DOI: 10.3745/KIPSTD.2008.15-D.3.429

형식개념분석기법을 이용한 사용자 질의 기반의 연관관계 추출 자동화지원도구의 개발

김 응 희[†] · 황 석 형[††] · 김 홍 기[†††]

요 약

형식개념분석기법(Formal Concept Analysis)은, 주어진 데이터로부터 공통속성을 갖는 객체들을 개념단위로 추출, 계층화하여 데이터에 내재된 개념들의 구조를 가시화 해주는 데이터분석기법으로써, 최근 다양한 분야에서 응용되고 있다. 본 연구에서는, 형식개념분석기법을 토대로, 사용자의 질의에 대한 함의관계(Implication)와 연관관계(Association rule)에 관한 정보추출과, 추출된 제반 정보들을 구조화하여 가시적으로 표현하기 위한 기법을 제안하고, 이를 지원하기 위하여, 함의/연관관계 추출 및 가시화 지원도구인 QAG-Wizard를 개발하였다. 본 연구결과는, 주어진 데이터의 속성을 기반으로 하는 사용자의 질의에 대하여, 데이터에 내재되어 있는 관계정보를 보다 다양하게 추출하고 직관적으로 표현 가능하므로, 데이터분석과 마이닝 뿐만 아니라, 질의기반의 정보검색분야 등에서 다양한 목적에 맞추어 활용될 수 있다.

키워드 : 형식개념분석기법, 개념격자, 사용자 질의, 함의관계, 연관관계

On Development of an Automatic Tool for Extracting Association Rules of a user query using Formal Concept Analysis

Eung-Hee Kim[†] · Suk-Hyung Hwang[††] · Hong-Gee Kim[†††]

ABSTRACT

Formal Concept Analysis (FCA) is a widely used methodology for data analysis, which extracts concepts and builds a concept hierarchy from given data. A concept consists of objects and attributes shared by those objects, and a concept hierarchy includes information on super-sub relations among the concepts. In this paper, we propose a method for extracting Implication and Association rules from a concept hierarchy given a query by a user. The method also describes a way for displaying the extracted rules. Based on this method, we implemented an automatic tool, QAG-Wizard. Because the QAG-Wizard not only elicits relation information for the given query, but also displays it in structured form intuitively, we expect that it can be used in the fields of data analysis, data mining and information retrieval for various purposes.

Key Words : Formal Concept Analysis, Concept lattice, User Query, Implication, Association Rules

1. 서 론

대용량 데이터를 저장하고 관리할 수 있는 데이터베이스 기술과 정보기술이 발전함에 따라, 개인 및 조직이 보유하고 접근할 수 있는 데이터의 양은 기하급수적으로 증가하고 있다. 그러나, 가용한 데이터 양이 증가함에 따라서, 대용량 데이터를 사용하는 도메인의 특성을 수월하게 파악하기 어렵다는 문제점이 발생하고 있다. 이러한 문제점을 해결하기 위하여, 가공되지 않은 데이터 내에 잠재되어있는 정보를 추출해내는 데이터마이닝 분야가 최근 각광을 받고 있으며, 특히 데이터들 사이의 함의관계(Implication)와 연관관계에 관한 규칙(Association Rules)은 데이터마이닝 및 데이터분석기법의 중요한 축으로 인식되고 있다. [3, 5]

데이터들간의 함의관계 및 연관관계와 관련된 정보는 다양한 분야에 적용되고 있다. 함의관계와 연관관계는 '속성 A를 갖는 객체 혹은 집단은 속성 B를 수반하는 경향이 있다'라는 일종의 패턴정보라 할 수 있으며 일반적으로 'A→B' 형식으로 표현한다. 이러한 패턴정보는, 데이터가 활용되고 있는 도메인의 현재 상황 및 경향 파악에 활용가능하며, 도메인의 향후 전망에 있어서 초석으로 활용될 수 있다. 추천 시스템(Recommend System)을 위한 기반지식(Knowledge base)로써 활용이 가능하며, 현재 다양한 연구와 개발이 진

※ 본 논문은 "보건복지부 용어 표준화지원도구 및 온톨로지 기반의 EHR 상호운용기술개발과제(과제번호: A05-0909-A80405-05N1-00050B)"와 "2007학년도 선문대학교 교수연구년제도"의 지원을 받아서 수행되었음.
† 준 회 원 : 서울대학교 치과대학 석사과정
†† 종신회원 : 선문대학교 컴퓨터공학부 부교수
††† 정 회 원 : 서울대학교 치과대학 부교수
논문접수 : 2008년 2월 28일
심사완료 : 2008년 4월 2일

행되고 있다. 그 대표적인 예로는 구매자의 상품 검색 패턴 및 구입 품목을 분석하여, 구매자가 관심 있을 법한 상품을 추천하는 Amazon.com의 상품추천(Product recommendation) 기능과 검색 시스템에서 제공하는 검색어 추천, 확장 시스템이 있다[10, 11].

한편, 본 연구의 기반이 되는 형식개념분석기법은 도메인 내의 다양한 데이터들로부터 객체(Object)와 속성(Attribute)들을 추출하고, 이들 사이의 포함관계를 파악하여 개념(Concept)을 생성하고 개념계층구조(Concept hierarchy)를 구축하기 위한 수학적 데이터마이닝 기법 중의 하나이다[1]. 형식개념분석기법을 기반으로 하여 데이터를 분류하고, 함의관계와 연관관계를 추출하기 위한 알고리즘들이 다수 제안되었고[2~6], 이를 기반으로 ConExp, Galicia, ToscanaJ 등과 같은 자동화지원도구들이 개발되었다[7~9]. 이러한 알고리즘들은 주어진 데이터 사이에 존재하는 모든 함의관계 및 연관관계를 보다 효율적으로, 중복된 정보 없이 일괄적으로 모두 추출하는 것을 목표로 하며, 자동화지원도구들 중 일부만이 이러한 기능을 제공한다. 그러나, 현존하는 기법과 도구들은, 사용자로부터 주어진 질의를 기반으로 임의의 특정한 상황에 적합한 함의관계 및 연관관계만을 추출하는 기능을 제공하고 있지 않으며<표 1>, 특히, 사용자질의를 기반으로 함의 및 연관관계를 추출하기 위해서는 추가적인 연산과정과 복잡한 가공작업을 필요로 하기 때문에 실제로 응용분야에서 활용하려는 경우에는 곤란하다.

따라서 본 연구에서는, 형식개념분석기법을 토대로, 사용자의 질의에 대한 함의관계와 연관관계를 추출하기 위한 새로운 기법을 제안하고 이를 지원하기 위한 자동화지원도구(QAG-Wizard)를 개발하였다. 기존의 많은 연구들이 구조화되어 있지 않은 데이터로부터 데이터 간의 관계를 추출하는 반면, 본 연구에서는 형식개념분석기법에서 제공하는 개념격자(Concept Lattice)가 내포하고 있는 정보를 적극 활용하여 사용자질의에 상응하는 함의 및 연관관계에 관련된 정보를 추출한다. 또한, 추출된 정보를 사용자질의기반관련그래프(Query-anchored Association Graph)로 구조화하고, 노드 간 순환이 없는 가중치가 부여된 유방향성그래프(Weighted directed acyclic graph)형태와 트리(Tree)형태로 표현하기 위한 제반 방법과 자동화지원도구를 제안한다. 데이터 분석가는 이를 활용하여 자신의 목적, 즉, 사용자 질의에 적합한 데이터 간의 관계 정보를 보다 직관적이면서도 수월하고 풍부하게 획득, 활용할 수 있다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 연구의 기초가 되는 형식개념분석의 기본적인 개념 및 제반 정의들을 소개하고, 제3장에서는 주어진 사용자 질의를 기반으로, 함의/연관관계 추출을 위한 제반 정의들 및 알고리즘과, 이를 토대로 개발한 자동화 지원도구를 소개한다. 제4장에서는 간단한 실험결과 및 연구결과의 활용방안을 기술하며, 5장에서는 결론과 향후 연구과제에 대해 설명한다.

2. 형식개념분석기법

형식개념분석기법[1, 2]은 개념격자라는 수학적 모델을 기반으로 하는 데이터분석기법의 일종으로서, 데이터분석과 지식처리분야의 제반 문제들에 대한 수학적 해법을 제공한다. 본 논문에서는 형식개념분석기법의 기본 개념과 제반 정의들을 간단한 예와 함께 소개한다.

