

객체기반 시소러스 시스템의 설계 및 구현: 반자동화 방식의 구축, 추상화 방식의 개념 브라우징 및 질의기반 참조

(Design and Implementation of an Object-Based Thesaurus System: Semi-automated Construction, Abstracted Concept Browsing and Query-Based Reference)

최재훈[†] 김기현[†] 양재동^{**}

(Jae-Hun Choi) (Ki-Heon Kim) (Jae-Dong Yang)

요약 본 논문에서는 객체지향 패러다임을 적용함으로써 도메인 종속적인 시소러스를 효율적으로 구축하고 관리할 수 있는 객체기반 시소러스 시스템을 설계하고 구현하였다. 이때, 객체지향 패러다임은 시소러스의 구축, 개념 브라우징 그리고 질의기반 참조 기능에 적용된다. 이 시스템에서 객체지향 패러다임의 상속 메커니즘은 시소러스에 표현된 개념들간의 관계를 구조적으로 파악할 수 있게 하여 전문가가 시소러스를 반자동 방식으로 구축할 수 있도록 지원한다. 특히, 방대한 시소러스를 여러 전문가들이 서로 다른 호스트에서 구축할 경우, 이 메커니즘에 의해 파악된 정보는 시소러스의 의미적 일관성을 유지시킬 수 있도록 도와주며, 전문가가 직접 개념들간의 관련 정도를 모두 명시해야하는 부담을 최소화할 수 있다. 객체기반 시소러스 시스템은 또한 질의기반 참조 기능과 추상화 방식의 개념 브라우징 기능을 제공한다. 이 기능들은 검색 질의에 이용될 시소러스 개념들을 사용자가 사전에 탐색해 봄으로써 쉽게 검증할 수 있게 한다. 특히, 이 질의 검증 과정은 높은 정확률을 요구하는 도메인에 적절히 이용될 수 있다.

Abstract In this paper, we design and implement a system for managing domain specific thesauri, where object-oriented paradigm is applied to thesaurus construction, concept browsing and query-based reference. This system provides an object-oriented mechanism to assist domain experts in constructing thesauri; it determines a considerable part of relationship degrees between terms by inheritance and supplies domain experts with information available from a thesaurus being constructed. This information is especially useful to enforce consistency between the hierarchies of a thesaurus, each constructed by different experts in different sites through cooperation. It may minimize the burden of domain experts caused from the exhaustive specification of individual relationship. This system also provides an abstracted browsing and a query-based reference, which allow users to easily verify thesaurus terms before they are used in usual boolean queries. The verification is made by actively searching for them in the thesaurus. Reference queries and abstracted browsing views facilitate this searching. The facility is indispensable especially when precision counts for much.

1. 서론

정보 검색 시스템에서 시소러스는 일반적으로 질의 처리 과정에 이용되어 검색 성능을 향상시킬 수 있게 한다[1,2,3]. 즉, 사용자 질의를 검색에 적합한 형태로 변형하거나 확장하기 위해 시소러스를 이용함으로써 시스템의 정확률과 재현률을 향상시키게 된다[4,5]. 일반적으로 시소러스는 노드와 링크로 구성된다. 여기서, 노

· 본 논문은 전북대학교 영상정보통신기술연구소의 지원으로 수행되었음.

† 비 회 원 : 전북대학교 컴퓨터과학과

jhchoi@cs.chonbuk.ac.kr

khhkim@cs.chonbuk.ac.kr

** 정 회 원 : 전북대학교 컴퓨터과학과 교수

jdyoung@jiri.chonbuk.ac.kr

논문접수 : 1999년 2월 10일

심사완료 : 1999년 12월 8일

드는 개념을 표현하며, 링크는 이 개념들 사이의 의미 관계와 그 관련 정도를 표현한다. 두 개념들 사이의 의미 관계는 주로 상의어(BT: Broader Term), 하의어(NT: Narrower Term) 그리고 관련어(RT: Related Term) 관계 등을 사용하여 나타낸다[1,2,3,4].

최근, 검색의 정확률 및 재현률이 검색의 속도 못지 않게 중요해짐에 따라 시소러스를 구축하고 효과적으로 이용할 수 있게 하는 시스템에 관한 많은 연구들이 진행되고 있다[6,7]. 특히, [2,5,8,9]에서는 응용 도메인에 관련된 문서들로부터 개념들의 동시 출현 빈도를 분석해서 네트워크 구조의 시소러스를 자동으로 구축하고 이들을 질의 처리에 이용하고 있으며, [3,4,10,11]에서는 도메인 전문가가 ISO(International Standardization Organization) 2788과 5964 표준에 따라 계층 구조의 시소러스를 직접 구축하고 지능적인 추론을 통해 사용자가 원하는 시소러스 개념들을 추출하여 검색에 이용하고 있다. 이 시스템들은 기본적으로 시소러스 구축 및 유지 기능, 개념 탐색 기능 그리고 검색 시스템과의 연동 기능을 제공한다[6,7].

전자의 시스템들에 의해 구축된 시소러스들은 매우 복잡한 구조를 가지는 반면에 개념들 사이의 정확한 의미 관계를 표현할 수 없다는 문제점을 가지고 있다[3,12]. 따라서, 사용자가 시소러스로부터 원하는 개념을 탐색하기 어려우며, 질의 처리 과정에서 단순히 개별적인 탐색어만을 고려하기 때문에 주변의 관련된 개념들을 전체적으로 연관시켜 이용할 수 없다는 단점이 있다. 이 단점을 보완하기 위해 [9,13,14]에서는 모든 탐색어들과 전체적으로 관련된 시소러스 개념을 높은 관련 정도로 질의에 참여시키는 방법이 연구되었다. 그러나, 네트워크 시소러스의 부족한 의미 표현력의 문제 때문에 만족할 정도의 성능 향상을 보이지 못하고 있다[4,12].

후자의 시스템들은 전문가가 시소러스를 구축하는 동안 일부 무결성을 제한적으로 유지시켜 준다[6,7]. 그러나 이들은 구축에 너무 많은 비용을 요구하며, 서로 연관된 개념들 사이의 일관성 있는 관련 정도를 명시하기 어려운 단점을 가지고 있다[5,12,15]. 이를 해결하기 위해 [15]에서는 객체지향 패러다임을 시소러스에 적용함으로써 도메인 지식을 보다 구조적으로 표현할 수 있는 객체기반 시소러스를 제안하였다. 특히, 이 시소러스의 구축 및 유지에 객체지향 패러다임의 상속 성질이 효과적으로 활용될 수 있도록 설계되었다. 즉, 객체기반 시소러스는 개념을 객체로 간주하고 기존에 BT/NT로 표현되었던 객체들 사이의 의미 관계를 일반화 또는 클래스화 관계로 정의하여 객체 계층을 형성하도록 하였다.

또한, 다양한 의미로 해석되던 RT를 집성화 관계나 연관화 관계로 세분화하여 정의하였으며, 객체 계층에서 상위 객체와 집성화/연관화 관계를 가지는 객체들은 속성 객체로 작용하여 하위 객체들로 상속되는 구조로 되어 있다.

본 논문에서는 이 객체기반 시소러스 시스템을 보다 상세히 설계하였으며, 이를 바탕으로 Solaris 2.5 상에서 구현하였다. 특히, 본 논문은 1)개념들 사이의 관계와 그 관련 정도를 기존의 방법보다 일관되고 정확하게 명시할 수 있는 구축 메커니즘과 2)사용자가 탐색을 원하는 시소러스 개념들을 직접 접근할 수 있는 질의기반 참조 메커니즘을 상세 설계하고 구현하였으며, 3)서로 연관된 개념들을 추상화된 관점에 따라 순차적으로 브라우징할 수 있는 기능 역시 구현하였다는 점에서 [15]와 차별화된다.

본 논문의 구성은 다음과 같다. 먼저, 2장에서는 기존의 시스템에서 사용하는 시소러스 구축과 참조 방법에 관한 기존의 연구들을 살펴보고, 3장에서는 객체기반 시소러스의 표현 방법과 이들의 성질을 설명한다. 4장에서는 객체들 사이에 존재하는 의미 관계들이 어떻게 구축되고 유지되는지를 보이며, 5장에서는 질의를 기반으로 시소러스를 직접 참조할 수 있는 접근 방법을 기술한다. 6장에서는 객체기반 시소러스 시스템의 구현 내용을 설명하며, 마지막으로 7장에서는 결론 및 향후 연구 방향에 대해 기술한다.

