

# RSS와 OLAP 큐브를 이용한 FOAF의 동적 관리 기법

손종수<sup>a</sup>, 정인정<sup>a</sup>

<sup>a</sup> 고려대학교 컴퓨터 정보학과

208, Seochangri, Yeongigun, Chungnam, 339-700, Korea

Tel: +82-41-860-1342, Fax: +82-41-864-0014, E-mail: {mis026, chung}@korea.ac.kr

## Abstract

웹2.0 기술이 소개된 이후 사회연결망 서비스는 미래 정보기술의 기초로서 중요하게 인식되고 있다. 이에, 웹2.0 환경에서 사회연결망을 구축하기 위하여 온톨로지 기반의 사용자 프로파일 기술언어인 FOAF를 활용하기 위한 다양한 연구가 이뤄지고 있다. 그러나 FOAF를 이용하여 사회연결망을 추출하는 대부분의 방법은 시간의 흐름에 따라 변화하는 사용자의 사회연결망을 자동적으로 반영하지 못하는 단점이 있다.

따라서 본 논문에서는 기존 FOAF를 이용한 사회연결망 추출방법의 한계를 극복하기 위하여 사용자 프로파일 기술 언어인 FOAF와 웹 저작물 출판 매커니즘인 RSS를 OLAP 시스템에 적용시켜 동적으로 FOAF를 갱신하고 관리하기 위한 방법을 제안한다. 본 논문에서 제안하는 방법은 수집한 FOAF와 RSS 파일들을 스타스키마로 설계된 데이터베이스에 넣어 OLAP 큐브를 생성한다. 그리고 OLAP 연산을 이용하여 사용자의 연결관계를 분석하고 FOAF에 그 결과를 반영한다. 본 논문에서 사용한 방법을 활용하면 시간의 흐름에 따른 동적인 사회 연결망도 추출이 가능하다.

## Keywords:

Ontology, Semantic web, Data mining, OLAP, FOAF, RSS

## 서론

시맨틱 웹은 컴퓨터가 웹 상의 다양한 정보자원들을 의미적으로 이해하고 처리할 수 있도록 고안된 웹을 의미한다. 시맨틱 웹은 2001년에 팀 버너스 리에 의해 제안이[1] 되었으며 시맨틱 웹이 제안된 이후 시맨틱 웹을 이용한 다양한 연구 및 응용이 제시되었다. 그러나 시맨틱 웹은 정보자원을 메타데이터로 기술하는 형식이 까다롭고 많은 양의 온톨로지를 생성해야한다는 점이 단점으로 제시되고 있다[2]. FOAF(Friend-Of-A-Friend)는 시맨틱 웹의 빠른 보급을 위하여 어휘를 단순화 시킨 OWL/XML 기반의 온톨로지 언어이다. FOAF는 사용자의

프로필(Profile)을 기술하며 주된 내용은 친구, 관심사, 이메일 주소, 홈페이지 주소 등이다[3]. 최근에 들어 FOAF는 라이브저널 (livejournal.com), 라이브도어(livedoor.com) 등을 포함한 많은 블로그(Blog)와 사회연결망서비스(Social Network Service)에서 받아들여지고 있다. 그리고 이를 통해 시맨틱 웹 응용이 좀 더 빠르게 접목될 수 있는 환경이 갖춰지고 있다[4].

이에 따라 FOAF를 이용하여 사회연결망을 추출하고 서비스에 활용하기 위한 많은 연구가 이루어지고 있다. 그러나 대부분의 연구에서는 FOAF의 활용방법에 대하여 주요한 초점을 맞추고 있어 이기종 분산환경하에서 사회연결망의 동적인 관리를 위한 연구는 상대적으로 미흡하다. 또한 FOAF는 사용자의 프로파일을 잘 표현하지만 시간의 흐름에 따라 변화한 사용자의 사회연결망을 자동적으로 반영하기 위한 연구가 부족하다. [5]에서는 현재 사회연결망 서비스의 단점을 다음과 같이 지적하였다.

- 통상적인 객체에 대한 표현이 부족하다. 어떤 사용자와 사용자가 연결된다고 했을 때 한 사용자의 관심사, 출신학교, 성별, 언어 등은 전혀 고려되지 않고 단순한 연결관계만을 표현한다.
- 사회연결망의 표현에 있어 주요한 연결관계 한가지만을 이용하여 사용자와 사용자간의 다양한 의미적 관계를 표현하지 못한다.
- 한번 생성된 사회연결망은 시간의 흐름에 따른 연결 관계의 변화를 반영하기 위하여 많은 연산을 필요로 한다.
- 이기종 분산 처리환경에서 효과적으로 대량의 메타데이터를 통합하고 처리할 수 있는 방법이 부족하다.

