

# 상황과 정보 집적도를 고려한 유사도 기반의 맞춤형 지식 생성프레임워크<sup>☆</sup>

## Customized Knowledge Creation Framework using Context- and intensity-based Similarity

손 미 애\*                      이 현 정\*\*  
Mye M. Sohn                  Hyun Jung Lee

### 요 약

정보의 출처와 형식이 다양해지고 정보의 양 또한 많아짐에 따라 소셜 웹에서의 맞춤형 지식 생성은 더욱 어려워지고 있다. RSS(Really Simple Syndication)가 정보 수집 방법의 개선에 일조했으나, 웹에 산재된 정보를 찾아 필요한 정보들만으로 구성된 맞춤형 지식을 생성하는 것은 여전히 사용자들의 몫으로 남아 있다. 본 논문에서는 맞춤형 지식 생성의 용이성을 제고하기 위해 상황 기반 유사도를 이용한 맞춤형 지식생성 프레임워크를 제안하였다. 본 프레임워크는 기본적으로 사례 기반 추론의 절차를 따르지만, 기존 사례 기반의 유사도 계산 방식이 문법적 추론에 기반했던 것과 달리, 온톨로지를 활용한 의미적 유사도를 이용한 사례 기반 추론을 활용한다. 또한 사용자 요구를 만족하는 유사사례의 보정을 위해 온톨로지를 활용한 정보 집적도 기반의 유사도 방법론을 제안하였다. 본 프레임워크에서는 첫째 비구조적인 웹 정보를 사례 형태의 구조적 정보로 변환하고, 둘째 사용자의 요구에 적합한 의미론적 유사사례를 찾은 후 셋째, 선택된 유사사례의 정보 집적도를 고려한 보정을 통해 맞춤형 지식을 생성하는 과정을 거친다. 본 논문에서는 유사도 계산에 일반적으로 활용되는 여러 방법론들과 비교를 통하여 제안한 온톨로지 기반 의미적 유사도 계산 방법론의 타당성을 입증하였다.

### ABSTRACT

As information resources have become more various and the number of the resources has increased, knowledge customization on the social web has been becoming more difficult. To reduce the burden, we offer a framework for context-based similarity calculation for knowledge customization using ontology on the CBR. Thereby, we newly developed context- and intensity-based similarity calculation methods which are applied to extraction of the most similar case considered semantic similarity and syntactic, and effective creation of the user-tailored knowledge using the selected case. The process is comprised of conversion of unstructured web information into cases, extraction of an appropriate case according to the user requirements, and customization of the knowledge using the selected case. In the experimental section, the effectiveness of the developed similarity methods are compared with other edge-counting similarity methods using two classes which are compared with each other. It shows that our framework leads higher similarity values for conceptually close classes compared with other methods.

☞ keyword : 맞춤형 지식생성(Knowledge Customization), 소셜 웹(Social Web), 상황기반 유사도(Context-based Similarity), 사례기반추론(Case-based reasoning), 온톨로지(Ontology)

\* 중신회원 : 성균관대학교시스템경영공학과 부교수  
myesohn@skku.edu

\*\* 정 회 원 : 서강대학교 연구교수(교신저자)  
hjlee5249@yahoo.com

[2011/04/15 투고 - 2011/04/27 심사(2011/07/05 2차) - 2011/08/08 심사완료]

☆ A preliminary version of this paper appeared in ICONI/APIC-IST 2010, Dec 16-20, Mactan Island, Philippines. This version is improved considerably from the previous version

## 1. 서 론

참여, 공유 및 분산을 핵심가치로 갖는 소셜 웹은 블로그, 위키 및 소셜 네트워크 등과 같은 서비스를 제공함으로써 정보제공자 또는 생산자와 정보사용자의 벽을 허물고 있으며, 이로 인해 정보원

by including new results and features.

과 이들이 포함하고 있는 정보의 양이 폭발적으로 증가되고 있다. 소셜 웹에 게시된 정보들의 대부분은 정보를 정의하는 수단 중의 하나인 태그 집합을 포함하며, 이들 태그집합은 정확한 검색에 활용된다. 그러나 정보 사용자가 태깅 등의 방법으로 소셜 웹에 게시된 유관 정보를 수집한 후, 자신이 필요로 하는 정보들로 재구성된 맞춤형 정보를 생성하는 것은 매우 어려운 일이다. 이러한 어려움을 완화시킬 수 있는 방법 중의 하나가 지능형 에이전트를 활용해 정보의 검색, 수집, 정제 및 맞춤화의 과정을 대행하게 하는 것이다. 그러나 웹 정보들이 대부분 비구조화된 형식의 텍스트 문서이기 때문에 지능형 에이전트를 활용하기 위해서는 웹 문서를 구조적으로 변환하는 것이 요구된다. 태깅 방법에 더해 비구조화된 형식의 텍스트 정보로부터 맞춤형 정보를 생성할 수 있는 방법인 RDF(Resource Description Framework)를 이용해 웹 정보에 주석을 달거나 온톨로지를 활용하는 방법이 제안된 바 있으나, 이들 방법만으로 사용자의 요구를 의미적으로 만족시킬 수 있는 정보를 생성하는 데 한계가 있다.

본 논문에서는 소셜 웹 환경에서 사용자 요구에 부합하는 맞춤형 정보를 생성해 주는 프레임워크를 제안하고자 한다. 본 프레임워크 설계의 주안점은, 수집된 이질적인 웹 정보를 구조적으로 가공하고 이를 이용해 사용자의 요구사항에 맞게 의미적으로 맞춤형 지식을 생성하는 것이다. 맞춤형 지식 생성을 위해 활용될 소셜 웹 정보는 RSS(Really Simple Syndication)를 통해 자동적으로 수집되는 것으로 가정하였다. 본 프레임워크에서는 맞춤형 지식의 효과적 생성을 위해 수집된 정보를 구조화된 사례로 변환하였으며, 사용자의 요구에 의미론적으로 부합하는 유사 사례를 찾기 위한 유사도 계산방법과 찾아진 유사사례가 사용자의 요구사항에 맞게 맞춤형 지식으로 변환하기 위한 정보 집적도 기반 유사도 계산방법을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서는 정보의 통합과 관련된 연구들의 한계를 살펴보고, 3

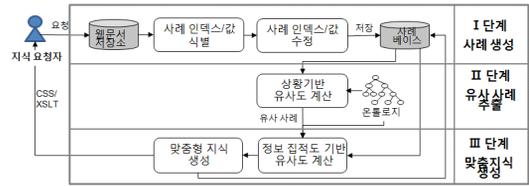
장에서는 본 논문에서 제안하는 맞춤형 지식 생성 프레임워크의 구조를 소개하였다. 4장에서는 문제 해결절차를 설명하기 위해 사례구조, 유사사례추출 및 맞춤형 지식 생성에 대해 상술하였다. 이와 관련하여 상황과 정보 집적도를 고려한 온톨로지 상에서의 유사도 계산 방법을 제안하였다. 또한 제안한 프레임워크에 대한 적용사례는 5장에 제시되었으며, 6장에서는 본 프레임워크의 우수성을 입증하기 위해 실험을 수행하여 그 결과를 도시하였다. 마지막으로 7장에서는 본 연구의 결론과 추후 연구 사항에 대해 약술하였다.