2. 관련 연구

정보 검색 시스템의 높은 재현률과 정확률을 위해 최근 시소러스를 효율적으로 구축하고 이용할 수 있게 하는 시소러스 시스템에 관한 많은 연구들이 수행되고 있다[1,2,3,16]. 이 시스템들에 의해 구축되는 시소러스는 그 구조에 따라 네트워크 시소러스와 계층화된 시소러스로 구분된다[8,9,17]. 각각에 대한 구축 시스템과 구축된 시소러스 특징은 다음과 같다.

네트워크 구조의 시소러스를 구축하는 대표적인 시스템으로는 “PhraseFinder”, “RBT” 그리고 “SSOM” 등이 있으며, 이들은 일반적으로 개념들 사이의 연관성을 관련 정도로 표현한다. 여기서, 관련 정도는 도메인 문서들로부터 개념들의 동시 출현 빈도에 대한 통계적 정보를 통해 자동 평가된다[8,12,14,18]. 특히, 동시 출현 빈도는 단어들간의 상관 관계 행렬 구성과 종속도 결정 등에 사용되며, 상관 관계 행렬과 단어 종속도는 확장된 벡터 공간 모델[19]과 확률 모델[20]에서 각각 유용한 것으로 알려져 있다. 그러나, 도메인 통계 정보에 의해

자동 구축되는 이 시소러스는 관련 정도만으로 개념들 간의 의미 관계를 정확히 명시할 수 없으며, 우리의 직관과 일치하지 않는 무의미한 관계들을 필연적으로 생성하게 된다[3,4,12,13]. 또한, 이 개념들 사이의 의미 관계를 사용자가 파악하여 원하는 개념을 참조하기란 매우 어렵다는 단점을 가지고 있다[10,11]. 이를 해결하기 위해 자연어 처리 방법을 이용하여 문서의 내용을 분석하고, 이를 통해 개념들의 의미 관계를 명확히 표현할 수 있는 시소러스에 관한 연구가 수행되고 있다. 그러나, 자유 어순으로 표현된 방대한 문서로부터 개념들 사이의 의미 구조를 추출하기란 매우 어려운 일이다[13, 17].

계층화된 구조의 시소러스를 구축하는 대표적인 시스템으로는 "IBM Thesaurus Administrator", "WordNet", "NCI Thesaurus System", "BEAT", "Hierarch", "MultiTes", "TCS" 그리고 "STRID"이 있으며, 시소러스에는 "NASA", "INSPEC" 그리고 "Roget Thesaurus" 등이 있다. 이 시소러스에서 서로 관련된 개념들의 의미 관계 및 관련 정도는 전문가의 수작업에 의해 직접 명시된다[3,4,8,16]. 따라서, 이 시소러스는 우리의 직관과 매우 일치할 수 있는 정확한 의미 구조를 가지기 때문에 현재 많은 도메인에서 상용화되어 사용되고 있다. 그러나, 방대한 시소러스를 전문가가 직접 구축하기 위해서는 많은 시간과 노력이 필요하게 된다. 또한, 정확한 의미 구조의 시소러스를 일관성 있게 구축하고 유동적 특성을 가진 개념들을 유지 보수하기 위해 전문가는 개념들 사이의 복잡한 의미 구조를 모두 파악해야 하는 어려움이 있다. 특히, 방대한 시소러스는 여러 전문가들에 의해 동시에 구축될 수 있기 때문에 구축 과정에서 상호 의미적 일관성을 유지시킬 수 있어야 하며, 구축된 시소러스 개념들을 쉽게 접근할 수 있는 기능을 포함해야 한다[6,7,10,11].

그러나, 계층화된 시소러스를 구축하기 위한 이 시스템들은 앞에서 설명한 구축 방법의 단점들을 해결하지 못하고 있을 뿐만 아니라 시소러스로부터 사용자가 원하는 개념을 탐색하기 위해 하이퍼링크 방식의 매우 단순한 순차 접근 방법만을 지원하고 있다. 이 방법의 문제점은 매우 복잡한 구조의 시소러스에서 사용자가 자신이 의도한 개념을 순차적으로 탐색할 경우, 탐색 방향성을 상실할 위험이 있다. 또한, 이 시스템들은 직접 접근 방법을 지원하고 있지 않기 때문에 여러 개념들과 복합적으로 연관된 사용자 관심 개념을 여러 단계의 순차적인 접근 과정을 통해 탐색해야 하는 단점을 가지고 있다.

따라서, 시소러스 시스템은 다음과 같은 조건들을 만족해야 한다. 첫째, 사용자의 직관과 일치할 수 있는 시소러스를 최소한의 비용으로 구축할 수 있도록 지원해야 한다. 둘째, 여러 전문가가 동시에 시소러스를 구축할 경우, 개념들 사이의 의미 구조를 시스템이 미리 파악하여 전문가들에게 제시함으로써 시소러스의 의미적 일관성을 유지시킬 수 있어야 한다. 셋째, 순차 접근 방법과 직접 접근 방법을 모두 지원하여 복잡한 시소러스로부터 사용자가 의도한 개념을 효과적으로 탐색할 수 있도록 해야 한다. 다음 장에서는 본 논문의 객체기반 시소러스 시스템이 이 요구들을 어떠한 방법으로 수용할 수 있는지를 자세히 기술한다.

3. 객체기반 시소러스

객체기반 시소러스는 기존의 시소러스에 객체지향 패러다임의 구조적인 특성을 적용한 시소러스이다. 즉, 모든 시소러스 개념은 객체로 간주되며, 일반적인 의미의 개념 객체는 보다 구체적인 의미의 개념 객체나 인스턴스 객체를 하위 객체로 가진다. 여기서, 인스턴스 객체는 개념 객체의 실제 예의 의미를 가진 객체로서 실제의 많은 개념들이 인스턴스로 분류될 수 있다. 따라서, 하나의 개념 객체는 자신이 소유한 모든 인스턴스들을 대표하게 된다. 특히, 시소러스 개념들은 그 응용 도메인에 따라 인스턴스 또는 개념 객체로 그 단위가 결정된다. 객체들 사이의 의미 관계는 기존의 시소러스 개념들 사이의 관계인 BT/NT 그리고 RT를 그 의미에 따라 일반화(super/sub-concept-of), 클래스화(owner/instance-of), 집성화(whole/part-of) 그리고 연관화(association-of) 관계로 재정의하여 표현한다. 따라서, 하나의 최상위 개념은 수직적으로 일반화/클래스화 관계에 의해 객체 계층을 형성한다. 동시에, 수평적으로는 다른 객체 계층의 객체와 집성화/연관화 관계를 가지며, 이 관계 정보는 하위 객체에 상속되는 성질을 가진다. 이 시소러스는 객체지향 패러다임의 메소드나 메시지 전달과 같은 객체의 행위적인 특성 보다 구조적인 특성에 의해 모델링되기 때문에 객체지향이 아닌 객체기반 시소러스로 정의된다. 이 장에서는 객체기반 시소러스의 표현과 성질에 대해 설명한다.

그림 1은 최상위 객체가 'Switching System'인 객체 계층을 포함하는 객체기반 시소러스의 예이다. 이 예에서 나타나는 시소러스 개념들은 "telecommunication software"에 관련된 문서에서 자주 사용되는 용어들로 구성되었다. 여기서, 노드는 객체를 의미하며, 링크는 객체들 사이의 관계를 나타낸다. 개념 레벨과 인스턴스 레

벨은 개념들 사이의 관계와 인스턴스들 사이의 관계를 각각 표현하며, 이 두 레벨은 클래스화 관계에 의해서 연결된다.