FOAF는 사용자의 프로필과 다른 사용자와의 연결관계를 잘 표현할 수 있도록 도구를 제공하고 있다. 그러나 FOAF는 사용자의 콘텐츠 자체는 포함하고 있지 않기 때문에 FOAF만을 사용하면

사용자의 관심사 변화에 따른 사회연결망의 동적인 반영이 어렵다. 따라서 본 논문에서는 FOAF의 동적인 관리를 위하여 RSS (RDF Site Summary)[9]를 함께 사용한다. RSS는 인터넷 저작물 출판 매커니즘으로써 웹 사이트와 콘텐츠를 RDF/XML 형식으로 배포할 수 있도록 하는 표준 어휘를 제공한다. RSS를 이용하여 서로 다른 서버에 위치한 웹 사이트나 응용에 콘텐츠를 배포할 수 있다[6].

본 논문에서는 동적인 사용자의 연결관계의 관리를 위하여 FOAF와 RSS 파일들을 수집하고 이를 OLAP 시스템에 넣어 처리한다. OLAP 시스템은 다차원 데이터의 분석 및 처리에 뛰어난 성능을 발휘한다. 다차원 데이터를 분석하고 처리하기 위하여 관계형 데이터베이스의 스키마를 스타스키마로 설계하는데 이를 통해서 대량의 데이터를 대상으로 빠르고 간단하게 분석할 수 있다.

본 논문의 목적은 FOAF와 RSS 파일들을 OLAP 시스템에 접목하여 동적으로 사용자의 연결관계를 관리하고 FOAF에 이를 반영하는 것이다. 이를 위하여 FOAF와 RSS의 각 요소(Entity)를 스타스키마로 설계된 데이터베이스에 삽입하고 이를 OLAP 데이터 큐브(Cube)로 변환하여 사회연결망을 분석한다. 본 논문에서 제시하는 방법은 XML/RDF 기반의 FOAF와 RSS를 사용함으로써 이기종 분산 컴퓨팅 환경에서 언어와 기종에 상관없이 자료를 수집 및 처리할 수 있으며 OLAP 시스템을 접목하여 다차원 사회연결망 분석 및 추출이 가능하다. 그리고 RSS를 사용하여 시간의 흐름에 따른 동적인 사회연결망 구축이 가능하다.

## 연구의 배경

### FOAF

다양한 사회 연결망 서비스의 사용자들을 통합하고 이들간의 관계를 설정하기 위하여 표준적인 방식의 사용자 프로필 기술 방법이 필요하다[7]. 이에 따라 FOAF 프로젝트는 시맨틱 웹 기술을 활용하여 사용자의 프로파일 및 사회 연결관계를 표현하는 방법을 제시하였다. FOAF를 통해 사회연결망 데이터를 공유하기 위한 잠재적 해결책을 도출할 수 있다.

FOAF는 사용자들과 그들의 사회적 연결에 관한 정보를 표현하기 위한 프레임워크이다. FOAF 어휘는 OWL을 기반으로 쓰여졌으며 개인 정보, 그룹에서의 멤버쉽과 사회적 연결을 포함한다. 그림 1[8]은 FOAF의 클래스 집합을 보이고 있다.



그림 1 - FOAF classes

그림 1에서 보이는 바와 같이 FOAF의 어휘는 다양하지만 모든 클래스가 사용되고 있지는 않다. FOAF 어휘 중에 주로 사용되는 사용자 프로필의 내용은 사용자명, 이메일, 홈페이지, 학교홈페이지, 직장 홈페이지, 전화번호, 별명, 호칭, 아는 사람 등이다. 이 중 아는 사람을 표현할 때 사용되는 foaf:knows는 foaf:Person에서 기술되는 내용을 똑같이 넣을 수 있으며 약식으로 중요한 몇 가지만 넣을 수도 있다. foaf:knows를 통해 해당 사용자와 관계가 있거나 아는 사람을 기술할 수 있다. 그러나 한번 작성된 FOAF는 작성자가 직접 수정하지 않으면 업데이트가 되지 않는 단점이 있으며 이와 맥을 같이하여 FOAF를 통해 만든 사회연결망 역시 정적인 단점이 있다. 이러한 문제를 해결하기 위하여 본 논문에서는 사용자가 콘텐츠를 배포함과 동시에 생성되는 RSS를 활용한다.