## 2. 관련 연구

정보 통합은 다수의 정보원으로부터 수집된 이질적인 내용과 형식을 가지고 있는 웹 정보를 의미적 구조적으로 모순 없는 하나의 통합된 정보로 가공하는 것으로 정의된다. 통합의 대상이 되는 웹 정보에는 XML로 표현된 구조적인 형식의 문서와 HTML로 표현된 비구조적인 문서가 있다. 구조적인 문서의 통합은 상이한 스키마를 갖는 XML 문서를 공통의 고유스키마로 변환하는 것을 의미하며, 이들을 통합하는 ETL(Extraction, Transform, Load) 기법이나 사전에 정의된 중립 스키마로의 변환 방법이 널리 활용된다[1, 2]. 이에 반해 비구조적 형식의 정보를 자동적으로 통합하기 위해서는 높은 수준의 자연어 처리기술 등이 필요하나, 현재의 자연어 처리 수준으로는 제한적인 자동화만이 가능하다. 따라서 현재는 사용자들이 검색된 웹 정보를 정제하고 통합하는 실정이다. 검색의 질을 높이기 위해 웹 정보에 포함되어 있는 메타데이터를 검색하거나 웹 문서가 포함하고 있는 하이퍼링크(hyperlink)를 고려하기도 하지만 정보의 통합을 통한 새로운 지식의 창출과는 거리가 있다. 최근에는 RDF를 이용해 웹 정보들 간의 의미적 관계를 모델링하기도 하지만, 온톨로지를 이용해 이질적으로 표현된 정보의 의미적 관계를 파악하여 정보를 통합하는 방법이 널리 활용되고 있다[3]. 온톨로지를

이용해 두 개념간의 의미적 유사도를 계산하는 대표적인 방법론에는 비교 대상인 두 개념 사이에 존재하는 링크의 수를 고려한 방법(edge counting), 특정 개념들의 출현 빈도를 고려한 방법 (information content), 출현빈도에 퍼지 개념을 접목한 방법 (fuzzy method) 및 이들을 조합한 하이브리드 방법 등이 있다[7, 8]. 두 개념 사이에 존재하는 링크의 수를 고려한 방법의 경우 온톨로지 상에서 비교 대상인 두 개념 사이의 거리를 계산하는 것으로서, 두 개념간의 최단거리를 도출해 유사도를 비교한다[9]. 이 방법은 계산이 용이하고 개념들 간의 유사도를 적절히 도출해 주는 장점으로 인해 널리 활용되고 있다[10]. 그러나 이 방법은 두 개념 간의 단순 거리만을 이용하여 유사도를 계산하기 때문에 비교 대상인 두 개념의 유사도는 항상 동일한 값을 갖는 한계가 있다. 즉, 비교 대상인 두 개념의 유사도가 이들이 처해 있는 상황, 예를 들어 온톨로지 상에서 비교대상의 개념이 서로 공유하는 개념의 정도 등에 따라 달라질 수 있음에도 불구하고 기존 방법은 이를 고려할 수 없다. 개념들 간의 단순거리를 계산하는 것과는 다르게 특정 문서에서 해당 개념들이 출현한 빈도를 고려해 유사도를 계산하는 방법이 제안되었으나, 이 경우 확률 값을 계산할 수 있는 어휘집이 존재해야만 한다는 전제 조건이 있기 때문에 사례기반추론에서의 유사도 계산에는 적용이 어렵다. 이에 본 논문에서는 사용자 요구사항과 사례 간의 유사도 계산을 위해 상황기반의 유사도 계산방법을 제안하였다. 상황기반의 유사도는 두 개의 비교 대상이 공유하는 개념이 무엇인지를 고려해 도출한다. 즉, 비교 대상인 두 개념의 유사도는 이들이 어떤 개념을 공유하고 있는지에 따라 달라지게 된다.

그러나 온톨로지를 활용할 수 있다고 하더라도 온톨로지가 추론해야 할 정보의 양이 지나치게 많은 현재의 웹 환경에서는 이를 보완할 수 있는 새로운 통합 방법에 대한 연구가 요구된다. RSS를 이용함으로써 처리해야 할 정보의 양은 일부 줄일 수 있으나, 이를 통해 전달되는 정보 역시 지능형 에



(그림 1) 맞춤형 지식생성 프레임워크

이전트가 처리할 수 없는 비구조적인 형식의 텍스트 문서가 대부분이기 때문에[4], 사용자는 정보의 정제를 통해 자신에게 필요한 맞춤형 정보의 생성에 여전히 많은 시간과 노력을 투입하는 실정이다 [5, 6].

이상의 논의를 종합해 볼 때, 웹에 산재한 정보들을 정제하고 통합함으로써 사용자가 즉시 활용할 수 있는 맞춤형 지식을 생성할 수 있는 새로운 프레임워크가 필요함을 알 수 있다. 이에 본 논문에서는 온톨로지와 사례기반추론 방법을 이용해 비구조적인 웹 정보들로부터 맞춤형 지식을 생성할 수 있는 프레임워크를 제안하고자 한다.

### 3. 맞춤형 지식 생성 프레임워크

소셜 웹으로부터 수집된 다양한 정보를 이용해 맞춤형 지식을 자동적으로 생성하기 위한 조건은 다음과 같다. 첫째, 지능형 에이전트가 통합의 대상이 되는 웹 정보를 이해하고 처리할 수 있어야 한다는 것이다. 둘째, 지능형 에이전트는 이 정보로부터 사용자가 필요로 하는 정보만을 자동적으로 추출할 수 있어야 한다. 셋째, 추출된 정보로부터 사용자의 요구를 의미적으로 만족시키는 동시에 구조적으로도 통합된 맞춤형 지식을 생성할 수 있어야 한다. 이러한 조건을 만족시키기 위해 제안되는 프레임워크는 다음과 같다.