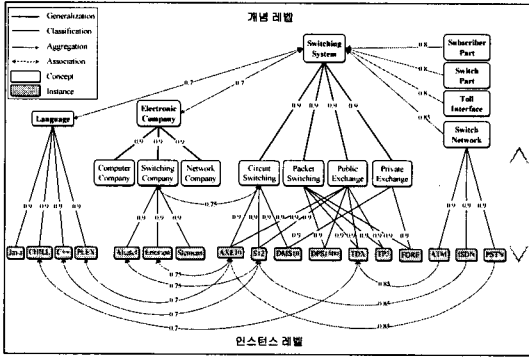


그림 1 객체기반 시소러스의 예

3.1 시소러스의 표현

객체기반 시소러스에서 개념과 인스턴스는 기본적으로 객체로 간주된다. 개념 객체는 수직적으로 많은 하위 개념들과 일반화 관계를 가지며, 인스턴스들과는 클래스화 관계를 가진다[21]. 또한, 하나의 객체와 수평적으로 집성화/연관화 관계를 가지는 객체들은 속성 객체로 정의되며, 상위 개념에 대한 속성 객체들의 정보는 하위 객체에 상속되는 성질을 가진다. 이러한, 정보들을 표현하기 위해 객체는 자신의 식별자와 관계 정보로 구성되며, 관계 정보는 (R, O₁, O₂, W_{O₁O₂})과 같은 관계 튜플들의 집합으로 표현된다. 여기서, O₁와 O₂는 객체이고 R은 이들의 의미적 관계를 나타내며, W_{O₁O₂}는 두 객체 사이의 관련 정도를 나타내는 0.5와 1 사이의 퍼지 값을 가진다.

하나의 객체에 표현된 관계 튜플들은 포함된 객체의 종류에 따라 3 가지 형태로 표현된다. 첫째, 개념들 사이의 관계를 정의한 튜플은 개념 레벨을 표현하며, 둘째, 인스턴스들 사이의 관계를 정의한 튜플은 인스턴스 레벨을 명시한다. 셋째, 개념과 인스턴스의 관계를 정의하는 튜플은 두 레벨을 연결할 수 있는 역할을 하며 클래스화 관계에 의해 표현된다.

관계 정보 튜플에서 관련 정도(W_{O₁O₂})는 시소러스 구축 과정에서 다음과 같은 두 규칙에 따라 자동 할당된다. 하나는 각 관계에 미리 정의된 묵시적 관련 정도를 할당하는 방법이다. 묵시적 관련 정도는 도메인에 따라 각각 다르게 정의될 수 있는데, 본 논문에서는 편의상 일반화/클래스화 관계에 0.9, 집성화 관계에 0.8 그리

고 연관화 관계에 0.7를 각각 할당하였다. 다른 방법으로 상위 객체들의 관련 정도를 3.3에서 설명될 상속 메커니즘에 따라 할당받는 방법이 있다. 이렇게 할당된 관련 정도는 도메인 전문가에 의해 적절한 값으로 다시 수정될 수 있으며, 보다 자세한 사항은 다음 장에서 설명한다.

3.2 전이적 성질

개념 레벨에서 하나의 개념이 다른 개념의 의미보다 일반적이거나 구체적일 때, 이 두 개념은 일반화 관계로 표현되며, 그림 2와 같은 전이적 성질을 가지게 된다. 즉, 개념 객체 c₁, c₂, c₃ 그리고 c₄에 대해, (sub-concept-of, c₁, c₂, w_{c₁c₂}), (sub-concept-of, c₁, c₃, w_{c₁c₃}), (sub-concept-of, c₂, c₄, w_{c₂c₄}) 그리고 (sub-concept-of, c₃, c₄, w_{c₃c₄})이면, (sub-concept-of, c₁, c₄, w_{c₁c₄})인 관계가 성립한다. 따라서, c₄는 c₁의 하위 개념이 된다. 이때, 이 두 개념 사이의 관련 정도 w_{c₁c₄}는 t-norm(Δ: triangular-norm)과 s-norm(∇: triangular-conorm)함수에 의해 계산된다. 일반적으로 이 norm 함수는 두 퍼지 값에 대한 and/or를 계산하기 위해 사용된다. 현재, 응용 도메인에 따라 많은 t-norm/s-norm 함수들이 사용되고 있으나[22], 본 논문에서는 가장 단순하며 널리 사용되는 min/max 함수를 사용한다.

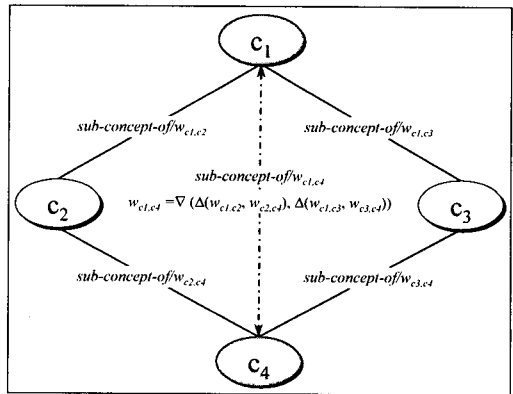


그림 2 일반화 관계의 전이적 성질

하나의 개념은 자신이 소유한 인스턴스들과 'instance-of' 관계를 가지며, 반대로 하나의 인스턴스는 자신이 속한 개념과 'owner-of' 관계를 가진다. 따라서, 하나의 개념에 속한 인스턴스는 의미적으로 이 개념의 모든 상위 개념에 대한 인스턴스가 된다[21]. 즉, 그림 3과 같이 (sub-concept-of, c₁, c₂, w_{c₁c₂}), (sub-

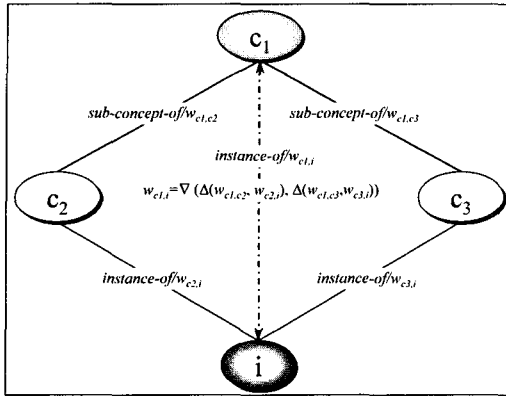


그림 3 클래스화 관계의 전이적 성질

concept-of, $c_1, c_3, w_{c1,c3}$, (instance-of, $c_2, i, w_{c2,i}$) 그리고 (instance-of, $c_3, i, w_{c3,i}$)이면, (instance-of, $c_1, i, w_{c1,i}$)이다. 이 성질은 하나의 개념이 가지는 많은 인스턴스를 효과적으로 분류할 수 있게 한다. 이때, c_1 과 i 사이의 관련 정도 역시 Δ 와 ∇ 에 의해 계산된다.

3.3 상속의 성질

하나의 개념이 의미적으로 다른 개념의 부분(part-of) 또는 전체(whole-of)를 나타낼 경우, 두 객체는 집성화 관계로 나타낸다. 연관화 관계는 두 객체의 의미적 관계가 명확히 표현될 수 없는 경우 사용되며, 이 관계의 의미적 방향성은 존재하지 않는다. 이 집성화/연관화 관계는 개념 레벨과 인스턴스 레벨에서 모두 이용된다. 하나의 개념에 대한 속성 객체 정보는 하위 객체에 상속되는 성질을 가지며, 그림 4와 5는 'part-of'와 'association-of'에 대한 상속 성질을 나타낸다. 이 두 관계에 대한 상속 성질은 같은 특성을 가지므로 여기에서는 편의상 'part-of'만을 설명한다.