### RSS

RSS[9]는 Rich Site Summary, RDF Site Summary 등의 약칭으로 뉴스 사이트나 블로그 사이트에서 주로 사용하는 콘텐츠 표현 방식이다. 웹 사이트 관리자는 RSS 형식으로 웹 사이트의 콘텐츠를 출판한다. 또한 태그를 통한 사용자 선택적 정보 수집이 가능하며 XML 포맷을 기반으로 제공되어 자동적인 연동이 쉽게 이루어진다. RSS가 등장하기 전에는 원하는 정보를 얻기 위해 해당 사이트를 직접 방문하여야 했으나 RSS 관련 프로그램이나 서비스를 이용하여 자동 수집이 가능해졌기 때문에 사용자는 각각의 사이트 방문 없이 최신 정보들만 골라 한 자리에서 볼 수 있는 특징이 있다. 새로운 글이 업데이트 되면 블로그, 위키 등의

서비스는 자동으로 콘텐츠 리스트가 XML/RDF 포맷으로 생성시키며 이를 RSS Feed라고 한다. RSS Feed는 채널에 해당하는 사이트의 정보를 담은 부분(Channel)과 개별 콘텐츠 아이템을 표시하는 부분(Item)으로 나뉘며, 콘텐츠의 제목과 URL 정보, 요약 내용을 포함하고 있다[9].

livejournal.com 이나 facebook.com과 같은 웹2.0 서비스들은 RSS와 FOAF를 시스템에서 자동으로 생성할 수 있는 도구를 제공하고 있다. 그러나 이 웹2.0 서비스들은 RSS를 이용하여 사용자들의 저작물을 출판하고 FOAF를 이용하여 사용자의 프로필을 의미적으로 표현하고 있으나 아직까지 이들의 사회연결망을 만들어내고 있지는 못하고 있다. 그리고 사회연결망의 정적 관리의 문제를 여전히 가지고 있다. 따라서 본 논문에서는 실시간으로 출판되는 RSS와 사용자의 프로필의 기술해주는 FOAF를 이용하여 동적으로 FOAF를 관리하는 방법을 제시한다.

### OLAP cubes

OLAP 큐브는 의사결정시스템이나 데이터마이닝 등의 분야에서 대량의 다차원 데이터들을 통합하여 데이터를 요약하고 가공하기 위하여 사용된다. 본 논문에서는 RSS와 FOAF 데이터를 수집하여 이를 다차원테이블에 삽입한다. 다차원테이블과 스타스키마를 기초로 OLAP 연산을 수행한다. 이를 통해 FOAF를 동적으로 관리한다. 본 논문에서 사용한 OLAP 큐브는 [10]에서 제시한 OLAP 질의를 사용하여 작성하고 처리하였다. OLAP 큐브를 만들기 위하여 FOAF와 RSS의 데이터를 각 요소별로 나누어 테이블에 넣을 수 있도록 전처리를 수행하였고 각 차원 테이블의 속성을 사실 테이블에 연결시켰다. OLAP 큐브를 만드는 방법은 크게 두 가지가 있는데 이는 ROLAP[11]과 MOLAP[12]이다. ROLAP은 실체화 뷰를 사용하고 MOLAP은 다차원 배열을 사용한다. 본 논문에서는 각 차원 테이블의 관계를 기초로 OLAP 큐브를 생성하는 ROLAP을 사용한다.

### 관련연구

최근 들어 사회연결망을 추출하고 분석하기 위한 연구가 활발해지고 있다. 그 중 [13]의 논문에서는 시맨틱 웹 커뮤니티에서의 온라인 사회연결망의 추출, 통합 및 시각화를 위한 시스템을 개발하였다. 이 연구에서는 기업과 학계에서 연구자를 추출하고 분석하여 608명의 연구자의 사회연결망을 구축하였다.

[14]은 POLYPHONET이라는 Co-occurrence of name을 사용하는 검색엔진을 구축 하였다. POLYPHONET은 Jaccard coefficient, Overlap coefficient, Cosine 등 다양한 Co-occurrence 방법을 사용하여 결과값을 비교하고 이를 통해 사회 연결망을 추출하였다. 그리고 [15]는

사용자의 사회연결망을 End-To-End 시스템으로 구축하였다. 이 시스템은 사용자의 이메일과 홈페이지를 유일 식별자로 사용하였고 이를 이용하여 관련 있는 사람들의 주소록을 생성하였다. 사회연결망을 추출하고 분석하기 위한 연구는 최근 들어 다양하게 진행되고 있으나 대부분의 사회연결망 연구는 정적인 사용자 프로필을 이용하고 있기 때문에 시간의 흐름에 따른 네트워크의 변경이 되지 않는 단점을 가지고 있다. 그리고 FOAF, SIOC 등 사회연결망을 지원하는 온톨로지 기반의 생성언어가 많이 사용되고 있으나 이들의 통합 및 관리를 위한 방법은 아직 미결과제로 남아 있다.

