

# Analyzing RDF Data in Linked Open Data Cloud using Formal Concept Analysis

Suk-Hyung Hwang\*, Dong-Heon Cho\*\*

## Abstract

The Linked Open Data(LOD) cloud is quickly becoming one of the largest collections of interlinked datasets and the de facto standard for publishing, sharing and connecting pieces of data on the Web. Data publishers from diverse domains publish their data using Resource Description Framework(RDF) data model and provide SPARQL endpoints to enable querying their data, which enables creating a global, distributed and interconnected dataspace on the LOD cloud. Although it is possible to extract structured data as query results by using SPARQL, users have very poor in analysis and visualization of RDF data from SPARQL query results. Therefore, to tackle this issue, based on Formal Concept Analysis, we propose a novel approach for analyzing and visualizing useful information from the LOD cloud. The RDF data analysis and visualization technique proposed in this paper can be utilized in the field of semantic web data mining by extracting and analyzing the information and knowledge inherent in LOD and supporting classification and visualization.

▶ Keyword: Linked Open Data, Resource Description Framework, SPARQL, Formal Concept Analysis

## 1. Introduction

최근 웹 분야에서는 웹기술의 진화발전과 더불어서 웹상에서 폭발적으로 증가하고 있는 다종다양한 정보자원들을 URI(Unified Resource Identifier)라는 고유한 이름과 RDF형식으로 정의하고, 이를 접근하기 위해 HTTP프로토콜을 이용하도록 하는 링크드데이터(Linked Data)형태로 작성되어 LOD클라우드(Linked Open Data Cloud)로서 웹에 공개 및 공유되고 있다[1]. RDF(Resource Description Framework)형식으로 작성된 글로벌한 대규모 시멘틱 링크드 오픈데이터(Semantic LOD)는, 미국, 영국 등을 비롯한 여러나라들이 2015년 현재 약196억개의 공공데이터를 일반에게 공개하고 있고, 국내에서는 2012년부터 정부, 지방자치단체, 공공기관들에서 일반에 공개하여 자유롭게 공유하고 활용할 수 있도록 추진하고 있다[2].

한편, 웹에 공개된 시멘틱 웹 데이터, 즉, LOD클라우드의 데이터들(이하, 본 논문에서는 LOD 또는 RDF데이터라고 부르기로 함)을 수집/분석하기 위해서 SPARQL이라는 RDF질의언어

가 주로 사용되고 있다[3]. SPARQL를 이용하여 질의결과로서 구조화된 데이터들을 간단하게 추출할 수는 있으나, SPARQL에서는 추출된 다량의 질의결과에 대한 사용자중심의 가시화 및 분석방법이 제공되지 않고 있다. 따라서, 본 논문에서는 LOD클라우드로부터 유용한 정보를 추출하고 가시화하기 위한 형식개념분석법[4]을 기반으로 하는 RDF데이터 분석 및 가시화기법을 제안한다. 본 논문에서 제안하는 RDF데이터 분석 및 가시화기법은, RDF데이터의 고유한 특성인 <주어 서술어 목적어>로 구성되는 삼원구조(Triple구조)를 기반으로 시멘틱 데이터마이닝의 핵심이 되는 분류 및 클러스터화를 위한 유용한 기법으로서 활용할 수 있다.

본 논문의 구성은 다음과 같다. 제2장에서는 RDF데이터의 분석 및 가시화방법에 대한 기존의 관련연구들에 대하여 요약한다. 제3장에서는 본 논문에서 필요한 기본적인 사항들로서, RDF와 SPARQL, 그리고 형식개념분석법에 대해서 설명하고, 제4장에서는 형식개념분석법을 기반으로 하는 RDF데이터의

• First Author: Suk-Hyung Hwang, Corresponding Author: Dong-Heon Cho

\*Suk-Hyung Hwang (shwang@sunmoon.ac.kr), Dept. of Global Software Engineering, SunMoon University

\*\*Dong-Heon Cho (jovision@hanmail.net), College of General Studies, SunMoon University

• Received: 2017. 05. 12, Revised: 2017. 06. 01, Accepted: 2017. 06. 19.

Table 1. Examples of RDF triples

	Subject(Resource)	Predicate(Property)	Object(Value)
$t_1$	cf:Museum-of-korean-embroidery	rdf:type	schema:Museum
$t_2$	cf:Museum-of-korean-embroidery	schema:name	"한국자수박물관"
$t_3$	cf:Museum-of-korean-embroidery	cf:entryFee	"무료"
$t_4$	cf:Museum-of-korean-embroidery	cf:dayOff	"토요일, 일요일, 공휴일"
$t_5$	cf:Museum-of-korean-embroidery	ad:gu	<http://lod.seoul.go.kr/id/ad/11230-Gangnam-gu>
$t_6$	<http://lod.seoul.go.kr/id/ad/11230-Gangnam-gu>	schema:name	"강남구"
$t_7$	<http://lod.seoul.go.kr/id/ad/11230-Gangnam-gu>	ad:population	565710

분석기법에 대하여 제안한다. 제5장에서는 지원도구의 개발과 실험결과를 제시하고, 마지막으로 제6장에서는 결론과 향후연구과제에 대해서 설명하도록 한다.

1. RDF

RDF(Resource Description Framework)는 인터넷과 웹 상의 메타데이터(데이터에 대한 의미, 정의, 설명 등과 같은 의미 정보)를 지원하기 위한 기반구조를 제공하기 위하여 W3C이 제정한 정보자원기술을 위한 국제표준체계이다[12]. RDF는 주어(Subject), 서술어(Predicate), 목적어(Object)와 같은 삼원(트리플, triple)구조를 토대로 문장(<주어 서술어 목적어> 또는 <리소스 속성 속성값>)을 구성함으로써, 정보자원의 의미 및 다른 정보자원과의 의미적 관계를 표현할 수 있는 메타데이터를 기술하여 시멘틱 웹 상의 정보자원을 표현하기 위한 유용한 방법을 제공하고 있다. RDF의 기술대상이 되는 리소스는 URI(Uniform Resource Identifier)에 의해 지정될 수 있는 모든 정보자원이며, 웹상의 문서, 그림, 동영상 뿐만 아니라, 사람, 시설, 리소스들 간의 의미까지도 리소스로서 취급된다. RDF는, 사람이 쉽게 읽고 이해할 수 있을 뿐만 아니라 기계적인 처리, 즉, 응용프로그램들이 웹에 표현된 정보들을 처리하기 용이하여, 검색엔진을 비롯하여 전자상거래, 사물인터넷, 그리고 지능형 소프트웨어 에이전트에 의한 지식공유와 교환 등, 다양한 어플리케이션 영역에서 사용될 수 있다.

RDF트리플(RDF Triple)과 RDF그래프(RDF Graph)는 시멘틱 웹에서 리소스(정보자원)의 RDF표현에 관한 기반이 된다. 즉, URI의 집합U, 공백노드들의 집합B, 그리고 리터럴(literal)들의 집합L이 주어졌을 때, RDF트리플t는,

$$t = (s, p, o) \in (U \cup B) \times U \times (U \cup B \cup L)$$

이다. 이때, s, p, o는 각각 주어(subject), 서술어(predicate), 목적어(object)를 나타낸다. 특히, RDF트리플의 주어, 서술어, 목적어를 각각 리소스(Resource), 속성(Property), 속성값(Property Value)이라고 부르기도 한다. RDF그래프  $G = (V, E) \in \Omega = (U \cup B) \times U \times (U \cup B \cup L)$ 는 RDF트리플들의 집합으로서, 임의의 RDF트리플t의 URI집합, 공백노드들의 집합, 리터럴들의 집합을 각각,  $uri(t), blank(t), literal(t)$ 로 표기한다.

표1에서는 7개의 RDF트리플을 예시하고 있다(표1에서는 서

술시 열린 데이터광장LOD(http://lod.seoul.go.kr)에서 규정한 URI의 축약표현으로서, schema:는 <http://schema.org/>, cf:는 <http://lod.seoul.go.kr/def/cf/>, ad:는 <http://lod.seoul.go.kr/def/ad/>그리고, rdf:는 <http://www.w3.org/1999/02/22-rdf-syntax-ns#>을 나타낸다). 표1의 RDF트리플( $t_1 \sim t_7$ )은, <http://lod.seoul.go.kr/def/cf/Museum-of-korean-embroidery>라는 URI로 참조되고 있는 리소스가 “한국자수박물관”이라는 이름을 갖는 박물관이며, 입장료는 무료이고 토요일, 일요일, 공휴일이 휴관, 소재지는 강남구이고, 강남구의 인구는 565,710명이라는 의미관계를 표현하고 있다.