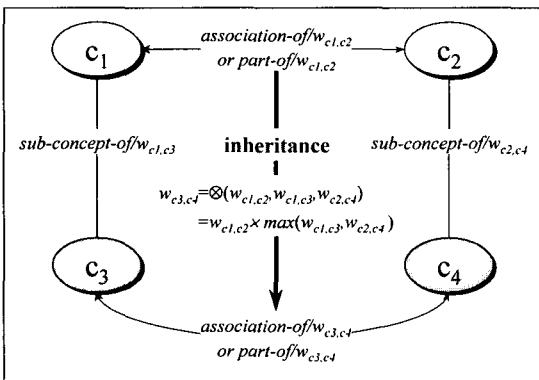


그림 4 개념 레벨에서의 상속 성질

그림 4는 개념 레벨에서 (part-of, $c_1, c_2, w_{c1,c2}$), (sub-concept-of, $c_1, c_3, w_{c1,c3}$) 그리고 (sub-concept-of, $c_2, c_4, w_{c2,c4}$)이면, (part-of, $c_3, c_4, w_{c3,c4}$)인 관계가 성립함을 나타내고 있다. 즉, 이 경우 c_4 는 c_3 의 속성 객체가 된다. 이 두 객체 사이의 관련 정도 $w_{c3,c4}$ 은 $w_{c1,c2}$, $w_{c1,c3}$ 그리고 $w_{c2,c4}$ 를 매개변수로 하는 상속함수에 의해 계산될 수 있다. 상속함수(\otimes) 역시 Δ / ∇ 함수와 같이 도메인에 따라 다양하게 정의될 수 있으며, 단지 그 결과가 $w_{c1,c2}$ 를 초과할 수 없다는 제약 조건을 만족하면 된다. 즉, 전문가에 의해 명시적으로 결정된 상위 객체들 사이의 'part-of' 관계로부터 상속되어 묵시적으로 'part-of' 관계를 가지는 하위 객체들 사이의 관련 정도는 매우 대략적으로 정해질 수밖에 없기 때문에 상위 객체들간의 관련 정도보다는 적은 값으로 설정되어야 한다. 예를 들어, 그림 1에서 'AXE10'은 (part-of, Switching System, Switch Network, 0.85)의 상속에 의해 'Switch Network'의 모든 인스턴스들과 묵시적으로 'part-of' 관계를 가진다. 그러나, 이 인스턴스들이 모두 'AXE10'과 직접적으로 관련이 있다고 볼 수는 없기 때문에 이들의 관련 정도는 상속되는 값 0.85 보다 적은 값으로 정해져야 한다. 본 논문에서는 상속함수를 위해 $w_{c_1,c_2} \times \max(w_{c_1,c_3}, w_{c_2,c_4})$ 을 사용하였으며, 보다 정확한 함수의 개발은 그 성능 분석을 포함하여 향후 과제로 한다.

상속되는 관계가 관련시켜주는 하위 객체들의 관계 일부는 오히려 상위 객체들 보다 매우 밀접한 관련을 가질 수 있다. 예를 들어, 'Switching System'과 'Switch Network' 사이의 집성화 관계 보다 이들의 인스턴스인 'AXE10'과 'PSTN' 사이의 관계가 더 밀접히 연관된다고 볼 수 있는데, 그 이유는 "AXE10 교환기"가 실제 "공중전화망(PSTN: Public Switched Telephone Network)"을 자신의 부품으로 이용하고 있기 때문이다. 따라서, 시스템은 이 관계를 정확히 명시하려는 도메인 전문가에게 묵시적으로 상속된 관련 정도 보다 더 크며, 집성화 관련 정도의 상계 값 0.9 보다 작은 값을 제시해야 한다. 본 논문에서는 이러한 값으로 상위 객체들의 관련 정도를 그대로 상속하여 사용하며, 전문가가 이 값을 필요에 따라 재정의할 수 있도록 하였다.

그림 5는 인스턴스 레벨에서 (part-of, $c_1, c_2, w_{c1,c2}$), (instance-of, $c_1, i_1, w_{c1,i1}$) 그리고 (instance-of, $c_2, i_2, w_{c2,i2}$)이면, (part-of, $i_1, i_2, w_{i1,i2}$)이 성립함을 나타내고 있다. 즉, 이 경우 i_2 는 i_1 의 속성 객체가 되고 그 관련 정도는 $w_{c_1,c_2} \times \max(w_{c_1,i_1}, w_{c_2,i_2})$ 에 의해 계산될 수 있다. 상

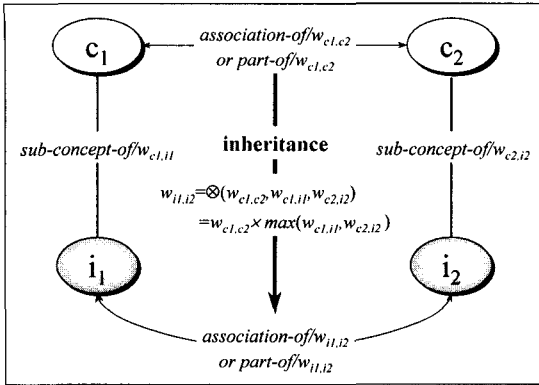


그림 5 인스턴스 레벨에서의 상속 성질

속 메커니즘에서 두 객체가 집성화/연관화 관계를 가지기 위해서는 이들의 상위 객체들이 이 관계를 가지고 있어야 한다. 따라서, 시스템은 이 메커니즘을 이용하여 객체들 사이의 복잡한 관계를 구조적으로 파악할 수 있기 때문에 전문가의 시소러스 구축 및 관리 비용을 감소시킬 수 있으며, 시소러스의 무결성을 유지시킬 수 있는 전략으로도 이용될 수 있다.

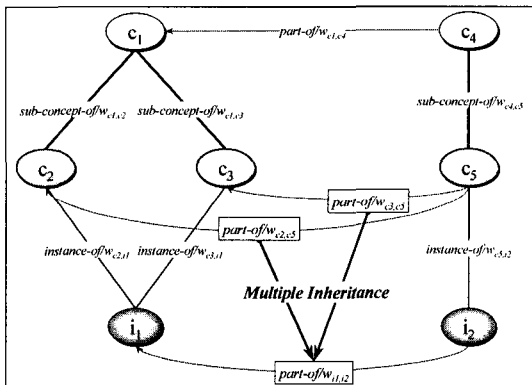


그림 6 다중 관련 정도의 상속

그림 6은 i_1 과 i_2 사이의 관련 정도 $w_{i1,i2}$ 가 \otimes 에 의해 $w_{c2,c5}$ 또는 $w_{c3,c5}$ 를 상속받을 수 있는 경우를 보여주고 있다. 이 경우 두 값 중에 하나가 다음 규칙에 의해 선택된다. 즉, 이들 중 전문가에 의해 재정의된 값이 목시적인 값보다 더 중요한 정보이므로 우선한다. 만약 이 두 값이 모두 목시적인 값이거나 재정의된 값이라면, " $w_{c2,c5}$ 또는 $w_{c3,c5}$ "의 의미가 \max 함수로 해석되어 이들 중 최대 값이 상속된다.

4. 객체기반 시소러스의 구축 및 유지

시소러스에 표현되는 객체들 사이의 복잡한 관계들은 도메인 전문가가 이들의 의미적 구조를 전체적으로 파악할 수 없게 하여, 일관성 있는 시소러스의 구축과 유동성 있는 관리를 어렵게 한다[6,7]. 객체지향 패러다임은 객체들 사이의 복잡한 관계를 의미적인 손실 없이 쉽게 모델링할 수 있게 하는 장점을 가지고 있다. 이 장에서는 객체기반 시소러스가 이 패러다임을 이용하여 복잡한 관계들을 모델링함으로써 쉽게 구축되고 관리될 수 있음을 보이도록 한다.

4.1 개념 레벨의 구축

하나의 최상위 개념은 다른 객체 계층의 최상위 개념과 집성화/연관화 관계를 가진다. 따라서, 개념 레벨을 구축하기 위해 먼저 전문가는 하나의 최상위 개념을 생성하고, 이 개념과 집성화/연관화 관계를 가질 수 있는 개념들을 선택함으로써 관계를 설정한다. 이때, 시스템은 미리 정의된 목시적 관련 정도를 각각의 관계에 자동으로 할당한다.

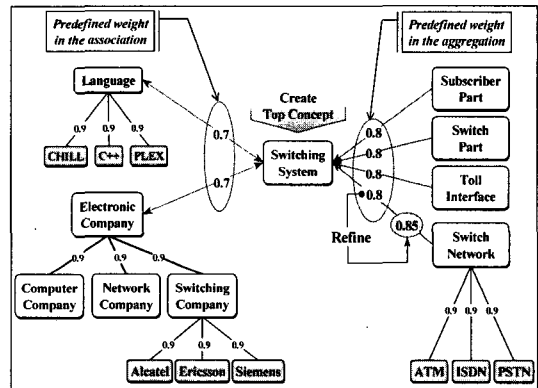


그림 7 최상위 객체 'Switching System'의 생성

그림 7에서 'Language', 'Electronic Company' 그리고 'Switch Network'의 객체 계층이 이미 여러 다른 전문가에 의해 구축되었다면, 하나의 최상위 객체 'Switching System'는 다음과 같은 과정에 따라 삼입된다. 먼저, 도메인 전문가가 'Switching System'을 생성한 다음, 연관화 객체로 'Language'와 'Electronic Company'를 선택하면 연관화 관계에 미리 정의된 관련 정도 0.7이 이들 사이의 관계에 자동으로 할당된다. 비슷한 방법으로, 집성화 객체 'Subscriber Part', 'Switch Part', 'Toll Interface' 그리고 'Switch Network'이 순차적으로 선택되면, 0.8이 자동으로 할당

된다. 자동으로 부여된 관련 정도는 전문가에 의해 재정의될 수 있으며, 이 예에서는 (part-of, Switching System, Switch Network, 0.8)의 관련 정도를 0.85로 재정의하였다.