[정의 1] Formal context [1]

Formal context K=(G, M, I)는 객체들의 집합 G와 속성들의 집합 M, 그리고 G와 M사이의 이항관계 $I \subseteq G \times M$로 구성된다. 즉, G와 M의 원소들은 각각 해당 formal context의 객체들과 각 객체들이 가질 수 있는 속성들을 나타낸다. 또한, 어떤 객체 g가 속성 m을 가지고 있을 경우, gIm 또는 $(g, m) \in I$ 로 나타내며, g는 m을 갖는다는 것을 의미한다. ■

Formal context는 형식개념분석기법의 기본이 되는 구조로서, 데이터 테이블 형태로 나타낼 수 있으며, 해당 표의 행과 열의 헤드부분은 각각 formal context를 구성하는 객체들과 속성들로 구성된다. 또한, 데이터 테이블의 각 셀에 대해서는, 해당 셀에 관련된 객체와 속성이 이항관계 I를 만족할 경우에는 X표시하고, 이외의 경우에는 빈 공란으로 남겨둔다. <표 2>는 컴퓨터공학부 학생을 객체 G로 과목을 속성 M으로 하여, 학생 별 수강한 과목을 나타낸 formal context의 예이다.

이와 같은 formal context로부터 동일한 속성을 갖는 객체들을 클러스터링(Clustering)하여, 정보의 기본단위로서 개념들(Concepts)을 추출할 수 있다. 각 개념들은 (O, A)와 같은 형태의 쌍(pair)으로 정의되며, 보다 정형적인 정의는 다음과 같다.

〈표 1〉 형식개념분석기법 자동화지원도구들의 기능 비교

	개념추출 및 개념격자 생성	함의관계추출		연관관계추출	
		모든 함의 관계 일괄추출	사용자 질의기반 함의 관계 추출	모든 연관 관계 일괄추출	사용자 질의기반 연관 관계 추출
ConExp[7]	O	O	X	O	X
Galicia[8]	O	X	X	X	X
ToscanaJ[9]	O	X	X	X	X
QAG-Wizard	O	X	O	X	O

〈표 2〉 학생 별 수강한 과목에 대한 formal context의 예

	t1	t2	t3	t4	T5	t6	t7	t8	t9	t10
s1	X	X			X					
s2	X	X	X			X	X			
s3	X			X				X	X	X
s4	X		X				X			
s5	X	X			X					
s6	X			X					X	
s7	X	X	X			X	X			
s8	X		X				X			
s9	X	X				X				
s10	X	X				X				
s11	X		X							
s12	X	X	X			X	X			
s13	X			X				X		
s14	X	X				X				
s15	X			X				X		

t1: C언어, t2: 자료구조, t3: Java언어, t4: 논리설계, t5: 운영체제, t6: 알고리즘, t7: 소프트웨어공학, t8: 계산기구조, t9: 컴퓨터구조, t10: 마이크로프로세서

〈표 3〉 표2의 context로부터 추출한 개념들

Concept	Extent	Intent
c1	Ø	{t1, t2, t3, t4, t5, t6, t7, t8, t9, t10}
c2	{s1, s5}	{t1, t2, t5}
c3	{s2, s7, s12}	{t1, t2, t3, t6, t7}
c4	{s1, s2, s5, s7, s9, s10, s12, s14}	{t1, t2}
c5	{s3}	{t1, t4, t8, t9, t10}
c6	{s1, s2, s3, s4, s5, s6, s7, s8, s9, s10, s11, s12, s13, s14, s15}	{t1}
c7	{s2, s4, s7, s8, s12}	{t1, t3, t7}
c8	{s3, s6}	{t1, t4, t9}
c9	{s2, s7, s9, s10, s12, s14}	{t1, t2, t6}
c10	{s2, s4, s7, s8, s11, s12}	{t1, t3}
c11	{s3, s13, s15}	{t1, t4, t8}
c12	{s3, s6, s13, s15}	{t1, t4}

[정의 2] Formal concept [1]

임의의 Formal context K=(G, M, I)에 대하여, O⊆G, A⊆M일 때, intent(O)=A ∧ extent(A)=O를 만족하는 (O, A)를 개념(formal concept)이라고 한다.

단, intent(O):={a∈M|∀o∈O:(o, a)∈I}=O', extent(A):={o∈G|∀a∈A:(o, a)∈I}=A'. ■

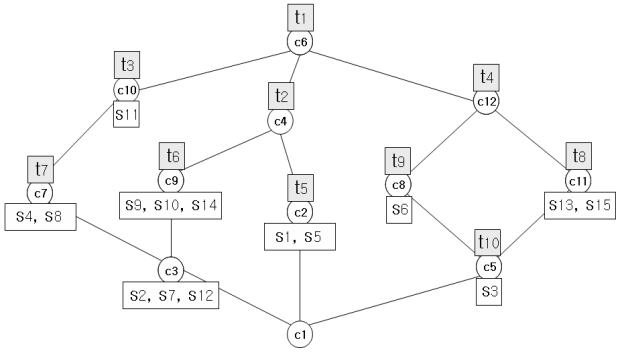
임의의 O⊆G에 대하여, intent(O)에 의해 O의 모든 객체들이 공통적으로 갖는 속성들의 집합을 구할 수 있다. 예를 들면, 위의 표 2의 context에 있어서, O={s1, s2}에 대하여, intent(O)={t1, t2}이다. 한편, 임의의 A⊆M에 대하여, extent(A)에 의해 A의 속성들을 갖는 객체들의 집합을 구할 수 있다. 예를 들면, A={t6, t7}에 대하여, extent(A)={s2, s7, s12}이다.

즉, 각 개념들은 (O, A)와 같은 형태의 쌍(pair)으로 정의되며 특히, 객체집합O는 속성집합A의 extent이며, 동시에, 속성집합A는 객체집합O의 intent가 된다. 이와 같은 방법으로 formal context K=(G, M, I)로부터 추출한 모든 개념들의 집합을 B(K):={(O, A)∈2^G×2^M | intent(O)=A ∧ extent(A) = O}로 나타낸다. 표 2의 formal context로부터 추출한 모든 개념들 B(K)는 표3과 같다. 이후 본 논문에서 기술되는 c1~12는 <표 3>에 표기된 개념을 나타낸다.

위와 같이 주어진 formal context로부터 추출된 개념들은 extent 또는 intent를 토대로 상위-하위개념관계를 정의할 수 있다.

[정의 3] 상하위관계(Superconcept-Subconcept relation) [1]

주어진 formal context의 임의의 개념 (O1, A1), (O2, A2)∈B(K)에 대하여, 상하위관계 (O1, A1)≤(O2, A2)는 일종의 반순서관계(partial order relation)로서 다음과 같이 정의된다.



(그림 1) 표2의 context에 대한 개념격자

$(O1, A1) \leq (O2, A2) \Leftrightarrow O1 \subseteq O2 (\Leftrightarrow A1 \supseteq A2)$. ■

[정의 4] 근접 하위 이웃(lower neighbor)과 근접 상위 이웃(upper neighbor) [1]

Formal context K=(G, M, I)에 존재하는 모든 개념 집합을 B(K)라 하고, 개념 (X_1, Y_1), $(X_2, Y_2) \in B(K)$에 대하여, $(X_1, Y_1) < (X_2, Y_2)$를 만족하고, $(X_1, Y_1) < (X_3, Y_3) < (X_2, Y_2)$를 만족하는 (X_3, Y_3)가 B(K)에 존재하지 않을 때, (X_1, Y_1)을 (X_2, Y_2)의 근접 하위 이웃이라 하고, (X_2, Y_2)를 (X_1, Y_1)의 근접 상위 이웃이라 하며 $(X_1, Y_1) \prec (X_2, Y_2)$라 표기한다. ■

Formal context K=(G, M, I)로부터 만들어진 모든 개념들 간의 상위-하위개념관계$\leq$는 일종의 반순서관계(partial order relation)에 해당하며, 개념들 사이의 상하위관계에 의해 만들어진 계층적 개념구조를 개념격자(Concept Lattice 또는 Galois Lattice)라고 부르고 $L:=(B(K), E_{\leq})$과 같이 표현한다. <표 2>에 기술된 formal context로부터 추출된 개념과 개념간의 상하위관계 정보를 담고 있는 개념격자를 Hasse Diagram을 사용하여 (그림 1)과 같이 가시화할 수 있다(원은 개념을, 각 원의 상위에 위치한 레이블은 개념의 intent를, 하위에 위치한 레이블은 개념의 extent를 각각 나타낸다).