(그림 1)은 비구조적인 웹 문서를 사례로 변환한 후 유사사례를 찾고 최종적으로 맞춤형 지식을 생성하는 프레임워크의 흐름도이다. 즉, 웹으로부터 추출된 문서로부터 사례를 생성하고, 생성된 사례와 사례베이스에 저장되어 있는 과거사례들을 대

상으로 상황기반 유사도를 적용해 유사사례를 발견한다. 발견한 유사사례가 사용자의 요구사항에 부합되지 않는 경우, 사례베이스에 저장되어 있는 과거 사례를 이용해 정보 집적도 기반 유사도계산을 수행함으로써 맞춤형 지식 생성을 한다.

**사례 생성:** 수집된 비구조적인 웹 정보들을 지능형 에이전트가 이해할 수 있는 구조적 형식을 갖는 사례로 변환한다. 변환의 대상이 일반화가 어려운 소셜 웹 정보라는 점을 감안해 규칙 형식의 정형화된 지식표현이 아닌 보다 포괄적 정보를 포함할 수 있는 사례형식의 구조적인 변환을 시도했다.

**유사사례 추출:** 사용자의 요구와 가장 유사한 정보를 내포하고 있는 유사사례를 찾는 단계이다. 유사사례를 추출하기 위해 사용자 요구사항과 사례간의 의미적 유사도를 계산하는 상황기반 유사도 계산방법(context-based similarity calculation method)을 제안하였다.

**맞춤형 지식 생성:** 전 단계에서 식별한 유사 사례가 사용자의 요구사항을 모두 만족시키지 못할 경우에는 유사사례에 대한 보정이 필요하다. 이를 위해 사용자의 요구사항에는 있으나 유사사례에는 포함되어 있지 않는 사용자의 요구사항이 있는지를 확인한다. 만약 이러한 요구사항이 있다면, 이 요구사항을 인덱스로 갖는 사례들을 추출한 후 이 들로부터 해당 인덱스와 값들을 추출한다. 이 과정이 종료되면 하나의 인덱스와 이 인덱스가 가질 수 있는 복수 개의 값이 생성된다. 이 값들 중에서 요구사항에 부합하는 최적 값을 찾기 위해 정보 집적도 기반의 유사도 계산방법(intensity-based similarity calculation)을 제안하였다.

**사례저장:** 생성된 맞춤 지식은 사례베이스에 저장되어 다음 사용자의 맞춤지식 생성에 재활용된다.

위와 같은 과정을 거쳐 사용자의 요구사항에 의 미론적으로 부합하는 최적의 정보를 생성한다. 맞

춤형 지식생성을 위한 각 과정의 세부 절차를 다음 절에서 살펴보고자 한다.

## 4. 맞춤형 지식생성 절차

### 4.1 사례 생성

본 연구에서 사례는 사례기반추론에서의 사례와 유사하게 인덱스와 값의 쌍으로 이루어진다. 인덱스는 사용자의 요구사항에 해당하며 값은 사용자의 요구사항을 만족하는 문장들로 이루어진다. 즉, 사례는 수집된 비구조적인 웹 정보를 지능형 에이전트가 이해하고 처리할 수 있도록 구조적 정보 표현방법으로 사용되었다. 지식은 규칙, 프레임 및 사례 등으로 표현될 수 있으나 본 논문에서는 사례표현방식을 채택하였다. 이는 사례표현 방식이 비구조적인 웹 문서를 구조적인 문서로 변환할 경우 발생할 수 있는 정보의 손실을 줄이고, 표현의 자유도가 상대적으로 크다고 판단했기 때문이다. 웹 문서를 사례로 변환하는 과정은 다음과 같다. 변환 대상인 웹 문서를 문장 단위로 파싱하는 과정에서 각 문장이 요구사항 키워드들의 부분집합을 포함하고 있으면, 그 부분집합을 사례의 인덱스로 하며 그 문장을 사례 인덱스의 값으로 한다.

예를 들어, 맞춤형 지식을 원하는 여행자의 요구사항이 ‘Seoul,’ ‘Korea’ 및 ‘cuisine’일 때, 웹 정보에 포함된 문장으로부터 ‘Seoul’ 과 ‘Korea’ 값을 식별했다면, 다음 문장의 인덱스들은 ‘Seoul’과 ‘Korea’가 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.’*

위와 같이 사용자 요구사항과 관련되어 수집된 모든 웹 문서들은 인덱스와 값들로 구성된 사례로 변환되며, 생성된 사례는 다음과 같이 표현된다. 즉 요구사항을 포함하지 않는 웹 문서는 추출되지 않으며 따라서 사례로 변화되지 않는다. 또한 기존의

사례베이스에서는 요구사항을 포함하지 않는 사례는 추출되지 않는다.

사용자 요구사항과 이를 이용해 검색된 하나의 웹 문서를 사례로 변환해 생성한 지식의 구조는 식 (1)과 같다. 식 (1)로부터 웹 문서가 인덱스와 값들의 쌍으로 구성됨을 알 수 있다.

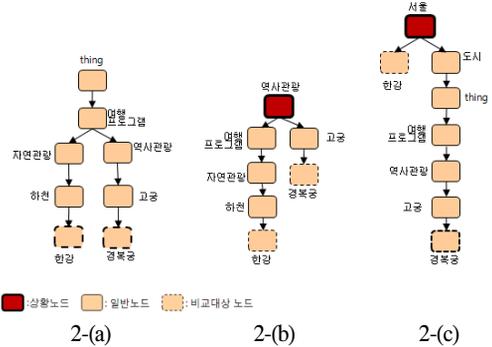
$$\begin{aligned}
 R &= \{z_h\} \\
 C_i &= \{(X_{ij}, V_{ij}) | X_{ij} \subset R, X_{ij} \neq \emptyset\} \\
 V_{ij} &= \{v_{ijk}\}
 \end{aligned}
 \tag{1}$$

- $R$ : 사용자 요구사항 집합
- $h$ : 집합의 원소들에 대한 색인
- $z_h$ : 사용자의  $h$ 번째 요구사항
- $i$ : 사례 색인
- $c_i$ :  $i$ 번째 사례
- $j$ : 집합들에 대한 색인
- $X_{ij}$ :  $i$ 번째 사례의  $j$ 번째 집합,  $X_{ij} \subset R$
- $V_{ij}$ :  $X_{ij}$ 를 포함하는 문장들의 집합
- $k$ : 집합의 원소들에 대한 색인
- $v_{ijk}$ :  $V_{ij}$ 의  $k$ 번째 원소

웹 문서로부터 정형화된 문서로 변환된 사례들과 사례베이스로부터 추출된 사례들 중 요구사항을 포함하는 사례들이 있으면 이들 중 사용자 요구사항에 최적으로 적합한 사례를 추출하는 것이 요구된다. 이를 위해 다음 절에서 상황기반 유사도 계산을 제안한다.