개념 레벨을 구축하기 위한 다음 단계로 'Switching System'에 하위 개념 'Circuit Switching'을 삽입하는 과정을 살펴보자. 이 과정에서 전문가가 삽입되는 개념인 'Circuit Switching'과 집성화 그리고 연관화 관계를 가지는 속성 객체들을 모두 파악하여 의미적으로 일관된 관계를 명시하기란 매우 어렵다. 이때, 객체기반 시소러스의 상속 메커니즘이 매우 유용하게 이용된다. 즉, 이 속성 객체들은 'Switching System'과 집성화/연관화 관계를 가지는 개념들의 하위 객체들 중에서 선택되기 때문에 이 후보 객체들은 시스템에 의해 쉽게 파악되어 구축 과정에 있는 전문가에게 제시될 수 있다. 구축 과정에서 이러한 정보들을 시스템이 파악하여 전문가에게 제시하는 메커니즘은 시소러스의 구축비용을 감소시킬 수 있을 뿐만 아니라 여러 전문가에 의해 구축되는 방대한 시소러스의 의미적 일관성을 유지시킬 수 있다.

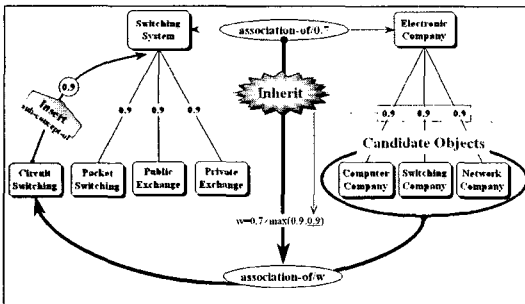


그림 8 'Circuit Switching'의 삽입과 목시적 관계 설정

그림 8과 같이 'Circuit Switching'을 삽입하면, 시스템은 생성된 관계 (sub-concept-of, Switching System, Circuit Switching, w)에 목시적 관련 정도 0.9를 할당한 다음, 이 개념과 집성화 연관화 관계를 가질 수 있는 후보 객체들을 파악한다. 즉, 'Circuit Switching'의 상위 개념 'Switching System'이 'Electronic Company'와 연관화 관계를 가지고 있기 때문에 'Electronic Company'의 하위 개념 'Computer Company', 'Switching Company' 그리고 'Network Company'이 후보 객체로 파악된다. 따라서, 'Circuit Switching'은 이 후보 객체들과 목시적으로 연관화 관계를 가지며, 각 관계는 객체 계층에서 정의된 상속 합

수에 의해 목시적 상속 관련 정도 0.63을 가지게 된다. 그러나, 이 목시적인 관련 정도는 하위 객체들 모두에게 적용시킬 수 있을 정도로 일반적인 값은 아니기 때문에 필요에 따라 전문가에 의해 개별적으로 재정의될 필요가 있다. 예를 들어, 전문가가 시스템에 의해 제시된 후보 객체들 중에 'Switching Company'를 선택하였을 경우, 시스템은 후보 객체들 중에서 'Switching Company'가 'Circuit Switching'과 개별적으로 보다 관련이 있는 개념으로 인식한다. 따라서, 그들의 관련 정도 역시 0.63보다 높은 값 0.7을 제시하게 되는데, 이 값은 앞에서 언급했듯이 'Circuit Switching'과 'Switching Company'의 상위 객체 'Switching System'과 'Electronic Company'의 관련 정도이다. 그림 9는 이 과정을 자세히 설명하고 있다. 시스템에 의해 제시된 관련 정도 역시 전문가에 의해 재정의될 수 있으며, 이 예에서는 0.7을 0.75로 재정의하였다.

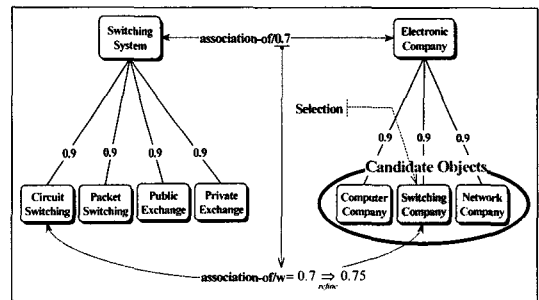


그림 9 'Circuit Switching'과 'Switching Company'의 관련 정도

4.2 인스턴스 레벨의 구축 및 유지

시소러스에서 인스턴스들은 가장 구체적인 의미의 객체들로 개념들 보다 빈번히 변화되며, 질의 평가에 많이 이용된다는 특징을 가지고 있다[3]. 따라서, 시소러스를 구축하고 유지할 때, 전문가는 인스턴스들의 관계를 정확히 명시해야 한다. 객체기반 시소러스에서 클래스화 관계는 많은 객체들이 인스턴스로 처리되는 도메인에서 시소러스를 효율적으로 유지할 수 있도록 한다. 즉, 개념 레벨의 구축 메커니즘을 클래스화 관계에 적용하여 인스턴스 레벨까지 확장함으로써, 전문가가 인스턴스 관계를 개별적으로 명시하기 위한 비용을 상당히 감소시킬 수 있다.

그림 10은 'AXE10'이 'Circuit Switching'과 'Public Exchange'의 인스턴스로 각각 0.93과 0.92의 관련 정도로 삽입되는 과정을 설명하고 있다. 이때, 전

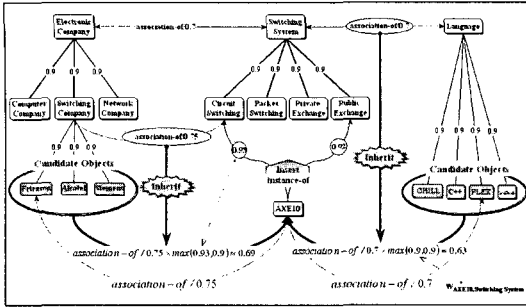


그림 10 'AXE10'의 삽입과 연관화 관계

문가에 의해 명시적으로 기술되는 이들의 관련 정도는 클래스화 관계에 미리 할당된 목시적 관련 정도 0.9 보다 우선한다. 한편, 개념 레벨에서와 같은 방법으로 전문가가 관계 (instance-of, Circuit Switching, AXE10, 0.93)과 (instance-of, Public Exchange, AXE10, 0.92)를 명시하면, 시스템은 상속 메커니즘을 이용하여 'AXE10'과 집성화/연관화 관계를 가지는 속성 객체에 대한 후보 인스턴스들을 구조적으로 파악하게 된다. 즉, 'AXE10'은 관계 (association-of, Circuit Switching, Switching Company, 0.75)에 의해 'Ericsson', 'Alcatel' 그리고 'Siemens'와 목시적으로 상속된 관계 'association-of/0.69'를 가진다. 이들 중에서 전문가가 'Ericsson'을 선택하면, 시스템은 목시적 상속 관련 정도 0.69 대신 0.75를 제시한다. 같은 방식으로 시스템은 관계 (association-of, Switching System, Language, 0.7)에 의해 'Language'의 인스턴스 'CHILL', 'C++', 'PLEX' 그리고 'Java'를 후보 객체들로 식별하며, 이들 중에서 'PLEX'이 선택되면, 'AXE10'과의 관계에 대한 관련 정도 0.7을 제시하게 된다.

후보 객체들 중 특정한 객체가 선택되지 않았을 경우에도 이들과 'AXE10'의 목시적 관련 정도는 여전히

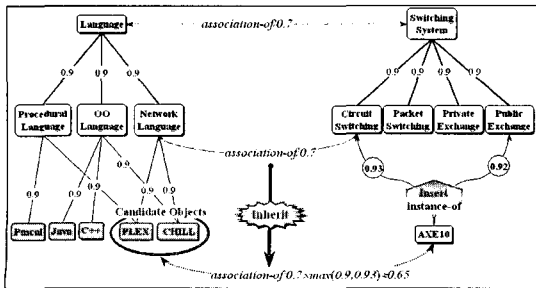


그림 11 의미 없는 관계의 축소

유효한 정보로 유지된다. 그러나, 이 과정에서 너무 많은 목시적인 상속 관계들이 생성된다면, 이들은 검색 성능을 저하시키는 요인이 될 수 있다. 예를 들어, 'AXE10'은 모든 'Language'의 인스턴스들과 관련이 있다고 보기는 어렵기 때문에 그림 10의 (association-of, AXE10, Java, 0.63)은 의미 없는 정보일 수 있다. 이러한 문제는 'Language'를 보다 구체적인 개념들로 세분함으로써 해결할 수 있다. 즉, 그림 11과 같이 'Language'를 'Procedural Language', 'OO Language' 그리고 'Network Language'로 세분하고, 개념 레벨에서 'Circuit Switching'과 'Network Language' 사이에 연관화 관계를 설정함으로써, 의미 없는 관계의 관련 정도를 현저하게 감소시킬 수 있다. 'AXE10'의 집성화 관계 역시 연관화 관계에서와 같은 방법으로 후보 객체를 자동으로 파악하여 그 관계를 설정할 수 있다.

