

# 사용자 질의 의미 결정을 위한 새로운 N-ary 개체 관계 디자인 패턴

김수경\*, 안기홍\*\*

\*한국정보통신대학교 공학부

\*\*한밭대학교 정보통신컴퓨터공학부

e-mail:ksk0314@icu.ac.kr, khahn@hanbat.ac.kr

## A New N-ary Entities Relation Approach for User Query Mean Decision

Su-Kyoung Kim\*, Kee-Hong Ahn\*\*

\*Dept of Engineering, Korea Information Communication University

\*\*Dept of Computer Engineering, Hanbat University

### 요 약

본 연구는 웹이나 정보 검색 환경에서 사용자로부터 입력되는 단순한 키워드 형태의 질의가 아닌 문장형태의 질의에 있어 문장이 내포하는 질의의 의미를 결정하여 더 정확한 검색 결과를 제공하기 위해 온톨로지 내 개념들 간의 속성간 연결을 위해 A-Box 기반의 관계 선언과 새로운 N-ary 개체 관계 방법을 제안한다. 특히 개념 개체들 간의 의미를 더 정확히 결정하기 위해 기존의 N-ary 개체 관계 방법이 갖고 있는 속성에 가중치를 포함하는 것이 아니라 가중치에 관련된 새로운 개체를 생성 패턴을 제시하여 특정 개념에 연관된 개념들의 관련성 결정의 성능을 높이도록 하였다. 본 연구의 실험에 있어 사용자가 입력한 병증의 문장을 결정하기 위해, A-Box 기반의 관계 선언과 N-ary 디자인 패턴에 결합하는 지식 도메인 온톨로지 등을 구축하였으며, 이를 통한 실험 결과 문장의 의미에 따른 더 정확한 결과를 보여주었다.

### 1. 서론

최근 웹이나 정보 검색은 시맨틱 웹의 기반 기술들 중 온톨로지의 의미론적 표현과 추론 기능을 이용하여 의미론적 처리와 조회, 검색이 가능한 시맨틱 서비스(Semantic Service)에 많은 관심이 집중되고 있다. 특히 온톨로지를 통해 응용 가능한 많은 기술들 중 언어가 갖는 동음이의적 그리고 이음동의적 어려움을 해결하여 의미론(Semantic)에 입각한 정보 검색 서비스 기술은 미래 인터넷(Future Internet) 분야에서 지능화된 정보 처리 기술의 선점에 있어 중요한 기술이다. 현재 의미론적 정보 검색 서비스 기술은 온톨로지를 기반으로 정보 검색의 정확도를 높이기 위해, 벡터 모델(Vector Model)이나 퍼지 모델(Fuzzy Model)과 같은 정보 검색 모델과 온톨로지를 접목하는 연구들이 진행되고 있고 W3C의 디자인 패턴을 이용해 개념들 간의 관계에 대해 더욱 정량적인 관련성을 부여하기 위해 개념간 속성에 가중치(priority) 주석을 적용하여 개념들 간의 유사도에서 가중치가 높은 것을 더 유사한 개념으로 선택하는 방법이 연구되고 있다.

그러나 W3C의 디자인 패턴을 이용한 속성의 가중치 주

석 적용 방법은 속성의 의미와 가중치의 정보를 n-ary 형

태로 적용함으로써 인해 실제 온톨로지내 표현된 구문을 직접 쿼리에 적용하는데 어려움이 있다. 또한 질의 응답 시스템이나 진단 시스템은 더욱 정확한 응답이나 진단을 제공하기 위해 사용자 질의를 문장형태로 제공하는 것은 기본적인 요소이다. 예를 들어 “배에 물이 차다”는 문장은, 탈 것의 ‘배’에 물이 차는 경우와 사람의 ‘배’에 물이 차는 경우가 있다. 만약 ‘배에 물이 차서 아프다’는 문장이 입력 되었을 때, 질의 또는 진단 시스템은 ‘배’라는 개념과 ‘물’이라는 개념의 연결 술어인 ‘차다’와 ‘아프다’라는 술어가 갖는 개념간의 관계를 추론하여 이 문장에서의 ‘배’는 사람의 배로 판단할 수 있다. 이같은 문장에서 각 단어의 의미 추론을 더욱 명확하게 하기 위해 본 연구는 이같은 A-Box 기반의 관계 추론 규칙에 추가하여 기존 n-ary 형태의 가중치 표현을 새롭게 디자인한 방법을 제안한다.

