

의미 수준이 다른 비즈니스 프로세스의 검색 방법  
A methodology for discovering business processes in different semantic levels

최영환, 채희권, 김광수  
포항공과대학교 산업공학과 Infosys Lab.  
{yhchoi, hkchae, kskim}@postech.ac.kr

Abstract

e-Transformation of an enterprise requires the collaboration of business processes to be suited to the business participants' purpose. To realize this collaboration, business processes should be implemented as components and the system developers could be able to reuse the components for their specific purpose. The first step of this collaboration is the discovery of exact components for business processes. A dilemma, however, is the fact that there are thousands or even millions of business processes which vary from one enterprise to another. Moreover, business processes could be decomposed into multiple levels of semantics and classified into several process areas. In general, discovery of exact business processes requires understanding of widely adopted classification schemes such as CBPC, OAGIS, or SCOR. To cope with this obstacle, business process metadata should be defined and managed regardless of specific classification schemes to support effective discovery and reuse of business processes components. In this paper, a methodology to discover business process components published in different semantic levels is proposed. The proposed methodology represents the metadata of business process components as topic maps stored in a registry and utilizes the powerful features of topic maps for process discovery. TM4J, an open-source topic map engine, is modified to support concept matching and navigation. With the implemented tool, application system developers can discover and publish the business process components effectively.

key words : e-Transformation, Business Process, Process Catalog, Classification, Semantic Similarity, Topic Map, Product Configuration

1. 서론

e-트랜스포메이션은 기업 전반의 경영활동을 e-비즈니스 환경에 적합하도록 변화시켜 확장된 기업(extended enterprise)으로 전환하는 총체적인 경영혁신 활동을 의미한다. 이를 위해서는 기업의 가치사슬에 참여하고 있는 다른 기업들과 서로 협력함으로써 공통의 사업 목표를 달성하기 위한 비즈니스 프로세스의 통합 및 동기화와 같은 비즈니스 프로세스의 혁신이 요구된다. e-트랜스포메이션을 지향하는 새로운 비즈니스 프로세스가 결정된 후에 시스템 개발자들은 각각의 비즈니스 프로세스 또는 서브프로세스의

특징 및 요구사항을 분석하여 해당 프로세스를 수행할 수 있는 소프트웨어 컴포넌트(또는 웹 서비스)들을 검색(discovery)하고, 검색된 컴포넌트들을 조합(composition)함으로써 실행 가능한 애플리케이션을 개발하게 된다[1].

시스템 개발 과정에서 가장 먼저 고려되어야 할 사항은 과연 특정 비즈니스 프로세스의 요구사항을 모두 만족하며 재사용 가능한 소프트웨어 컴포넌트가 어디에 존재하는가 하는 문제이다. 비즈니스 프로세스는 기업 고유의 특수한 상황에 따라 요구사항 및 표현방법이 매우 다양하고, 이와 마찬가지로 소프트웨어 컴포넌트 개발업체도 고유의 사업 전략에 따라 제공하는 서비스의 내용과 표현방법이 다양하기 마련이다. 따라서 프로세스 요구자와 서비스 제공자가 의사소통 과정에서 의미의 혼동 없이 프로세스나 서비스에 대한 메타데이터(metadata)를 교환하기 위해서는 그림 1의 (a)와 같은 m:n 종류의 매핑 규칙이 필요하다.

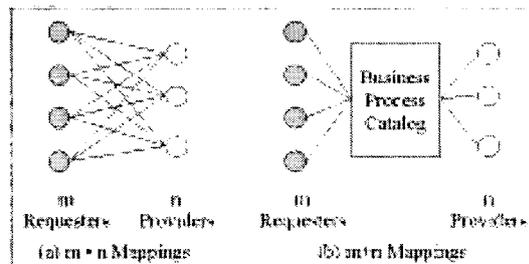


그림 1. 요구자와 제공자 사이의 매핑

웹 환경에서는 요구자와 제공자가 다수 존재하고 새로운 기업의 시장 진입에 의해 매핑규칙의 수가 급격히 증가하게 되므로 m:n 매핑방법은 실제 적용하기에는 어려움이 있다. 이 문제를 해결하기 위하여 그림 1의 (b)와 같이 비즈니스 프로세스의 메타데이터에 대한 표현방법을 명시한 온톨로지(ontology)를 기반으로 프로세스의 메타데이터를 카탈로그의 형태로 만들어 공지하는 방법이 제안되었다. 프로세스 카탈로그를 중립 형태의 매개체로 정하고 자신과 카탈로그 사이의 매핑규칙만을 구축해 놓으면 다른 어떤 요구자나 제공자하고도 의사소통이 가능해지므로 매핑규칙의 수가 현저히 감소하는 효과가 있다. 그러나 CBPC[2], OAGIS[3], SCOR[4]와

같이 프로세스 카탈로그에 관한 다양한 연구[5,6,7]에도 불구하고 아직 비즈니스 프로세스의 모든 메타데이터를 완벽하게 표현할 수 있는 유일한 카탈로그가 공표되지 않아 다양한 프로세스 카탈로그 고유의 분류체계(classification scheme)에 따른 프로세스의 계층적 구조에 대한 이해가 필수적이라는 문제가 있다.

또한 하나의 프로세스는 여러 개의 서브프로세스로 분해되어 서로 다른 의미수준(semantic level)에서 표현될 수 있고, 요구자와 제공자는 동일한 프로세스에 대한 관점과 지식이 서로 다르기 때문에 해당 프로세스의 메타데이터가 프로세스 카탈로그의 계층적 구조 상에서 어떤 개념(concept)을 지칭하는지 판단하기가 쉽지 않다(그림 2 참조). 따라서 일반적으로 요구사항은 "anything"으로 질의되고, 서비스 내용은 "everything"으로 광고되는 경향[8]이 있다. 동일한 프로세스에 대한 메타데이터를 요구자와 제공자가 서로 다른 의미수준에서 표현할 경우에는 비즈니스 프로세스 카탈로그의 저장정보에 대한 "EXACT" 키워드매칭은 실제로는 사용 가능한 대안들을 제외시킬 수 있기 때문에, 특정 분류체계에 제한받지 않으면서 프로세스 메타데이터 사이의 유사도에 근거한 "SIMILAR" 매핑방법이 필요하다.