[정의 5] 함의관계(Implication) [2]

주어진 Formal context K=(G, M, I)의 임의의 두 속성 Q, R$\subseteq$M 이, extent(Q)$\subseteq$extent(R)를 만족하는 경우, Q는 R을 함의한다 라고 부르며, Q$\Rightarrow$R로 표기한다. ■

임의의 Q, R$\subseteq$M에 대하여, extent(Q)와 extent(R)에 의해서 속성 Q과 R을 갖는 개체들을 각각 구할 수 있다. 예를 들어, <표 2>의 context에 있어서, Q={t10}, R={t9}이라 하면 extent(Q)={s3}, extent(R)={s3, s6}을 각각 구할 수 있으며, 이는 extent(Q)$\subseteq$extent(R)을 만족하므로, Q$\Rightarrow$R, 즉 {t10}$\Rightarrow${t9}가 성립함을 알 수 있다. 다시 말해, 표2에 주어진 context의 데이터로부터 '마이크로프로세서를 수강한 학생은 컴퓨터구조 과목 역시 수강한다'라는 추가적인 정보를 추출해 낼 수 있다.

[정의 6] 연관관계(Association rule) [2]

주어진 Formal context K=(G, M, I)의 임의의 두 속성 Q, R$\subseteq$M 이,

$\frac{|\text{extent}(Q \cup R)|}{|G|} \geq$minsup와 $\frac{|\text{extent}(Q \cup R)|}{|\text{extent}(Q)|} \geq$minconf를 만족하는 경우

Q는 R과 연관된다 라고 하며, Q$\rightarrow$Rminsup,minconf로 표기한다. 단 minsup, miconf$\in$[0, 1]. ■

정의 6에서 언급된 $\frac{|\text{extent}(Q \cup R)|}{|G|}$와 $\frac{|\text{extent}(Q \cup R)|}{|\text{extent}(Q)|}$는, 각각 연관관계 Q$\rightarrow$R의 지지도(support)와 확신도(confidence)라고 부르며, 특히, minsup(Minimum Support: 최소 지지도)와 minconf(Minimum Confidence: 최소 확신도)는 집합 Q가 R과 갖는 연관관계를 바라보는 분석가의 주관적인 경계값(Threshold)이다. 먼저 minsup는 Q와 R 사이에 존재하는 관계가 전체 객체들 중 적어도 얼마나 되는 객체 사이에서 성립할 경우에, 이 두 집합 간의 연관관계를 인정할 것인가에 대한 경계 값이다. 그리고 minconf는 Q라는 속성을 갖는 개체 중 적어도 얼마나 되는 개체가 R이라는 속성 역시 가질 때, Q는 R과 연관관계에 있다고 판단 할 것인지에 대한 경계 값이다. 예를 들어, <표 2>의 context에 있어서, Q={t3}, R={t7}이라 하고, minsup=0.3, minconf=0.6이라 하면, $|\text{extent}(\{t3, t7\})|/|G| \fallingdotseq 0.33 \geq$minsup, $|\text{extent}(\{t3, t7\})| / |\text{extent}(\{t3\})| \fallingdotseq 0.83 \geq$minconf 이므로, Q$\rightarrow$R 즉, 30%이상의 지지도와 60%이상의 확신도에서 {t3}$\rightarrow${t7}은 성립한다고 할 수 있다. 이를 문맥상 'Java언어를 수강하는 학생 중 60% 이상이 소프트웨어공학을 수강하는 경향을 보이며, 이러한 현상은 전체 학생 중 30% 이상의 학생으로부터 관찰된다' 라고 해석할 수 있다. 특히, minsup=0, minconf=1인 경우의 연관관계는 곧 함의관계를 나타내므로, 함의관계는 연관관계의 특수한 경우라 할 수 있다.

3. 질의기반관련그래프의 추출 및 자동화 지원도구

3.1 질의기반관련그래프의 추출

본 절에서는 형식개념분석기법에서 제공하는 개념격자를 활용하여 사용자의 질의를 기반으로 함의관계 및 연관관계를 추출하고, 가중치가 부여된 비순환 유방향성 그래프 및 트리 형태로 가시화하여 표현하기 위한 제반 정의들과 알고리즘을 설명한다. 즉, 구체적으로는, 주어진 질의Q에 대해서, 적합한 함의관계 Q$\Rightarrow$R과 연관관계 Q$\rightarrow$Rminsup, minconf를 만족하는 'R'에 관련된 제반 정보를 개념격자로부터 추출하여 그래프 및 트리 형태로 표시하기 위한 제반 정의들과 알고리즘을 제안한다.

[정의 7] 질의에 적합한 개념 및 질의노드

Formal context K=(G, M, I)에 대한 질의Q$\subseteq$M가 주어졌

을 때, Q⊆Y를 만족하는 임의의 개념(X, Y)∈B(K)에 대해서, {(X1, Y1)∈B(K)|(X, Y)≺(X1, Y1)∧Q⊆Y1}=Ø인 경우, 개념(X, Y)를 질의 Q에 적합한 개념이라고 부르고, (Q, (X, Y))를 질의 Q에 대한 질의노드라고 정의한다. ■

예를 들어 표2의 context에 대하여, 질의 Q={t2}라 할 때, (X, Y)=({s2, s7, s12}, {t1, t2, t3, t6, t7})=c3라 하면, Q⊆Y는 만족하지만, {(X1, Y1)∈B(K)|c3≺(X1, Y1)∧Q⊆Y1}={c9}≠Ø이므로 c3는 질의 {t2}에 적합한 개념이라 할 수 없다. 반면 (X, Y)=({s1, s2, s5, s7, s9, s10, s12, s14}, {t1, t2})=c4라 하면, Q⊆Y를 만족하고 {(X1, Y1)∈B(K)|c4≺(X1, Y1)∧Q⊆Y1}=Ø이므로, 개념 c4를 질의 {t2}에 적합한 개념이라 할 수 있으며 ({t2}, ({s1, s2, s5, s7, s9, s10, s12, s14}, {t1, t2}))를 질의 {t2}에 대한 질의노드라 할 수 있다.

[정의 8] 함의노드

Formal context K=(G, M, I)에서, 임의로 주어진 질의 Q⊆M에 대한 질의노드(Q, (X, Y))에 대하여, Q≠Y ∧ $\frac{|X|}{|G|}$≥minsup∈[0, 1]인 경우, (Y\Q, (X, Y))를 질의 Q에 대한 함의노드라 정의한다. ■

함의관계를 표현하는 함의노드를 정의함에 있어 경계 값인 minsup을 반영하는 이유는, 2장에서도 언급했듯이, 함의관계는 연관관계의 특수한 경우이기 때문이다. 즉 고유한 의미의 함의관계를 반영하는 함의노드는 minsup=0인 경우라 할 수 있다. 반면 함의관계를 표현하는 함의노드를 정의함에 있어, minconf를 반영하지 않은 이유는 다음과 같다. 질의 Q에 적합한 개념을 (X, Y)∈B(K)라 할 때, Q≠Y가 성립하여 함의노드가 존재할 경우, 정의 2에 의하여, 항상 extent(Q)⊆extent(Q\Y), 다시 말해 Q⇒Q\Y를 만족하므로, 즉 minconf=1이므로 고려하지 않는다.

예를 들어 <표 2>의 context에 대하여, 질의 Q={t2}, minsup=0.4일 때, 질의 Q에 적합한 개념은 c4이며, 질의노드는 ({t2}, c4)이다. 또한 질의 {t2}가 c4의 intent인 {t1, t2}와 같지 않으며, |{s1, s2, s5, s7, s9, s10, s12, s14}|/|G|≒0.53≥minsup이므로, 질의 {t2}에 대한 함의노드는 ({t1}, ({s1, s2, s5, s7, s9, s10, s12, s14}, {t1, t2}))라고 할 수 있다.