한편, RDF트리플은 그림1과 같이, 비순환유향그래프(DAG: Directed Acyclic Graph)를 이용하여 표현할 수 있다. 단, DAG의 정점과 변은 각각 리소스와 리소스들 사이의 관계를 나타내며, 타원은 리소스를, 사각형은 리터럴을 각각 나타낸다. RDF그래프의 정점들과 변들은 상호연결되어 RDF트리플의 형태로 RDF문장에 상응하는 정보를 나타낸다.

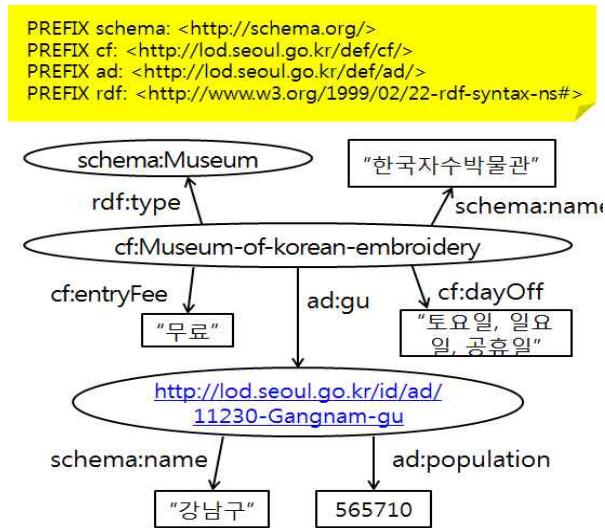


Fig. 1. RDF Graph representation for RDF triples in Table 1

2. SPARQL

SPARQL(SPARQL Protocol and RDF Query Language)은 시멘틱웹을 위한 데이터 프로토콜 및 RDF를 대상으로 하는 W3C표준 질의언어로서, RDF데이터에 대한 질의언어와 원격

지의 RDF데이터에 대한 질의처리에 필요한 프로토콜을 제공하고 있다[3]. SPARQL질의를 처리하기 위해서는, 질의로부터 그래프패턴(BGP : Basic Graph Pattern)을 구축하여, 이 패턴과 일치하는 리소스를 RDF집합으로부터 추출한다. W3C에서 최신판으로 공개한 SPARQL 1.1에서는, 단순한 선택뿐만 아니라, BGP에 적합한 질의결과로부터 새로운 RDF를 생성하는 기능과 집약연산 등이 제공되고 있다.

SPARQL질의에서, 접두어 선언부에서는 여러개의 이름공간(name space)들을 선언할 수도 있다. SPARQL에서 변수선언은 “?”문자로 시작한다. 이러한 변수를 사용하여 추출하는 그래프패턴은 트리플형식으로 WHERE절에 기술하고, 선택조건을 SELECT절에 기술한다. 또한, WHERE선언부에서는 각각의 변수들에 대해서 FILTER를 사용하여 조건을 기술할 수 있다. 조건의 종류로는, 존재(EXISTS, NOT EXIST)와 비교연산자(“=”, “>=” 등)를 이용할 수 있다.

그림2의 질의문(a)에서는, PREFIX에 의해 질의에서 사용될 어휘들(schema, ad, cf, rdf)을 지정하고, URL을 기술하고 있다. 또한, 검색결과로 얻고자하는 값의 변수들(?name, ?gu, ?entryFee, ?dayOff)을 SELECT절에 지정하였다. WHERE절에는 질의패턴이 기술되어있고, 질의결과에 대한 제약사항(10건까지 합치되는 질의결과를 추출)이 지정되었다. 예를들어, 서울 열린 데이터광장(http://lod.seoul.go.kr)에서 제공하는 RDF 데이터에 대하여, 그림2의 질의문(a)를 적용하여 실행하면 그림2(b)와 같은 결과가 도출된다.

```

이름공간 접두어 선언
PREFIX schema: <http://schema.org/>
PREFIX ad: <http://lod.seoul.go.kr/def/ad/>
PREFIX cf: <http://lod.seoul.go.kr/def/cf/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
질의형식 지정
SELECT ?name ?gu ?entryFee ?dayOff
WHERE {
    ?s rdf:type <http://schema.org/Museum>.
    ?s schema:name ?name .
    ?s ad:gu ?t .
    ?s cf:entryFee ?entryFee .
    ?s cf:dayOff ?dayOff .
    ?t schema:name ?gu.
    FILTER (lang(?name)='ko')
}
질의결과 제약사항 지정
LIMIT 10
    
```

(a)

	name	gu	entryFee	dayOff
M1	경기여고 경운박물관	강남구	무료	일요일, 월요일, 공휴일
M2	한국지수박물관	강남구	무료	토요일, 일요일, 공휴일
M3	김치박물관	강남구	유료	월요일, 신정, 구정추석연휴
M4	코리아나 화장박물관	강남구	유료	일요일, 명절
M5	서울대학교 박물관	관악구	무료	일요일, 국정공휴일, 개교기념일
M6	호림박물관	관악구	유료	월요일 1월1일 설날당일
M7	세종대학교 박물관	광진구	무료	토요일, 일요일, 공휴일
M8	건국대학교 박물관	광진구	무료	토요일, 일요일, 공휴일, 방학기간
M9	평강성서유물박물관	구로구	유료	월요일
M10	육군사관학교 박물관	노원구	유료	월요일, 신정, 설, 추석

(b)

Fig. 2. SPARQL query example(a) and result(b)

3. LOD

LOD(Linked Open Data)는 웹 상의 데이터 상호간에 의미적으로 연결시켜 거대한 데이터베이스처럼 활용할 수 있는 기술로서, Tim Berners-Lee에 의해 2006년 최초로 제안되었다

[1]. LOD는, 웹상의 데이터들을 URI로 식별 및 연결하여 RDF 형식으로 표현하고, SPARQL로 질의하고 HTTP로 유통하기 위해 구축된 데이터집합으로서, LOD개념이 등장한 이후 다양한 데이터집합을 RDF형태로 웹에 공개/공유함으로써 시멘틱웹을 구축하고자하는 노력이 지속적으로 이루어지고 있다. 특히, 웹상에 공개된 LOD에 대해서 SPARQL을 실행하여 질의를 처리할 수 있으며, 이와같은 서비스를 SPARQL엔드포인트(SPARQL endpoint)라고 부른다. 참고문헌[2]에 의하면, 2017년 4월 현재 전 세계에는 약 557개의 SPARQL엔드포인트가 존재하고 있으며, 국내의 대표적인 SPARQL Endpoint[13]로는 국립중앙도서관LOD(http://lod.nl.go.kr서울열린데이터광장LOD(http://lod.seoul.go.kr), 한국과학기술정보원LOD(http://lod.ndsl.kr), 특허청 산업재산권LOD(http://lod.kipo.kr), 한국사LOD(http://lod.koreanhistory.or.kr) 등이 있다.

4. FCA

FCA(Formal Concept Analysis, 형식개념분석법)[4]는 개념격자(Concept Lattice)라는 수학적 모델을 기반으로 하는 데이터분석기법의 일종으로서, 개념적인 데이터분석과 지식기반 처리분야의 제반문제들에 대한 수학적 해법을 제공하고 있다. 형식개념분석법에서는, 분석대상데이터를 객체와 속성, 그리고 이들 사이의 포함관계를 이진데이터테이블(Binary Data Table)이라고 부르는 데이터테이블로 나타낸다. 즉, 이진데이터테이블  $K := (G, M, I)$ 는 객체들(Objects)의 집합  $G$ 와 속성들(Attributes)의 집합  $M$ , 그리고  $G$ 와  $M$ 사이의 이항관계  $I \subseteq G \times M$ 으로 구성된다. 이때, 어떤 객체  $g \in G$ 가 속성  $m \in M$ 을 가지고 있을 경우,  $gIm$  또는  $(g, m) \in I$ 로 나타내며, 객체  $g$ 는 속성  $m$ 을 갖는다는 것을 의미한다. 해당 셀에 관련된 객체와 속성이 이항관계  $I$ 를 만족할 경우에는 표2와 같이  $\times$ 표시하고, 이외의 경우에는 빈 공간으로 남겨둔다.

