

온톨로지 기반 소셜 네트워크 분석을 이용한 전문가 추천 시스템 (An Expert Recommendation System using Ontology-based Social Network Analysis)

박 상 원 [†] 최 은 정 [†]
(Sangwon Park) (Eunjeong Choi)

박 민 수 [†] 김 정 규 [†]
(Minsu Park) (Jeonggyu Kim)

서 은 석 [†] 박 영 택 ^{**}
(Eunseok Seo) (Youngtaek Park)

요 약 시맨틱 웹 기반의 소셜 네트워크는 다양한 분야에서 높은 활용성을 가진다. 본 논문에서는 FOAF 기반 소셜 네트워크에 대하여 다양한 분석을 수행하고, 이를 바탕으로 네트워크 내의 전문가를 추천하는 시스템을 제안한다. 분석 시스템은 SparQL, RDFS 추론, 그리고 시각화 도구를 이용하여 온톨로지 기반 소셜 네트워크에 대한 효과적인 활용 방안을 제시한다. 그리고 이러한 분석 시스템을 실제 소셜 네트워크에 적용시켜 Centrality, Small world, Scale free 특성 등의 다양한 분석을 수행하고, 특정 분야에 대한 전문가를 분석하는 방법을 제시한다. 이러한 활용

- 본 연구는 숭실대학교의 지원을 받았습니니다.
- 2008년도 Business Intelligence 강의를 수강한 학생들에게 감사합니다.
- 이 논문은 제35회 추계학술대회에서 '온톨로지 기반 소셜 네트워크 분석을 이용한 전문가 추천 시스템'의 제목으로 발표된 논문을 확장한 것이다

[†] 학생회원 : 숭실대학교 컴퓨터학과
hudoni@gmail.com
lovej0918@naver.com
yfoottim@yahoo.co.kr
yalshyaish@naver.com
erin214@gmail.com

^{**} 종신회원 : 숭실대학교 컴퓨터학과 교수
park@ssu.ac.kr
논문접수 : 2008년 12월 19일
심사완료 : 2009년 3월 1일

Copyright©2009 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제15권 제5호(2009.5)

방법은 마케팅, 조직 관리, 지식 경영 시스템 등 다양한 분야에서 이용될 것으로 기대한다.

키워드 : 온톨로지, 온톨로지 추론, 소셜 네트워크, 소셜 네트워크 분석, SparQL, 추천

Abstract The semantic web-based social network is highly useful in a variety of areas. In this paper we make diverse analyses of the FOAF-based social network, and propose an expert recommendation system. This system presents useful method of ontology-based social network using SparQL, RDFS inference, and visualization tools. Then we apply it to real social network in order to make various analyses of centrality, small world, scale free, etc. Moreover, our system suggests method for analysis of an expert on specific field. We expect such method to be utilized in multifarious areas - marketing, group administration, knowledge management system, and so on.

Key words : Ontology, ontology reasoning, social network, SNA(social network analysis), SparQL, recommendation

1. 서 론

소셜 네트워크에 대한 연구는 과거에는 사회 과학의 범주였지만 최근에는 웹을 통해 소셜 네트워크를 구축하고 활용하는 형태로 진행되고 있다. 소셜 네트워크는 각 개인에 대한 정보보다는 사람들 간의 관계와 상호작용을 바탕으로 형성된 연결망에 초점을 맞춘 연구이다 [1]. 연구를 통해 얻은 결과는 마케팅, 지식 관리, 조직 개발, 경제학과 사회학에 대한 연구 등 매우 다양한 분야에서 활용되고 있다. 이번 연구에서는 학생들의 FOAF (Friend Of A Friend)[2] 문서를 바탕으로 구축된 소셜 네트워크에 대해서 다양한 분석을 수행하여 학생 네트워크의 특징과 구성을 확인한다. 이번 연구는 숭실대학교 학생 47명이 직접 작성한 실제 FOAF 문서를 기반으로 진행되었다.