### FOAF의 동적 관리 기법

#### 시스템 개관

본 논문에서 제시하는 FOAF의 동적 관리 기법은 온톨로지 기반 사용자 프로필 기술언어인 FOAF와 실시간 콘텐츠 출판 언어인 RSS를 기초로 사용한다. FOAF와 RSS는 스타스키마로 설계된 데이터베이스에 각 항목별로 삽입한다. 그리고 만들어진 스타스키마를 기초로 OLAP 큐브를 생성한다. OLAP 큐브를 이용하여 사용자의 연결관계를 갱신하고 FOAF를 재 작성한다. 본 시스템은 크게 자료의 입력과 연결관계 갱신으로 나뉘질 수 있다. 자료의 수집 및 입력부터 연결관계 갱신까지는 네 가지 단계의 작업이 필요한데 이는 다음과 같다.

- Step 1) FOAF와 RSS의 각 요소를 이용하여 스타스키마를 작성한다.
- Step 2) FOAF와 RSS를 수집하여 데이터베이스에 자료를 삽입한다.
- Step 3) 스타스키마에서 관심 분야에 대한 차원을 추출하여 OLAP 큐브를 만든다.
- Step 4) OLAP 큐브에서 OLAP 연산을 실행하여 사용자의 연결관계를 갱신한다.

#### 스타스키마의 설계 및 자료의 입력

FOAF파일과 RSS파일을 이용하여 OLAP 큐브를 만들기 위하여 다차원 테이블을 만들어야 한다. 본 논문에서는 6개의 차원테이블과 1개의 사실테이블을 가진 스타스키마를 구성하였다. 그림 2에서 보는 바와 같이 FOAF 파일과 RSS 파일들은 각 항목별로 나누어 여섯 개의 차원 테이블에 입력된다. 그리고 유일식별자로 만들어진 각 테이블의 기본키들은 사실테이블에 연결된다. 사실테이블의 각 항목은 6개의 차원테이블의 유일식별자와 연결된다. 본 논문에서는 시스템에 입력되는 FOAF와 RSS의 주소를 가지고 있는 것을 가정으로 한다. 그리고 실시간으로 RSS와 FOAF를 수집하여 그림 3과 같은 데이터베이스에 자료를 입력한다.

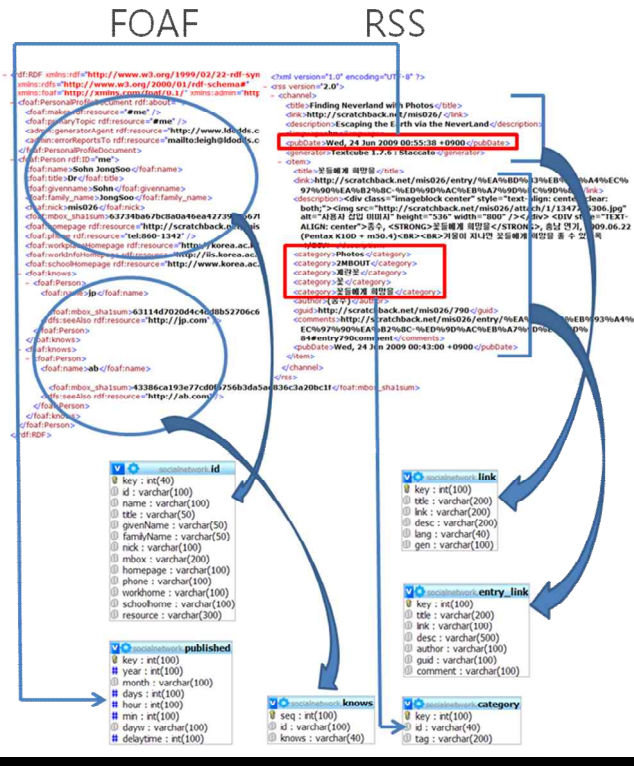


그림 2 - 스타스키마의 설계

본 스타스키마의 6개 차원 테이블은 아래와 같이 6개로 구성된다.

