

객체지향 시소러스에서의 참조 질의 조건 완화 기법

김정애⁰ 박종민 김원중 양재동
전북대학교 컴퓨터정보학과
(jakim,jmpark,wjkim,jdyang)@cs.chonbuk.ac.kr

Relaxing Reference Queries in the Object-Oriented Thesaurus

Jung-Ae Kim⁰ Jong-Min Park Won-Jung Kim Jae-Dong Yang
Dept. of Computer Science, Chonbuk National University

요 약

현재의 정보검색 시스템에서 사용자는 관련 분야의 지식을 가지고 정확한 질의를 작성해야만 원하는 결과를 얻을 수 있다. 이 문제를 해결하기 위하여, 본 논문에서는 객체지향 시소러스에서의 참조 질의 조건 완화를 이용한 지능적인 질의 처리 기법을 제안한다. 참조 질의 조건 완화 기법은 객체지향 시소러스의 구조적인 특징을 이용하여 질의 조건을 일반화 시킨다. 따라서 사용자는 정확한 질의를 제공하지 않아도 정확한 결과 뿐만 아니라 유사하거나 관련 있는 정보를 얻을 수 있다.

1. 서 론

정보검색 시스템에서 사용자는 원하는 정보를 얻기 위해 질의를 작성하지만, 항상 사용자 요구조건을 만족하는 답을 얻지는 못한다. 그 주된 이유의 하나는, 현재의 질의 처리 과정에서 사용자는 관련 분야의 지식을 정확한 질의를 작성해야만 원하는 답을 얻을 수 있기 때문이다. 사용자의 이러한 부담을 덜어주기 위하여 질의 자동생성 도구들을 사용하지만 주어진 질의를 관련 분야의 지식과 데이터 구조에 적합하게 조정하는 기능만을 수행한다. 그러므로, 사용자 질의에 포함된 질의어가 적합하지 않을 때 기존 질의 처리 과정과 마찬가지로 사용자에게 유의한 다른 정보를 제공하지 못한다. 이러한 문제를 해결하기 위하여 사용자의 질의를 분석하여 질의의 조건을 일반화 시키는 질의 완화(Query Relaxation) 기법[2]이 필요하다.

본 논문에서는 객체지향 시소러스에서의 참조 질의 조건 완화를 이용한 지능적인 질의 처리 기법을 제안한다. 객체지향 시소러스란 객체지향 패러다임을 적용하여 시소러스를 효율적으로 구축하고 관리할 수 있으며, 상속 메커니즘을 통해 개념들간의 관계를 구조적으로 파악할 수 있도록 해준다. 본 논문에서는 객체지향 시소러스의 상속 성질에 착안하여, 사용자 질의에 포함된 질의 조건을 더 일반적인 질의 조건으로 완화하여 질의를 재 정의 한다. 따라서 사용자는 정확한 질의를 제공하지 않아도 정확한 결과 뿐만 아니라 유사하거나 관련 있는 정보를 얻을 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구에 관하여 기술하고, 3장에서는 객체지향 시소러스에서의 참조 질의에 대해 간략히 설명하고, 4장에서는 본 논문에서 제안한 참조 질의 완화 기법에 대해 구체적으로 설명한다. 그리고 5장에서는 데이터베이스에서의 질의 조건 완화 방법에 대해 알아본다. 마지막으로 6장에서는 결론 및 향후 과제를 제시한다.

2. 관련 연구

질의 조건 완화를 지원하기 위해서는 관련분야 데이터베이스의 데이터와 추상적인 개념사이의 의미적인 관계를 나타내기 위한 지식베이스의 지식 표현 방법이 중요한 역할을 한다. 따라서, 지식 표현 방법에 따른 질의 완화 기법에 대한 연구들에 대해 살펴보도록 한다.

규칙 기반 접근 방법(Rule-based approach)[3,4]은 1차 술어 논리로 데이터 사이의 의미적 관계와 무결성 제약을 표현한다. 질의 처리는 경합 해결(Conflict resolution)이나 추론 메커니즘을 통해 얻게 되고, 질의 조건 완화는 무결성 제약을 통해 조정한다. 그러나 이 방법은 질의 조건 완화 처리 과정에 체계적인 구조가 없으며 직관적이지 못하다는 단점이 있다.

의미 거리 기반 접근 방법(Semantic distance approach)[5,6]은 데이터들 간의 유사도를 표현하기 위하여 수학적 거리를 정의한다. 이 방법은 데이터 간의 양적인 거리 계산이 쉽기 때문에 질의 조건 완화 알고리즘 개발이 쉽지만, 양적 또는 질적인 데이터를 일관된 형태의 양적인 수치 값으로 변화 시키는 것이 어렵고 다양한 데이터 간의 의미적인 유사성을 평가하기 위한 객관적인 기준이 부족하다.