## 4.2 유사사례 추출

사용자의 요구사항에 가장 부합하는 사례를 추출하기 위해서는 상황기반의 유사도 계산 방법을 적용한다. 상황기반 유사도 계산의 기본 전제는 "두 개념의 유사도는 이들이 공유하고 있는 개념에 따라 달라질 수 있다"는 것이다. 두 개념 사이에 존재하는 아크의 수, 즉 거리를 고려해 유사도를 계산하는 기존 방법의 경우 비교 대상인 말단 노드가 바뀌지 않는 한 항상 동일한 유사도 값을 갖는다.



(그림 2) 상황기반 유사도 계산에 활용되는 그래프

예를 들어, 기존 방법에서는 사용자 요구사항과 사례가 공유하는 개념이 ‘서울’인지 ‘역사여행’인지에 관계없이 ‘경복궁’과 ‘한강’의 유사도는 항상 같은 값을 갖는다. 그러나 ‘서울’이라는 도시 관점에서, 즉 ‘경복궁’과 ‘한강’이 공유하는 개념이 ‘서울’인 경우의 유사도와 여행 프로그램 관점인 ‘역사여행’인 경우의 ‘경복궁’과 ‘한강’ 간의 유사도가 전자의 같을 수 없다. 이에 본 논문에서는 유사도 계산의 대상인 사용자 요구사항과 사례 간의 공통적인 키워드를 ‘상황(context)’으로 정의하고 유사도 계산에서 비교대상들이 포함하고 있는 상황을 고려하기로 한다.

상황기반 유사도를 계산하기 위해, ‘상황’으로 정의된 인덱스를 제외한 사용자 요구사항 집합과 사례로부터 역시 ‘상황’ 인덱스 제외한 나머지 인덱스들 중에서 인덱스를 하나를 선택한다. 선택된 두 개의 인덱스를 말단 노드로 갖는 그래프를 온톨로지에서 추출한다. 이 그래프는 온톨로지의 최상위 노드로부터 비교 대상인 두 개념까지를 연결하는 최단거리 경로의 조합이다. 예를 들어, 온톨로지에 포함된 ‘한강’과 ‘경복궁’의 유사도 계산을 위해 추출한 그래프는 이 온톨로지의 최상위 노드인 ‘thing’과 ‘한강’, ‘thing’과 ‘경복궁’을 연결하는 각 최단거리 경로의 조합으로 생성한다(그림 2-(a)). 다음 단계에서는 추출한 그래프를 ‘상황’으로 정의된 인덱스를 기준으로 변형한다. 이때 ‘상황’으로 정의된 인덱스가 그래프 상의 원소라면, (그림 2-(b))와

같이 그래프의 최상위 노드를 ‘상황’ 노드로 변경한다. 그러나 ‘상황’으로 정의된 인덱스가 온톨로지에는 존재하나 그래프 상의 원소가 아니라면, ‘상황’으로 정의된 인덱스를 최상위 노드로 하고 최상위 노드로부터 말단 노드까지의 최단 거리를 연결해 그래프를 변형한다 (그림 2-(c)).

이상의 논의를 통해, 두 개념의 유사도를 계산하는 데 활용할 그래프가 ‘상황’에 따라 달라질 수 있음을 확인하였다. 이에 본 연구에서는 사용자 요구사항과 사례가 공유하는 상황에 따라 유사도가 동적으로 변화되는 상황 기반의 유사도 계산방법을 제안하였다. 상황 기반의 유사도는 ‘사용자 요구사항과 사례간의 공유하는 인덱스의 수’를 의미하는 정적 상황 유사도와 ‘사용자 요구사항과 사례간의 공유하는 인덱스를 제외한 나머지 인덱스들간의 유사도’를 동적 상황 유사도라 정의한다. 동적 상황 유사도는 사용자 요구사항과 사례가 공유하는 개념에 따라 나머지 개념들의 유사도가 달라짐을 의미한다. 동적으로 상황 기반의 유사도 계산 절차는 다음과 같다.

**1 단계.** 사용자 요구사항  $R$ 과  $i$ 번째 사례인  $C_i$  간의 정적 상황 유사도  $SIM_s(R, C_i)$ 를 계산한다. 이를 위해 사용자 요구사항  $R$ 과 사례  $C_i$ 가 공통적으로 포함하고 있는 상황을 식 (2)과 같이 정의한다.

$$CX_i = \left\{ cx_{il} | cx_{il} \in ((\cup_j X_{ij}) \cap R), \right. \\ \left. 1 \leq l \leq L_i \right\} \quad (2)$$

이상에서 정의한  $CX_i$ 를 이용해 도출한  $SIM_s(R, C_i)$ 는 식 (3)과 같다.

$$SIM_s(R, C_i) = L_i \quad (3)$$

**2 단계.**  $CX_i$ 를 고려한 동적 상황 유사도를 계산한다. 이를 위해, (그림 2-(b)) 또는 (그림 2-(c))와 같이  $CX_i$  원소 중의 하나를 최상위 노드로 하고, 유사도 비교대상인 두 개념을 말단 노드로 갖는 그래

프를 추출한다. 이때 그래프를 구성하는 말단 노드 중의 하나는  $cx_{il}$ 이 아닌 사용자 요구사항의 원소 ( $cx_{ip}^c$ )이고 다른 하나는  $cx_{il}$ 이 아닌 사례의 원소 ( $cx_{ijk}^c$ )이다. (그림 2-(b))에서 본 바와 같이 유사도 계산을 위해 도출되는 부분 온톨로지의 수는  $cx_{il}$  원소 수와 같다. 그래프의 말단에 위치한 두 노드 간의 유사도는 식 (4)와 같이 계산된다.

$$X_{ij}^c \equiv X_{ij} - CX_i = \{x_{ijk}^c\}$$

$$Sim(C_i, cx_{ip}^c, X_{ij}^c) = \frac{\sum_l \sum_k (sim(cx_{ip}^c, x_{ijk}^c | cx_{il}))}{L_i} \quad (4)$$

$$Sim(C_i, cx_{ip}^c) = \max_j \{Sim(C_i, cx_{ip}^c, X_{ij}^c)\}$$

$sim(cx_{ip}^c, x_{ijk}^c | cx_{il})$  : 사용자 요구사항과 사례 간의 공유하는 상황  $cx_{il}$ 가 주어진 경우,  $cx_{ip}^c$ 와  $x_{ijk}^c$  사이의 유사도