[정의 9] 연관노드

Formal context K=(G, M, I)에 대한 주어진 질의 Q⊆M에 적합한 개념이 (X, Y)∈B(K) 일 때, 임의의 (X1, Y1), (X2, Y2)∈B(K)가, (X1, Y1)<(X, Y) ∧ (X2, Y2)≤(X, Y) ∧ (X1, Y1)≺(X2, Y2)를 만족하고, $\frac{|X1|}{|G|}$≥minsup, $\frac{|X1|}{|X|}$≥minconf 이면, (Y1\Y2, (X1, Y1))를 질의Q에 대한 연관노드이다 라고 정의한다. 단 minsup, minconf∈[0, 1] ■

주어진 질의 Q에 적합한 개념이 (X, Y)라면, (X1, Y1)<(X, Y) 관계에 있는 개념 (X1, Y1)의 extent인 X1은, 정의 3에 의해, Y⊃Q를 속성으로 가지면서 Y1\Y를 추가 속성으로 갖는 객체의 집합이라 할 수 있다. 이때, 정의 3(상하위관계)과 정의 4(근접하위이웃 및 근접상위이웃)를 활용하여, 주어진 질의Q와 연관관계에 있는 속성을 도출해 낼 수가 있다. 예를 들어 <표 2>의 context에 대하여, 질의 Q={t2}, 경계 값 minsup=0.1, minconf=0.25이라 하면, 질의 {t2}에 적합한 개념은 c4가 된다. 이때 (X1, Y1)=c3, (X2, Y2)=c9라 하면, c3<c4, c9≤c4, c3≺c9를 만족하며, |X1|/|G|=0.2≥minsup, |X1|/|X|≒0.38≥minconf가 성립하므로, Y1\Y2={t1, t2, t3, t6, t7}\{t1, t2, t6}={t3, t7}이 된다. 즉, Q={t2}이고, minsup=0.1, minconf=0.25인 경우, ({t3, t7}, ({s2, s7, s12}, {t1, t2, t3, t6, t7}))를 연관노드라 할 수 있다.

지금까지 정의한, Formal context K(G, M, I)에 주어진 질의 Q⊆M, 각 경계 값 minsup, minconf에 대한, 질의노드, 함의노드 및 연관노드는, (Y*, (X, Y))와 같은 형태로 일반화하여 나타낼 수 있다(단, (X, Y)∈B(K), Y*⊆Y). 이 때, Y*를 노드의 레이블(Label), (X, Y)를 노드의 개념(Concept)이라 하며, 모든 노드들의 집합을 N(KQ)minsup,minconf 라 표기하며, 이를 질의기반관계노드(QAN: Query-anchored Association Node)라 한다. <표 2>의 context에 대하여, N(K{t2})0.1,0.25의 원소를 각 구성요소 별로 정리하여 나타내면 <표 4>과 같다.

[정의 10] 노드 간 관계

Formal context K=(G, M, I)에 주어진 질의 Q⊆M, minsup, minconf에 관한 두 노드 n_1=(Y1*, (X1, Y1)), n_2=(Y2*, (X2, Y2))∈N(KQ)minsup,minconf 에 대하여 다음과 같은 조건을 만족하면 n_1은 n_2와 질의 Q에 대해 연관관계에 있다고 정의하고 r(n_1, n_2)로 표기한다.

$$r(n_1, n_2)\begin{cases}(X2, Y2)\prec(X1, Y1)\wedge Y1^*\cap Y2^*=\emptyset\wedge Y1^*\neq Q \\ \quad \text{혹은 } (X1, Y1)=(X2, Y2)\wedge Y1^*=Q \\ \text{(함의노드가 } N(K_Q)\text{minsup,minconf 내에 존재할 경우)} \\ \text{or} \\ (X2, Y2)\prec(X1, Y1)\wedge Y1^*\cap Y2^*=\emptyset \\ \text{(함의노드가 } N(KQ)\text{minsup,minconf 내에 존재하지 않는 경우)}\end{cases}$$

단, R⊆N(KQ)minsup,minconf × N(KQ)minsup,minconf ■

〈표 4〉 표 2의 context에 대한 N(K{t2})0.1,0.25

노드	유형	레이블	개념
n1	질의노드	{t2}	c4
n2	함의노드	{t1}	c4
n3	연관노드	{t5}	c2
n4	연관노드	{t6}	c9
n5	연관노드	{t3, t7}	c3

[정의 11] 관계 가중치

Formal context K=(G, M, I)에 주어진 질의 Q⊆M, minsup, minconf에 관한 두 노드 n_1=(Y1*, (X1, Y1)), n_2=(Y2*, (X2, Y2))∈N(KQ)minsup,minconf에 대하여, 관계 r=(n_1, n_2)∈R이 존재할 때, 관계 r의 가중치를 W(r) = ($\frac{|X2|}{|G|}$, $\frac{|X2|}{|Q'|}$)라 정의하며, $\frac{|X2|}{|G|}$를 관계 r의 지지도(support), $\frac{|X2|}{|Q'|}$를 관계 r의 확신도(confidence)라 각각 정의한다. ■

<표 4>을 예로 들어 설명하면, 이는 함의노드가 N(K{t2}) 0.1,0.25 내에 존재하는 경우에 속한다. 그러므로 n_1=({t2}, c4), n_2=({t1}, c4)∈N(K{t2})0.1,0.25 에 대하여, (n_1의 개념 = n_2의 개념) ∧ (n_1의 레이블 = Q)이므로, n_1은 n_2와 질의 Q에 대해 연관관계에 있다고 할 수 있으며, 관계 r=(n_1, n_2)의 가중치는 (0.53, 1)이라 할 수 있다. 또한 n_4=({t6}, c9), n_5=({t3, t7}, c3)∈N(K{t2})0.1,0.25 에 대하여, c3≺c9 ∧ {t6} ∩ {t3, t7}= Ø ∧ {t6}≠Q 이므로, n_4는 n_5와 질의 Q에 대해 연관관

〈표 5〉 표 4에 존재하는 노드 간의 관계 집합 R

관계	Source	Target	Support	Confidence
r1	n1	n2	0.53	1
r2	n2	n3	0.13	0.25
r3	n2	n4	0.4	0.75
r4	n4	n5	0.2	0.38

* Support와 Confidence는 각각 소수점 아래 셋째 자리에서 반올림한 결과이다

질의기반관련그래프 생성 알고리즘

Input: Formal Context K=(G, M, I)의 개념격자 L=(B(K), $E_{\leq}$), 사용자 질의Q⊆M, minsup, minconf∈[0, 1]
Output: 질의기반관련그래프 (N(KQ)minsup,minconf, R)

```
1. N(KQ) minsup,minconf:=Ø, R:=Ø,  nextLevel:=Ø
2. for each (X, Y)∈B(K)
3.     if Q⊆Y then
4.         Super:=Find all (X1, Y1)∈C such that (X, Y)≺(X1, Y1) in E≤
5.         if Q⊄Y1 of any concept (X1, Y1) in Super then
6.             N(KQ) minsup,minconf:={(Q, (X, Y))}
7.             if |X|/|G|≥minsup then
8.                 if Y\Q≠Ø and then
9.                     N(KQ) minsup,minconf:=N(KQ) minsup,minconf ∪{(Y\Q, (X, Y))}
10.                    nextLevel:={(Y\Q, (X, Y))}
11.                    Add edge (Q, (X, Y))→(Y\Q, (X, Y)) to R
12.                    (Q, (X, Y))→(Y\Q, (X, Y)).weight=(|X|/|G|, 1)
13.                else then
14.                    nextLevel:={(Q, (X, Y))}
15.            break loop;
16. if nextLevel≠Ø then
17.    addAssociationNode(nextLevel)
```

function *addAssociationNode(nextLevel)*

```
1. currentLevel:=nextLevel
2. nextLevel:=Ø
3. for each (Y*, (X, Y))∈currentLevel
4.     Sub:=Find all (X1, Y1)∈C such that (X1, Y1)≺(X, Y) in E≤
5.     for each (X1, Y1)∈Sub
6.         if |X1|/|G|≥minsup and |X1|/|Q'|≥minconf then
7.             if ({Y1\Y}, (X1, Y1))∉N(KQ) minsup,minconf then
8.                 N(KQ) minsup,minconf:=N(KQ) minsup,minconf ∪{({Y1\Y}, (X1, Y1))}
9.                 nextLevel:=nextLevel∪{({Y1\Y}, (X1, Y1))}
10.                Add edge (Y*, (X, Y))→({Y1\Y}, (X1, Y1)) to R
11.                (Y*, (X, Y))→({Y1\Y}, (X1, Y1)).weight=(|X1|/|G|, |X1|/|Q'|)
12. if nextLevel≠Ø then
13.    addAssociationNode(nextLevel)
```

계에 있다고 할 수 있으며, 관계의 가중치는 (0.2, 0.38)라 할 수 있다. <표 4>에 표기된 노드 사이에 존재하는 모든 관계 및 관계의 가중치를 기술하면 <표 5>와 같다.