Table 2. Example of Binary Data Table

	a1	a2	a3	a4
o1	×		×	
o2	×	×		
o3		×	×	×
o4		×	×	×

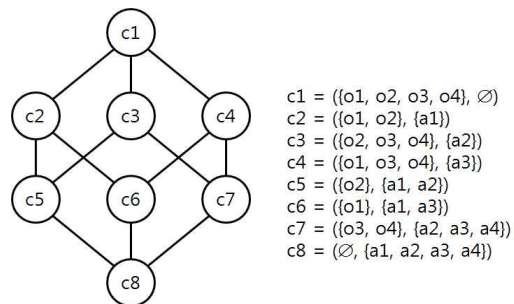


Fig. 3. Concept lattice for Table 2

**Algorithm** *Solution*( $Q, D, G$ )

```

//Input : SPARQL Graph Pattern  $Q$ , RDF Dataset  $D = \{G, \langle u_1, G_1 \rangle, \dots, \langle u_n, G_n \rangle\}$  (where,  $n \geq 0$ ), RDF Graph  $G$ 
//Output : Query Results (Mapping set  $\mu$ )
1: if  $Q = t$  then return  $\{\mu \mid \text{dom}(\mu) = \text{var}(t) \wedge \mu(t) \in G\}$ 
2: if  $Q = (Q_1 \text{ AND } Q_2)$  then
3:   return  $\text{join}(\text{Solution}(Q_1, D, G), \text{Solution}(Q_2, D, G))$ 
4: if  $Q = (Q_1 \text{ UNION } Q_2)$  then
5:   return  $\text{union}(\text{Solution}(Q_1, D, G), \text{Solution}(Q_2, D, G))$ 
6: if  $Q = (Q_1 \text{ OPTIONAL } Q_2)$  then
7:   return  $\text{join}(\text{Solution}(Q_1, D, G), \text{Solution}(Q_2, D, G)) \cup \text{diff}(\text{Solution}(Q_1, D, G), \text{Solution}(Q_2, D, G))$ 
8: if  $Q = (Q_1 \text{ FILTER } R)$  then
9:   return  $\{\mu \in \text{Solution}(Q_1, D, G) \mid \mu(R) = \text{true}\}$ 
10: if  $Q = (\text{GRAPH } u \{Q_1\})$  then
11:   return  $\text{Solution}(Q_1, D, \text{graph}(u)_D)$ 
12: if  $Q = (\text{GRAPH } ?X \{Q_1\})$  then
13:   return  $\bigcup_{u \in \text{names}(D)} \text{join}(\text{Solution}(Q_1, D, \text{graph}(u)_D), \{\mu \mid \text{dom}(\mu) = \{?X\} \wedge \mu(?X) = u\})$ 

```

이진데이터테이블  $K := (G, M, I)$ 의  $O \subseteq G$ 와  $A \subseteq M$ 에 대하여, 2종류의 개념구성연산자,

$$\text{intent}(O) := \{a \in M \mid \forall o \in O: (o, a) \in I\},$$

$$\text{extent}(A) := \{o \in G \mid \forall a \in A: (o, a) \in I\}$$

를 토대로,  $(\text{intent}(O) = A) \wedge (\text{extent}(A) = O)$ 를 만족하는  $(O, A)$ 를 형식개념(formal concept) 또는 개념이라고 한다. 또한, 이진데이터테이블  $K := (G, M, I)$ 로부터 추출한 모든 개념들의 집합을  $B(K) := \{(O, A) \in 2^G \times 2^M \mid \text{intent}(O) = A \wedge \text{extent}(A) = O\}$ 와 같이 정의한다. 즉, 형식개념  $(O, A)$ 는 객체집합  $O$ 의 모든 원소들(객체들)에 공통적인 속성집합이 집합  $A$ 와 같고, 속성집합  $A$ 의 모든 원소들(속성들)을 공통적으로 갖는 객체들의 집합이 집합  $O$ 와 같음을 의미한다. 또한, 임의의 형식개념  $(O_1, A_1)$ ,  $(O_2, A_2)$ 에 대하여,  $O_1 \subseteq O_2$  ( $\Leftrightarrow A_1 \supseteq A_2$ ) 라면,  $(O_1, A_1)$ 은  $(O_2, A_2)$ 의 하위개념, 또는,  $(O_2, A_2)$ 는  $(O_1, A_1)$ 의 상위개념이며,  $(O_1, A_1) \leq (O_2, A_2)$ 로 표현한다.

이진데이터테이블  $K := (G, M, I)$ 로부터 추출된 모든 개념들  $B(K)$ 과 그들 사이의 상하위개념관계  $\leq$ 로 이루어진 계층구조  $L(K) := (B(K), \leq)$ 를 개념격자(Concept Lattice)라고 부른다. 개념격자에서는, 각 개념들과 이들 사이의 상하위관계가 링크에 의해 표시되며, 특히, 개념들 간의 링크에 의해 만들어지는 경로에 의해 상위개념으로부터 하위개념으로 속성들이 상속되며, 하위개념으로부터 상위개념으로 해당 객체들이 전파된다. 표4의 이진데이터테이블로부터 총 8개의 개념들( $c1, \dots, c8$ )이 추출되어 그림3의 개념격자가 구성되었다. 이와 같은 방법을 사용함으로써, 주어진 데이터집합의 객체들과 이들이 갖는 속성들을 기반으로, 외연 및 내포적인 관점에서 개념들을 추출하고 격자구조를 구성함으로써 수월하게 개념계층 및 분류체계를

구축할 수 있다.

## IV. FCA-based approach for analyzing RDF Datasets

본 논문에서 제안하는 형식개념분석법을 기반으로 하는 RDF데이터 분석은 그림4와 같이 3개의 서브모듈들(SPARQL질의처리모듈, 변환처리모듈, 형식개념분석처리모듈)로 구성되며, (1)SPARQL질의 처리, (2)RDF데이터테이블의 구성, (3)데이터변환처리(RDF데이터테이블의 이진화), (4)형식개념분석처리, (5)계층구조화된 결과(Concept Lattice) 제공, 등과 같은 일련의 과정으로 처리된다.

### 1. SPARQL Query Progressing

SPARQL질의는 4종류의 형식(SELECT, ASK, CONSTRUCT, DESCRIBE)이 제공되며, 대부분의 SPARQL질의는 SELECT형식으로 수행되고 ASK, CONSTRUCT, DESCRIBE형식의 질의유형은 SELECT질의를 응용하여 적용할 수 있으므로[14], 본 논문에서는 논의대상을 SELECT형식의 SPARQL질의에 한정시키기로 한다.

SPARQL SELECT질의문은 다음과 같이 정의된다.

$$\text{SELECT } \overrightarrow{\text{Var}} \text{ FROM } u \text{ WHERE } Q$$

단,  $u$ 는 질의대상이 되는 RDF데이터집합의 URI,  $Q$ 는 SPARQL그래프패턴,  $\overrightarrow{\text{Var}}$ 는  $Q$ 에 사용되는 변수들의 튜플(tuple)이다. SPARQL질의문에서 사용되는 변수들은 “?”기호를 변수이름앞에 붙여서 사용한다. 변수들의 집합을  $\text{Var}$ 로 표

기하며(예를들면,  $Var = \{?x, ?y, ?z, \dots\}$ ), 이때  $Var \cap (U \cup B \cup L) = \emptyset$ 이다. 참고문헌[3]을 토대로 SPARQL 그래프패턴을 다음과같이 정의한다.

[정의1](SPARQL graph pattern)

SPARQL그래프패턴  $Q$ 는 다음과같이 재귀적으로 만들어지는 표현식이다.