$sim(cx_{ip}^c, x_{ijk}^c | cx_{il})$ 의 값을 도출하는 과정은 다음과 같다. 유사도 계산의 대상인 두 개념이 공유하는 상황노드를 정점으로 생성된 모든 부분 온톨로지를 이용해 두 개념의 유사도를 계산한다. 유사도는 부분 온톨로지에 포함된 두 개념 사이의 링크(edge)의 수나 깊이를 고려해 유사도를 계산하는 방법 [10]인 Wu와 Palmer 방법, Leacock과 Chodorow 방법 및 Li의 방법 중의 하나를 이용해 계산한다. 본 논문에서 제안한 방법의 우수성을 입증하기 위해, 이들 방법을 이용해 계산한 유사도와 본 방법을 적용한 유사도 계산 결과를 비교하였다. 사용자 요구사항과 사례에 포함된 모든 인덱스와의 조합에 대해  $SIM_s(R, C_i)$ 를 계산한 후 식 (5)와 같이 동적 상황 유사도  $SIM_d(R, C_i)$ 를 계산한다.

$$SIM_d(R, C_i) = \frac{\sum_p Sim(C_i, cx_{ip}^c)}{n(CX_i^c)} \quad (5)$$

상황 기반의 유사도  $SIM_{crt}(R, C_i)$ 는 식 (6)과 같이 정적 유사도와 동적 유사도의 합으로 도출된다.

$$SIM_{crt}(R, C_i) = SIM_s(R, C_i) + SIM_d(R, C_i) \quad (6)$$

사용자 요구사항과 가장 유사한 사례는 상황 기반의 유사도가 가장 큰 사례이다. 이때 유사도가 가장 높은 사례가 사용자의 요구사항을 모두 포함하고 있지 않으면 이를 충족시키기 위해 4.3절의 집적도 기반의 유사도를 계산하여 사례가 사용자의 요구사항 모두를 포함한 지식을 생성할 수 있도록 수정한다.

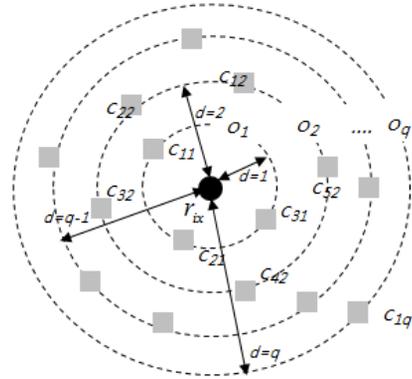
### 4.3 맞춤형 지식 생성

4.2절에서 소개된 상황 기반 유사도 계산을 통해 식별된 유사사례가 요구사항의 일부 정보만 포함하여 사용자를 만족시키지 못할 경우, 만족하는 해를 찾기 위한 정보 집적도 기반의 유사도 계산을 제안한다. 집적도 기반의 유사도 계산과정은 다음과 같다.

**1 단계.** 사용자의 요구사항에는 있으나 유사사례에는 포함되어 있지 않는 사용자의 요구사항과, 이 요구사항에 대한 의미적 포함관계에 있는 인덱스를 포함하는 모든 사례들을 추출하기 위하여 온톨로지로부터 의미적 포함관계에 있는 인덱스들 ( $val_{alt_i}$ )을 추출한다.

**2 단계.** 사용자의 요구사항에는 있으나 유사사례에는 포함되어 있지 않는 사용자의 요구사항과, 이 요구사항에 대한 의미적 포함관계에 있는 각 인덱스를 포함하고 있는 온톨로지의 구성 요소(클래스 및 인스턴스)들을 식별해서 (그림 3)과 같이 동심원에 배치한다. (그림 3)에 포함된 기호의 의미는 다음과 같다.

$r_{ix} (= cx_{ip}^c)$ :  $i$ 번째 사례의 사용자 요구사항 중 값이 대응되지 않은  $x$ 번째 요구사항



(그림 3) 사용자 요구사항과 이와 관련된 값이 포함하고 있는 구성요소들 간의 거리 판별방법

$o_q$ :  $q$ 번째 동심원 ( $q \equiv$ 거리( $d$ ))

$c_{pq}$ :  $q$ 번째 동심원에 위치한  $p$ 번째 클래스

**3 단계.** 사례들로부터 추출한 값들 중에서 사용자 요구사항과 의미적으로 가장 일치하는 값을 결정하기 위해, 즉 하나의 요구사항과 2 단계에서 식별한 클래스들 간의 유사도 계산을 위해 집적도 개념을 도입하였다. (그림 3)에서 도식화한 바와 같이, 집적도는 사례에서 대응되는 각 값들로부터 추출한 클래스들이 사용자 요구사항을 중심으로 어느 정도나 집중 또는 분산되어 있는 지를 측정하는 개념이다. 특정 값에 포함되어 있는 클래스들의 대다수가 사용자 요구사항에 집중되어 있다면 그렇지 않은 값보다 사용자 요구사항을 잘 만족한다고 판단하며, 이를 집적도 기반의 유사도라 한다.

집적도  $I(r_{ix})$ 는 사용자 요구사항에는 있으나 유사 사례에는 포함되어 있지 않는 사용자 요구사항  $r_{ix}$ 을 중심으로 이의 후보 값( $val_{alt_i}$ )으로부터 추출된 클래스  $c_{pq}$ 들이 어느 정도나 집중되어 있는 지를 측정하는 지표로서 식 (7)과 같이 계산된다.  $val_{alt_i}$ 는 사용자 요구사항  $r_{ix}$ 을 인덱스로 갖는 사례에 포함되어 있는 값을 의미한다. 여러 사례들로부터 찾은 이 값들 중의 하나를 사용자 요구사항  $r_{ix}$ 의 값으로 취한다.

$$I(r_{ix}) = \sum_{pq} N(c_{pq} | r_{ix}, o_q) \times w_q \quad (7)$$

$N(c_{pq} | r_{ix}, o_q)$  :  $i$ 번째 사례의  $x$ 번째 요구사항  $r_{ix}$ 을 중심으로 갖는 동심원에서  $q$ 번째 동심원에 위치한 클래스의 수

$w_q$  :  $q$ 번째 동심원의 가중치 ( $w_q = 1/q$ )

$r_{ix}$ 에 대응되는 값의 대안이 될 수 있는 모든 값에 대해 식 (7)과 같은 집적도를 계산한 후,  $r_{ix}$ 이 아닌 즉, 사례가 이미 포함하고 있는 사용자 요구사항  $r_{ix}^c$ 과  $c_{pq}$ 간의 집적도를 역시 식 (7)을 이용해 계산한다. 이 계산 결과를 이용해  $R$ 에 대해 집적도를 다음과 같이 계산한다.