- 1) 사용자 프로필을 저장하는 id dimension table
- 2) 콘텐츠 발행 시간을 저장하는 published time dimension table
- 3) FOAF의 knows를 저장하는 knows dimension table
- 4) 웹 사이트 정보를 저장하는 site link dimension table
- 5) 콘텐츠를 저장하는 entry\_link dimension table
- 6) 태그와 카테고리를 저장하는 category dimension table

이 5개의 차원테이블의 키는 사실테이블에 저장이 되며 사용자들로부터 받은 추천 횟수가 저장된다. FOAF파일과 RSS파일의 각 요소들을 위에서 설명한 5개의 차원테이블과 사실테이블에 삽입한다. 그림 3은 모든 6개의 차원테이블을 사용한 스타스키마의 예제를 보인다.

### OLAP cube의 생성

다차원 데이터 분석을 위하여 OLAP 시스템에서는 OLAP 큐브를 생성하는데 이는 스타스키마로 설계된 데이터베이스를 기반으로 빠르고 쉽게 생성할 수 있다. 본 논문에서는 그림 3에서 보인 스타스키마를 기반으로 OLAP 큐브를 생성한다.

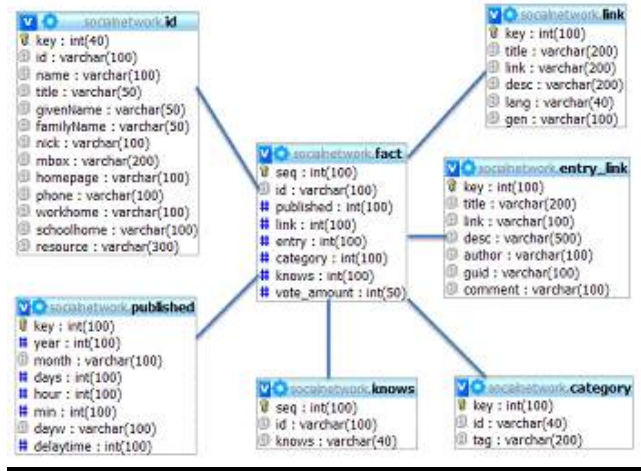


그림 3 - 스타스키마

본 논문에서 제시하는 동적인 FOAF의 관리 기법은 차원테이블에서 사용하고자 하는 차원을 선택하고 이를 큐브로 만드는 것에서 시작한다. 본 논문에서는 일정 시간대에 특정 사용자와 가장 비슷한 관심사를 가지고 있는 사용자를 찾아 그 사용자의 FOAF 파일에 foaf:knows를 추가하는 예제를 든다. 이를 위하여 id, time 그리고 category 이렇게 세 개의 차원을 가진 OLAP 큐브를 생성한다. 본 논문에서 제시하는 방법으로 FOAF 파일들을 동적으로 관리하기 위하여는 다음의 두 가지를 먼저 결정하여야 한다.

- 사용할 차원 테이블을 선택한다. 예를 들어 같은 언어를 사용하면서 동일한 관심사를 갖는 유저들을 대상으로 사회연결망을 추출하고자 한다면 link table, category table 그리고 id table을 차원으로 갖는 큐브를 생성한다.
- 평가값으로 사용할 항목을 결정한다. 본 논문에서는 사용자들의 추천횟수를 활용한다. 그 외에 사용할 수 있는 평가값은 foaf:knows의 개수나 조회수, 페이지랭킹 등이 있다.

사용할 차원의 선택은 최소 1개에서 최대 차원테이블의 개수까지 가능하다. 그리고 평가값은 OLAP 연산을 실행할 때 사용되며 이를 통해 연관관계를 갖는 사용자를 찾는다. OLAP 큐브를 만들어 OLAP 연산을 수행하려면 개념간의 상하위 관계를 정의해주어야 한다. 본 논문에서는 시간대 별로 사용자의 관심사를 추출하여야 하므로 그림 4와 같이 published time에 대하여 개념의 상하위관계를 정의하였다.

그리고 그림4에서 보이는 스타스키마를 기초로 category, time 그리고 id를 각각 하나의 차원으로 설정하고 사실테이블의 평가값으로써 사용자의 추천 횟수(vote amount)를 사용하면 그림 5와 같은 OLAP 큐브를 생성된다.

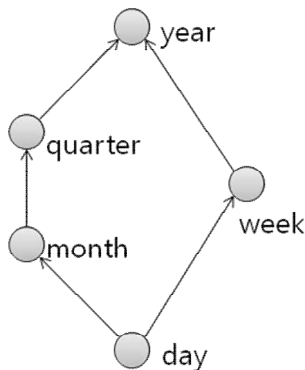


그림 4 - 시간에 관한 개념 상하위관계

본 논문에서 제시하는 방법은 시간이나 관심사에 따라 변할 수 있는 사용자간의 연결관계를 빠르고 효율적으로 반영하고 관리하는데 그 목적이 있다. 따라서 OLAP 큐브를 생성할 때 time 차원 테이블과 id 차원 테이블은 필수적으로 포함이 되고 하나 혹은 그 이상의 찾고자 하는 연결 관계에 해당하는 차원 테이블을 큐브에 포함시킨다.