1. 표현식  $()$ 은 그래프패턴이다(empty graph pattern).
2. 트리플패턴  $t \in (U \cup Var) \times (U \cup Var) \times (U \cup L \cup Var)$ 는 그래프패턴이다.
3.  $Q_1, Q_2$ 가 그래프패턴이라면,  $Q_1 \text{ AND } Q_2, Q_1 \text{ UNION } Q_2, Q_1 \text{ OPTIONAL } Q_2$ 도 그래프패턴이다.
4.  $Q$ 가 그래프패턴이고  $R$ 이 필터조건인 경우,  $Q \text{ FILTER } R$ 는 그래프패턴이다.
5.  $Q$ 가 그래프패턴이고,  $u$ 가 임의의 URI,  $?X$ 가 임의의 변수인 경우,  $(\text{GRAPH } u \{Q\})$ 와  $(\text{GRAPH } ?X \{Q\})$ 는 그래프패턴이다.■

SPARQL질의문에 대한 질의결과(solution)는, 변수들의 집합으로부터 RDF용어집합으로의 매핑  $\mu: Var \rightarrow (U \cup B \cup L)$ , 즉, 변수들에게 RDF용어( $U \cup B \cup L$ )를 할당하는 일종의 부분함수(partial function)를 기반으로 구할 수 있다(이러한 매핑들의 전체집합을  $\Omega$ 으로 표기한다). 또한, 매핑  $\mu$ 의 도메인을  $dom(\mu)$ 로 표기하며,  $dom(\mu) \subseteq Var$ 이다. 2개의 매핑  $\mu_1, \mu_2$ 는, 서로 공통인 모든 변수들  $?x \in dom(\mu_1) \cap dom(\mu_2)$ 에 대하여  $\mu_1(?x) = \mu_2(?x)$ 인 경우에 호환가능하다(compatible)라고 하고,  $\mu_1 \sim \mu_2$ 로 표기한다. 매핑  $\mu_1, \mu_2$ 가 호환가능한 경우(즉,  $\mu_1 \sim \mu_2$ ),  $\mu_1 \cup \mu_2$ 도 또한 매핑이다. 또한, 트리플패턴  $t$ 에 출현하고 있는 모든 변수들을  $vars(t)$ 로 표기하며,  $\mu(t)$ 는 트리플패턴  $t$ 에 출현하고 있는 모든 변수들  $?x \in dom(\mu) \cap vars(t)$ 을  $\mu(?x)$ 로 치환하여 얻어지는 트리플패턴을 나타낸다.

SPARQL질의문에 대한 질의결과는, 질의에 대한 가능한 모든 응답을 내포하고 있는 매핑들의 집합으로 표현된다. 따라서, SPARQL질의에 대한 질의결과는 매핑집합  $\Omega$ 에 적용되는 연산들을 기반으로 정형적으로 정의될 수 있다. 따라서, SPARQL관련 매핑집합에 관한 몇가지 연산들을 정의한다.

[정의2](SPARQL 매핑집합에 관한 연산)

임의의 매핑집합  $\Omega_1, \Omega_2$ 에 대하여, 결합(join), 합집합(union), 차집합(difference)을 다음과 같이 정의한다.

$$(1) join(\Omega_1, \Omega_2) = \{\mu_1 \cup \mu_2 \mid \mu_1 \in \Omega_1 \wedge \mu_2 \in \Omega_2 \wedge \mu_1 \sim \mu_2\}$$

$$(2) union(\Omega_1, \Omega_2) = \{\mu \mid \mu \in \Omega_1 \vee \mu \in \Omega_2\}$$

$$(3) diff(\Omega_1, \Omega_2) = \{\mu_1 \in \Omega_1 \mid \forall \mu_2 \in \Omega_2 : \mu_1 \not\sim \mu_2\} \quad \blacksquare$$

[정의3](RDF Dataset) RDF데이터집합  $D$ 는 다음과 같은 요소들로 구성되는 집합이다.

$$D = \{G, \langle u_1, G_1 \rangle, \dots, \langle u_n, G_n \rangle\},$$

단,  $G$ 는 디폴트로 지정된 RDF그래프, 각  $G_i$ 는 RDF그래프,

$u_i$ 는 해당 RDF그래프의 URI이다.■

이때, 각  $\langle u_i, G_i \rangle$ 를 명명된 그래프(named graph)라고 부르며, 이와 관련하여 2개의 술어 name과 graph를 각각  $name(G_i)_D := u_i$ 와  $graph(u_i)_D := G_i$ 로 정의한다. 또한, RDF데이터집합  $D$ 에 포함된 모든 URI들의 집합을  $names(D) := \{u_1, u_2, \dots, u_n\}$ 과 같이 나타낸다. 특히, 모든 RDF데이터집합에는 반드시 1개의 디폴트 RDF graph  $G$ 가 포함되며, 명명된 그래프는 포함되지 않을 수도 있다. 또한,

$$\forall u_i, u_j \in names(D) [u_i \neq u_j] \text{ 이고,}$$

$$blank(G_i) \cap blank(G_j) = \emptyset \text{ (for } i \neq j \text{)이다.}$$

RDF 데이터 집합  $D = \{G, \langle u_1, G_1 \rangle, \dots, \langle u_n, G_n \rangle\}$  (단,  $n \geq 0$ )의 RDF그래프  $G$ 에 대한 그래프패턴  $Q$ 의 질의결과는 알고리즘  $Solution(Q, D, G)$ 과 같이 재귀적으로 정의된다. 단, 이고,  $Q_1, Q_2$ 는  $Q$ 를 구성하는 그래프패턴,  $R$ 은 필터조건,  $t$ 는 임의의 트리플패턴이다.

서울 열린 데이터 광장LOD에서 제공하는 문화시설 SPARQL엔드포인트(<http://lod.seoul.go.kr/endpoint/cf/>)에 대하여, 그림2(a)의 SPARQL질의를 수행하면 그림2(b)와 같이 서울시에 있는 10건의 박물관의 정보(박물관이름, 소재지, 입장료(유료 또는 무료), 휴무일)를 질의결과로 얻을 수 있다.

## 2. Construction of RDF Data Table

SPARQL질의결과로부터 그림2(b)와 같이 여러 가지 다양한 속성값을 갖는 데이터테이블을 구성할 수 있다. 다양한 값을 갖는 속성을 다치속성(many-valued attribute)이라고 부르고, 다치속성을 갖는 데이터들로 구성된 데이터테이블(다치데이터테이블, Many-valued Data Table)은 다음과 같이 정의한다.

[정의 4] 다치데이터테이블  $MDT = (G, M, W, I)$ 는 객체들의 집합  $G$ 와 다치속성들의 집합  $M$ , 속성값의 집합  $W$ , 그리고  $G$ 와  $M, W$ 사이의 관계  $I \subseteq G \times M \times W$ 로 구성된다.■

즉,  $G, M, W$ 의 각 원소들은 각각 해당 다치데이터테이블의 객체들과 각 객체들이 가질 수 있는 다치속성들, 그리고 그 속성의 값들을 나타낸다. 또한, 어떤 객체  $g \in G$ 가 속성  $m \in M$ 을 가지고 있고 그 속성의 값이  $w \in W$ 인 경우,  $(g, m, w) \in I$  또는  $m(g) = w$ 로 나타내며, 객체  $g$ 는  $w$ 값을 가지는 속성  $m$ 을 갖는다는 것을 의미한다. 그림2(b)의 질의결과에 대하여, 속성  $name$ 의 값을 객체로 지정하면, 다음과 같은 요소들을 갖는 다치데이터테이블  $MDT = (G, M, W, I)$ 로 구성된다. 즉,  $G = \{\text{경기여고경운박물관, 한국자수박물관, \dots, 육군사관학교박물관}\}$ ,  $M = \{gu, entryFee, dayOff\}$ 이며, 속성값의 집합  $W$ 와 삼항관계  $I \subseteq G \times M \times W$ 는 표3과 같다.

## 3. Binarization of Many-Valued Data Table

데이터변환처리모듈에서는, 다치데이터테이블 형태로 수집된 SPARQL질의결과를 이진데이터테이블로 변환하는 이진화처리가 수행된다. 즉, 다치데이터테이블에 대하여 개념분석을

수행하여 개념들을 추출하고 개념격자구조를 구성하여 가시화 하기 위해서는 특정한 규칙에 따라 다치데이터를 이진데이터로 변환할 필요가 있으며, 이와 같이 변환과정을 이진화처리 (Binarization)라고 한다. 즉, 이진화처리에서는, 다치데이터 테이블의 각 다치속성들을 이진화규칙표(Binarization Rule Table)를 토대로 해석하여 이진속성(Binary-Valued Attribute)로 변환해서 이진데이터테이블(Binary Data Table)이 구성된다. 이진화처리에 관한 제반정의들은 다음과 같다.