$$I(R) = \sum_{r_{ix}} I(r_{ix}) \times w_{ix} + \sum_{r_{ix}^c} I(r_{ix}^c) \times w_{ix}^c \quad (8)$$

$$R = \{r_{ix}\} \cup \{r_{ix}^c\}$$

$w_{ix}$  : 집적도  $I(r_{ix})$ 에 대한 가중치

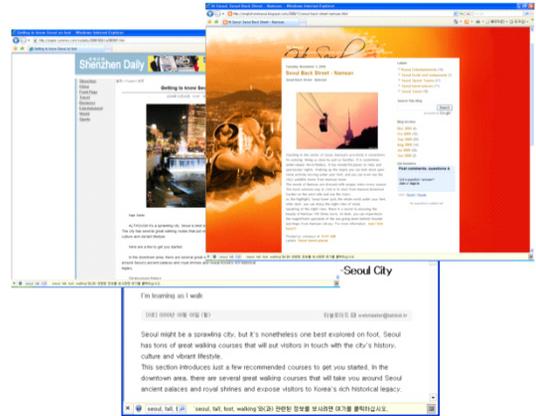
$w_{ix}^c$  : 집적도  $I(r_{ix}^c)$ 에 대한 가중치

$$w_{ix} + w_{ix}^c = 1$$

식(8)에 대한 정규화를 통해 사용자 요구사항  $R$ 과  $val_{alt_i}$ 간의 집적도 기반 유사도  $SIM_{int}(R, val_{alt_i})$ 를 계산한다. 이때  $val_{alt_i}$ 는 유사사례에는 포함되어 있지 않는 사용자의 요구사항이 취할 수 있는 값중의  $i$ 번째 대안을 의미한다.

$$SIM_{int}(R, val_{alt_i}) = \frac{(\sum_l I(r_{ix}) \times w_{r_{ix}} + \sum_{r_{ix}^c} I(r_{ix}^c) \times w_{r_{ix}^c})}{w_{r_{ix}} + w_{r_{ix}^c} = 1} \quad (9)$$

**4 단계.** 2 단계와 3 단계를 거친 후  $SIM_{int}(R, val_{alt_i})$  값이  $val_{alt_i}$ 를 사용자 요구사항의 값으로 취한다.



(그림 4) 사용자 요구사항 R과 연관된 웹 문서

이상의 과정을 거치면 사용자 요구사항을 만족하는 맞춤형 사례, 즉 맞춤형 지식이 생성된다.

#### 4.4 사례 저장

생성된 맞춤형 지식은 사례베이스에 저장된다. 즉, 생성된 사례는 유사한 요구사항을 가진 사용자에게 효과적이고 효율적으로 정보를 제공하기 위해 사례베이스에 저장된다.

### 5. 적용 사례

RSS 리더를 통해 수집된 웹 정보 중에서 여행자의 요구사항  $R=\{\text{Seoul, Fall, Foot, Walking}\}$ 과 관련된 정보가 (그림 4)와 같이 검색되었다면, 이들 정보에 해당하는 HTML 문서를 생성해 문서저장소에 저장한다.

사용자 요구사항 ‘Seoul’, ‘Fall’, ‘Foot’, ‘Walking’를 고려해 문서저장소에 저장된 웹 문서를 사례로 변환하면 (그림 5)와 같다. (그림 5)의 사례에 포함된 인덱스 중에서 ‘Seoul’과 ‘Fall’은 사용자 요구사항을 기반으로 생성된 인덱스이고, ‘Seoul’, ‘foot’는 이들의 조합으로 생성된 인덱스이다.

구성된 사례는 하나 이상의 인덱스를 포함하고 있으며, 또한 각 인덱스는 하나 이상의 값들을 포함하고 있다. 이를 활용해 사용자의 요구사항에 따

```
// Getting to know Seoul on foot
{{{is-a}(case)
  {(Case-id)(1)}
  {(url)(http://paper.sznews.com/szdaily/20081020/ca2903931.htm)}
  {(Seoul)(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]}
  {(Deoksugung) [*** **]}

// Seoul's back streets:
{{{is-all}(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 Seoulactively 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]}
  {(Namsan) [*** **]}

// 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)}}
```

(그림 5) 사례로 변환된 웹 정보(부분 예)

른 맞춤형 지식을 생성한다. 첫 단계로 사용자 요구사항과 사례가 공유하는 개념, 즉 상황을 고려해 정적 상황 유사도  $SIM_s(R, C_i)$ 와 동적 상황 유사도  $SIM_d(R, C_i)$ 을 계산한 후 상황 기반의 유사도

를 계산한다. 사례  $C_1, C_2$  및  $C_3$ 이 포함하고 있는 인텍스와 사용자 요구사항을 비교해 계산된  $SIM_s(R, C_i)$ 은 다음과 같다.

$$\begin{aligned} CX_1 &= \{Seoul, Fall, Foot\} \\ CX_2 &= \{Seoul, Foot, Walking\} \\ CX_3 &= \{Seoul, Walking\} \\ SIM_s(R, C_1) &= 3 \\ SIM_s(R, C_2) &= 3 \\ SIM_s(R, C_3) &= 2 \end{aligned}$$

3 개의 사례 중  $C_3$ 은 사례의 인텍스가 사용자 요구사항에 모두 포함되므로 상황 기반 유사도 계산에서 동적 상황 유사도를 계산할 필요가 없다. 따라서  $C_1$ 과  $C_2$ 를 대상으로 동적 상황 유사도  $SIM_d(R, C_i)$ 을 계산한다.

$$\begin{aligned} SIM_d(R, C_1) &= (sim(a,b) | Seoul) + \\ &\quad (sim(a,b) | fall) + \\ &\quad (sim(a,b) | foot) = 0.66 \end{aligned}$$

이때  $a$ 는  $C_1$ 의 Deoksugung이고,  $b$ 는  $R$ 의 walking을 의미하며,  $a$ 와  $b$ 의 유사도 값은 Wu과 Palmer의 방법을 적용한 결과이다. 다른 방법을 이용한 유사도 계산 예는 6장에 서술하였다.

$$\begin{aligned} SIM_d(R, C_2) &= (sim(c,d) | Seoul) + \\ &\quad (sim(c,d) | foot) + \\ &\quad (sim(c,d) | walking) = 0.51 \end{aligned}$$

이때  $c$ 는  $C_2$ 의 'Namsan'이고  $d$ 는  $R$ 의 'fall'을 의미하며,  $c$ 과  $d$ 의 유사도 값은 Wu과 Palmer의 방법을 적용하였다.