시맨틱 웹 기술을 적용하여 표현된 소셜 네트워크는 자동화된 처리가 가능하고, 다양한 온톨로지 기반 추론이 가능하다는 점에서 효과적인 활용성을 제공한다. 본 논문에서는 OWL(Web Ontology Language)[3]로 표현한 온톨로지와 소셜 네트워크에 대한 RDFS(Resource Description Framework Schema)[4] 추론을 수행하고, SparQL(SparQL Protocol And RDF Query Language)[5]을 이용하여 시맨틱 소셜 네트워크의 활용성을 높인다. 그리고 이를 기반으로 실제 소셜 네트워크를 분석하고 전문가 추천 시스템을 제안한다. 이러한 소셜 네트워크의 활용 기법은 다양한 분야에서 활용될 것으로 기대된다.

2. 관련 연구

Flink[7]는 다양한 방법으로 FOAF 기반의 소셜 네트워크를 분석하여 시각적으로 보여준다. 관심 주제와 위치에 따른 네트워크 그리고 소셜 네트워크 분석을 지원한다. Flink는 웹 정보, FOAF 프로파일, E-mail 전송 기록, 출판 기록의 정보를 RDF 데이터 저장소인 Sesami [8]에 저장하고, JUNG[9] 네트워크 API를 이용하여 분석한 결과를 웹 어플리케이션을 이용하여 보여준다. 이러한 Flink 어플리케이션의 구성은 소셜 네트워크 분석 시스템에 대한 하나의 예를 제시한다. 이외에도 온톨로지 기반 소셜 네트워크 분석에 대한 연구는 다양하게 진행되고 있다.

3. 소셜 네트워크 분석

소셜 네트워크의 구조적 특성을 이해하기 위해서는 관계의 밀도(Density)나 중심성(Centrality)에 대한 분석이 필요하다. 일반적으로 소셜 네트워크는 Small world와 Scale free 특성을 갖기 때문에 이에 대한 분석을 통해 네트워크의 특징을 파악할 수 있다.

Degree centrality는 네트워크를 구성하는 하나의 노드가 다른 노드들과 가지는 직접적인 연결 관계의 정도를 나타내며, Betweenness centrality는 다른 노드들을 연결시키는 매개성을 나타낸다. 높은 Betweenness를 갖는 노드는 네트워크의 흐름에 중요한 영향을 미치는 중개자 역할을 한다. 그리고 Closeness centrality는 노드가 네트워크 안에서 지리적으로 중심부에 위치하는지를 나타낸다. Closeness가 높은 노드는 네트워크에서 발생되는 상황을 잘 관찰할 수 있다. 또한 Characteristic path length(L)는 그룹 구성원들 간의 연결 거리가 길고 짧은지를 나타내며, Clustering coefficient(C)를 통해 구성원들의 응집력을 파악하는 것이 가능하다. L과 C는 다음과 같이 계산된다.

$$L = \text{모든 노드 쌍의 최단거리의 평균}$$

$$C = \frac{\text{이웃 노드 간 실제 링크의 수}}{\text{이웃 노드 간 연결 가능한 링크의 수}}$$

소셜 네트워크는 Small world의 특징을 갖는다. Small world network[10]의 핵심적인 논리는 가장 대표적인 Stanly Milgram의 실험[11]과 함께 많은 연구를 통해 정립되었다. Small world network은 L은 낮고 C는 높은 특징을 갖는다. Small world의 대표적인 데이터 모델인 α -model [12]과 β -model[13]을 통해 특성을 알 수 있다. 또한 Alpha와 Beta 네트워크는 각각의 노드가 모두 평균에 근접한 수의 연결을 가진다[14]. 즉, 각 노드의 Degree가 Poisson distribution[15] 형태를 나타내지만, 실제 소셜 네트워크는 Power-law distribution 형태를 갖는다. 이러한 Scale free network에는 매우 중요한 역할

을 하는 Hub가 존재한다.

4. FOAF 기반 전문가 추천 시스템

FOAF 기반의 소셜 네트워크에서 전문가를 분석하기 위해서는 시맨틱 웹 기술을 활용한 소셜 네트워크 분석 시스템이 필요하다. 본 논문에서는 OWL 온톨로지, RDFS 추론, SparQL 등을 활용하는 시스템을 제안한다.

4.1 시스템 구조

다음 그림 1은 온톨로지 기반 소셜 네트워크 분석 시스템의 구조를 나타내고 있다. 본 논문에서 제안하는 시스템은 OWL 온톨로지를 처리하는 기능을 제공하는 Jena Framework를 하부 구조로 두고 있다. Jena는 RDF/RDFS, OWL을 모두 지원하는 자바 기반의 프레임워크로써, 규칙 기반 추론 기능을 통해 논리 처리를 지원하며, Triple 형태로 저장한 데이터 모델에 대해서 SparQL을 이용한 접근을 지원한다[16].