제한한 방법의 실험을 위해 사람의 신체와 관련된 병증 그리고 관련된 진단 지식을 온톨로지로 구축한 뒤, 다양한 언어학적 관점에 입각한 병증을 입력하여 그에 대한 진단과 처방을 제공하는 실험시스템을 구축하였으며, 관련 문장들의 실험 결과 본 제안 방법이 의미 결정에 있어 더욱 정확한 결과를 제공할 수 있음을 증명하였다.

본 연구의 구성은 다음과 같다. 2장은 현재 온톨로지 기반의 다양한 유사도 연구에 대해 살펴보고 3장은 본 연구

\* 이 연구는 2007학년도 한국학술진흥재단의 “유망여성과학자지원사업”의 지원으로 연구되었음

\* 정 회 원 : 한국정보통신대학교 연구교수

\*\* 중 심 회 원 : 한밭대학교 컴퓨터공학전공 교수

논문접수 : 2008년9월19일, 심사완료 : 2008년10월16일

가 제안하는 A-Box 형태의 관계 속성 기술 방법과 새로운 n-ary 형태의 가중치 표현 디자인을 제안하고 4장은 제안된 방법에 의한 온톨로지 구축과 그를 기반으로 한 실험시스템을 구현하여 제안된 방법의 성능을 실험하고 5장은 연구 결과 및 연구 방향에 대해 기술한다.

## 2. 관련연구

[1]은 용어들 의미론적 유사도와 거리를 측정하기 위한 방법으로 개념의 정보 내용과 코퍼스로 주어진 지식 베이스의 정보 내용을 추출한 CommSpec에 패스 거리(path length)를 이용한 유사도 측정 방법을 제안하였고, [2]는 유사도 계산과 OWL 온톨로지 정렬을 위해 벡터 공간 모델을 제안하였으나 개념들 간의 유사도를 matrix 기반으로 계산함으로써 필요이상의 계산이 요구되는 경우가 있다. 특히 근래는 Vector Space를 이용하여 온톨로지 유사도를 계산하는 연구가 많이 진행되며 [3], [4]는 온톨로지 정렬과 온톨로지 정합을 위해 Vector space를 사용하였다. [5]는 OWL 속성을 이용하여 온톨로지 간의 의미 유사도 측정 방법을 제안하였다. 두 엔티티가 가지는 오브젝트의 집합에 각 술어에 대한 부분 유사도를 적용하는 방법을 제안하였으나 술어에 대한 명확한 관련성의 명시와 일대일 관계만을 정의하여 한 개체가 갖는 다양한 개체와의 관련성을 명시하는 데는 어려움이 있다.

## 3. A-Box 추론 기술 및 새로운 N-ary 디자인 패턴

### 3.1 A-Box 기반의 의미 추론 규칙

본 연구는 의미 추론 규칙의 적용을 위해, 언어학적 특징 중 하나인 동음이의어에 관련된 문장들을 분석한 결과 단어 '배'는 결합 동사나 목적어에 따라 다양한 의미가 포함됨에 주목하였다. '배'는 ① 탈것의 배, ② 신체의 배 ③ 과일인 배 ④ 술이나 음료의 잔 ⑤ 허리를 굽히는 인사 ⑥ 기타 이름 성씨 등을 표현하며, 관련되는 동사(술어) 또한 '먹다', '타다', '갈다', '하다', '울리다', '자르다', '부서지다', '구입하다'.... 등등 상대 개체에 따라 결합되는 동사(술어)가 다양한 특징이 있다.