정적 및 동적 상황기반 유사도를 모두 고려한 사례들 과 요구사항들 간의 상황 기반의 유사도  $SIM_{ct}(R, C_i)$ 는 다음과 같다.

$$\begin{aligned} SIM_{ct}(R, C_1) &= 3 + 0.66 = 3.66 \\ SIM_{ct}(R, C_2) &= 3 + 0.51 = 3.51 \\ SIM_{ct}(R, C_3) &= 2 \end{aligned}$$

```
// Customised 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 Seoulactively moving under your feet, and
    you can even see the city's satellite towns from
    Namsan tower)]
  [(Seoul, Foot)(As the highlight, Seoutower puts the
    whole world under your feet)]
  [(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)] }
```

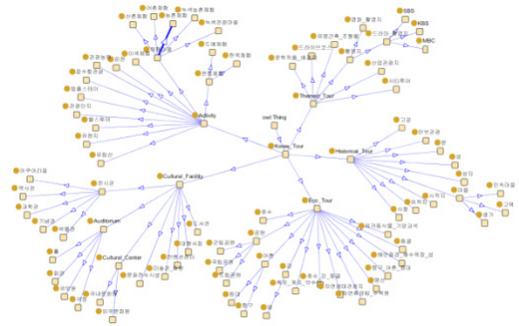
(그림 7) 사용자 요구사항에 대한 맞춤형 지식(부분 예)

$SIM_{crt}(R, C_i)$ 의 값이 가장 큰  $C_1$  사례를 유사 사례로 취한다. 그러나  $C_1$  사례가 사용자 요구사항 중에서 ‘Walking’에 대한 정보를 포함하고 있지 않음으로 집적도 기반의 유사도 계산을 통해 ‘Walking’을 인덱스로 하는 값을 찾는다. 이를 위해 사례베이스에 저장되어 있는 사례들 중에서 ‘Walking’을 인덱스로 하는 값을 포함하고 있는 사례 모두를 찾는다. 이로부터 추출한 ‘Walking’을 인덱스로 하는 값들 중에서 사용자 요구사항을 가장 잘 만족하는 값을 선택하기 위해 각 값들에 포함된 온톨로지의 클래스들을 찾는다. 만약 ‘Walking’의 값을 포함하고 있는 사례들 중에서  $val_{alt_i}$  값이 포함하고 있는 온톨로지의 클래스가 10개이고, 이중 2개는 (그림 3)의 2번째 동심원에, 3개는 3번째 동심원에 위치하며 나머지는 4번째 동심원에 존재한다면  $I(r_{1x})$ 는 다음과 같이 계산된다.

$$I(r_{1x}) = 2 \times w_2 + 3 \times w_3 + 5 \times w_4 = 3.25$$

각 동심원의 가중치는 식(7)에 의해  $w_2$ 이 1/2,  $w_3$ 가 1/3 및  $w_4$ 가 1/4이므로  $I(r_{1x})$ 는 3.25가 된다.

다음 단계는 ‘Walking’을 제외한 나머지 요구사항들  $r_{1x}^c$ 과  $val_{alt_i}$ 에 포함된 10개의 클래스와 집적도 기반 유사도  $I(r_{1x}^c)$ 를 계산한다.



(그림 8) 인스턴스를 제외한 여행 온톨로지 (일부)

$$\begin{aligned} I(r_{1x}^c) &= (1 \times w_2 + 5 \times w_3 + 4 \times w_8) \\ &\quad + (3 \times w_3 + 2 \times w_4 + 5 \times w_8) \\ &\quad + (7 \times w_4 + 2 \times w_8 + 1 \times w_9) \\ &= 6.68 \end{aligned}$$

$w_2$ 이 1/2,  $w_3$ 가 1/3 및  $w_4$ 이 1/4이며,  $w_8$ 이 1/8,  $w_9$ 은 1/9로  $I(r_{1x}^c)$ 는 6.68로 계산된다.

이때  $w_{r_{1x}} = 0.7$ ,  $w_{r_{1x}^c} = 0.3$ 이라 하면 최종적인  $SIM_{int}(R, val_{alt_i})$ 은 4.23이다. 모든  $val_{alt_i}$ 에 대해 위 과정을 반복한 후 가장 큰  $SIM_{int}(R, val_{alt_i})$ 를 갖는  $(R, val_{alt_i})$ 를 이용해 유사 사례를 수정한다. 이러한 과정을 거쳐 생성한 맞춤형 지식을 (그림 7)에 도시하였다. (그림 7)의 사례는 사례베이스의 학습을 위해 사례베이스에 저장된다.

사용자 요구사항  $R$ 과 연관되어 검색된 3개의 웹 정보 중 사용자의 요구를 모두 만족시키는 웹 문서를 찾을 수는 없었다. 그러나 본 프레임워크에서 제안한 방법론을 적용한 결과, (그림 7)과 같이 사용자의 요구를 모두 만족하는 새로운 맞춤형 지식이 생성되었다. 사용자에게 통합된 하나의 지식을 제공함으로써 정보의 추출, 정제와 관련된 사용자의 부담을 크게 줄일 수 있다.

## 6. 실험

본 논문에서 제안한 유사도 계산 방법의 타당성을 입증하기 위해, 본 연구진은 온톨로지 개발 도

(표 1) 사용자 요구사항과 변환된 사례의 인덱스

요구사항(R)	Seoul, T-Tour, Dongdaemun (D), Changgyeonggung (C)
사례 1(C <sub>1</sub> ) 인덱스	Seoul, Han-river (H), Korean-house (K)
사례 2(C <sub>2</sub> ) 인덱스	T-Tour, Han-river (H), Korean-house (K)

(표 2) 유사도 계산결과

		$Sim_{WP}$	$Sim_{WP}^{cut}$	$Sim_{LC}$	$Sim_{LC}^{cut}$	$Sim_{Li}$	$Sim_{Li}^{cut}$
(R <sub>C<sub>1</sub></sub> )	D,H	2.62	2.71	2.48	2.60	2.44	2.45
	D,K	2.14	2.14	2.04	2.21	2.06	2.06
	C,H	2.17	2.17	2.05	2.25	2.09	2.09
	C,K	2.46	2.63	2.30	2.48	2.29	2.30
(R <sub>C<sub>2</sub></sub> )	D,H	2.62	2.15	2.48	2.25	2.44	2.07
	D,K	2.14	2.15	2.04	2.20	2.06	2.07
	C,H	2.17	2.67	2.05	2.447	2.09	2.37
	C,K	2.46	2.78	2.30	2.727	2.29	2.55

구인 프로테제를 이용해 185개의 클래스와 2,884개의 인스턴스로 구성된 여행 온톨로지를 구축하였다. 여행 온톨로지의 구조를 프로테제의 플러그인 프로그램인 Jambalaya로 도식화하면 (그림 8)과 같다.