개념 클러스터링 기반 접근 방법(Conceptual clustering approach)[7]은 데이터 사이의 추상화 방법을 적용하여 클러스터링 계층(Clustering hierarchy)을 구성하고 클러스터링 계층을 이용하여 질의 조건을 일반화 시킨다. 이 방법은 클러스터링 계층을 구축하고 관리하는데 있어 유연하지 못하다.

3. 객체지향 시소러스와 참조 질의

객체지향 시소러스에서 개념과 인스턴스는 기본적으로 객체로 간주된다. 개념 객체는 수직적으로 많은 하위 개념들과 일반화 관계를 가지며, 인스턴스들과는 클래스화 관계를 가진다. 또한, 하나의 객체와 수평적으로 집성화/연관화 관계를 가지는 객체들은 속성 객체로 정의되며, 상위 개념에 대한 속

성 객체들의 정보는 하위 객체에 상속되는 성질을 가진다[1].

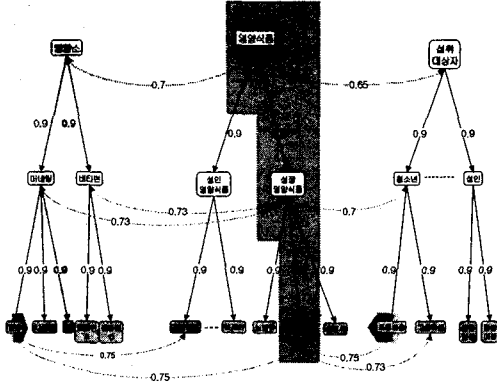


그림 1 객체지향 시소러스와 참조질의

객체지향 시소러스에서의 참조 질의는 $Q = [Main-Query, Restriction-Query]$ 와 같이 두 개의 부분 질의로 구성된다. 'Main-Query'는 객체기반 시소러스의 여러 개념 계층들 중 특정 계층의 일부분으로 사용자의 브라우징 범위를 한정하며, 'Restriction-Query'는 'Main-Query'의 범위를 더욱 제한하기 위해 사용된다[6]. 예를 들어, 그림 1의 시소러스에서, 참조 질의 [영양식품, 칼슘 AND 초등학생]은 '초등학생'의 성장에 도움을 주는 '칼슘' 성분을 함유한 '영양식품'과 관련된 객체를 사용자가 요구하고 있음을 의미한다. 질의에 대한 추론은 '영양식품'에 포함되는 모든 객체들 중에서 '칼슘' 그리고 '초등학생' 성분과 동시에 관계를 가지는 '광동기엔지'를 얻는다. 하지만, 'Restriction-Query'에서 사용되는 객체들이 'Main-Query'범위 안의 객체들과 관계가 없는 경우에 사용자는 어떤 결과도 얻지 못하는 문제점이 있다.

4. 객체지향 시소러스에서의 참조 질의 조건 완화

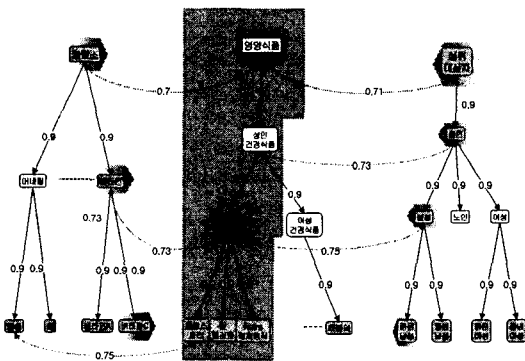


그림 2 [영양식품, 비타민C AND 젊은 남성]질의조건완화 그래프

이 장에서는 사용자가 원하는 결과를 제시하기 위한 객체지향 시소러스에서의 참조 질의 조건 완화에 대해 설명한다. 그림 2의 시소러스에서, 참조 질의 $Q = [영양식품, 비타민C \text{ AND } 젊은 \text{ 남성}]$ 은 '젊은 남성'들에게 '비타민C'를 보충해주는 '영양식품'을 사용자가 요구하고 있음을 의미한다. 그러나 객체 '비타민C'와 객체 '젊은 남성'과 관계를 가지는 객체가 없으므로 질의에 대한 결과는 없다. 위

의 질의를 참조 질의 완화 방법을 이용하면 다음과 같은 과정을 통해 원하는 결과나 관련 있는 결과를 추론해 낼 수 있다.