<표 1> 병증 표현과 관련 의미

병증표현	설명
속이 거북하다	몸이 지루트트하고 괴도워 움직임이 자연스럽지 못하거나 자유롭지 못하다
속이 아프다	몸의 어느 부분이 다치거나 맞거나 자극을 받아 괴로움을 느끼다.
속이 더부룩하다	소화가 잘 안되어 배 속이 거북하다.
속이 메스껍다	먹은 것이 되뉘어 올 것같이 속이 몹시 울렁거리는 느낌이 있다.
메시껍다	[형용사] '메스껍다'의 잘못된 표현.
메시껍다	[형용사] '메스껍다'의 잘못된 표현. 속이 울렁거리다
속이 메스꺼리다	먹은 것이 되뉘어 올 것같이 속이 몹시 울렁거리는 느낌이 있다.
속이 끝다	소화가 잘 안되어 배 속이 편안하지 아니하다
속이 니글니글하다	먹은 것이 내려가지 아니하여 곧 게을 뜻이 속이 매우 메스껍다

특히 문장형 사용자 질의에 대한 의미 결정 후 결과를 제공하기 위해 신체의 '배'와 관련된 병증을 중심으로 구문 규칙과 의미 추론 규칙을 연구한다. 다음 <표1>은 '배'와 유의어인 '속'을 포함한 병증 표현을 정리한 것이다. <표 1>의 의미에서 보여주듯이 '배'와 관련하여 유사한 증상과 의미가 다양하게 표현될 수 있음을 알 수 있다. 이 같은 내용에 따라 증상표현에 따른 용어와 유사한 의미 그리고 관련 신체 부위가 다를 수 있다. <표 3>은 조사에 의해 증상과 신체 부위 그리고 진단에 관련된 용어들의 유사어들에 대한 관계를 정의하는 ABox 기반의 선언적 형식 규칙 정의이다.

<표 2> 유사 의미 관계의 A-Box 선언 형식

선언적 형식	Domain	Range	의미
isSimBody	신체	신체 증상	신체에 대한 유사한 표현
isSimSymptoms	증상	증상	증상에 대한 유사한 표현
hasRelationSim	신체	증상	신체 부위별 관련 증상
isSimDescription	서술문 <sub>1</sub>	서술문 <sub>1</sub>	서술문과 서술문은 유사한 표현
hasDiagnose	서술문	진단문	서술문에 해당되는 진단문

<표 2>의 선언적 형식 중 '배'와 '속'이 유사한 관계임을 선언한 규칙은 [정의1]과 같다.

[정의1]  $isSimBody(신체:배, 신체:속) \dots$

[정의1]은 이행적 속성을 이용하여 다음과 같이 [정의2]와 [정의3] 같은 규칙을 정의할 수 있다.

[정의2]  $isSimBody(신체:배, 신체:속) \wedge isSimBody(신체:속, 신체:위) \rightarrow isSimBody(신체:배, 신체:위).$

[정의3]  $isSimSymptoms(증상:기침, 증상:해수) \wedge isSimSymptoms(증상:해수, 증상:기침) \rightarrow isSimSymptoms(증상:기침, 증상:기침)$

모든 증상은 신체의 부위와 관련되어 있다. 따라서 사용자가 질의문을 '배가 메스껍다'로 서술하였을 때나 '속이 메스꺼리다'를 하였을 때도 유사한 증상으로 결정하여 동일한 진단을 하여야 하므로, 먼저 [정의4]로 신체부위와 관련 증상을 정의하였다.

[정의4]  $hasRelationSim(신체:배, 증상:메스껍다)$

[정의4]의 선언적 규칙은 <표 2>에서 기술한 isSimDescription을 통해 [정의5]와 같이 확장되고 '배가 메스껍다'는 문장은 '속이 메스꺼리다'와 유사하다는 결정을 할 수 있다.

[정의5]  $hasRelationSim(신체:배, 증상:메스껍다) \wedge isSimBody(신체:배, 신체:속) \wedge isSimSymptoms(증상:메스껍다, 증상:메스꺼리다) \rightarrow hasRelationSim(신체:속, 증상:메스꺼리다) \vee isSimDescription(서술, 서술)$

증상에 대한 진단을 위한 추론 규칙은 [정의6]과 같다. 여기에서 "메스꺼리다"는 증상 또는 다른 증상들을 'or' 조건으로 연결하여 조건이 성립하면 "서술", "진단"의 개념으로 연결된다.