(그림 8)에서 도식화한 온톨로지를 활용해, 비교되는 두 개념이 온톨로지 상에서 상황(공유하는 개념이 무엇인지)에 따라 어느 정도나 유사한 지에 대한 평가를 수행하였다. 이를 위해, 사용자의 요구사항과 관련이 있는 웹 정보를 두 개로 제한하였으며, 이 들 웹 정보로부터 (표 1)과 같은 사례를 생성하였다.

(표 1)로부터 R과 C<sub>1</sub>이 공유하는 상황(CX<sub>1</sub>)은 ‘Seoul’이며, R과 C<sub>2</sub>가 공유하는 상황(CX<sub>2</sub>)은 ‘T-Tour’임을 알 수 있다. 온톨로지에 포함된 두 개념 간에 존재하는 링크(edge)의 수나 깊이를 고려해 유사도를 계산하는 Wu와 Palmer ( $Sim_{WP}$ ), Leacock과 Chodorow( $Sim_{LC}$ ) 및 Li ( $Sim_{Li}$ )방법을 적용할 경우, 비교 대상인 두 개념의 유사도는 항상 동일하다

(표 2). 그러나 본 논문에서 제안한 방법을 적용할 경우, 비교해야 할 두 값이 같음에도 불구하고 사용자 요구사항과 사례 1 및 사례 2가 공유하는 상황이 다르므로 인해 유사도 값이 상이하게 계산됨을 알 수 있다.

(표 2)에서는  $Sim_{WP}$ ,  $Sim_{LC}$  및  $Sim_{Li}$ 에 본 논문에서 제안한 컨텍스트를 고려한 유사도 계산 방법을 적용한 유사도 결과를  $Sim_{WP}^{cut}$ ,  $Sim_{LC}^{cut}$  및  $Sim_{Li}^{cut}$ 로 나타냈다.

## 7. 결 론

인지능력의 한계와 시간적인 제약을 가진 사용자가 소셜 웹으로부터 관련된 모든 정보를 수집한 후 자신이 필요로 하는 정보의 조각들 만을 추출해 지식을 생성하는 것은 가능하기는 하지만 매우 어렵고 부담스러운 일이다. 이러한 어려움을 완화하기 위한 도구로 온톨로지가 널리 활용되고는 있으나, 통제된 개념들 간의 관계만을 모델링 해주는 온톨로지만으로는 사용자들이 기대하는 수준의 맞춤형 지식을 생성하는 것은 용이하지 않다. 이에 본 연구에서는 사례기반 추론과 온톨로지를 활용한 상황기반의 유사도 계산방법이 탑재되어 있는 맞춤형 지식 생성 프레임워크를 제안하였다. 제안한 프레임워크를 활용함으로써 사용자의 요구를 만족시킬 수 있는 맞춤형 지식의 생성에 성공하였다. 그러나 본 논문에서 수행한 실험은 단지 상황기반 유사도 계산 방법이 보다 타당할 수 있음을 입증하였다.

그러므로 본 논문에서 제안한 프레임워크의 우수성을 입증하기 위해서는 본 프레임워크의 적용을 통해 생성된 맞춤형 지식에 대한 사용자 만족도 평가와 함께 집적도 기반의 유사도 계산 방법에 대한 평가가 추가적으로 수행되어야 한다. 또한 상황기반 유사도 계산 방법과 기존방법과의 차별화를 공고히 하기 위해 알고리즘에 대한 개선이 필요하다.

## Acknowledgement

본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다. (UD110006MD)

## 참 고 문 헌

- [1] A. P. Bernstein and L. Haas, "Information Integration in the Enterprise," *Communications of the ACM*, vol. 51, no. 9, 2008, pp. 72-79.
- [2] V. Alexiev, et al., "Information Integration with Ontologies: Experiences from an Industrial Showcase," Wiley, 2005
- [3] Y. A. Halevy, et al., "Enterprise information integration: challenges and controversies," *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, 2005, pp. 778-787.
- [4] M. Philippe and P. W. Eklund, "Knowledge Retrieval and the World Wide Web," *IEEE Intelligent Systems*, 2000 (May/June), pp.18-25.
- [5] I. Mani and E. Bloedorn, "Machine learning of Generic and User-Focused Summarization," *In proceedings of the National Conference on Artificial Intelligence (AAAI)*, 1998, pp. 820-826.
- [6] S. Teufel and K. Moens, "Sentence extraction as a classification task," *Workshop of Intelligent and scalable text summarization, ACL/EACL 1997*, 1997, pp.58-65.
- [7] T. Slimani, B. B.Yaghlane, and K. Mellouli, "A New Similarity Measure based on Edge Counting," *World Academy of Science, Engineering and Technology 23 2006*, 2006, pp 34-38.
- [8] L. Song, et al., "Fuzzy Semantic Similarity Between Ontological Concepts," *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, 2007, pp 275-280.
- [9] R. Rada, Mili, H., Bicknell, E., Blettner, M., "Development and application of a metric on semantic nets," *IEEE Transactions on Systems, Man, and Cybernetics* vol. 19, no. 1, 1989, pp. 17-30
- [10] A. Hliaoutakis, Varelas G., Voutsakis E., Petrakis E. G. M., Milios E., "Information Retrieval by Semantic Similarity", *International Journal on Semantic Web & Information Systems*, vol. 2, 2006, pp. 55-73
- [11] Resnik P., Using information content to evaluate semantic similarity in a taxonomy, *Proc. 14th International Joint Conference on Artificial Intelligence*, vol. 1, 1995, pp. 448-453

## ◎ 저 자 소 개 ◎



### 손 미 애 (Mye M. Sohn)

1985년 성균관대학교 산업공학과 학사

1988년 KAIST 산업공학과 석사

2002년 KAIST 경영정보공학박사

1988년~2004년 한국국방연구원

2004년~현재 성균관대학교 시스템경영공학과 부교수

관심분야 : 온톨로지, 시맨틱웹, 시맨틱웹서비스, 사례기반추론, 엔터프라이즈아키텍처



### 이 현 정 (Hyun Jung Lee)

1991년 이화여자대학교 물리학과 학사

1994년 이화여자대학교 전자계산학과 석사

2004년 KAIST 경영공학 박사

2007년~현재 서강대학교 경영대학 연구교수

관심분야 : 의사결정지원시스템, 시맨틱웹, 온톨로지관련연구, 에이전트시스템