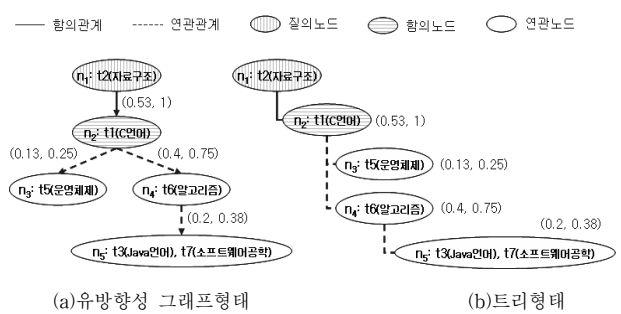
Formal context K=(G, M, I), 사용자 질의 Q⊆M, minsup, minconf∈[0, 1] 일 때 추출된 N(K_Q)minsup,minconf과 R에 의해 생성되는 구조를 '질의기반관련그래프(QAG: Query-anchored Association Graph)'라 부르고 (N(KQ)minsup,minconf, R)과 같이 표현한다. 질의기반관련그래프는 개념격자로부터 다음과 같은 방법을 사용하여 추출해 낼 수 있다.

위와 같은 방식으로 추출된 질의기반관련그래프를 토대로, 본 논문에서는 2가지 형태(비순환형 유방향성그래프형태와 트리형태)로 가시화하는 방법을 제안한다. (그림 2)는, <표 2>의 context에 대해, 질의 Q={t2}, minsup=0.1, minconf = 0.25 일 때의 질의기반관련그래프를, 노드의 레이블 중심의 유방향성 그래프(그림2(a) 참조)와 트리(그림2(b) 참조)형태로 각각 가시화 한 예이다.

[정의 12] 질의기반관련그래프의 해석 경로

Formal context K=(G, M, I), Q⊆M, minsup, minconf∈[0, 1]에 대한 질의기반관련그래프 (N(K_Q)minsup,minconf, R)에 대하여, 임의의 순서집합(Ordered set)을 P={n_1, n_2, n_3, …, n_m}⊆N(K_Q)minsup,minconf라 할 때, n_1∈P의 레이블이 Q이고, 1≤k≤m-1를 만족하는 k에 대하여, 모든 n_k, n_{k+1}∈P에 있어, r(n_k, n_{k+1})∈R를 만족할 때, 순서집합 P를 질의기반관련그래프의 해석 경로라 정의하며, 'n_1의 레이블→ n_2의 레이블∧...∧n_m의 레이블' 이라 해석한다. 또한 해석 경로 P의 가중치는 r(n_{m-1}, n_m)∈R 의 가중치라 정의한다. ■

(그림 2)에 표현된 질의기반관련그래프의 해석 경로의 한 예를 살펴보면 다음과 같다. 순서집합 P={n_1, n_2, n_4}⊆N(K{t2})0.1,0.25라 하면, 표5에서도 알 수 있듯이, 집합P의 원소간의 관계 r(n_1, n_2), r(n_2, n_4)∈R이므로, 순서집합 P는 질의기반관련그래프의 해석 경로라 할 수 있으며, 't2→t1∧t6'라 해석할 수 있다. 이를 보다 문맥에 맞게 해석하면, '자료구조를 수강하는 학생은 C언어와 알고리즘을 수강하는 경향이 있다' 라고 해석할 수 있다. 또한 수치정보인 가중치 정보를 부과하여 해석하면 '자료구조를 수강한 학생 중 25% 이상(정확히 25%)의 학생이 C언어와 알고리즘을 수강하는 경향이 있고, 이러한 현상은 전체 학생 중 10% 이상(정확히 13%)의 학생들로부터 관찰된다' 라고 해석할 수 있다.



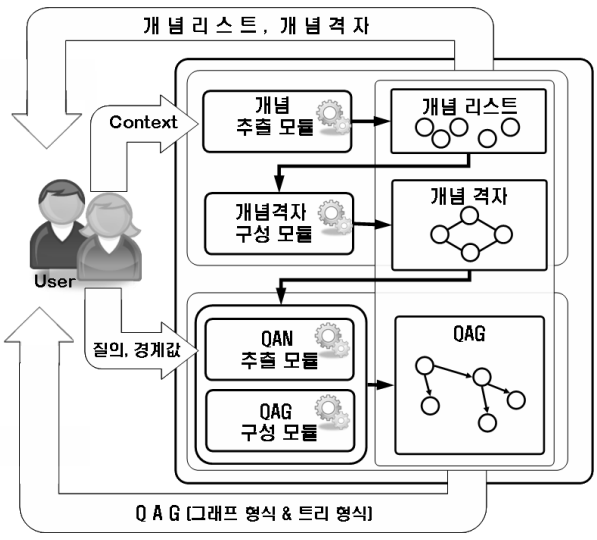
(a)유방향성 그래프형태 (b)트리형태

(그림 2) 표2의 context K에 대한 (N(K{t2})0.1,0.25, R)의 가시화

3.2 질의기반관련그래프 자동화지원도구의 개발

형식개념분석기법을 지원하는 자동화지원도구들[7~9]이 다양하게 개발되고 있다. 그러나, 대부분의 지원도구에서는 함의관계 및 연관관계를 추출하고 가시화하는 기능이 매우 빈약하며, 특히, <표 1>과 같이 사용자의 질의에 즉시적으로 대응하는 연관관계를 추출하는 기능은 지원하지 못하고 있어서, 사용자의 요구에 부응하는 관련정보를 추출함에 어려움이 있다. 따라서, 본 연구에서는 형식개념분석기법의 핵심기본 요소라 할 수 있는 formal context, 개념 및 개념격자의 추출 및 가시화와 더불어서, 질의기반관련그래프 추출 및 가시화 기능 등을 지원하는 QAG-Wizard를 개발하였다. (그림 3)은 QAG-Wizard의 전체적인 구성요소들과 각 기능을 나타내고 있다.

본 연구에서 개발한 QAG-Wizard는 형식개념분석기법의 제반 정의들(formal context, 개념, 개념격자 등)과 관련 알고리즘들[2]을 기반으로 개념 추출 및 개념격자 구성 모듈을 각각 구현하였다. 개념 추출 모듈은 사용자로부터 입력 받은 formal context를 기반으로, formal context 내에 존재하는 모든 개념을 추출해내는 모듈이며, 개념격자 구성 모듈은, 추출된 개념 간의 근접하위, 근접상위 이웃 관계를 도출하여 개념격자를 구성한다. JTable[14] 및 JPowerGraph[12]를 통해 개념 리스트 및 개념격자를 각각 가시화하였다. 또한, QAN(질의기반관련노드) 추출 모듈은 사용자로부터 입력 받은 질의 및 경계 값과 개념격자를 입력으로 하여, 질의기반관련노드를 추출하며, QAG(질의기반관련그래프) 구성 모듈은 생성된 질의기반관련노드 간의 관계를 도출하여, 질의기반관련그래프를 구성한다. 본 연구에서 제안한 질의기반관련그래프 추출기법 및 관련 알고리즘을 토대로 질의기반관련그래프 추출모듈을 구현하였고, JPowerGraph와



(그림 3) QAG-Wizard의 구성요소 및 기능

JTree[13]를 사용하여 각각 유방향성 그래프와 트리 형식으로 가시화하였다.

(그림 4)는 QAG-Wizard의 사용자 인터페이스(User interface)의 모습이다. (그림 4)의 ①영역은 formal context를 편집하는 기능을 지원하며, ②영역은 편집된 formal context 내에 존재하는 개념들의 리스트를, ③영역은 개념격자를 각각 가시화하며, ④영역은 사용자의 질의 및 경계 값을 입력 받고, 질의기반관련그래프를 가시화하는 기능을 지원한다. 질의기반관련그래프는 탭(Tab)을 활용하여, 그래프 및 트리 형식으로 가시화된다. 또한 QAG-Wizard는 사용자에게 동적인 사용자 인터페이스를 제공한다. 그림 5에서 확인할 수 있듯이, 사용자는 ①~④ 영역의 조합을, 선택적으로 활용, 제어할 수 있는 환경을 제공받는다. 특히 수집된 대량의 데이터를 처리해야 할 경우 이를 CSV(Comma Separated Values) 파일 형태로 저장, QAG-Wizard에서 쉽게 읽어 들여 처리할 수 있도록 구축되었다.