[정의 5] 다치데이터테이블  $MDT = (G, M, W, I)$ 의 임의의 속성  $m \in M$ 에 대한 이진화규칙표는  $B_m = (G_m, M_m, I_m)$ 으로 정의된다. 단,  $m(G) \subseteq G_m$ 이고,  $M_m$ 은 이진화속성들의 집합이며  $I_m \subseteq G_m \times M_m$ 이다. 이진화규칙표의 객체들은 이진화값 (binarization values), 속성들은 이진화속성(binanzation attributes)이라고 부른다.■

본 논문에서는 개념분석의 목적 및 사용자의 관점, 속성의 특성 등에 따라서, 표4~6과 같은 이진화규칙표들 ( $\Delta_{MDT} = \{B_{gu}, B_{entryFee}, B_{dayOff}\}$ )을 이진화처리에 적용한다.

주어진 다치데이터테이블  $MDT = (G, M, W, I)$ 에 대하여, 각 속성  $m_i \in M$ (단,  $1 \leq i \leq |M|$ )의 특성에 맞게 작성된 이진화규칙표들을 토대로 이진화변환처리를 수행하기 위한 알고리즘 Binarization( $MDT, \Delta_{MDT}$ )을 정의한다.

```

Algorithm Binarization( $MDT, \Delta_{MDT}$ )
// Input : Many-valued Data Table  $MDT = (G, M, W, I)$ ,
// Binarization rule set  $\Delta_{MDT}$ 
// Output : Binary Data Table  $K = (G_K, N, J)$ 
1:  $G_K = G$ ;
2:  $N = \emptyset$ ;
3:  $J = \emptyset$ ;
4: for each  $m_i \in M (1 \leq i \leq |M|)$ 
5:   for each  $g \in G$  do
6:     if  $\exists w \in W, n \in M_{m_i} [(g, m_i, w) \in I \wedge (w, n) \in I_{m_i}]$ 
7:       then
8:          $N = N \cup \{(m_i, n)\}$ ;
9:          $J = J \cup \{(g, (m_i, n))\}$ ;
10:      end if
11:   end for

```

표4~6의 이진화규칙표들( $\Delta_{MDT} = \{B_{gu}, B_{entryFee}, B_{dayOff}\}$ )을 토대로 본 논문에서 제안하고 있는 이진화변환처리 알고리즘 Binarization( $MDT, \Delta_{MDT}$ )을 표3의 다치데이터테이블에 적용하면, 정보손실없이 표8의 이진데이터테이블로 변환된다.

#### 4. Concept Analysis

주어진 이진데이터테이블로부터 개념을 추출하고 개념격자를 구축하는 형식개념분석처리 알고리즘(ConceptAnalysis)은 다음과 같다.

Table 3. Example of Many-Valued Data Table

	gu	entryFee	dayOff
경기여고 경운박물관	강남구	무료	일요일, 월요일, 공휴일
한국자수박물관	강남구	무료	토요일, 일요일, 공휴일
김치박물관	강남구	유료	월요일, 신정, 구정추석연휴
코리아나 화장박물관	강남구	유료	일요일, 명절
서울대학교 박물관	관악구	무료	일요일, 국정공휴일, 개교기념일
호림박물관	관악구	유료	월요일 1월1일 설날당일
세종대학교 박물관	광진구	무료	토요일, 일요일, 공휴일
건국대학교 박물관	광진구	무료	토요일, 일요일, 공휴일, 방학기간
평강성서유물 박물관	구로구	유료	월요일
육군사관학교 박물관	노원구	유료	월요일, 신정, 설, 추석

Table 4. Binarization Rule Table  $B_{gu}$

	강남구	관악구	광진구	구로구	노원구
강남구	×				
관악구		×			
광진구			×		
구로구				×	
노원구					×

Table 5. Binarization Rule Table  $B_{entryFee}$

	무료	유료
무료	×	
유료		×

Table 6. Binarization Rule Table  $B_{dayOff}$

	토요일	일요일	월요일	공휴일	개교 기념일	방학 기간
토요일	×					
일요일		×				
월요일			×			
공휴일				×		
개교기념일					×	
방학기간						×
신정				×		
구정, 설				×		
추석				×		
명절				×		
국정공휴일				×		
1월1일				×		



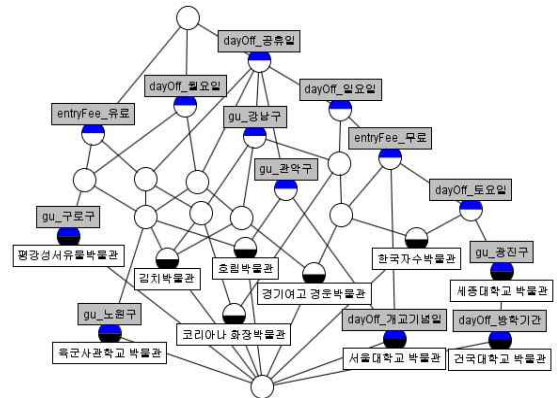
```

Algorithm ConceptAnalysis( $K$ )
    // Input : Binary Data Table  $K := (G, M, I)$ 
    // Output : Concept Lattice  $L(K) := (B(K), E_{\leq})$ 
1   $B(K) \leftarrow \emptyset; E_{\leq} \leftarrow \emptyset;$ 
2  for each  $g \in G$  do
3     $B(K) \leftarrow B(K) \cup (extent(intent(g)), intent(g));$ 
4  end for
5  for each  $(O, A) \in B(K)$  do
6    for each  $g \in (G - extent(A))$  do
7       $X \leftarrow extent(A) \cup \{g\};$ 
8      if  $(extent(intent(X), intent(X)) \notin B(K))$  then
9         $B(K) \leftarrow B(K) \cup (extent(intent(X)), intent(X));$ 
10     end if
11   end for
12 end for
13 for each  $c1 \in B(K)$  do
14   for each  $c2 \in (B(K) - \{c1\})$  do
15     if  $(c1 \leq c2) \wedge$ 
16        $(\nexists c3 \in (B(K) - \{c1, c2\}) [(c1 \leq c3) \wedge (c3 \leq c2)])$ 
17       then
18          $E_{\leq} \leftarrow E_{\leq} \cup \{(c1, c2)\};$ 
19       end if
20   end for
21 end for
22 return  $L(K) := (B(K), E_{\leq}).$ 
    
```

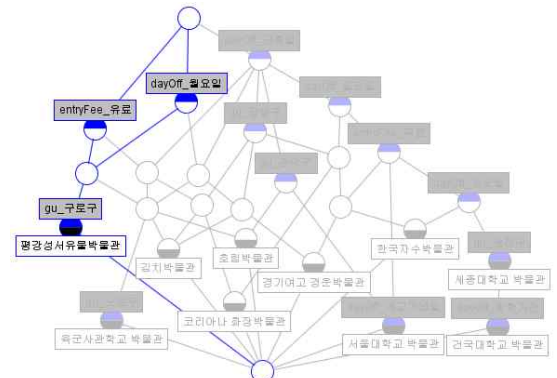
형식개념분석처리 알고리즘ConceptAnalysis( $K$ )에서는, 개념구성연산자( $intent, extent$ )를 토대로 주어진 이진데이터테이블 $K := (G, M, I)$ 로부터 개념들( $B(K)$ )을 추출하고(1~12행), 개념들사이의 상하위개념관계( $\leq$ )를 파악하여(13~19행), 최종적으로 개념격자 $L(K) := (B(K), E_{\leq})$ 를 구성한다(20행).