5. 질의기반 참조

시소러스는 탐색어들로 표현된 질의의 검색 결과가 사용자의 요구를 만족시키지 못할 경우, 탐색어와 의미적으로 유사한 시소러스 개념을 질의에 추가함으로써 검색 효율을 높일 수 있다[45]. 그러나, 사용자가 의도하지 않은 개념들까지 질의에 추가된다면 시스템의 검색 성능이 오히려 더 저하될 수도 있을 것이다. 따라서, 정확률을 높이기 위해서는 질의에 추가될 시소러스 개념들이 미리 사용자에게 의해 검증될 필요가 있다. 이 검증 과정을 용이하게 하기 위해 질의기반 참조 메커니즘은 질의를 통해 사용자가 정확히 의도한 관점의 개념들을 직접 브라우징할 수 있게 한다. 이때, 참조 질의는 객체기반 시소러스에 대한 사용자의 브라우징 범위를 줄여주기 위해 사용된다.

5.1 참조 질의의 표현

참조 질의는 [Main-Query, Restriction-Query]와 같이 두 개의 부분 질의로 구성되며, 각각은 표 1의 연산자들로 표현된다. 'Main-Query'는 객체기반 시소러스의 여러 객체 계층들 중 특정 계층의 일부분으로 사용자의 브라우징 범위를 한정하며, 'Restriction-Query'는 'Main-Query'의 범위를 더욱 제한하기 위해 사용된다. 이때, 두 부분 질의의 연산자들은 부울 연산자와 다른 의미를 가진다. 예를 들어, 참조 질의 [Switching System, PSTN AND Ericsson]는 교환기 회사 "Ericsson"에서 교환 네트워크 "PSTN"를 이용하여 개발한 교환기(switching system)와 관련된 객체들을 사용자가 요구하고 있음을 의미한다. 따라서, 이 질의들

통해 사용자는 'AXE10' 및 이와 연관된 시소러스 객체들을 바로 브라우징해 볼 수 있을 것이다.

표 1 참조 부분 질의의 연산자

Main-Query의 연산자		Restriction-Query의 연산자	
종류	의미	종류	의미
∨	일반화 연산자	OR	선택적 한정 연산자
∧	구체화 연산자	AND	동시 한정 연산자

참조 질의의 의미를 파악하기 위해 최상위 객체 'top'을 가지는 하나의 객체 계층을 부분 순서 집합(POSet: Partial Ordered Set), $H_{top} = \langle O, \leq \rangle$ 으로 정의한다. 여기서, O는 H_{top} 에 속하는 모든 객체들의 집합이며, $o_1, o_2 \in O$ 에 대해, $o_1 \leq o_2$ 은 o_2 가 o_1 보다 더 일반적인 의미의 객체임을 나타낸다. 즉, o_1 은 o_2 의 하위 객체이다. 따라서, 하나의 집합 $\{o_i \in O, i=1, 2, \dots, n\}$ 에 대한, 상계 집합(UBS: Upper Bound Set)과 하계 집합(LBS: Lower Bound Set)은 다음과 같이 구할 수 있다.

$$UBS = \{o \in O \mid o_i \leq o \text{ for } i=1, 2, \dots, n\}$$

$$= \bigcap_{i=1}^n \{Sup(o_i)\} \text{ 이고, 여기서 } Sup(o_i) = \{o \in O \mid o_i \leq o\}$$

$$LBS = \{o \in O \mid o \leq o_i \text{ for } i=1, 2, \dots, n\}$$

$$= \bigcap_{i=1}^n \{Sub(o_i)\} \text{ 이고, 여기서 } Sub(o_i) = \{o \in O \mid o \leq o_i\} \quad (23)$$

객체기반 시소러스의 객체 계층은 여러 개의 최소 상계(LUB: Least Upper Bound)나 최대 하계(GLB: Greatest Lower Bound)를 가지기 때문에 격자(lattice) 구조로 표현될 수 없다. 따라서, 최소 상계 집합(LUBS: Least Upper Bound Set)과 최대 하계 집합(GLBS: Greatest Lower Bound Set)을 다음과 같이 정의한다.

$$LUBS = UBS - \{o_2 \mid o_1 \leq o_2 \text{ 이고, 여기서 } o_1, o_2 \in UBS\}$$

$$GLBS = LBS - \{o_1 \mid o_1 \leq o_2 \text{ 이고, 여기서 } o_1, o_2 \in LBS\}$$

하나의 집합 $\{o_i \in O \mid i=1, 2, \dots, n\}$ 에 대해, LUBS와 GLBS는 각각 일반화 연산자와 구체화 연산자를 이용하

여 $\bigvee_{i=1}^n o_i$ 와 $\bigwedge_{i=1}^n o_i$ 같이 표현할 수 있다. 이에 대한 예는 다음 절에서 보이도록 한다.

5.2 참조 질의의 의미 분석

참조 질의는 사용되는 연산자에 따라 3 종류의 기본 형식으로 표현된다. 이 절에서는 이들 질의가 의미하는 객체들을 시소러스로부터 추론하는 메커니즘을 설명한다. 시스템은 이 추론된 객체들을 모두 이용하여 자동으로 검색 질의를 구성할 수 있다. 그러나, 이러한 질의는 너무 많은 검색 결과를 생성하기 때문에 필요한 경우 사용자는 추론된 객체들 중 자신이 의도한 객체들만을 선택적으로 이용하여 높은 정확률의 질의를 구성할 수 있다.

5.2.1 일반화/클래스화되는 질의

참조 질의 $[\bigvee_{i=1}^n o_i, *]$ 는 사용자가 이미 알고 있는 구체적인 의미의 객체(o_i)들을 모두 포괄할 수 있는 일반적인 의미의 상위 객체들을 추론하기 위해 사용된다. 이 질의로부터 시스템은 모든 객체 o_i 를 하위 객체로 가지는 개념들 중에서 가장 구체적인 개념을 추론한다. 즉, 추론된 개념은 모든 o_i 에 대한 LUBS로 표현될 수 있다. 이때, 이 참조 질의에 의해 브라우징되는 사용자의 관심 객체들은 참조 그래프로 시각화되며, 이 참조 그래프는 LUBS의 각 원소에 대한 모든 상·하위 객체들로 구성된다.

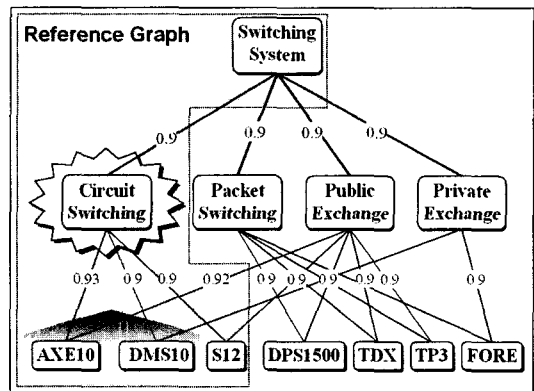


그림 12 [AXE10 ∨ DMS10, *]의 참조 그래프

예를 들어, 참조 질의 $[AXE10 \vee DMS10, *]$ 에 대해 추론된 시스템은 연산자 '∨'의 의미를 분석하여 인스턴스 'AXE10'과 'DMS10'을 동시에 하위 객체로 가지는 개념 'Circuit Switching'을 추론할 것이다. 추론 과정은 먼저 두 인스턴스 'AXE10'과 'DMS10'에 대한 각각의 상위 객체 집합을 구한다. 즉, 그림 12에서 이

라우징할 수 있는 특징을 가지고 있다.