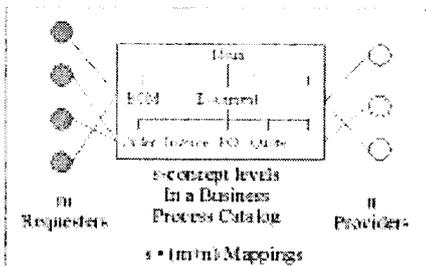


그림 2. 의미 수준이 다른 경우의 매핑

따라서 본 논문에서는 e-트랜스포메이션을 위한 시스템 개발의 첫 번째 단계인 프로세스 명세서로부터 사용가능한 소프트웨어 컴포넌트나 서비스를 유사도에 근거하여 검색하는 새로운 방법을 제안한다. 즉, 비즈니스 프로세스의 요구자와 서비스 제공자가 서로 다른 의미수준에서 프로세스나 서비스에 대한 메타데이터를 표현할 경우에, 메타데이터 사이의 유사도를 계산하여 요구되는 프로세스에 가장 적합한 서비스를 검색하는 방법을 소개하며 주요 내용은 다음의 항목으로 요약된다.

- 의미 수준이 다른 프로세스의 표현 방법 (3 장)
- 비즈니스 프로세스의 유사도 계산 방법 (4 장)
- 비즈니스 프로세스의 검색 방법 (5 장)

비즈니스 프로세스의 메타데이터는 대부분의 프로세스 카탈로그에서 제시하는 항목들을 선정하였으며 topic map[9]을 이용하여 표현하였다. 그 이유는 topic map 이 웹의 정보자원에 대한 인덱싱 표준(ISO/IEC 13250)으로서 데이터 구조의 변동에 대한 융통성이 크다는 장점이 있기 때문이다. 따라서 본 논문에서 제시하는 유사도를 이용한 검색 방법은 topic map 의

데이터 구조를 기반으로 하고 있으며, 비즈니스 프로세스의 요구자와 서비스 제공자가 동일한 개념분류체계(concept taxonomy)를 이용하여 프로세스나 서비스의 메타데이터를 topic map 으로 표현함을 전제 한다.

## 2. 관련 연구

### 2.1 비즈니스 프로세스의 표현

프로세스는 시스템의 행위를 구성하는 기본적인 요소이므로 오래 전부터 프로세스의 표현방법에 대한 다양한 연구가 수행되어 왔다[10]. 그러나 비즈니스 프로세스가 수행되는 환경이 off-line 에서 웹으로 이동되었고, 비즈니스 프로세스의 실제 수행 과정에서는 multi-party, multi-step conversation, loosely coupled, asynchronous 한 특성들이 요구되므로 이와 같은 복잡한 상황들에 대한 표현력이 뛰어난 BPEL[11], EDOC[12]에 대한 관심이 높아지고 있다.

상기한 프로세스의 표현방법과는 별도로 가치사슬에 참여하고 있는 기업들이 서로 협력해가며 비즈니스 프로세스를 수행하기 위하여 기업 내외부의 애플리케이션들을 통합하려는 연구가 추진되었고, 그 노력의 일환으로 중립형태의 비즈니스 프로세스 카탈로그를 제정하여 공표하였다. 대표적인 예로는 ebXML 의 CBPC(Common Business Process Catalog)[2], OAG 의 OAGIS(Open Applications Group Integration Specification)[3], SCC 의 SCOR(Supply Chain Operations Reference) model[4] 등이 있다. 그러나 이와 같은 프로세스 카탈로그들은 각각의 고유한 분류체계를 가지고 있으며 프로세스를 표현하는 abstraction level 이 서로 다르기 때문에 이를 이용하여 프로세스를 검색하기 위해서는 해당카탈로그의 프로세스 분류체계와 프로세스 메타데이터의 계층적 구조에 대한 이해가 요구된다.

### 2.2 비즈니스 프로세스의 메타데이터 표준

특정 분류체계와는 독립적으로 프로세스에 대한 의미를 표현하기 위한 방법은 semantic web 이 지향하는 바와 같다. 즉, 어떤 사물에 대한 assertion 을 자연어로 서술하기보다는 온톨로지에 기반한 제한적인 표현방법인 controlled vocabulary 를 이용하여 구조적인 문서(structured document)로 표현하는 것이다. 이를 구현하기 위한 방법은 크게 RDF 를 기반으로 하는 방법과 topic map 을 기반으로 하는 방법으로 분류된다. RDF(Resource Description Framework)[13]는 (resource)-(property)-(property value)로 구성되는 서술문을 이용하여 assertion 을 표현하며 정보자원에 대한 구조적인 메타데이터와 논리적인 추론의 지원용 목표로 한다[14]. Topic map[9]은 RDF resource 에 상응하는 topic 에 name, occurrence, association 특성(topic characteristics)을 부여함으로써 assertion 을 표현하며 웹에 분포되어 있는 정보자원의 검색을 위한 인덱싱 지원을 목표로 한다[14]. RDF 로 표현된 내용은 복잡하지만 내용에 대한 엄격한 제한이 가능한 반면, topic map 으로 표현된 내용은 간단한 구조로 되어 있어 이해하기 쉽고 구조변경에 대한 융통성이 높지

만 엄격한 제한은 어렵다. 따라서 RDF 는 machine 을 위한 온톨로지 위주의, topic map 은 인간을 위한 개념분류 위주의 semantic 을 제공한다고 보는 견해도 있다[15].