### 5. Visualization of Concept Hierarchies

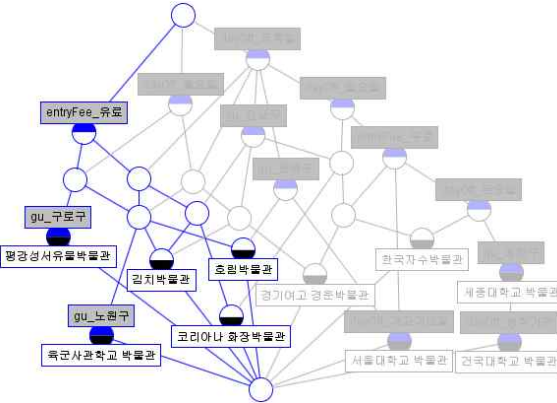
형식개념분석처리 알고리즘ConceptAnalysis( $K$ )에 의해서 구성된 개념격자 $L(K) := (B(K), E_{\leq})$ 는, 그림5(a)와 같이 28개의 개념들을 정점으로 하고, 각 개념들사이의 상하위개념관계를 변으로 하는 Line Diagram으로 나타낼 수 있다. Line Diagram의 각 정점(개념)에는 해당 개념을 구성하는 객체들과 속성들이 레이블로 부착되어 있다. 예를들면, 그림5(a)의 가장 왼쪽의 정점은 ( $\{평강성서유물박물관\}, \{구로구, 유료, 월요일\}$ )로 구성되는 개념으로서, 그림5(b)와 같이, 상위개념들로부터 속성들(“유료”, “월요일”)을 상속받아서, “구로구에 위치하고 있으며, 입장료는 유료이고, 월요일에 휴관하는 평강성서유물박물관”을 나타내고 있다. 또한, 형식개념분석처리에 의해서 최종적으로 구축되는 개념격자에는, 주어진 데이터들의 공통속성을 기반으로 하는 분류체계를 포함하고 있다. 예를들면, 그림5의 (c)와 (d)는 각각 유료 및 무료 박물관들의 분류체계를 나타내고 있다.



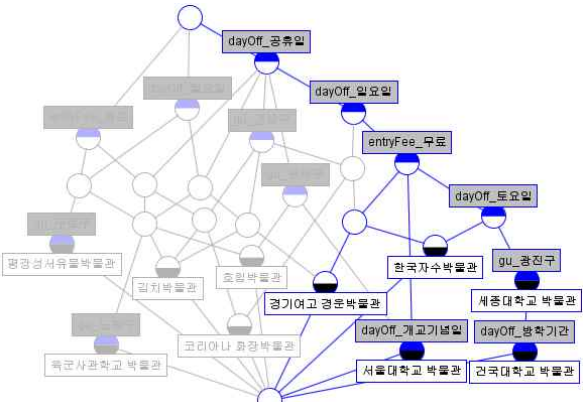
(a) Concept Lattice for Table 8



(b) Inheritance Hierarchy of Concept Lattice



(c) Concept Hierarchy for “유료박물관”



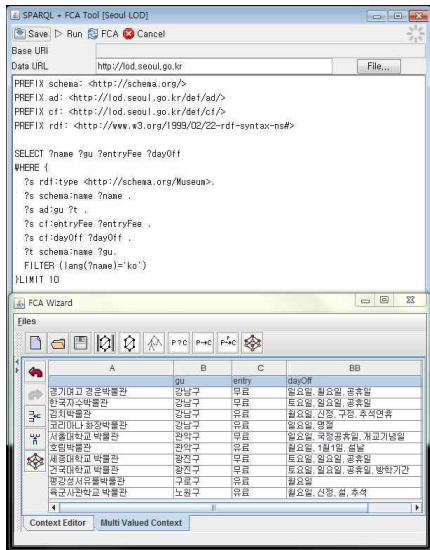
(d) Concept Hierarchy for “무료박물관”

Fig. 5. An Example of Concept lattice

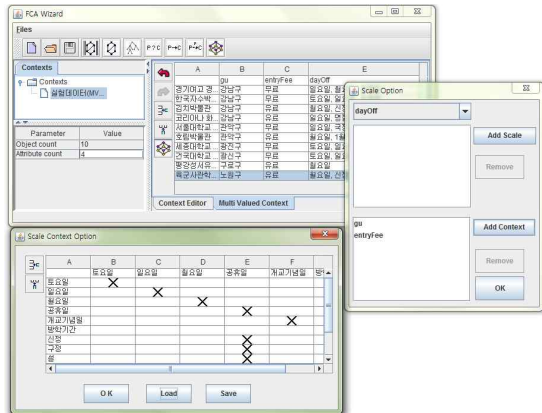
## V. FCA-based RDF Data Analysis Tool and Experiment

### 1. FCA-based RDF Data Analysis Tool

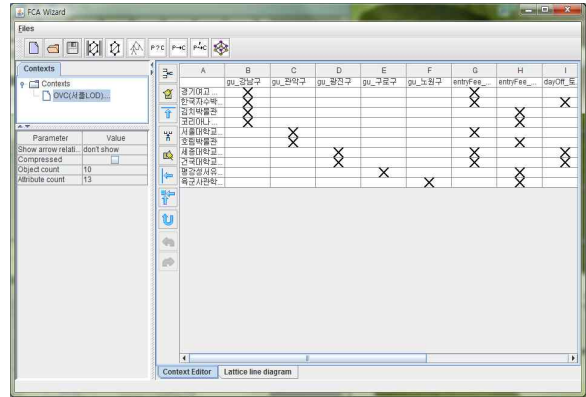
앞서 제4장에서 논의된 내용을 토대로, 본 논문에서는 SPARQL질의처리 및 개념분석처리를 위한 자동화도구를 개발하였다. 개발된 자동화도구의 처리과정은 그림6의 (a)~(e)와 같다. 먼저, SPARQL질의처리를 위하여, 질의대상이 되는 LOD(서울열린데이터광장LOD)의 URI를 지정하고 SPARQL질의문을 입력(그림6(a)의 왼쪽)하여 실행하면, 질의결과가 다치 데이터테이블 형태로 변환된다(그림6(b)의 오른쪽). 이어서, 이진화규칙표를 토대로 이진화처리를 수행하면(그림6(b)), 그림6(c)와 같은 이진데이터테이블 형태로 변환된다. 이후에는, 개념추출과 개념격자구성을 실시하여(그림6(d)), 최종적으로 다양한 형태의 개념분류체계를 구축할 수 있다(그림6(e)).



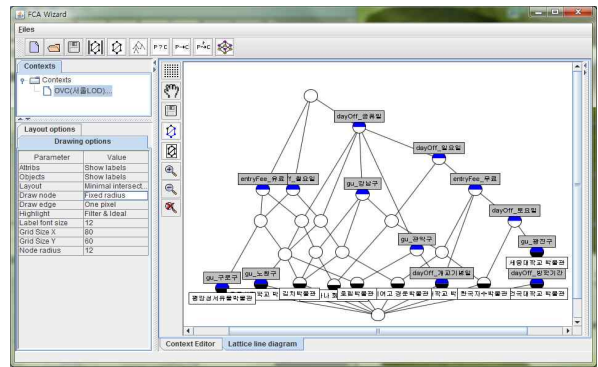
(a) SPARQL query processing and RDF data table construction



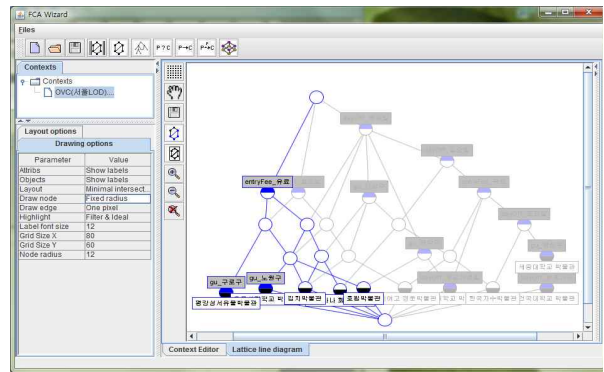
(b) A screen capture of Binarization



(c) Result of Binarization



(d) Construction of Concept Lattice



(e) A display of Concept Hierarchy

Fig. 6. Screenshots of Tool for SPARQL query and Concept analysis

### 2. Experiment

본 연구에서 제시하는 RDF데이터 분석 및 가시화기법의 검증을 위해서 서울 열린데이터광장 LOD(<http://lod.seoul.go.kr>)에서 제공하는 다양한 RDF데이터들 중에서 임의로 “서울시 문화시설현황”을 대상으로 실험을 진행하였다. 먼저, RDF데이터 수집을 위하여, 본 연구에서 개발한 SPARQL질의처리 모듈을 사용하여 그림7과 같이 서울시에 있는 박물관의 이름과 소재지(구이름), 입장료, 휴관일에 대한 SPARQL질의문을 처리하여, 표7과 같이 총 76건의 구체적인 실험대상 RDF 데이터집합이 수집되었다. 표7의 데이터집합을 토대로 RDF데이터테이블을