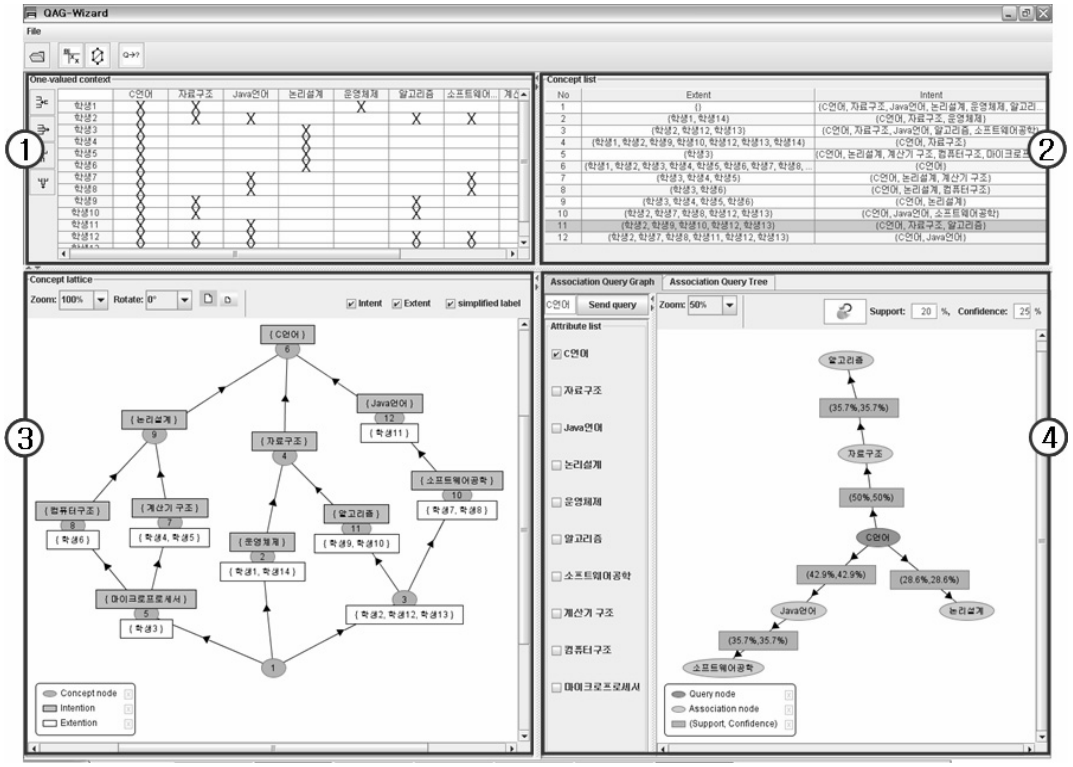
뿐만 아니라, QAG-Wizard는 사용자의 편의를 위해 다양한 옵션을 지원한다. 먼저 ②영역과 ③영역을 동기화 시켜, ③영역의 개념격자에서 임의의 개념을 선택하였을 경우, ②영역의 개념 리스트에서 선택된 개념을 자동으로 강조해주는 기능을 구현하였다(그림4, 5 참조). 또한 개념격자가 복잡한 구조를 형성할 때, 사용자의 보다 직관적인 개념격자의 구조 파악을 위하여, 개념의 intent, extent를 선택적으로 가시화하는 옵션과, intent 및 extent를 간략화(Simplified)하여 가시화하는 옵션을 제공한다(그림 6 참조).

질의기반관련그래프를 추출하고 가시화하는 ④영역은, 추출된 관계를 그래프 및 트리로 각각 가시화하는 기능을 제공하며, 그래프 기반의 가시화 모듈에서는 표현된 관계에 속하는 객체들을 가시화할 수 있는 옵션을 지원한다(그림7 참조). 뿐만 아니라 트리 기반의 가시화 모듈에서는, 트리의 특정 노드를 선택하였을 경우, 질의기반관련그래프의 해석경로(정의 12)를 강조하여, 사용자가 보다 직관적으로 파악할 수 있는 기능을 제공한다(그림 8 참조). 특히, 트리 기반의 규칙 가시화 모듈은 트리의 루트인 질의노드로부터 단계적으로 질의기반관련그래프를 생성하므로, 그래프 기반의 가시화 모듈과 비교하여 대용량의 데이터 처리에 있어서 보다 효율적인 성능을 지원한다.

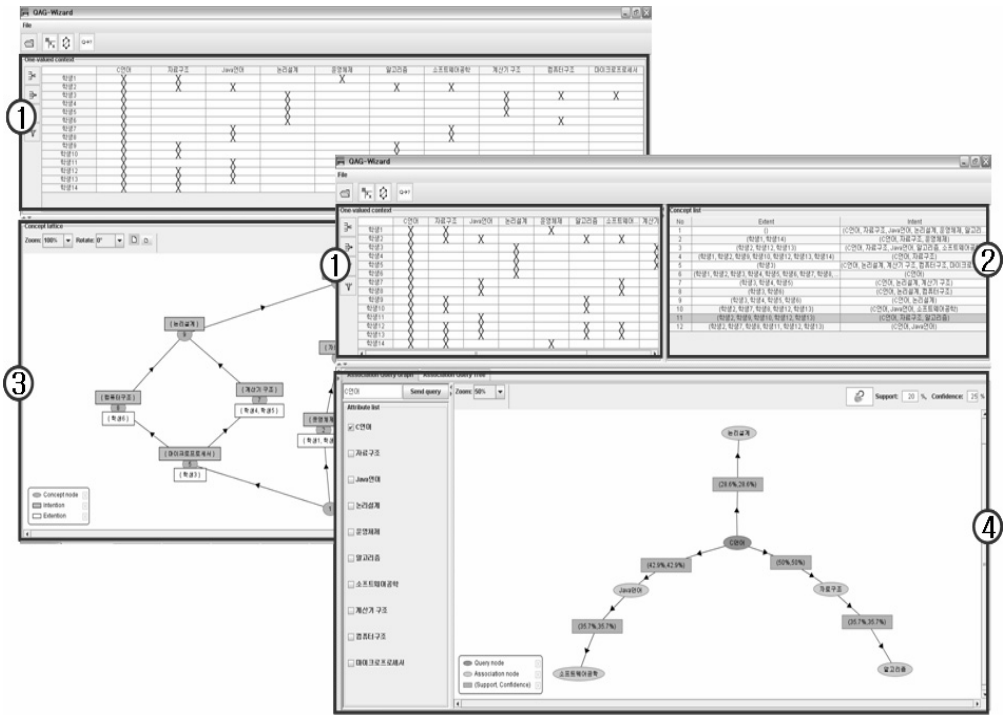
4. 실 험

본 논문에서 제안한 기법과 개발한 지원도구의 유용성을 살펴보기 위하여, <표 2>에 주어진 '학생들이 수강한 과목에 관한 formal context'를 QAG-Wizard를 사용하여 형식개념분석을 수행하고, 사용자의 질의에 기반한 관련규칙을 추출하는 실험을 실시해보았다. 구체적으로는, 본 연구에서 개발한 자동화지원도구(QAG-Wizard)를 사용하여, 질의 Q={C언어}와 minsup=0.2, minconf=0.25로 각각 임의의 값을 지정하여 함의관계 및 연관관계를 추출하고, 기존의 관계 추출방법론[2] 및 ConExp[7]에서 지원하는 기능과 비교한다.