비즈니스 프로세스의 의미를 표현하기 위한 메타 데이터 표준의 대표적인 예로는 웹 서비스를 공지하는 표준인 DAML-S[5], WSDL[16]이 있다. 이들은 모두 RDF 기반으로 구성되어 있으며, 서론에서 언급한 중립형태 표준의 문제점을 그대로 내포하고 있다.

### 2.3 프로세스 유사도를 이용한 검색 방법

사물의 개념에 대한 의미의 유사도(semantic similarity index)를 판별하는 방법은 PSI 방법과 개념분류 방법으로 분류된다. PSI (Published Subject Indicator) 방법[17]은 특정 사물의 개념에 대하여 고유 판별자(UID)를 부여하여 공용 라이브러리에 등록하고 동일한 UID 를 가리키는 개념들은 서로 동일한 대상으로 간주하는 방식으로서 서로 다른 온톨로지의 맵핑 과정에 적용되는 방법이다. 이 방법은 검색과정에서 질의에 대한 반응이 EXACT 한 경우에만 적용할 수 있다는 문제가 있다. 개념분류 방법은 superclass-subclass 관계에 있는 두 개념은 서로 유사하다고 간주하는 방법이다. Paolucci[8]는 질의(query)와 반응(response)에 사용되는 개념들을 class 로 표현하고, superclass-subclass 관계를 고려하여 EXACT, PLUG IN, SUBSUMES, FAIL 의 네가지 카테고리 값으로 유사도를 부여하는 방법을 제시하였다. 이 방법은 질의에 대한 반응이 EXACT 하지 않은 경우에도 검색할 수 있다는 장점이 있으나, 비교하는 개념의 특징이 프로세스의 input 과 output 만으로 제한되어 있다. 또한 직접 연결된 superclass 와 subclass 만을 고려하고 이 관계가 반복되는 경우는 고려하지 않는 문제가 있다. Lee[1]는 Paolucci 의 검색방법을 확장하여 다양한 비교기준에 대한 유사도를 계산하는 방법을 제시했는데, 유사도 계산방법은 Paolucci 의 방법과 동일하며 각 비교기준마다 요구되는 유사도의 한계값을 미리 설정해야 한다는 어려움이 있다.

본 논문에서는 topic map 의 특성을 이용하여 비즈니스 프로세스의 검색을 위한 PSI 방법과 개념분류 방법의 문제를 해결하고 두 방법의 장점을 반영한 검색방법을 제시한다.

### 3. 의미 수준이 다른 프로세스의 표현

동일한 개념분류체계를 사용하더라도 시스템 개발자와 서비스 제공자는 관점과 지식의 차이로 인하여 동일한 비즈니스 프로세스의 동일한 항목에 대하여 서로 다른 개념을 부여할 수 있다. 이와 같은 경우의 예가 그림 3 에 나타나 있다. 즉, 프로세스의 요구자와 제공자는 항목에 대한 개념의 계층적 구조에 따라 상위 개념이나 하위 개념을 선택하여 해당 프로세스의 특징을 명시하게 된다. 따라서 질의에 대한 반응이 EXACT 하지 않은 경우라도 두 개념 사이의 관계가 유의하다면 검색 결과에 포함시켜야 한다. 이와 같이 시스템 개발자나 서비스 제공자가 프

로세스의 메타데이터를 직접 결정해 주어야 하므로 RDF 보다는 topic map 을 사용하는 것이 바람직하다.

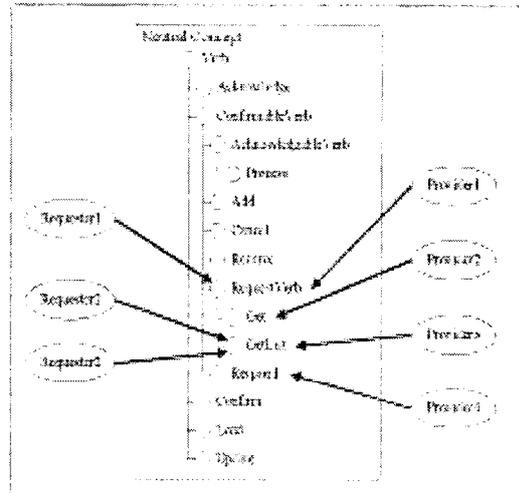


그림 3. 의미 수준이 다른 질의 및 반응

본 논문에서는 프로세스 메타데이터의 항목을 표현하는 개념의 계층구조를 topic map 으로 표현하기 위하여 개념 사이의 관계를 equivalence, inclusion, exclusion 의 세 가지로 구분하였다. Equivalence 는 두 개념이 완전히 서로 일치하는 동의어(synonym) 관계임을 의미하며, 이는 topic map 에서 다른 topic 을 지칭하는데 사용되는 태그인 <subjectIdentity>를 사용하여 간단히 표현된다. Inclusion 은 어떤 특수한 개념(subclass)이 이것보다 일반적인 다른 개념(superclass)에 포함되는 관계임을 의미하며, topic map 에서는 <instanceOf> 태그를 사용하여 표현할 수 있다. Exclusion 은 두 개념이 상반된 의미를 가진 관계임을 명시하는데, topic map 에는 이를 표현하기 위한 기본 태그가 제공되지 않는다. 따라서 exclusion 관계는 exclusion 이라는 topic 을 생성하고 이에 대한 association 인스턴스를 만든 후에 두 개념을 지칭하는 topic 들을 member 로 연결시켜 줌으로써 표현할 수 있다. 이와 같이 세 가지의 개념 사이의 관계를 이용하면 프로세스 메타데이터에 대한 모든 개념들을 하나의 계층구조로 구성할 수 있다.