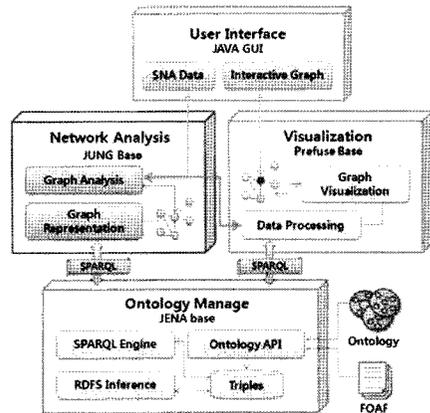


그림 1 소셜 네트워크 분석 시스템 구조도

시스템은 크게 온톨로지 처리, 네트워크 분석, 시각화 그리고 사용자 인터페이스의 네 부분으로 구성된다. 첫째로 가장 하단에 위치하는 온톨로지 처리 부분은 Jena 프레임워크 기반으로 학생들의 FOAF 파일과 온톨로지를 로드하고, 기본 모델과 RDFS 추론된 모델을 기반으로 SparQL을 이용한 질의 및 결과 집합을 제공한다. 두 번째는 네트워크 분석 부분으로 JUNG(Java Universal Network/Graph Framework)에서 지원하는 그래프 모델 및 분석 API를 사용하였다. SparQL 질의를 통해 Triple 저장소로부터 가져온 학생들의 관계를 다양한 그래프로 구성하고 이를 분석한다. 세 번째 구성은 Prefuse 프레임워크를 활용한 시각화 부분으로, 분석 결과를 쉽게 확인할 수 있도록 도와주며, 마지막 구성 요소는 JAVA 라이브러리 기반의 사용자 인터페이스이다.

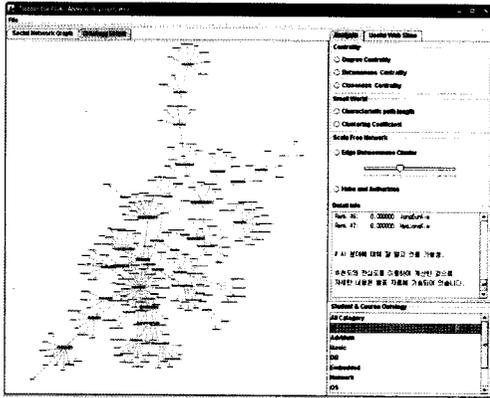


그림 2 프로그램 실행 화면

4.2 FOAF 기반 온톨로지

FOAF를 활용하면 확장성과 유연성을 가지는 소셜 네트워크를 구성할 수 있다. 본 논문은 각각의 학생들이 작성한 FOAF 정보를 이용하기 때문에 전문가를 분석하기 위한 추가적인 정보로써 교과목 온톨로지를 활용하였다. 교과목 온톨로지는 숭실대학교 컴퓨터학부의 교과목 트리를 기반으로 구성하였다. Course 클래스의 인스턴스들은 서로 교과 과정에 정의된 필수선수과목과 권장선수과목의 관계를 가지며, AI, OS, Network 등 하나 이상의 Category 인스턴스들과 belongsTo 관계를 갖는다. 이러한 교과목 인스턴스들은 학생들이 FOAF 문서에 자신의 관심 과목을 기록할 수 있도록 제공되었으며, 그림 4의 각 Category에 대한 추천 속성을 이용하여 해당 분야에 대해 실력 있는 친구들을 추천하도록 요청했다. 그림 3은 하나의 예로써 다음과 같은 의미를 가진다. 신일용, 이유미 학생이 인공지능 분야에 대해서 추천한 김진호 학생은 '지능형 시스템 실습'과 '인공지능' 과목에 관심이 있다. 그리고 이들 과목은 권장 선수과목 관계이므로 서로 밀접한 관계에 있으며, 인공지능과 고급 수학 분야에 속해있다. 이러한 정보를 활용하면 각 분야에 대한 학생들의 추천도와 관심도를 유추해 내는 것이 가능하다. 특정 분야에 대해서 추천도와 관심도가 높은 학생은 그 분야에 대해 전문성을 나타낼 것으