[정의6]  $hasRelationSim(\text{신체:속, 증상:배속거리다}) \vee isSimDescription(\text{서술, 서술}) \rightarrow hasDiagnose(\text{서술, 진단})$

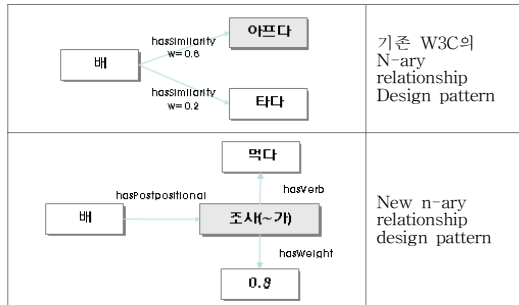
[정의7]  $L_q(C_x, C_y) = \sqrt[q]{\sum_i |C_x - C_y|^q}$

개념간 유사도의 정량적 거리 결정을 위해 본 연구에서 새롭게 제시된 n-ary의 가중치 속성값은 벡터 거리 (Vector distance)를 적용한다. [정의 7]에서 개념  $C_x$ 와 개념  $C_y$ 는 두 개념간 상·하위 또는 선언적 형식으로 정의된 속성이며 각 개념의 거리는 1씩 할당하고  $isSim[C_x, C_y:weight_i]$  곱을 합한 결과로 나타낸다. 개념간의 거리가 계산되었으면 이를 바탕으로 의미 유사도를 결정한다. 의미 유사도는  $L_q$  또는 *Minkowski* 거리 계산법에 의해 두 개념을 벡터로 판단하여 두 점들 간의 끝 거리를 결정한다.

[정의8]  $hasPostpositional[x, \text{활용어온톨로지.조사}] \wedge hasVerb[\text{활용어온톨로지.조사, 활용어온톨로지.동사}] \rightarrow hasWeight[y]$

[정의 8]은 신체 온톨로지와 관련된 활용어의 정확한 선택을 위해 정의한 선언적 규칙으로 입력한 문장에서 조사가 제거되었을 때 관련되는 활용어(중상)를 더욱 명확하게 결정하게 된다.

3.2 새로운 N-ary 디자인 패턴



(그림 1) N-ary 관계 디자인 패턴 비교

(그림 1)은 기존 W3C의 N-ary 관계 디자인 패턴과 새롭게 제안한 n-ary 관계 디자인 패턴을 비교한 그림이다. 위에 '배'는 '아프다'와 '타다'가 동일한 속성명인 'hasSimilarity'에 속성의 주석으로 가중치를 부여하는데 단순히 '아프다'나 '타다'가 어느 단어와 관련하여 그같은 가중치를 갖는지에 대한 명시적인 표현이 없기 때문에 유사도의 결정의 애매함을 갖게 된다. 그러나 새로 제안한 n-ary 관계는 '배'와 관련된 조사들 중 '~가'가 연결될 때 hasVerb 속성은 '먹다'이고 hasWeight 속성은 0.9가 됨을 표현할 수 있다.

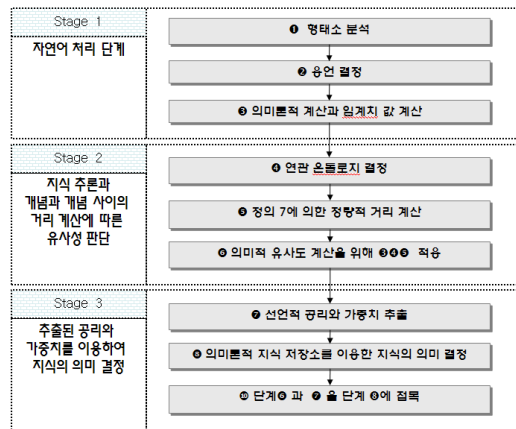
<표 3>은 각 활용어(중상)의 의미유사도를 연결할 기준으로 기본적으로 제시되는 의미추정치와 유사도이다. 각 개념의 유사도는 다음에 새롭게 제시된 n-ary 디자인 패