이와 같이 비즈니스 프로세스의 표현에 사용되는 다양한 개념들이 topic map 을 이용하여 계층화된 구조로 구성되고, 이를 바탕으로 요구자와 제공자가 비즈니스 프로세스와 서비스를 표현한다고 가정할 경우에, 비즈니스 프로세스의 의미를 topic map 으로 표현하는 절차는 다음과 같다.

- 1) 해당 프로세스를 지칭하는 topic P 를 생성
- 2) 명시하고자 하는 프로세스의 특정항목(k)마다 이를 지칭하는 topic 에 대한 association 인스턴스  $A_k$ 를 생성
- 3)  $A_k$  의 member 들을 연결: subject 역할을 하는 topic 으로서 P 를 연결하고, object 역할을 하는

topic 으로서 개념의 계층구조 상에서 가장 적절하다고 판단되는 개념에 대한 topic 을 연결

예를 들어 "request 라는 프로세스는 output 으로서 BillOfMaterial 과 Invoice 를 출력한다"라는 assertion 은 그림 4 와 같은 topic map 으로 표현된다. 그림 4 에 나타난 바와 같이 assertion 을 구성하는 개념들은 association 인스턴트인 association\_001 을 제외하고 모두 개념의 계층구조를 구성하고 있는 topic 들이다. 프로세스의 특징 항목인 output 은 association 인스턴트에 의하여 연결되어 있고, request 라는 프로세스는 subject 역할을 하고 있으며, value 에 해당하는 각 개념의 topic 들은 association 에 대한 object 역할로서 연결되어 있다. Input 과 같은 프로세스의 다른 특징 항목에 대해서도 동일한 방법이 적용된다.

이중 BOM 은 BillOfMaterial 과 동의어 관계임을 의미하는데 이는 다른 프로세스의 특징을 표현하는 assertion 들 중에서 BOM 으로 표현된 assertion 은 BillOfMaterial 로 표현된 assertion 과 일치함을 의미한다.

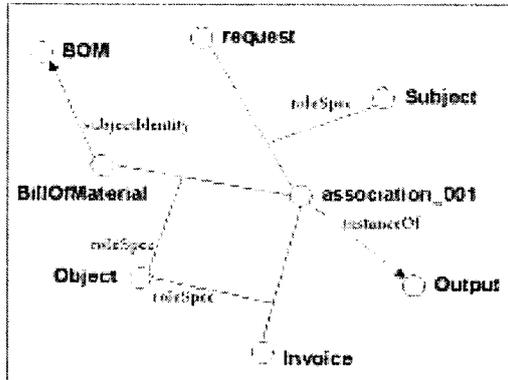


그림 4. Topic map 을 이용한 프로세스 특징의 표현

그림 4 에서는 output 이라는 한가지 특징항목에 대하여 두가지의 개념이 표현되었다는 점에 주목할 필요가 있다. 기존의 방법[1,8]에서는 한가지 특징항목에 대하여 한가지 개념밖에는 표현할 수 없다는 제약이 있는데, topic map 에서는 하나의 association 에 참여하는 member 의 수에 제한이 없으므로 assertion 을 생성하는 사람의 지식이나 관점에 따라 필요한 만큼의 개념을 간단한 작업으로 연결시킬 수 있으며, 이는 불충분한 정보에 대한 검색과정에서 매우 유용하게 작용한다.

#### 4. 비즈니스 프로세스의 유사도 계산

프로세스 카탈로그의 분류체계는 일반적으로 프로세스의 functionality 에 기반하고 있으며 DAML-S[5] 의 service profile 에서는 프로세스의 functionality 를 "serviceName"이라는 항목으로 정의하고 있다. 그러나 프로세스의 이름은 프로세스의 특징을 표현하는 다양한 항목들 중 하나일 뿐이며, 두 프로세스의 이

들이 동일하다고 해서 동일한 프로세스라고 판단하기에는 무리가 따른다. 또한 프로세스의 검색을 위한 질의 및 반응에 대한 개념들이 계층구조에서 서로 다른 의미 수준에 위치할 경우에 기존의 "EXACT" 키워드매칭은 한계가 있다. 따라서 본 논문에서는 topic map 의 개념 구성 방식의 특성을 활용하여 PSI 방법과 taxonomy 방법의 장점을 모두 기대할 수 있는 새로운 방법을 제시한다.

#### 4.1 Semantic Distance 의 결정

계층구조를 구성하고 있는 개념들 중에서 상대적으로 가까이 위치한 개념들은 서로 유사성이 높고, 상대적으로 멀리 떨어진 개념들은 유사성이 낮다고 판단하는 것이 본 연구의 기본 발상이다. 즉, 프로세스의 특징을 표현하는 항목들에 대하여 부여된 개념들 사이의 semantic distance 를 측정하고, 이를 바탕으로 프로세스 사이의 유사도 값(similarity index)을 계산하여 검색결과와 순위 결정하는 것이다. Paolucci[8]가 제안한 프로세스 유사도 값은 프로세스의 input 과 output 을 명시하는 개념들의 superclass-subclass 관계를 고려하여 포함 여부를 나타내는 네 가지 카테고리 값으로만 제한되어 있다. 그러나 개념의 계층구조를 따라서 두 개념 사이의 superclass-subclass 관계는 증가될 수 있으며, 이 관계가 증가할수록 두 개념 사이의 유사성은 떨어진다고 보는 것이 합리적이다. 따라서 본 연구에서는 프로세스의 특징 항목들에 대하여 부여된 개념들 사이에 존재하는 superclass-subclass 관계의 개수를 기준으로 개념들 사이의 semantic distance (SD)를 결정하였다.