첫째, 개념 추상화를 통한 브라우징은 일반적인 의미의 객체로부터 구체적인 객체를 점진적으로 브라우징하는 방법이다. 특히, 이 기능은 하나의 최상위 개념으로부터 파생되는 많은 하위 개념들 중에서 사용자가 의도한 방향에 있는 개념들만을 순차적으로 브라우징할 수 있게 한다. 예를 들어, 그림 17과 같이 일반적인 의미의 'Switching System'으로부터 'Circuit Switching' 그리고 'AXE10'를 점진적으로 브라우징할 수 있다. 다른 의미 방향에 있는 객체들은 '+'기호로 표시되며, 이 방향의 객체들 역시 마우스 클릭을 통해 하위 객체들을 시각화함으로써 브라우징할 수 있다.

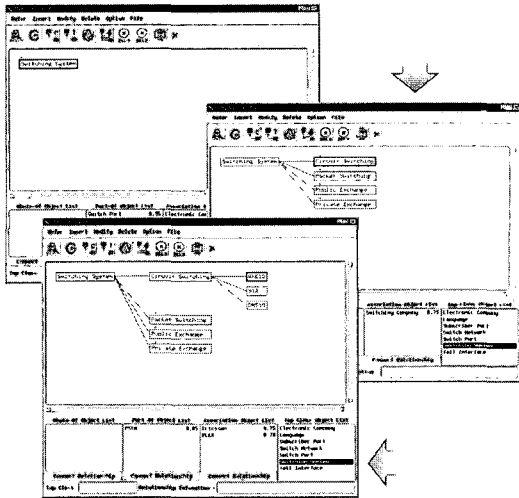


그림 17 개념 추상화를 통한 브라우징

둘째, 인스턴스 소속성 추상화를 통한 브라우징은 하나 이상의 개념들에 동시에 속한 인스턴스들을 브라우징하는 방법이다. 일반적으로 하나의 개념은 많은 인스턴스들을 가지며, 하나의 인스턴스는 여러 개념에 속하게 된다. 이때, 이 소속성 추상화를 통한 브라우징은 사용자가 인스턴스의 소속 관계를 명확하게 판단할 수 있게 한다. 예를 들어, 그림 18과 같이 'Circuit Switching'과 'Public Exchange'에 동시에 속하는 인스턴스들의 브라우징을 원하는 경우, shift key와 마우스 버튼을 동시에 이용하여 이 개념들을 순차적으로 클릭하면, 두 개념이 동시에 소유하는 인스턴스 'AXE10'과 'S12'만을 시각화하여 브라우징할 수 있다. 이 기능은 특히 방대한 양의 인스턴스들을 가지는 응용 환경에 적합한 기능이다.

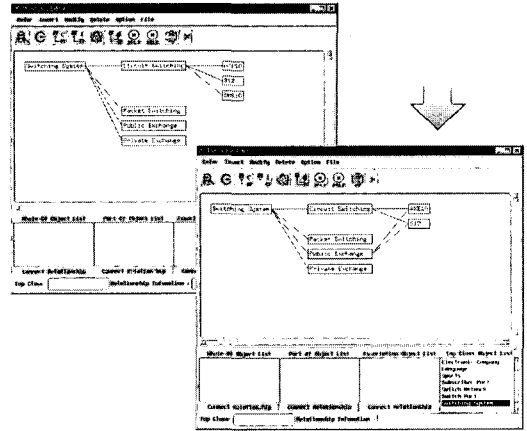


그림 18 인스턴스 소속성 추상화를 통한 브라우징

셋째, 관계 추상화를 통한 브라우징은 다양한 관계로 표현된 하나의 객체로부터 사용자가 원하는 관계를 가지는 객체들만을 선별적으로 브라우징하는 방법이다. 예를 들어, 그림 19에서 사용자는 'AXE10'의 집성화/연관화 객체를 시각화하고, 이 둘 중 집성화 방향에 있는 객체 'PSTN'으로 브라우징할 수 있다.

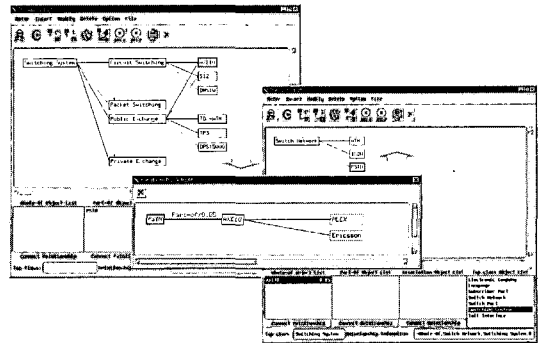


그림 19 관계 추상화를 통한 브라우징

6.3 질의기반 참조 기능

질의기반 참조는 5장에서 언급했듯이 방대한 양의 시소러스로부터 질의를 통해 사용자가 의도한 객체들을 선별적으로 참조할 수 있는 기능을 제공한다. 특히, 이 기능은 사용자가 의도한 적합한 개념을 찾기 위해 대용량의 복잡한 시소러스를 여러 단계 브라우징해야 하는 단점을 보완할 수 있는 직접 참조 방법을 지원한다. 즉, 사용자가 자신이 의도한 정확한 객체들을 찾기 위해 연관되는 개념들을 이용하여 참조 질의를 표현하면, 시스

템은 이 질의와 의미적으로 일치하는 객체들을 추론하고, 이 객체들로부터 검색 질의를 자동으로 또는 선택적으로 구성할 수 있게 한다. 따라서, 이 방법으로 구성된 질의는 높은 정확률의 검색을 보장할 수 있을 것이다.

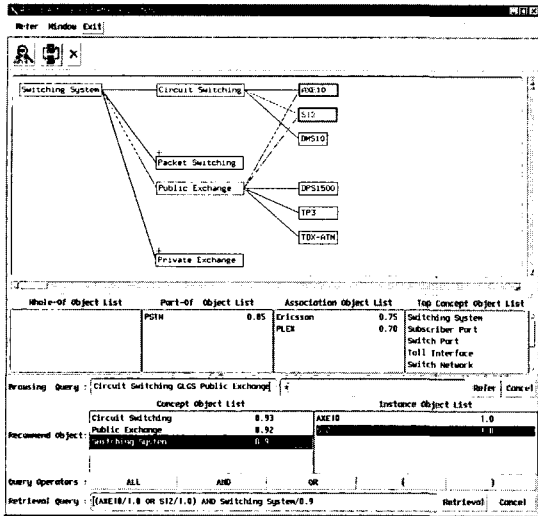


그림 20 참조 질의를 통한 'AXE10'과 'S12'의 추론과 선택적 질의의 구성

그림 20은 참조 질의의 [Circuit Switching ^ Public Exchange, *]과 의미적으로 일치하는 객체를 추론한 화면이다. 이 화면에서 'Main-Query'의 연산자 '^'는 키보드로 입력될 수 없기 때문에 'GLBS'를 대신 사용하였다. 이 참조 질의는 {AXE10, S12}를 추론하며, 이 객체들을 포함하는 참조 그래프는 적색으로 시각화된다. 이 참조 그래프에 포함된 개념들과 인스턴스들은 'Recommend Object' 영역에 참조 질의와의 관련 정도에 따라 순위화되어 나타난다. 이들은 "(AXE10/1.0 OR S12/1.0) AND Switching System/0.9"와 같이 확장된 부울 질의의 구성에 효과적으로 이용될 수 있다.

그림 21은 참조 질의의 [Switching System, PSTN AND Ericsson AND PLEX]을 처리한 결과이다. 여기서, 사용자가 추천된 객체들의 의미를 정확히 파악하지 못할 경우에는 시스템에게 자동으로 검색 질의 "AXE10/1.0 OR Circuit Switching/0.93 OR Public Exchange/0.92 OR Switching System/0.9"을 구성하도록 요청할 수도 있다. 이 질의는 특히 일정한 정확률을 보장할 수 있어야 하며 동시에 높은 재현률을 요구하는 도메인에 적절히 이용될 수 있을 것이다.