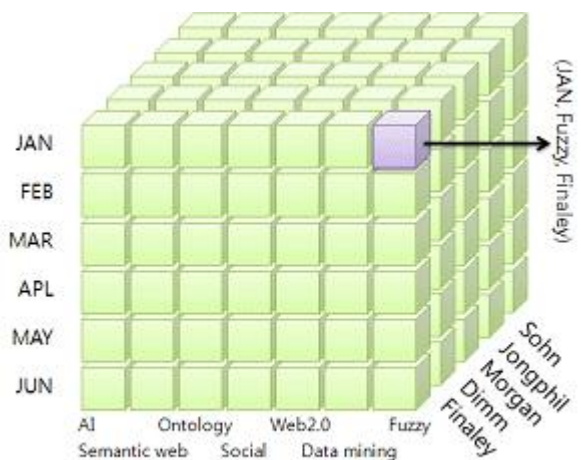


그림 5 - 3-D OLAP cube

### OLAP을 이용한 동적 FOAF의 관리

연결관계를 추출하고자 하는 사용자와 가장 비슷한 관심사를 가진 사람들을 찾고 그들과의 연결을 만들기 위해서는 두 가지 단계가 필요하다. 첫 번째로 해당 사용자의 주요 관심사를 추출하는 것이다. 이를 위하여 그림 5와 같은 OLAP 큐브를 대상으로 몇 번의 OLAP 연산을 실행한다. 두 번째로 해당 사용자와 가장 비슷한 관심사를 갖는 사용자를 추출한다. 이 두 번의 단계를 거치면 해당 사용자의 주요 관심사를 추출할 수 있게 되며 그 관심사와 가장 유사한 추출하여 연결관계를 만들 수 있다. 표 1은 사용자들의 연결관계를 찾고 FOAF를 갱신하기 위한 알고리즘을 보인다. 본 논문에서는 OLAP 연산의 실행을 표현하기 위하여 [16]에서 제시한 데이터마이닝 질의 언어인 DMQL을

사용하여 질의를 표현한다. DMQL은 OLAP 큐브를 만들고 OLAP 연산을 실행하기 위하여 만들어진 질의 언어로서 SQL과 비슷한 언어 구조를 갖고 있으며 표현이 직관적이다.

표 1 - 동적 FOAF 관리 알고리즘

알고리즘 1. extracting related users	
input :	<i>selected dimensions, id, period of time, number of keywords</i>
output :	<i>updated FOAF</i>
1	<b>load_cube</b> ( <i>time, id, selected_dimension</i> ) // 선택된 차원들과 id, time으로 구성된 cube를 불러옴
2	<b>while</b> (number of selected dimensions <i>j</i> ) //선택된 차원의 수만큼 반복
3	<b>find_id_interest</b> ( <i>dimension[j], id, period of time, number of keywords</i> )
4	<b>find_relation</b> ( <i>dimension[j], id, period of time, most voted category</i> ) //j차원에 대한 id의 관계찾기
5	<b>end while</b>
6	<b>save related user IDs</b> //k 명의 관계된 유저를 저장
7	<b>return</b> FOAF of id // FOAF 반환
function - find_id_interest() input : <i>dimension j, id, period of time</i> output : <i>most voted category</i>	
1	<b>DMQL</b> <- slice for id = "id"
2	<b>DMQL</b> <- roll up on time (from month to quarter)
3	<b>select</b> number of keywords of <i>most voted category</i>
4	<b>return</b> <i>most voted category</i>
function - find_relation() input : <i>dimension j, id, period of time, most voted category</i> output : <i>related user IDs</i>	
1	<b>DMQL</b> <- dice for ( <i>dimension j</i> )
2	<b>DMQL</b> <- roll up on time(from date to period of time)
3	<b>select</b> <i>related users</i>
4	<b>return</b> <i>related user IDs</i>