주어진 두 가지 개념 사이의 semantic distance 결정은 다음의 방법을 따른다. PSI 에 의하여 개념이 서로 일치하는 경우에는 0 의 값을 부여하고, 개념의 계층구조를 따라서 어느 한 개념으로부터 다른 개념으로 찾아가는 최단경로에서 superclass-subclass 관계가 나타날 때마다 1 씩 추가하는 방식으로 semantic distance 의 값을 증가시킨다. 두 개념 사이의 최단경로 도중에 한번이라도 exclusion 관계가 나타나면 두 개념 사이의 semantic distance 의 값을  $\infty$ 로 부여한다.

유사도를 비교하기 위하여 주어진 두개의 프로세스 중 어느 한쪽 프로세스의 특징항목에 대한 개념이 부여되어 있지 않는 경우에는 두 개념 사이의 semantic distance 는 0 의 값을 부여한다. 왜냐하면 유사도를 이용한 검색 방법은 매칭 가능성이 낮은 반응이라도 가급적 검색 결과에 포함시킴으로써 대안의 선택폭을 넓히려는 의도를 가지고 있기 때문이다. 또한 두개의 프로세스가 모두 동일한 특징항목에 대하여 다수의 개념들을 부여한 경우에는, 질의와 반응에 해당하는 각 개념의 조합에 대하여 semantic distance 를 모두 구한 후에, 질의에 부여된 개념들 각각에 대하여 작은 semantic distance 를 갖는 반응 개념의 쌍(pair)들을 선택하여 두 프로세스의 해당 특징항목에 대한 semantic distance 로 각각 부여한다.

Topic map 은 새로운 개념의 도입이나 삭제, 또는 위치의 이동과 같은 계층 구조의 변동에 대하여 매우 유연하게 대처할 수 있으므로 상기한 유사도 산출과정을 구현하는데 적합하다. Topic map 으로 표현

된 프로세스의 개념들에 대한 semantic distance 의 결정 방법을 topic map 의 태그에 각각 대응하면 다음과 같다.

- <subjectIdentity> : SD = 0
- <instanceOf> : SD += 1
- <member> of exclusion association : SD = ∞

#### 4.2 의미 유사도의 계산

주어진 두개의 프로세스  $P_1$  와  $P_2$  에 대하여 N 개의 특징항목 중에서 k 번째의 특징항목의 semantic distance 가  $SD_k$  로 결정되었다면 두 프로세스의 유사도 SI(similarity index)는 다음 식으로 계산된다.

$$SI(P_1, P_2) = 1, \text{ if } \square SD_k = 0 \text{ or } 1$$

$$= 0, \text{ if } \square SD_k = \infty$$

$$= N / \sum_k SD_k, \text{ otherwise}$$

where,  $k = 1, 2, \dots, N$

위 식의 첫 번째 경우는 특징항목별로 비교되는 개념들의 쌍이 모두 서로 일치하거나 적어도 모든 개념들의 쌍이 direct superclass-subclass 관계에 있음을 의미하며, 이는 Paolucci[8]가 제안한 EXACT 매칭 조건에 정확히 부합된다. 특징항목의 개념 쌍에 대한 비교에서 exclusion 관계가 나타나면 질의에 대한 요구조건을 만족시키지 못하는 특징항목이 존재함을 의미하므로 검색 결과에서 제외되거나 혹은 질의에 부여한 개념을 수정할 필요가 있음을 의미한다. 개념들의 쌍에 대한 비교에서 exclusion 관계가 존재하지 않는다면 일반적인 유사도에 기반한 검색 방법들이 도출하는 유사도와 같이 0 과 1 사이의 연속적인 형태의 값으로 유사도를 보여줄 수 있으므로 검색 결과에 대한 ranking 과 같은 후처리 과정에서 유용하게 사용될 수 있다. 단, 동일한 프로세스의 특징항목에 대하여 다수의 개념들을 부여한 경우에는 질의에 부여된 개념마다 각각 semantic distance 를 부여함으로써 질의의 내용이 검색과정에 충분히 반영되도록 해야 한다.

### 5. 비즈니스 프로세스의 검색

새로운 비즈니스 프로세스가 설계되어 프로세스에 대한 명세서가 주어졌을 때 시스템 개발자들은 이 프로세스의 요구사항을 만족시키며 실제 업무에서 실행시킬 수 있는 서비스 또는 소프트웨어 컴포넌트를 검색해야 한다. 본 논문에서는 요구되는 비즈니스 프로세스의 검색에 사용되는 질의(query)와 실행 가능한 서비스의 반응(response)에 대한 표현에 모두 topic map 을 이용하였다. 그리고 topic map 의 merging 기능을 이용하여 개별 프로세스나 서비스에 대한 topic map 들을 하나의 topic map 으로 통합한 후에 프로세스와 서비스의 유사도를 계산하거나 또는 이들을 표현하는 topic 들의 연결 링크를 따라 navigation 함으로써 요구사항에 가장 적합한 서비스를 찾는 방법을 제시하고 있다. 본 논문에서 제시하는 비즈니스 프로세스의 검색 방법은 그림 5 와 같은 과정을 거치게 된다.