먼저, 'Restriction-Query'에 포함된 객체, '비타민C'와 '젊은남성'의 일반화 객체를 구하여 각각 S_1, S_2 라 하자.

$$S_1 = Super(비타민C, \alpha) = \{비타민C/1.0, 비타민/0.9, 영양소/0.8\}$$

$$S_2 = Super(젊은남성, \alpha) = \{젊은남성/1.0, 남성/0.9, 성인/0.81, 섭취대상자/0.73\}$$

여기서, α 는 일반화 임계 값을 말한다.

주어진 질의 Q 를 S_1, S_2 를 이용하여 조건 완화하여 재정의된 질의 Q' 는 다음과 같이 정의할 수 있다.

$$Q' = [영양식품, S_1 \text{ AND } S_2]$$

일반 객체지향 시소러스에서의 질의 처리 과정을 이용하여 S_1, S_2 의 관계 객체 R_1, R_2 를 구한다.

$$R_1 = Relation(S_1) = \{남성건강식품/0.66, 영양식품/0.57\}$$

$$R_2 = Relation(S_2) = \{남성건강식품/0.68, 성인건강식품/0.59, 영양식품/0.52\}$$

마지막으로, R_1, R_2 에 AND 연산자를 적용하면 질의 조건이 완화된 질의 Q' 는 {'남성건강식품'/0.65, '영양식품'/0.51}을 결과로 얻게 된다. 여기서, '영양식품'은 '남성건강식품'의 상위 객체이므로 조건 완화 과정에서 좀 더 구체적인 정보를 담고 있는 하위 객체인 '남성건강식품'을 선택한다. 따라서 '남성건강식품'을 중심으로 한 {남성건강식품, 콜레스 프리, 단텔리온, Natra 발효 생식, 성인건강식품}이 결과로 추론 된다. '남성건강식품'을 추론하기 위해 사용되는 관련 정도 값의 계산은 객체지향 시소러스에서 기본적으로 사용되는 알고리즘을 사용한다.

다음은 위의 방법을 알고리즘 형태로 기술한 내용이다.

```

MODULE RelaxQuery (Qm, Qr, α, β);
/*
Qm: Main-Query, Qr: Restrict-Query
α: Restrict-Query에서 일반화 임계 값
β: 질의 결과 객체 임계 값
*/
BEGIN
/* Qr 질의 문에 명시된 객체 집합 */
T = T1, T2, ..., Tn ∈ Qr;

/* 임계 값 α 이내의 Ti의 일반화 객체 */
WHILE(Ti) DO
    Si = Super(Ti, α);
END; /* WHILE */

/* 일반화된 객체 Si를 이용한 질의 재정의 */
Q' = SubstituteQuery(Qm, T, S);
/* Restrict-Query 처리 */
resultQ' = ProcessRestrictQuery(Q');
    
```

```

/* 시소러스 계층을 이용한 불필요한 객체 제거 */
resultQr = removeRedundant ( Qr );

/* Main-Query 처리 */
resultQm = ProcessMainQuery ( Qm );

/* 임계 값 β 이내의 최종 결과 */
Solution = EvaluateQuery ( resultQm, resultQr, β );

END; /* RelaxQuery */
    
```

그림 3 질의 조건 완화 알고리즘

여기서, *Super*, *ProcessRestrictQuery*, *ProcessMainQuery*, *EvaluateQuery* 은 기존 객체지향 시소러스에서 질의 처리에 사용되는 모듈을 그대로 사용한다.

5. 데이터베이스에서의 질의 조건 완화

이 장에서는 제안한 객체지향 시소러스에서의 질의 조건 완화 기법을 이용해서 실제 데이터베이스에서 데이터를 지능적으로 처리할 수 있는 방법에 대해 설명한다. 먼저, 그림 4 는 영양식품에 대한 정보를 담고 있는 테이블과 테이블의 각 필드에 해당하는 객체지향 시소러스의 계층도를 나타낸다.