턴에 의해 의미유사도 판단 임계치와 비교하여 의미유사성을 판단한다. 이때 동사 온톨로지의 각 개념들은 신체 온톨로지나 음식온톨로지 또는 수송기관 온톨로지 등의 각 개념들과 중복되어 사용되는 경우가 많이 있다. 이 때 정확한 의미 판단을 위해 (그림1)에서 제시한 새로운 n-ary 관계에 의해 활용어(동사)가 선택되고 그의 의미를 추정하게 된다.

<표 3> 활용어별 의미추정치/우선도

활용어	조사	의미추정치/의미		우선도			
		신체 Ont	음식 Ont	수송 Ont	기타		
먹다 먹었다 먹어 먹으니 먹는	를 / 을	1	0.9		1		
	가		0.8	1			
	에		0.7				1
	고르다 고파 고르니	가	1		1		1
타다 탔다	에 / 람	1		0.8		1	
	를	1					1
	을	0.8					1
	이	0.6					1
아프다 아파 아프니 아팠다	가/ 이	1		1			
	이	1		1			
쓰리다 쓰러 쓰리니	가/ 이	1		1			
	이	1		1			
	이	0.8					1

(그림 2)는 앞에서 제안한 각 정의의 단계와 N-ary 관계 를 기반으로 <표 3>의 임계치를 이용해 의미 유사도를 결정하는 전체적인 흐름이다.

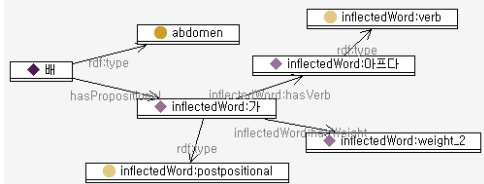


(그림 2) 의미유사도 결정 과정

4. 실험

제안된 방법의 실험을 위해 지식영역을 '배'를 기준으로 5개의 영역으로 구분한 뒤, 각각의 온톨로지를 구축한 뒤 실험시스템을 구현하였다. 먼저, 각 지식영역은 신체온

톨로지, 음식온톨로지, 탈것온톨로지, 활용어온톨로지, 진단온톨로지)로 구성되며, (그림3)은 (그림1)에 의해 제안된 n-ary 관계로 표현된 온톨로지의 일부이다. 또한 (그림4)는 SPARQL을 이용해 (그림 3)에 표현된 관계를 추론한 결과로 ‘배’는 활용어온톨로지의 ‘아프다’는 동사와 연결될 때 가중치는 ‘0.8’임을 cascade inference로 추론하였다.



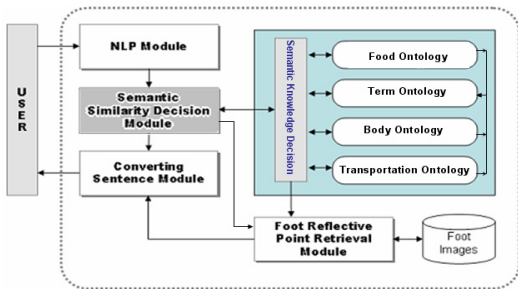
(그림 3) 제안된 방법에 의한 온톨로지 표현

```
SELECT ?x ?z
WHERE {
  ?x :hasPropositional ?y
  ?y inflectedWord:hasVerb ?z }
  [x] z
  배 아프다

SELECT ?x ?k
WHERE {
  ?x :hasPropositional ?y
  ?y inflectedWord:hasWeight ?z .
  ?z rdfs:label ?k }
  [x] k
  배 0.8
```