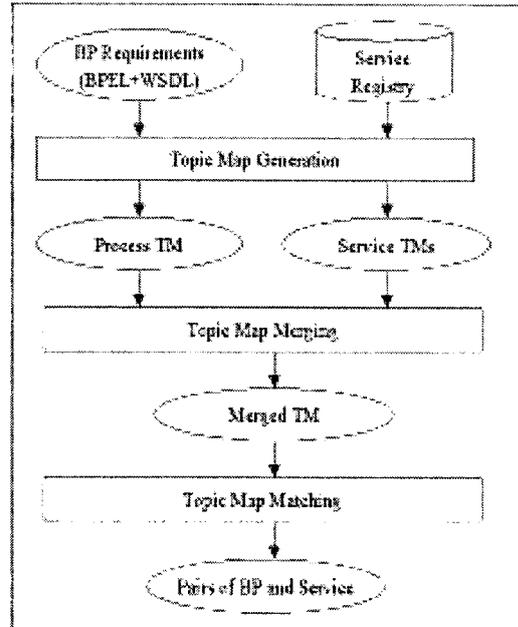


그림 5. 비즈니스 프로세스의 검색 과정

DAML-S[5]의 service profile 에서는 웹 서비스에 대한 name, input, output, precondition, effect, 등과 같이 프로세스의 특징을 표현하는 다양한 항목들이 정의되어 있다. 그리고 웹 서비스의 등록소인 UDDI [7]는 TModel 이라는 데이터 구조의 확장 메커니즘을 제공하기 때문에 DAML-S service profile 의 항목들은 UDDI 의 BusinessEntity 와 BusinessService 의 항목들과 1:1 매핑이 가능하다[18]. 따라서 비즈니스 프로세스의 검색과정에서 프로세스에 대한 질의와 반응 사이의 유사성을 보다 정확히 판단하기 위해서는 확장된 UDDI 와 같은 서비스 저장소에 등록되어 있는 정보를 최대한 활용할 수 있어야 한다.

#### 5.1 Topic Map 작성

비즈니스 프로세스의 요구사항은 프로세스 설계자가 작성한 BPEL 또는 EDOC 문서에서 참조하는 WSDL 문서의 <portType> 태그 안에 담겨 있다. 이 내용을 분석한 후에 프로세스의 특징항목에 대한 개념을 계층구조에서 선택하여 topic map 으로 작성해야 한다. 프로세스의 특징항목을 표현하고 있는 WSDL 문서와 이로부터 추출된 프로세스의 특징항목을 표현한 topic map 문서의 일부 예가 각각 그림 6 과 그림 7 에 나타나 있다.

```

<portType name="purchaseOrderPT">
  <operation name="sendPurchaseOrder">
    <input message="pos.FOMessage"/>
    <output message="pos.InvMessage"/>
    ...
  </operation>
</portType>
  
```

그림 6. WSDL 문서의 프로세스의 특징항목 표현 예

```

<Topic id="asnPurchaseOrder">
  <baseName>
    <baseNameString>asnPurchaseOrder</baseNameString>
    <baseName>
  </baseName>
  <Topic>
    <Characterization>
      <InstanceOf><TopicRef link href="#procMainInput"></InstanceOf>
    </InstanceOf>
    <Member>
      <CodeSpec><TopicRef link href="#procMainObject"></CodeSpec>
      <TopicRef link href="#asnPurchaseOrder">
    </Member>
    <Member>
      <CodeSpec><TopicRef link href="#procMainOutput"></CodeSpec>
      <TopicRef link href="#procMainInputBaseOrder">
    </Member>
    <Association>
      <Characterization>
        <InstanceOf><TopicRef link href="#procMainOutput"></InstanceOf>
      </InstanceOf>
      <Member>
        <CodeSpec><TopicRef link href="#procMainObject"></CodeSpec>
        <TopicRef link href="#asnPurchaseOrder">
      </Member>
      <Member>
        <CodeSpec><TopicRef link href="#procMainObject"></CodeSpec>
        <TopicRef link href="#procMainInvoice">
      </Member>
    </Association>
  </Topic>
</Topic>
  
```

그림 7. 프로세스의 특징항목의 Topic Map 작성 예

### 5.2 Topic Map Merging

Topic map 의 중요한 특징 중의 하나는 여러 개의 topic map 들이 하나의 topic map 으로 merge 되는 메커니즘을 제공한다 는 것이다. 즉, 서로 다른 분야의 지식을 담고 있는 topic map 들이 통합되면서 웹에 퍼져 있는 다양한 정보자원에 대한 인텍싱 역할을 할 수 있는 것이다[14]. Topic map 에 포함된 topic 들이 동일한 사물을 지칭하고 있다면 새로운 topic 을 하나 생성하여 해당 topic 들이 표현하고 있던 정보를 새로운 topic 에 저장한 후에 해당 topic 들을 제거 하는 것이 topic merging 의 원리이다. Topic map merging 이전과 이후의 topic map 구성 상태에 대한 예를 각각 그림 8 과 그림 9 에 나타내었다.

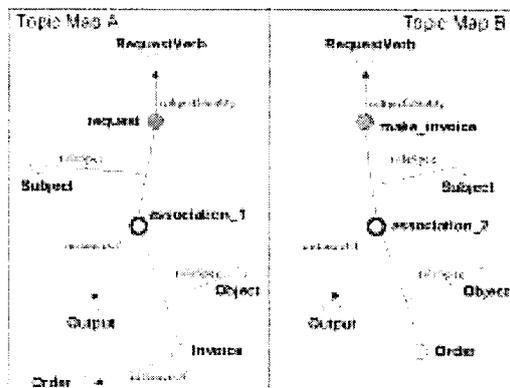


그림 8. Merge 이전의 Topic Map

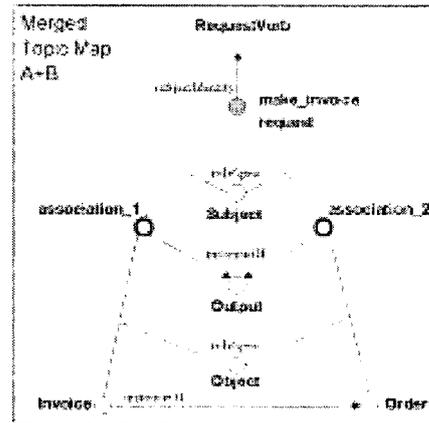


그림 9. Merge 이후의 Topic Map

### 5.3 Topic Map Matching

Topic map engine 은 topic map 을 parsing 한 후 그 정보를 topic map 구성요소의 종류에 따라 별도의 테이블로 관리하므로, merging 을 통하여 프로세스와 서비스에 대한 topic map 들을 하나의 merged topic map 으로 재구성하면 프로세스의 유사도 계산 과정을 효과적으로 처리할 수 있다.

