vol.23(2006), 34~38.
- Song, L., J. Ma, H. Liu, L. Lian, and D. Zhang, "Fuzzy Semantic Similarity Between Ontological Concepts", *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, 2007, 275~280.

Abstract

Framework for Information Integration and Customization Using Ontology and Case-based Reasoning

Hyun Jung Lee* · Mye Sohn**

The requirements of knowledge customization have increased as information resources have become more various and the numbers of the resources are increased. Even if the method for collecting the information has improved like Really Simple Syndication (RSS), information users are still struggling for extracting and customizing the required information through the Web. To reduce the burden, we offer the dynamic knowledge customization framework by using ontology-based CBR. The framework consisting of three phases is comprised of the conversion phase of web information as a machine-accessible case, the extraction phase to find a case appropriate for information users' requirements, and the case customization phase to create knowledge depending on information user's requirements. Newly, the dynamic and intensity-based similarity is adopted to support timely dynamic change of users' requirements. The framework has adopted to create traveler's knowledge to the level users wanted.

Key Words : Case-based Reasoning, Dynamic Knowledge Management, Ontology, Dynamic Similarity, Intensity-based similarity

* Korea University Business School

** Department of System management Engineering, Sungkyunkwan University

저자 소개



이현정

이화여자대학교 물리학사 및 전자계산학과 석사를 받았으며, 한국과학기술원(KAIST) 테크노경영대학원 경영공학 박사학위를 취득하였다. 현재 고려대학교 BK21 연구교수로 재직 중이다. 주요 연구분야는 지능정보시스템, 의사결정지원시스템, 유비쿼터스 컴퓨팅 및 시멘틱 웹 서비스 등이다.



손미애

성균관대학교 산업공학과를 졸업하고 한국과학기술원에서 경영정보공학으로 박사 학위를 취득한 후, 한국국방연구원에 근무했으며, 현재 성균관대학교 시스템경영공학과 부교수로 재직중이다. 주요 관심분야는 시멘틱 웹, 시멘틱 웹 서비스, BPMS 및 Enterprise Architecture/Architecture Framework 등이며, 현재 온톨로지상에서의 유사도 계산과 SBA를 위한 EA 구축연구를 수행하고 있다. 한국 경영공학회와 ITA/EA학회의 임원으로 활동 중이며, CACM을 비롯한 국내외 우수 저널에 논문을 발표하고 있다.