(그림 4) ‘배’의 연관 동사와 가중치 추론

온톨로지를 이용한 실험시스템은 다음의 시나리오에 의해 구현하였다. “사용자가 병증에 대한 문장을 입력하면 시스템은 그 문장의 의미를 결정하여 그에 대한 처방을 출력하는 시스템이다. 이 시스템은 사용자로부터 입력된 질의 문장을 파싱하는 “자연어 처리 모듈”과 추론 규칙과 n-ary 디자인 패턴에 의해 온톨로지내 유사 의미들을 추출하는 “추론 엔진” 그리고 각 온톨로지 간의 지식 전달 모듈과 결정된 진단에 따른 “발 반사구 이미지 검색 모듈”, 마지막으로 온톨로지에서 전달 받은 진단 내용을 사용자가 이해할 수 있는 문장 형태로 변환하는 “진단 문장 제공 모듈”로 구성된다. 실험 시스템의 구조는 (그림5)와 같다.



(그림 5) 실험시스템 전체 구조

(그림5)에서 사용자로부터 입력된 증상 문장을 처리하기 위한 “자연어 처리 모듈”의 구문 분석은 크게 4가지 언어의 특성으로 구분되며, 다음과 같다.

- ① 일반적인 증상문 처리 : 속이 답답하다, 머리가 아프다

- ② 유의어 처리 : 속이 도올하다(배=속, 도올=답답)
- ③ 모호한 의미 : 위가 쓰리다. 눈 위가 아리다.
- ④ 복합 증상문 처리 : 머리가 아프고 눈이 쑤신다, 속이 쓰리고 가슴이 답답하다 등

언어 특성에 따라 구분된 각각의 문장을 따로 실험한 결과, 모호하거나 동음이의어의 특징을 가진 문장에 있어서도 온톨로지내 표현된 각각의 개념들과 연결되는 조사와 동사들에 부여된 의미도에 따라 입력된 사용자의 질의를 가중치별로 파악하여 가장 적합하게 해당하는 목적어 개념으로 연결하였으며 사용자의 의도를 정확하게 파악함을 보여주었다.

### 5. 결론

본 연구는 웹이나 정보 검색 환경에서 사용자로부터 입력되는 단순한 키워드 형태의 질의가 아닌 문장형태의 질의에 있어 문장이 내포하는 질의의 의미를 결정하여 더 정확한 검색 결과를 제공하기 위해 온톨로지 내 개념들 간의 속성간 연결을 위해 A-Box 기반의 관계 선언과 새로운 N-ary 개체 관계 방법을 제안한다. 특히 개념 개체들 간의 의미를 더 정확히 결정하기 위해 기존의 N-ary 개체 관계 방법이 갖고 있는 속성에 가중치를 포함하는 것이 아니라 가중치에 관련된 새로운 개체를 생성 패턴을 제시하여 특정 개념에 연관된 개념들의 관련성 결정의 성능을 높이도록 하였다. 본 연구의 실험에 있어 사용자가 입력한 병증의 문장을 결정하기 위해, A-Box 기반의 관계 선언과 N-ary 디자인 패턴에 결합하는 지식 도메인 온톨로지 등을 구축하였으며, 이를 통한 실험 결과 문장의 의미에 따른 더 정확한 결과를 보여주었다. 차후 연구에 대한 더 정확한 성능 보장을 위해 실험 문장을 확대할 필요가 있다고 사료된다.

### 참고문헌

- [1] P.Selvi, N.P.Gopalan, "The New Semantic Similarity Measure Using Ontology and Corpus", The Icfai Journal of Computer Sciences, Vol.2, No.1, pp.29-37, Jan. 2008
- [2] Zahra Eidoon, Nasser Yazdani, Farhad Oroumchian, "Ontology Matching Using Vector Space", ECIR 2008, LNCS 4956, pp.472-481, 2008
- [3] Ce Zhang, Yu-Jing Wang, Bin Cui, Gao Cong, "Semantic Similarity Based on Compacc Concept Ontology", WWW2008, April, 2008, Beijing, China
- [4] Rajesh Thiagarajan, "Computing Semantic Similarity Using Ontology", ISWC 2008.
- [5] 안우식, 박정은, 오경환, "OWL 속성을 이용한 온톨로지 간 의미 유사도 측정 방법", 2006 한국컴퓨터종합학술대회논문집 Vol.33, No.1(B)