질의에 대한 프로세스 topic map 과 반응에 대한 서비스 topic map 이 통합되어 merged topic map 이 만들어지면 우선 요구사항이 담긴 프로세스 topic map 에서 subject 역할을 담당하는 프로세스 subject topic(이하 PST)을 찾는다. PST 가 발견되면 서비스 subject topic(이하 SST)에 대하여 개념 사이의 semantic distance 를 결정하고 유사도를 계산한다. 유사도 계산이 모두 끝나면 유사도 값에 따라 SST 들의 순위를 부여하여 프로세스 요구자에게 검색 결과를 보여주게 된다. 이때 검색 결과도 역시 topic map 으로 만들어지므로 서비스의 input, output, precondition, effect 등과 같은 프로세스의 특징항목들이나 이에 부여된 개념들이 연결되어 있는 링크를 따라 merged topic map 을 navigation 해 가면서 다양한 대안들을 검토할 수 있다는 장점이 있다. 또한 본 논문에서 제시하는 검색 방법은 Lee[1]의 검색방법에서 요구하는 프로세스의 특징항목에 따른 유사도의 한계값의 입력이 불필요하다.

## 6. 구현

본 논문에서는 OAGIS[3] release 8.0 에 나타나있는 개념 중 verb(그림 3 참조)와 noun(그림 10 참조)에 대한 계층구조의 일부를 이용하여 비즈니스 프로세스와 서비스를 표현하였다. 비즈니스 프로세스의 검색을 위한 프로토타입 시스템은 J2EE[19]를 플랫폼으로 그림 11 과 같은 3-tier로 구성되어 있다.

애플리케이션 중 topic map parsing 과 merging 모듈은 open source 로 제공되는 topic map engine 인 TM4J[20]를 수정하여 subject-based merging rule 을 적용하였으며, 프로세스의 유사도 계산을 수행하는 Topic Map Matchmaker 와 Taxonomy Reasoner 는 java

로 구현되었다. 프로세스 topic map 은 시스템 개발자가 BPELWS 과 WSDL 문서의 요구사항을 분석하여 작성하고 서비스 topic map 은 서비스 제공자가 작성하는 것으로 가정하였다.

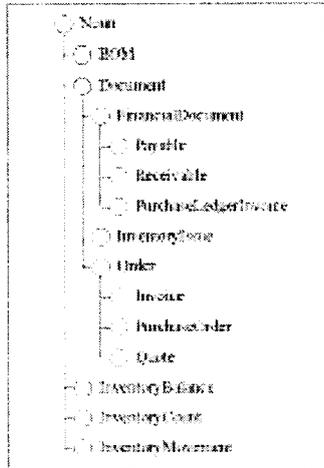


그림 10. 개념 계층구조의 예

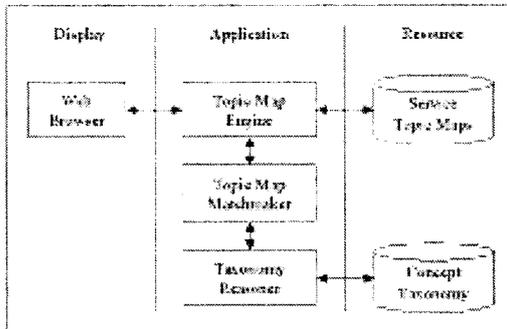


그림 11. 시스템 아키텍처

그림 10 에 예시된 비즈니스 프로세스에 관련된 noun 의 계층구조를 참조하여 input 과 output 항목에 대한 프로세스 요구사항과 이를 실행할 수 있는 서비스에 대한 검색결과가 그림 12 에 예시되어 있다.

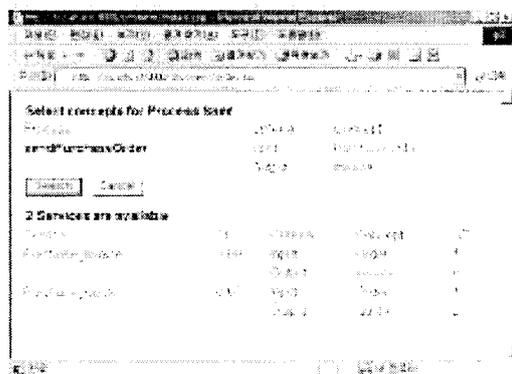


그림 12. 프로세스 검색 결과 예

본 논문에서 제시하는 검색방법은 다음과 같은 장점이 있다. 우선 키워드 매칭에 의한 EXACT 한 검색결과 뿐만 아니라 프로세스를 표현하는 개념 사이의 유사도를 비교하여 CLOSE 한 검색결과를 보여주기 때문에 대안 선택의 범위를 넓혀준다. 또한 기존의 방법들은 프로세스의 특정항목에 대한 개념이 한 가지일 경우만을 고려하고 있는데 제안된 검색방법은 특정항목에 대한 개념이 다수 존재할 경우에도 유사도 계산이 가능하다.

Topic map 을 이용하여 프로세스의 특정항목별 요구자와 제공자가 모두 참조 가능한 개념을 연결시켜 프로세스를 표현함으로써 프로세스 카탈로그 고유의 분류체계에 따른 프로세스의 계층적 구조를 이해하지 못한 상태에서 프로세스에 대한 검색이 가능하다. Topic map 으로 작성된 개념에 대한 계층구조는 개념의 추가, 삭제, 이동과 같은 구조변경에 대하여 유연한 대처가 가능하다. 예를 들어, 상하운동, 수륙양용자동차와 같이 exclusion 관계에 있던 하위 개념들이 통합되어 새로운 의미수준을 생성할 경우 topic map 에서는 이를 지칭하는 topic 을 만들고 <instanceOf>의 연결 상태를 조정해주면 충분하다. 또한 topic map 은 불완전한 정보가 주어졌다 하더라도 사용 가능한 정보를 최대한 활용할 수 있는 구조를 가지고 있으므로 프로세스 검색을 위한 질의 구성이나 신규 서비스의 등록 과정에 편의성을 제공할 수 있다. 검색 결과 화면에 표시된 개념들의 연결링크를 따라가는 navigation 에 의한 정보의 검색방법은 사용자의 관심영역에 보다 집중할 수 있는 기능을 제공해 준다.