구성하고 표4~6의 이진화규칙표를 토대로 이진화처리를 수행하여 표8과 같은 이진데이터테이블로 변환하였다.

```

PREFIX schema: <http://schema.org/>
PREFIX ad: <http://lod.seoul.go.kr/def/ad/>
PREFIX cf: <http://lod.seoul.go.kr/def/cf/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?name ?gu ?entryFee ?dayOff
WHERE {
  ?s rdf:type <http://schema.org/Museum>.
  ?s schema:name ?name .
  ?s ad:gu ?t .
  ?s cf:entryFee ?entryFee .
  ?s cf:dayOff ?dayOff .
  ?t schema:name ?gu.
  FILTER (lang(?name)='ko')
}
    
```

Fig. 7. SPARQL Query

Table 7. SPARQL Query results

	name	gu	entryFee	dayOff
1	경기여고 경운박물관	강남구	무료	일요일 월요일 공휴일
2	한국자수 박물관	강남구	무료	토요일 일요일 공휴일
3	김치 박물관	강남구	유료	월요일 신정 구정 추석연휴
4	코리아나 화장박물관	강남구	유료	일요일 명절
5	서울대학교 박물관	관악구	무료	일요일 국경공휴일 개교기념일
6	호림박물관	관악구	유료	월요일 1월1일 설날 당일
7	세종대학교 박물관	광진구	무료	토요일 일요일 공휴일
8	건국대학교 박물관	광진구	무료	토요일 일요일 공휴일 방학기간
9	평강성서유물 박물관	구로구	유료	월요일
10	육군사관학교 박물관	노원구	유료	월요일 신정 설 추석
11	응기민속 박물관	도봉구	유료	월요일 1월1일 설날 추석연휴
	⋮	⋮	⋮	⋮
72	지구촌 민속박물관	중구	무료	월요일 명절
73	한국영화 박물관	마포구	무료	월요일 신정 설 추석
74	허준박물관	강서구	유료	월요일 1월1일 설날 추석
75	서울역사 박물관	종로구	유료	1월1일 월요일 휴관일
76	시몬느 핸드백 박물관	강남구	유료	월요일

이어서, 본 연구에서 개발된 형식개념분석처리 모듈을 사용하여 표8의 이진데이터로부터 개념들을 추출하고 개념계층구조를 가시화하는 실험을 수행하였다. 실험대상이 되는 총76건의 RDF데이터로부터 추출되는 개념들의 개수와 개념계층구조를 비교분석하기 위하여, 점증적으로 객체의 개수를 증가시켜서 실험을 진행하였고, 실험결과물(개념계층구조)를 그림8에 제시하였다.

본 실험으로부터 얻어진 각 개념계층구조의 구성요소들(변의 개수와 계층구조의 높이)에 관한 변화추이는 표9 및 그림9와 같다. 즉, 객체의 개수가 증가하면서 추출되는 개념들의 개수와 개념계층구조 상의 변들의 개수가 비례하여 증가하지만, 개념계층구조의 깊이(Lattice Height)는 객체의 개수가 15개인 경우부터는 7로 고정되어 변화하지 않음을 알 수 있다.

이상의 실험으로부터 본 연구에서 제안하고 있는 RDF데이터 분석도구에서 제공하는 제반기능들(SPARQL질의처리, 이진화변환처리, 형식개념분석 및 가시화처리)을 사용하여 LOD에서 공개하고 있는 RDF데이터를 수집하여 형식개념분석법을 기반으로 하는 분석을 수행함으로써 다양한 개념분석계층구조를 구축하고 유의미한 정보를 획득할 수 있음을 확인할 수 있다.

#### IV. Conclusions

최근 웹 분야에서는 컴퓨터에 의한 처리에 적합한 다종다양한 링크드데이터(LOD : Linked Open Data)가 RDF를 기반으로 작성되어 LOD클라우드 형태로 공개되고 있다. 그러나, 대부분의 LOD클라우드에서 제공하고 있는 SPARQL엔드포인트 서비스에서는 데이터를 수집/분석/활용하기 위해서 SPARQL를 이용하여 질의결과로서 구조화된 데이터들을 간단하게 추출할 수는 있으나, 추출된 다량의 질의결과에 대한 사용자중심의 분석 및 가시화방법이 제공되지 않고 있다.

따라서, 본 논문에서는 LOD클라우드로부터 유용한 정보를 추출하고 가시화하기 위하여, 형식개념분석법을 기반으로 하는 RDF데이터 분석 및 가시화기법을 제안하고 지원도구를 개발하여 실험을 통해서 그 유용성을 확인하였다. 본 논문에서 제안하는 RDF데이터 분석 및 가시화기법 및 지원도구는 LOD에 내재되어 있는 정보와 지식을 추출하고 분석하여 분류체계화 및 가시화를 지원함으로써, 시멘틱웹 데이터마이닝 분야에서 유용하게 활용할 수 있다. 최근에는 사물인터넷의 발전에 따라서 새롭게 공개되는 LOD클라우드가 등장함에 따라서, 향후에는 본 연구결과의 데이터분석기법을 사물인터넷과 관련된 LOD영역으로 확장시킨 후속연구가 진행될 예정이다.