### 예제

본 장에서는 본 논문에서 사용하는 방법을 예제를 들어 설명한다. 먼저 RSS와 FOAF를 수집하여 스타스키마 데이터베이스에 삽입한다. 그리고 OLAP 큐브는 category, month, id를 갖는 세 개의 차원테이블과 투표 값을 매개로 하는 사실 테이블을 구성한다. 본 예제에서 선택한 차원은 category 차원 테이블이지만 사용자가 어떤 차원을 선택했는가에 따라 OLAP 큐브의 구성은 달라질 수 있다. 그리고 표 2와 같이 사실은 세 차원 열과 투표값 열을 갖는 테이블로 구성된다고 가정한다. 표 2의 가장 우측의 vote amount 는 사용자들의 추천횟수로서 협업 기반 필터링을 위하여 중요하게 사용된다[17][18][19]. 본 논문에서는 키워드와 사용자의 관련성을 측정하기

위한 지표로 vote amount를 사용한다.

표 2 - 사실테이블

Dimensions			vote amount
month	category	id	
JAN	AI	Sohn	98
JAN	Web2.0	Morgan	45
FEB	Social	Dimm	3
FEB	Fuzzy	Dimm	21
MAR	Semantic web	Morgan	87
MAY	AI	Sohn	56
MAY	AI	Morgan	32
MAY	Semantic web	Sohn	57
MAY	Ontology	Morgan	12
JUN	Social	Finaley	0
JUN	Social	Jongphil	45
JUN	Semantic web	Sohn	34
JUN	Fuzzy	Finaley	11

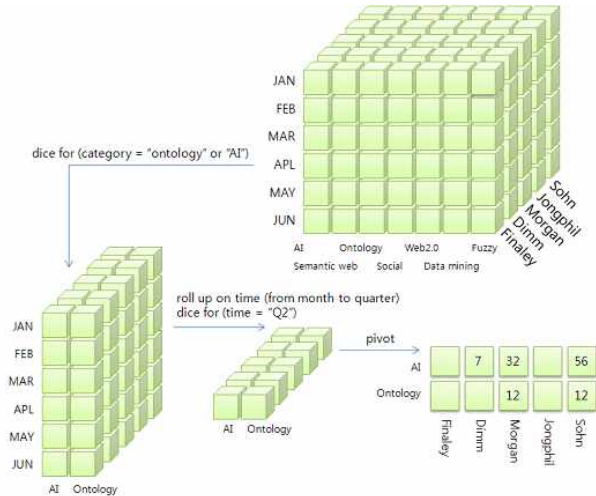


그림 7 - 연결관계 찾기

위 그림 7의 하단 오른쪽에서 보이는 바와 같이 Morgan과 관련이 있는 사용자는 Sohn 과 Dimm 임을 알 수 있다. 다음 단계로 Sohn과 Dimm을 대상으로 그림 5와 같이 find\_relation 함수를 실행한다. 본 알고리즘이 마지막까지 모두 실행되고 나면 (ontology, AI)와 관련된 사용자가 추출된다.

### 결론

사회연결망 서비스는 웹2.0의 등장과 함께 중요하며 미래 인터넷 기술로 각광받고 있다. 따라서 사회연결망을 나타내고 추출하기 위한 많은 연구가 진행되어왔으나 아직 다양한 관점에서 사회연결망을 추출하고 동적으로 업데이트하기 위한 방법은 제시되지 않았다.

본 논문은 사회연결망 추출 및 관리의 문제점을 해결하기 위한 방법을 제시하였다. 본 논문에서 제시한 방법은 FOAF와 RSS를 기초 자료로 한 OLAP 시스템을 이용한 방법이다. FOAF를 이용하면 상호호환성을 유지하면서 개인의 프로필 및 다른 사람과의 관계를 표현할 수 있다. RSS는 실시간 웹 콘텐츠 출판 언어로서 블로그나 위키 등의 미디어 사이트에서 새로 올라온 콘텐츠를 배포할 수 있는 방법을 제공한다. 본 논문에서는 FOAF를 통해 제공되는 사용자의 프로필과 RSS를 통해 제공되는 콘텐츠 정보를 OLAP 시스템에 넣어 가공한다.

본 논문에서 제시하는 방법을 사용하면 사용자의 관심사 뿐 아니라 언어, 출신학교 등에 따른 연결관계를 FOAF를 통해 표현할 수 있으며 실시간으로 배포되는 RSS를 이용하여 변화하는 사용자의 관심사를 동적으로 FOAF에 동적으로 반영할 수 있다.

id, category, time 차원을 기반으로 3차원 OLAP 큐브를 만들면 그림 6의 상단과 같다. 연결관계를 찾고자 하는 사용자가 “Morgan”이라고 할 때 Morgan의 최대 관심사를 추출한다.