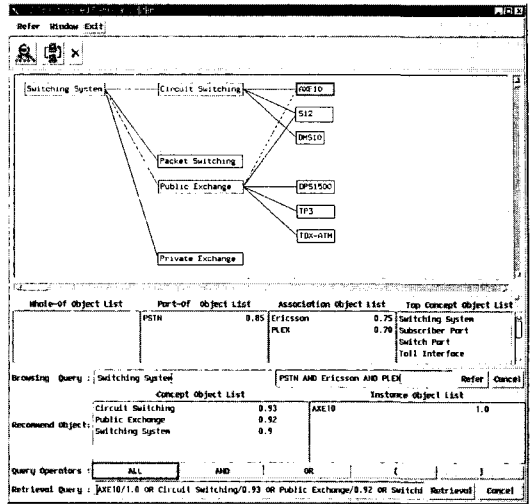


그림 21 참조 질의를 통한 'AXE10'의 추론과 자동 질의의 구성

7. 결론 및 향후 연구 과제

시소러스 시스템은 시소러스를 보다 효과적으로 구축하고 유지할 수 있게 하며, 구축된 시소러스를 검색 시스템이나 사용자들이 쉽게 이용할 수 있게 하여 검색 성능을 향상시킨다는 점에서 매우 중요한 의미를 가진다. 이를 위해 본 논문에서는 객체기반 시소러스 시스템을 설계하고 구현하였다. 이 시스템은 크게 시소러스를 구축하는 기능, 추상화 기법을 이용하여 관련된 개념들을 순차적으로 브라우징하는 기능 그리고 참조 질의를 통해 원하는 개념들을 직접 참조하는 기능으로 구성되었다. 구축 과정에 적용된 객체지향 패러다임은 객체들 사이의 관계를 정확하게 표현할 수 있게 하며, 전문가의 직관과 일치하는 시소러스를 구축할 수 있게 한다. 특히, 이 패러다임의 상속 메커니즘에 의해 구조적으로 파악된 정보는 반자동 방식의 시소러스 구축을 지원하며, 방대한 시소러스를 여러 전문가들이 서로 다른 호스트에서 구축할 때, 시소러스의 의미적 일관성을 유지시킬 수 있으며 구축 비용 역시 감소시킬 수 있게 한다. 브라우징 기능과 질의기반 참조 기능은 사용자가 의도한 개념들을 순차적으로 또는 직접 탐색할 수 있게 하여 정확한 검색 질의를 구성할 수 있도록 하였다. 특히, 이 질의의 구성 방법은 높은 재현률을 요구하는 도메인에 유용하게 이용될 수 있으며, 대표적인 응용 도메인으로 소프트웨어 재사용 부품 검색 시스템을 들 수 있다.

향후 연구 과제로, 객체기반 시소러스를 이용한 검색 시스템의 성능을 정량적으로 평가해야 하며, 이 평가를 통해 적합한 목시적 관련 정도를 분석해야한다. 또한, 객체지향 패러다임의 행위적 측면을 고려한 객체지향 정보검색 시스템의 개발이 요구된다.

참 고 문 헌

[1] H. Chen, T. Yim and D. Fye, "Automatic Thesaurus Generation for an Electronic Community System," Journal of the American Society for Information Science, Vol. 46, No. 3, pp. 175-193, 1995.

[2] C. J. Crouch, "An Approach to the Automatic Construction of Global thesaurus," Information Processing and Management, Vol. 26, No. 5, pp. 629-640, 1990.

[3] J. Y. Nie and M. Brisebois, "An Inferential Approach to Information Retrieval and its Implementation using a Manual Thesaurus," Artificial Intelligence Review, Vol. 10, No. 5, pp. 409-439, 1996.

[4] M. Hancock-Beaulieu, M. Fieldhouse and T. Do, "An Evaluation of Interactive Query Expansion in an Online Library Catalogue with a Graphical User Interface," Journal of Documentation, Vol. 5, No. 3, pp. 225-245, 1995.

[5] M. Hancock-Beaulieu and S. Walker, "An Evaluation of Automatic Query Expansion in an Online Library Catalogue," Journal of Documentation, Vol. 48, No. 4, pp. 406-421, 1992.

[6] J. Ganzmann, "Criteria for the Evaluation of Thesaurus Software," International Classification, Vol. 17, No. 3/4, pp. 148-157, 1990.

[7] J. L. Milstead, "Specifications for Thesaurus Software," Information Processing and Management, Vol. 27, No.2/3, pp. 165-175, 1991.

[8] Y. Jing and W. B. Croft, "An Association Thesaurus for Information Retrieval," Proceedings of the RIAO 94, C.I.D., Paris, pp. 146-160, 1994.

[9] Y. Qiu and H. P. Frei, "Applying a Similarity Thesaurus to a Large Collection for Information Retrieval," Technical Report. Dept. Computer Science, Swiss Federal Institute of Technology (ETH), Jan. 1995.

[10] S. Jones, M. Gatford, S. Robertson, M. Hancock-Beaulieu, J. Secker and S. Walker, "Interactive Thesaurus Navigation: Intelligence Rules OK?," Journal of the American Society for Information Science, Vol. 1, No. 46, pp. 52-59, 1995.

[11] M. P. Smith, A. S. Pollitt and C. S. Li, "Evaluation of Concept Translation through Menu Navigation in the MenUSE Intermediary System," Proceedings of the 14th BCS IRSG Research Colloquium on

Information Retrieval, University of Lancaster, pp. 38-54, April, 1992.

[12] H. J. Peat and P. Willett, "The Limitation of Term Co-occurrence Data for Query Expansion in Document Retrieval System," Journal of the American Society for Information Science, Vol. 42, No. 5, pp. 378-383, 1991.

[13] R. Rada and B. K. Martin, "Augmenting Thesauri for Information Systems," ACM Transaction on Office Information System, Vol. 5, No.4, pp. 378-392, 1987.

[14] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and Application of a Metric on Semantic Nets," IEEE Transaction on Systems, Man and Cybernetics, Vol. 19, No. 1, pp. 17-30, 1989.

[15] 최재훈, 한종진, 박종진, 양재동, "구조적인 시소러스 구축을 지원하는 객체기반 정보 검색 모델," 한국정보과학회 논문지(B), Vol. 24, No. 11, pp. 1244-1256, 1997.

[16] J. H. Lee, "Thesaurus-based Document Ranking for Boolean Retrieval Systems," KAIST, ph. D. Thesis, 1993.

[17] G. Salton, *Automatic Text Processing*, Addison-Wesley, 1989.

[18] H. L. Larsen and R. R. Yager, "The Use of Fuzzy Relational Thesauri for Classificatory Problem Solving in Information Retrieval and Expert Systems," IEEE Transactions on Systems, Man and Cybernetics, Vol. 23, No. 1, pp. 31-41, 1993.

[19] S. K. M. Wong, W. Ziarko, V. V. Raghavan and P. C. N. Wong, "On Modeling of Information Retrieval Concepts in Vector Space," ACM Transactions On Database Systems, Vol. 12, pp. 299-321, 1987.

[20] S. K. M. Wong and Y. Y. Yao, "A Generalized Binary Probabilistic Independence Model," Journal of the American Society for Information Science, Vol. 41, No. 5, pp. 342-329, 1990.

[21] W. Kim, *Introduction to Object-Oriented Databases*, The MIT Press, 1990.

[22] M. M. Gupta and J. Qi, "Theory of t-norms and Fuzzy Inference Methods," Fuzzy Sets and Systems, Vol. 40, No. 3, pp. 431-450, 1991.

[23] A. Doerr and K. Levasseur, "Applied Discrete Structures for Computer Science," Macmillan Publishing Company, 1989.



최 재 훈

1994년 2월 전북대학교 전자계산학(학사). 1996년 2월 전북대학교 전산통계학과(석사). 1996년 3월 ~ 현재 전북대학교 전산통계학과 박사 과정. 관심분야는 정보검색, 객체지향 데이터베이스, 소프트웨어 공학, 비디오 검색, 전자상거래 에이전트, 퍼지 이론, 지식기반 시스템.



김 기 현

1997년 2월 전북대학교 전자계산학(학사). 1999년 2월 전북대학교 전산통계학과(석사). 1999년 3월 ~ 현재 전북대학교 전산통계학과 박사 과정. 관심분야는 객체지향 데이터베이스, 정보검색, 운영체제, 전자상거래 에이전트, 분산 시스템



양 재 동

1983년 2월 서울대학교 컴퓨터공학과(학사). 1985년 2월 한국과학기술원 전산학과(석사). 1991년 2월 한국과학기술원 전산학과(박사). 1995년 1월 ~ 1996년 1월 Univ. of Florida, Visiting Scholar. 현재는 전북대학교 컴퓨터학과 부교수, 영상·정보 신기술 연구소 연구원. 관심분야는 OODBs, Expert System, CASE