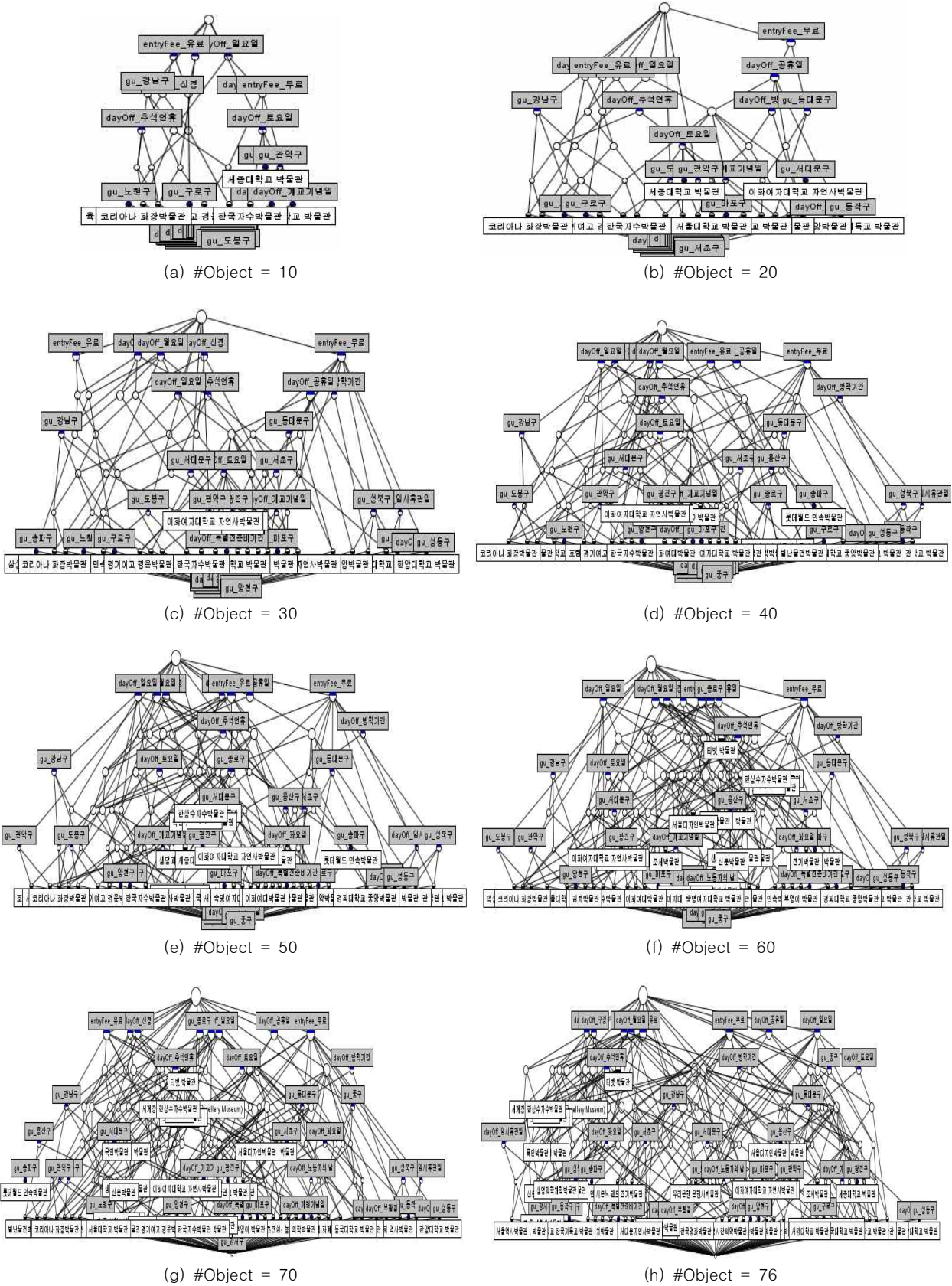


Fig.8. Experiment results(Concept Hierarchies)

Table 8. Experiment Results(Binary Data Table)

	gu_강남구	gu_관악구	gu_광진구	gu_구로구	gu_노원구	gu_도봉구	gu_마포구	gu_중구	gu_영등포구	gu_강서구	ent_yF_e무	ent_yF_e유	da_Off_토일	day_Off_일요일	da_Off_월요일	da_Off_화요일	da_Off_수요일	da_Off_목요일	da_Off_금요일	da_Off_토요일	da_Off_일요일	...
경기여고경운박물관	X									X			X	X								...
한국자수박물관	X								X		X	X										...
김치박물관	X										X			X				X	X	X		...
코리아나화장박물관	X										X		X					X	X	X		...
서울대학교박물관		X					...	...			X		X			...	X	X				...
호림박물관		X									X			X				X	X			...
세종대학교박물관			X								X	X	X					X				...
건국대학교박물관			X								X		X	X				X				...
평강성서유물박물관				X							X			X								...
육군사관학교 박물관					X						X			X				X	X	X		...
응기민속박물관						X					X			X				X	X	X		...
...							...	...														...
지구촌민속박물관									X		X			X				X	X	X		...
한국영화박물관							X			X				X				X	X	X		...
허준박물관									X		X			X				X	X	X		...
서울역사박물관								X			X			X				X			X	...
시몬느핸드백 박물관	X										X			X								...

Table 9. Experiment results(number of Objects, Concepts, Edges and Lattice Height)

실험	#Object	#Concept	#Edge	Lattice Height
1	1	1	1	1
2	2	4	4	2
3	3	7	9	3
4	4	10	15	4
5	5	14	22	5
6	6	19	34	5
7	7	21	38	5
8	8	22	39	6
9	9	25	44	6
10	10	28	49	6
11	11	29	51	6
12	12	31	56	6
13	13	35	65	6
14	14	37	69	6
15	15	41	77	7
⋮	⋮	⋮	⋮	⋮
72	72	157	379	7
73	73	159	384	7
74	74	160	386	7
75	75	166	405	7
76	76	168	410	7

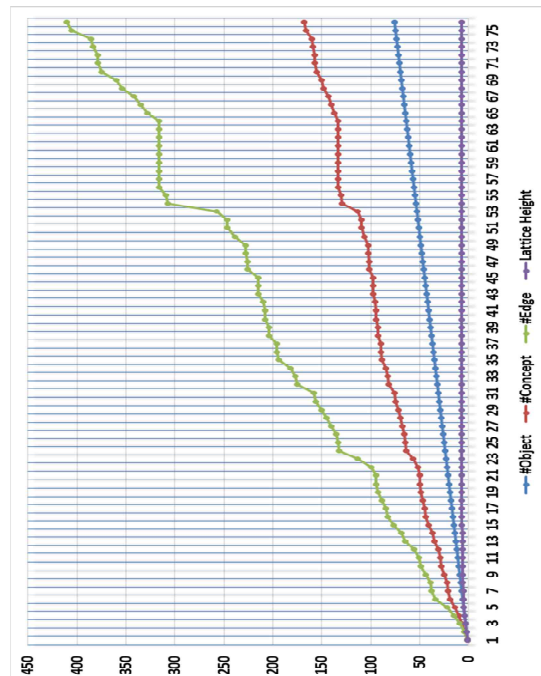


Fig. 9. Experiment results(number of Objects, Concepts, Edges and Lattice Height)

## REFERENCES

- [1] C. Bizer, T. Heath, T. Berners-Lee, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems*, Vol. 5, No. 3, pp. 1-22, March 2009.
- [2] "2014 Case Study of Korean Linked Open Data," National Information Society Agency, 2014.
- [3] SPARQL 1.1 Query Language, <http://www.w3.org/TR/sparql11-query/>
- [4] B. Ganter, R. Wille, "Formal Concept Analysis: Mathematical Foundations," Springer-Verlag, 1999.
- [5] P. Teufl, and L. and Günther Lackner. "Knowledge Extraction from RDF Data with Activation Patterns," *Journal of Universal Computer Science*, Vol. 17, No. 7, pp. 983-1004, January 2011.
- [6] J. Du, H. Wang, Y. Ni, Y. Yu, "HadoopRDF : A Scalable Semantic Data Analytical Engine," *Intelligent Computing Theories and Applications, Lecture Notes in Computer Science* Vol. 7390, pp. 633-641, July 2012.
- [7] D. Colazzo, F. Goasdoué, I. Manolescu, A. Roatis, "RDF analytics: lenses over semantic graphs," *Proceedings of the 23rd international conference on World Wide Web Conferences*, pp. 467-478, 2014.
- [8] E. Ruckhaus, O. Baldizán, M. Vidal, "Analyzing Linked Data Quality with LiQuate," *Proceedings of the 12th International OnTheMove (OTM 2013) Conferences*, pp. 629-638, 2013.
- [9] M. d'Aquin, E. Motta, "Extracting relevant questions to an RDF dataset using formal concept analysis," *Proceedings of the 6th international conference on Knowledge capture(K-CAP '11)*, pp. 121-128, 2011.
- [10] S. Sundara, M. Atre, V. Kolovski, S. Das, Z. Wu, E. Chong, J. Srinivasan, "Visualizing large-scale RDF data using Subsets, Summaries, and Sampling in Oracle," *Proceedings of the 26th international Conference on Data Eng.*, pp. 1048-1059, 2010.
- [11] P. Bellini, P. Nesi, A. Venturi, "Linked open graph: Browsing multiple SPARQL entry points to build your own LOD views," *Journal of Visual Languages & Computing*, Vol. 25, No. 6, Author Pierfrancesco Bellini, Year 2014, Vol 25, Issue 6. pp.703-716, December 2014.
- [12] Resource Description Framework, <https://www.w3.org/RDF/>
- [13] SPARQL Endpoints Status, <http://sparqls.okfn.org/>
- [14] M. Arias, J. Fernández, M. Martínez-Prieto, P. Fuente, "An empirical study of real-world sparql queries," *Proceedings of the 20th international conference companion on World Wide Web*, pp.305-306, 2011.

## Authors



Suk-Hyung Hwang received the B.S degree in Computer Science from Kangwon National University in 1991, the M.E. and Ph.D. degrees in Information and Computer Science from Osaka University, Japan, in 1994 and 1997, respectively. Dr. Hwang joined the faculty of the Department of Computer Science and Engineering at SunMoon University, Korea, in 1997. He is currently a Professor in the Department of Global Software Engineering, SunMoon University. His research interests include Object-Oriented Software Engineering, Data Science, Semantic Web, etc.



Dong Heon Cho received the B.S., M.S. in Electricity Engineering from ChungNam National University in 1992 and 1995 respectively and the Ph.D. degrees in Computer Science and Engineering from HoSeo University in 2005. He is a professor at College of General Studies, SunMoon University, Korea. His current research interests are medical engineering, Educational Evaluation and Education Curriculum.