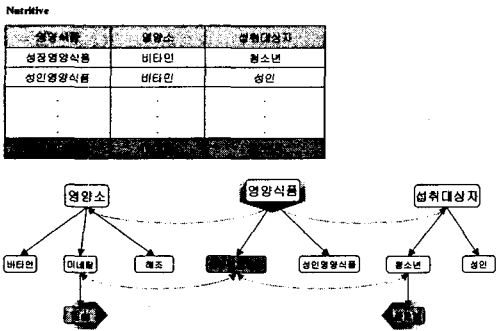


그림 4 영양식품 테이블과 객체지향 시소러스 계층도

다음과 같은 질의가 주어졌을 경우, 기존 데이터베이스로는 사용자에게 원하는 결과를 보여주지 못한다.

```

SELECT 영양식품
FROM Nutrive
WHERE 영양소 = '칼슘' AND 섭취대상자 = '중학생';
    
```

여기서, 객체지향 시소러스를 이용한 참조 질의와 관련 시켜 보면, 'FROM' 질은 Main-Query, 'WHERE' 질은 Restriction-Query 로 볼 수 있다. 따라서, 참조 질의 조건 완화 기법을 적용시키면, 조건 질의 [영양소 = '칼슘']은 [영양소 = '칼슘' OR '미네랄']로 재작성 될 수 있고, [섭취대상자='청소년']은 [섭취대상자 = '중학생' OR '청소년']으로 확장되어진다. 결국, 레코드 (성장영양식품, 미네랄, 청소년)을 결과로 사용자에게 제시할 수 있다.

다음으로, 일반적인 데이터베이스에서의 제약사항도 질의 조건 완화 방법을 통해 좀 더 유연하게 처리될 수 있다. 예를 들어, 그림 5 에서 영양식품 테이블의 영양소 필드와 섭취대상자 필드는 각각

영양소 테이블의 영양소 필드, 연령별 테이블의 대상자 필드와 참조제약관계가 있다. 영양식품 테이블에 레코드 (성장영양 식품, 칼슘, 초등학교)를 삽입하면 기존 데이터베이스에서는 참조 무결성에 의해 레코드간의 용어를 비교하여 정확히 일치하지 않으므로 삽입할 수가 없다. 이런 제약조건을 시소러스계층을 통해 초등학교를 청소년으로, 칼슘을 미네랄로 조건을 완화하면 보다 직관적이고 유연하게 제약사항을 조정할 수 있다.

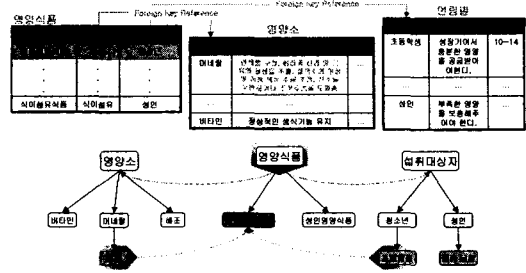


그림 5 영양식품 테이블에서의 참조 무결성 완화

6. 결론 및 향후 연구 과제

본 논문에서는 객체지향 패러다임을 적용한 객체지향 시소러스에서의 참조 질의 조건 완화를 이용한 지능적인 질의 처리 기법을 제안하였다. 사용자 질의에 포함된 질의 조건을 더 일반적인 질의 조건으로 완화하여 질의를 재 정의하여 정확한 질의를 제공하지 않아도 정확한 결과 뿐만 아니라 유사하거나 관련 있는 정보를 얻을 수 있다.

향후 연구 과제로, 더 많은 실험을 통하여 객체지향 시소러스에서의 참조 질의 조건 완화에서 사용되어지는 관련 정도 계산 알고리즘을 검토 및 재 작성해야 하며, 실제 데이터베이스와 통합하기 위한 정확한 방법론이 연구되어야 한다.

참고문헌

- [1] 최재훈, 김기현, 양재동, "객체기반 시소러스 시스템의 설계 및 구현: 반자동화 방식의 구축, 추상화 방식의 개념 브라우징 및 질 의기반 참조 " 정보과학회 논문지(데이터베이스), Vol. 27, No. 1, pp.64-78, March, 2000.
- [2] Gaasterland, T. "Cooperative answering through controlled query relaxation", IEEE Expert [see also IEEE Intelligent Systems], Volume: 12 Issue: 5, Sept.-Oct. 1997
- [3] L. Cholvy, R. Demolombe, Queryng a Rule Base, in: Proceedings of the First International Conference on Expert Database System, 1986, pp.621-643.
- [4] A. Hernerly, M. Casanova, A.Furtado, (Exploiting user models to avoid misconstruals), Nonstandard Queries and Nonstandard Answers, Oxford, 1994, pp73-98
- [5] T. Ichikawa, ARES: a relational database with the capability of performing flexible interpretation of queries, IEEE Transactions on Software Engineering SE-12 (5) 1986, pp624-634.
- [6] A. Motro, User interface to relational database that permits vague queries, ACM Transactions on Office Information Systems 6 (3) 1988, pp187-214.
- [7] W.Chu, Q.Chen, Structured approach for cooperative query answering, IEEE Transactions on Knowledge and Data Engineering 6 (5) 1994 pp738-749.