Topic map 을 이용한 검색방법의 문제점으로는 질의와 반응에 대한 topic map 을 프로세스 명세서 또는 서비스 저장소의 정보로부터 생성해 주어야 한다는 점이다. 이 과정에는 특정 개념에 대한 의미의 해석 및 개념 사이의 매핑이 필요한데 이를 자동으로 처리해 주는 방법론이 아직 개발되어 있지 않다. 또한 topic map 은 association 에 참여하는 topic 의 최대 개수, 역할에 따라 참여 가능한 topic 의 타입, 등에 대한 제한이 없으므로 topic map 의 일관성을 보장하기 위한 TMCL(Topic Map Constraint Language)이나 DAML+OIL 같은 온톨로지를 적용한 확장이 필요하다[15].

## 7. 결론 및 추후 연구과제

새로운 비즈니스 프로세스에 대한 명세서가 주어졌을 때 시스템 개발자들은 프로세스의 요구사항을 만족시키는 재사용 가능한 서비스 또는 소프트웨어 컴포넌트를 효과적으로 검색할 수 있어야 한다. 요구되는 비즈니스 프로세스의 효과적인 검색방법을 구현하기 위하여 본 논문에서는 프로세스와 서비스의 메타데이터를 모두 topic map 으로 표현하였다. 개념의 분류체계에 따른 semantic distance 를 판단하고 이로부터 유사도를 계산하는 검색방법을 사용함으로써 CLOSE 한 검색결과를 제시할 수 있으며, 검색결과도 topic map 으로 표현되기 때문에 화면에 표시된 topic 들의 연결 링크를 따라 navigation 함으로써 프

로세스의 요구사항에 가장 적합한 서비스를 찾는 것이 가능하다.

Topic map 은 지식표현을 위한 세 가지 assertion 방법을 제공하므로 이를 활용하면 의미적으로 보다 근접한 검색결과를 기대할 수 있다. 또한 유사도 비교 기준에 대한 가중치 부여 방법, 사용자 고유의 개념 분류체계를 사용할 수 있도록 priority 를 고려한 topic map 의 merging 방법, 등에 대한 연구가 필요하다.

### 참고 문헌

1. Juhnyoung Lee, "Matching Algorithms for Composing Business Process Solutions with Web Services", *IBM technical report*, 2002.
2. "Common Business Process Catalog-Technical Specification, Version 0.90", UN/CEFACT eBTWG, [http://www.ebtwg.org/projects/documentation/architecture/eBTWG\\_CBPC\\_0.90.pdf](http://www.ebtwg.org/projects/documentation/architecture/eBTWG_CBPC_0.90.pdf), August 2002.
3. "OAGIS: Open Applications Group Integration Specification, Inc. Release 3.0", Document Number 20020405.
4. "SCOR (Supply Chain Operations Reference) Model, Version 5.0", Supply-Chain Council (SCC), July 2001.
5. "DAML-S 0.5 Draft Release", <http://www.daml.org/services/daml-s/2001/05/>, May 2001.
6. Dieter Fensel, Ying Ding and Borys Omelayenko, "Product Data Integration in B2B E-Commerce", *IEEE Intelligent Systems*, July/August 2001.
7. "UDDI (Universal Description, Discovery and Integration), Version 3.0 Published Specification", <http://uddi.org/pubs/uddi-v3.0.0-published-0020719.htm>, July 2002.
8. Massimo Paolucci, Takahiro Kawamura, Terry R. Payne, and Katia Sycara, "Semantic Matching of Web Services Capabilities", *Proceedings of the 1st International Semantic Web Conference (ISWS2002)*, June 2002.
9. "XML Topic Maps (XTM) 1.0", TopicMaps.Org, <http://www.topicmaps.org/xtm/1.0/xtm1-20010806.html>, August 2001.
10. "Process Specification Language: An Analysis of Existing Representations", NISTIR 6160, National Institute of Standards and Technology, Gaithersburg, MD, 1998.
11. "Business Process Execution Language for Web Services, Version 1.0", <http://www-106.ibm.com/developerworks/library/ws-bpel/>, July 2002.
12. "UML Profile for Enterprise Distributed Object Computing Specification", OMG, Feb. 2002.
13. "Resource Description Framework(RDF) Model and Syntax Specification", W3C Recommendation, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>, February 1999.
14. Lars Marius Garshol, "Living with topic maps and RDF", *Ontopia, forthcoming in proceedings of XML Europe 2003*.
15. Eric Freese, "Using DAML+OIL as a Constraint Language for Topic Maps", *XML2002 Proceedings*, December 2002.
16. "Web Services Description Language (WSDL) 1.1", <http://www.w3.org/TR/2001/NOTE-wsdl-20010315>, W3C, March 2001.
17. Marc de Graauw, "Business Maps: Topic Maps Go B2B", <http://www.xml.com/pub/a/2002/08/21/topicmapb2b.html>, August 2002.
18. Massimo Paolucci, Takahiro Kawamura, Terry R. Payne, and Katia Sycara, "Importing the Semantic Web in UDDI", *Workshop on Web Services, e-Business and the Semantic Web: Foundations, Models, Architecture, Engineering and Applications*, Toronto, Ontario, Canada, May 2002.
19. "Java TM 2 Platform, Enterprise Edition", <http://java.sun.com/j2ee/>.
20. "TM4J-Topic Maps for Java", <http://www.tm4j.org/>, September 2001.