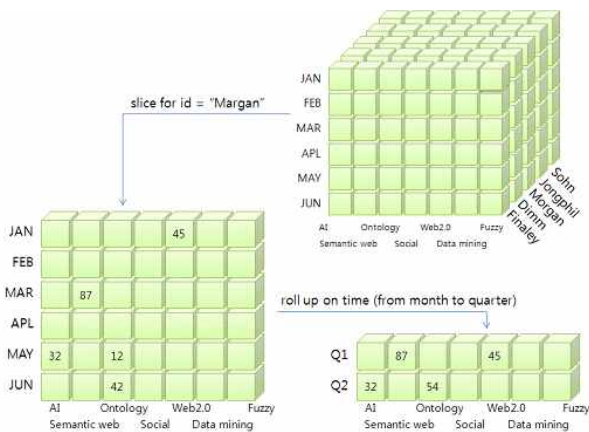


그림 6 - 사용자의 관심사 찾기

다음으로 사용자와 가장 관련이 깊은 사용자들을 찾는다. 이 과정은 표 1에서 보이는 바와 같이 OLAP 연산 중에서 dice와 roll up을 이용하며 그림 7에서 보이고 있다.

## References

- [1] Berners-Lee, T., Hendler, J., AND Lassila, O. (2001) "The Semantic Web". Scientific American, May
- [2] Paolillo, J. C., Wright, E.: Social Network Analysis on the Semantic Web: Techniques and Challenges for Visualizing FOAF. Visualizing the semantic Web Xml-based Internet And Information (2006).
- [3] Brickley, D., Miller, L. (2003). FOAF vocabulary specification. Technical report, RDFWeb FOAF Project.
- [4] Herring, S., Scheidt, L., Bonus, S., Wright, E. (2004). Bridging the gap: A genre analysis of Weblogs. Proceedings of the 37th Hawaii International Conference on SystemSciences (HICSS-37). IEEE Computer Society, Los Alamitos, California.
- [5] John Breslin, Stefan Decker, "The Future of Social Networks on the Internet: The Need for Semantics," IEEE Internet Computing, vol. 11, no. 6, pp. 86-90, November/December, 2007.
- [6] D. Gruhl, D. N. Meredith, J. H. Pieper, A. Cozzi, and S. Dill. The web beyond popularity: a really simple system for web scale rss. In WWW '06: Proceedings of the 15th international conference on World Wide Web, pages 183-192, New York, NY, USA, 2006. ACM.
- [7] Goldbeck, J., Rothstein, M.: Linking social Networks on the web with FOAF AAA08 (2008)
- [8] FOAF Specifications, <http://xmlns.com/foaf/spec/>
- [9] RSS Specifications, <http://www.rssboard.org/rss-specification>
- [10] J. Gray, A. Bosworth, A. Layman, and H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-total. In Proc. of Int'l Conf. on Data Engineering (ICDE), pages 152-159, 1996.
- [11] K. Morfonios and Y. Ioannidis. CURE for cubes: Cubing using a ROLAP engine. In Proc. of Very Large Data Bases (VLDB), pages 379-390, 2006.
- [12] Y. Zhao, P. Deshpande, and J. F. Naughton. An array-based algorithm for simultaneous multidimensional aggregates. In Proc. of ACM SIGMOD, pages 159-170, 1997.
- [13] P. Mika. Flink: Semantic web technology for the extraction and analysis of social networks. Journal of Web Semantics, Vol. 3, No. 2, 2005.
- [14] Y. Matsuo, J. Mori, M. Hamasaki, H. Takeda, T. Nishimura, K. Hasida, and M. Ishizuka. POLYPHONET: An advanced social network extraction system. In Proc. WWW 2006, 2006.
- [15] A. Culotta, R. Bekkerman, and A. McCallum. Extracting social networks and contact information from email and the web. In CEAS-1, 2004.
- [16] J. Han, Y. Fu, W. Wang, K. Koperski, and O. Zaiane. Dmql: A data mining query language for relational databases. In SIGMOD'96 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'SG), Montreal, Canada, 1996.
- [17] Robertson S.E., Sparck Jones K., Relevance Weighting of Search Term, Journal of the America Society for Information Science, May-June, 1976.
- [18] L. Greg , S. Brent , and Y. Jeremy. Amazon.com Recommendations, Item-to-Item Collaborative filtering, IEEE Internet Computing, 2003
- [19] Cotter, P., and Smyth, B., PTV: Personalised TV Guides. In Proceedings of the 12th Conference on Innovative Applications of Artificial Intelligence, IAAI 2000, Austin, Texas, 2000.