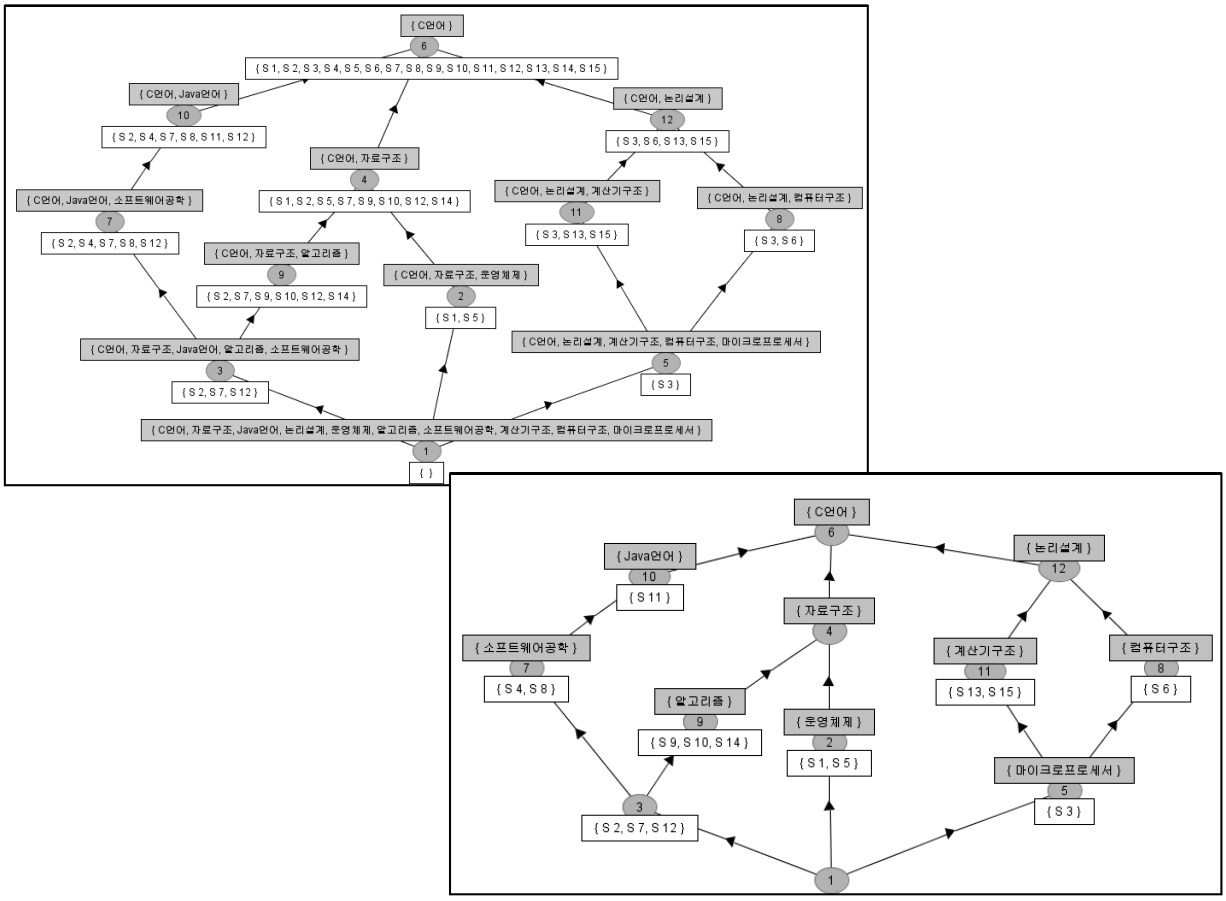
기존의 관계 추출방법론을 이용하여 함의관계 및 연관관계를 추출한 후에 추출된 규칙의 가중치를 기반으로 순서화(ordering)하면 <표 6>과 같다.



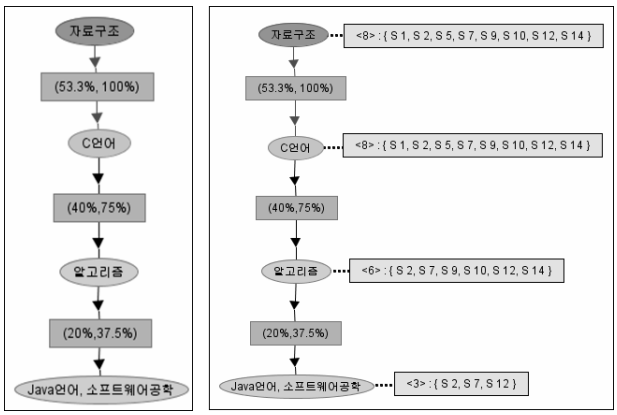
(그림 4) QAG-Wizard의 사용자 인터페이스 화면



(그림 5) QAG-Wizard의 가시화 영역 선택 옵션 지원 화면



(그림 6) QAG-Wizard의 개념격자 가시화 옵션 지원 화면



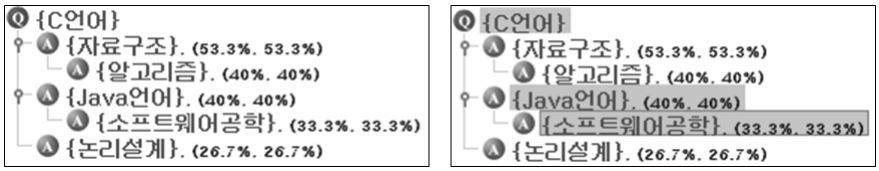
(그림 7) QAG의 그래프 기반 가시화 옵션 화면

<표 6>에 표현된 규칙들을 기반으로 {C언어}라는 수업을 수강한 학생에게 다른 수업을 추천하는 추천 시스템을 가정해보면, 규칙의 가중치를 기반하여 정해진 우선 순서대로, 자료구조 수업이 가장 상위에 추천되고, 우선순위가 낮은 논리설계 수업은 가장 하위에 추천될 수 있다. 하지만 <표 2>의 context에서 C언어라는 수업을 수강한 학생들의 속성을 살펴보면, 알고리즘을 수강한 학생은 반드시 자료구조를 수강하는 패턴을, 소프트웨어공학을 수강한 학생은 반드시 Java언어를 수강하는 패턴을 지녔음을 알 수 있다. 이를 보다 문맥에 적합하게, "C언어를 수강한 학생이 알고리즘을 수강함에 있어 요구되는 선수과목은 자료구조이다", "C언어를 수강한 학생이 소프트웨어공학을 수강함에 있어 요구되는 선수과목은 Java언어이다"라고 각각 해석할 수 있다. 반면 C언어를 수강한 학생이 논리설계를 수강하는 것은, 다른 수업들과는 독립적이라는 사실을 파악할 수 있다.

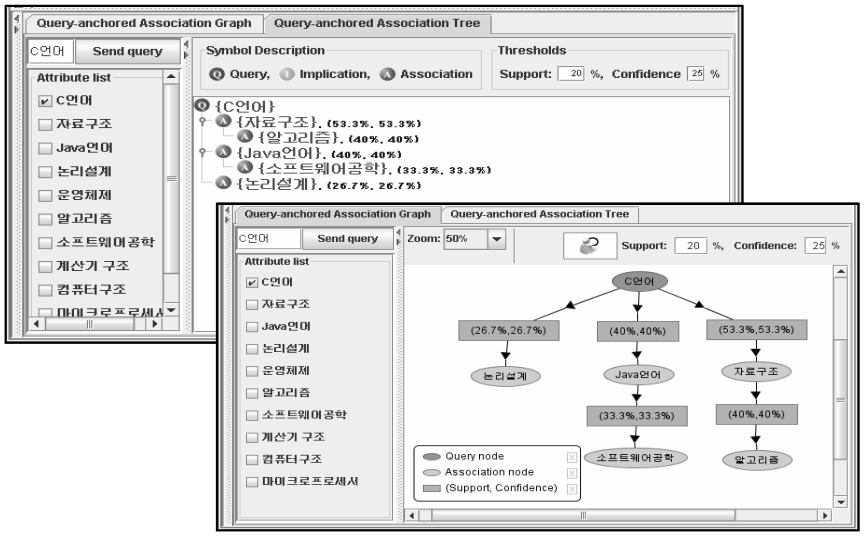
〈표 6〉 표 2의 context에서, 질의 {C언어}에 대한 연관관계의 가중치 기반 순서화

Priority	Rule	Support	Confidence
1	{C언어}→{자료구조}	0.53	0.53
2	{C언어}→{알고리즘}	0.4	0.4
2	{C언어}→{자료구조, 알고리즘}	0.4	0.4
2	{C언어}→{Java언어}	0.4	0.4
5	{C언어}→{소프트웨어공학}	0.33	0.33
5	{C언어}→{Java언어, 소프트웨어공학}	0.33	0.33
7	{C언어}→{논리설계}	0.27	0.27

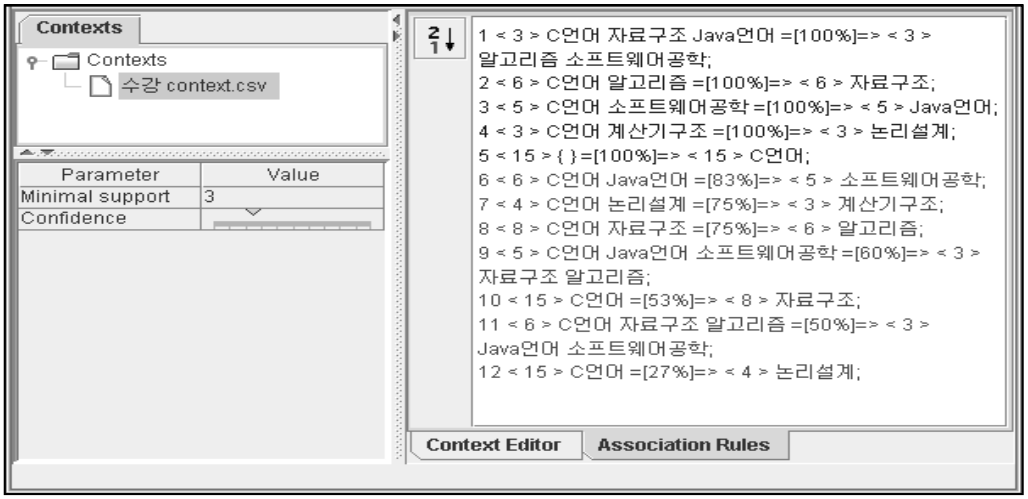
이러한 정보는 본 논문에서 제안한 질의기반관련그래프에서 손쉽게 파악할 수 있다. (그림 9)는 <표 2>의 context에 대해 질의={C언어}, minsup=0.2, minconf=0.25 일 때 QAG-Wizard에 의해 도출된, 질의기반관련그래프의 레이블 중심 표현이다. (그림 9)에 표현된 질의기반관련그래프를 살펴보면, 질의노드인 'C언어'로부터 직접적인 경로가 존재하는 노드는 '자료구조, Java언어, 논리설계'라 할 수 있으며, 이는 곧, 비록 논리설계라는 과목은, 알고리즘 및 소프트웨어공학과 비교하여 C언어라는 과목과의 관계에 있어 가중치는 낮지만, 선수과목 없이 바로 수강할 수 있는 과목이라는 것을



(그림 8) QAG의 트리 기반 가시화 옵션 화면



(그림 9) QAG-Wizard를 통한 (N(K{C언어})0.2,0.25, R)의 추출 및 가시화



(그림 10) ConExp[7]를 통한 규칙 추출 및 가시화

시사한다. 반면, 알고리즘과 소프트웨어공학 수업은, 논리설계보다 가중치는 높지만, 반드시 자료구조, Java언어를 각각 선수강해야 하므로, C언어만을 수강한 학생에게 추천할 만한 과목은 아니라는 것을 의미한다.

본 논문에서 제안된 방법론은, C언어를 듣는 학생들의 성향을 분석함에 있어, 단순한 가중치 기반 분석이 아닌, 패턴 중심의 분석을 수행한다 할 수 있다. 또한, 본 연구에서는 형식개념분석기법에 의해 생성되는 개념격자구조가 내포하고 있는 정보를 바탕으로, 사용자의 질의에 대응하는 함의관계 및 연관관계를 추출하여 가시화하고 있다. 이와 같은 연구결과는, 함의관계 및 연관관계가 기반지식으로 활용되는 추천시스템, 도메인분석 시스템, 의료분야의 진단시스템 등과 같은 분야에 활용 가능할 것으로 기대할 수 있다. 한편 QAG-Wizard와 ConExp의 함의관계 및 연관관계 추출/가시화 기능을 비교하면, ConExp의 경우 질의를 입력하는 기능이 부재하며, 추출된 규칙의 가시화 기법 역시 세련화되지 못해, QAG-Wizard가 사용자 질의 중심의 규칙 추출 및 가시화에 있어, 보다 직관적이며, 뛰어난 기능을 제공함을 확인할 수 있다(그림 9, 10 참조).

5. 결론 및 향후 과제

본 논문에서는 형식개념분석기법을 기반으로, 사용자 질의에 적합한 함의관계 및 연관관계를 구조화된 형태로 추출해내는 기법을 제안하고, QAG-Wizard를 개발하였다. 또한, 제안한 기법과 자동화지원도구의 유용성을 검증하기 위하여, 간단한 데이터를 대상으로 실험을 수행하였고, 그 결과, 기존에 제시된 관계 추출 방법론 및 도구에 비해, 보다 유용하고 풍부한 정보를 추출하여, 보다 직관적으로 표현할 수 있음을 알 수 있었다.

본 논문에서 제안한 기법과 도구의 장점을 요약하면 다음과 같다.

1. 사용자의 경계 값(minimum support, minimum confidence)이 반영된 질의에 대한 관계(implication, association rule)를 추출할 수 있다.
2. 추출된 규칙정보를 그래프/트리 형태로 구조화하여 가시화하는 기능을 제공함으로써 사용자는 직관적으로 관련정보를 파악/획득할 수 있다.
3. 형식개념분석기법(Formal Concept Analysis)에 의해 생성된 개념격자로부터 다양한 정보(개념들간의 상하위관계, 근접상하위이웃관계 등)를 활용하여 규칙을 추출하므로, 규칙들간의 패턴분석이 가능하다.

향후 연구과제로서, formal context가 표현하는 이진관계(Binary Relation)에 비해 보다 뛰어난 표현력을 지원하는 many-valued context[1] 및 scale[1]과의 융합, 대용량 데이터 처리성능 평가 및 고도화 등을 개발해 갈 예정이다.

참 고 문 헌

[1] B. Ganter, R. Wille, 'Formal Concept Analysis Mathematical Foundations,' 1st ED, Springer-Verlag, 1999.

[2] C. Carpineto, G. Romano. 'Concept Data Analysis, Theory and Applications,' 1st ED., Italy, Wiley, 2004.

[3] Nicolas Pasquier, Yves Bastide, Rafik Taouil, Lotfi Lakhal, "Efficient mining of association rules using closed itemset lattices," Information Systems, 24, 1, pp.25-46, 1999.

[4] Nicolas Pasquier, Yves Bastide, Rafik Taouil, Lotfi Lakhal, "Discovering Frequent Closed Itemsets for Association Rules," Lecture Notes in Computer Science, 1540, pp.398-416, 1999.

[5] Keyun Hu, Yuchang Lu, Lizhu Zhou, Chunyi Shi, "Integrating Classification and Association Rule Mining: A Concept Lattice Framework," Lecture Notes in Computer Science, 1711, pp.443-447, 2004.

[6] Gerd Stumme, Rafik Taouil, Yves Bastide, Nicolas Pasquier, Lotfi Lakhal, “Intelligent structuring and Reducing of Association Rules with Formal Concept Analysis,” Lecture Notes in Computer Science, 2174, pp.335-350, 2001.
[7] Concept Explorer : http://sourceforge.net/projects/conexp
[8] Galicia: http://www.iro.umontreal.ca/~galicia/
[9] Toscana J: http://tockit.sourceforge.net/toscanaj
[10] Yo-Ping Huang, Chia-Ann Tsai, Frode Eika Sandnes, “Using Association Rules for Expanding Search Engine Recommendation Keywords in English and Chinese for Search Engine Queries,” the 8th IASTED International Conference on Intelligent Systems and Control ISC 2005, Cambridge, USA, 2005, pp.465-470, 2005.
[11] Peter Bodík, Armando Fox, Michael I. Jordan, David Patterson, “Advanced Tools for Operators at Amazon.com,” First Workshop on Hot Topics in Autonomic Computing (HotAC’06), 2006.
[12] JPowerGraph: http://sourceforge.net/projects/jpowergraph/
[13] JTree: http://java.sun.com/j2se/1.4.2/docs/api/javax/swing/JTree.html
[14] JTable: http://java.sun.com/j2se/1.4.2/docs/api/javax/swing/JTable.html

김 응 희

e-mail : eungheekim@snu.ac.kr
2007년 선문대학교 컴퓨터정보학부 (이학사)
2007년~현 재 서울대학교 치과대학 석사과정(의생명지식공학전공)
관심분야 : Formal Concept Analysis, Data Mining

황 석 형

e-mail : shwang@sunmoon.ac.kr
1991년 강원대학교 전자계산학과 조기졸업 (이학사)
1993년 일본 오사카대학교 대학원 정보공학과(공학석사)
1997년 일본 오사카대학교 대학원 정보공학과(공학박사)
1997년~현 재 선문대학교 컴퓨터공학부 전임강사, 조교수, 부교수
2007년 1월~2008년 1월 아일랜드국립대학교 DERI 객원연구원
관심분야 : 소프트웨어공학, 객체지향, 온톨로지공학, Formal Concept Analysis, 시멘틱 웹

김 홍 기

e-mail : hgkim@snu.ac.kr
1985년 고려대학교 심리학(학사)
1993년 University of Georgia, Artificial Intelligence(MS)
1996년 University of Georgia, Philosophy AI (Ph.D.)
1997년~1998년 Fellow, University of Georgia 인공지능센터
1998년~2005년 단국대학교 경상학부 부교수
2005년~현 재 서울대학교 치과대학 부교수
관심분야 : Ontology, Semantic Web, 지식표현, 인공지능 등