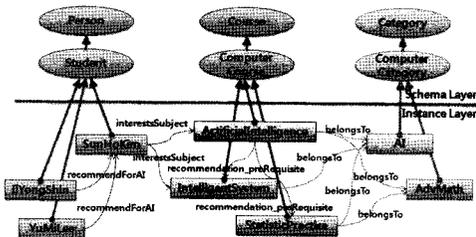


그림 3 FOAF 기반의 교과목 온톨로지 표현

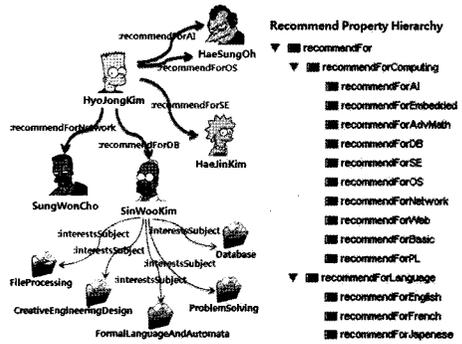


그림 4 추천 관계의 일부분과 추천 속성의 계층

로 기대할 수 있다.

본 논문에서는 학생 네트워크라는 특성에 따라서 교과목 온톨로지를 사용했지만, 소셜 네트워크의 분석에는 네트워크의 특징과 분석 목적에 따라서 다양한 온톨로지가 활용될 수 있다.

4.3 RDFS 추론 및 SparQL 활용

RDFS(Resource Description Framework Schema)는 RDF 데이터 모델에 대하여 자원의 타입, 클래스와 속성의 계층 관계를 기술할 수 있는 방법을 제공한다. RDFS 추론이란 RDF Semantics[17]에서 제안한 13가지 RDFS 함의 규칙을 바탕으로 RDF 문서에 기술되지 않은 새로운 정보를 확장하는 과정이다. 그림 4는 학생들의 추천 관계 중 김효중 학생이 컴퓨터 분야에 대해서 추천한 사람들과 추천 속성의 일부분을 표현한 것이다. 추천 관계는 그림과 같은 계층을 갖는 다양한 속성으로 표현된다. 그림 5는 RDFS 추론을 수행하여 제 7번 RDFS 함의 규칙에 따라서 추가된 새로운 Triple들을 보여준다. 이외에도 다양한 RDFS 함의 규칙을 활용하면 효과적인 규칙 기반 추론이 가능하다.

SparQL은 RDF 그래프에 접근하고 검색 할 수 있도록 개발된 질의 언어로 이전의 RDQL(RDF Data Query Language)에서 발전된 형태이다. SparQL은 Subject, Predicate, Object 구조의 Triple로 표현되는 RDF 데

RDFS Inference

Rule Name	If E contains:	then add:
Rdfs7	aaa rdfs:subPropertyOf bbb uuu aaa yyy	uuu bbb yyy

그림 5 추천 관계에 대한 RDFS 추론

이더 대해서 패턴 매치를 통해 효과적인 검색을 가능하게 한다. 그림 4에 표현된 학생 네트워크에 대해서 “김효종 학생이 데이터베이스에 대해서 추천한 학생의 관심 과목은 무엇인가?”의 SparQL 질의는 다음과 같다.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX ns: <http://www.mindswap.org/2003/owl/foaf#>
SELECT ?expert ?subject WHERE {
    ns:HyoJongKim ns:recommendForDB ?expert .
    ?expert rdf:type ns:Student .
    ?expert ns:interestsSubject ?subject
}
```

5. 소셜 네트워크 분석 결과

본 논문에서는 학생 47명이 Protégé를 이용하여 직접 작성한 FOAF 문서 기반의 소셜 네트워크를 분석한다. 이 FOAF 문서들은 숭실대학교 2008년도 1학기 ‘비즈니스 인텔리전스’ 수업의 일환으로 작성된 것이다.

구현한 시스템을 통해 분석한 결과는 다음과 같다. 표 1은 Centrality 분석 결과를 1위부터 5위까지 나타낸 것이다. 이를 통해 활동이 왕성한 학생, 브로커 역할을 하는 학생, 그리고 상황을 잘 관찰할 수 있는 학생들을 파악할 수 있다. 그림 6은 Shortest path length, Clustering coefficient, Degree distribution 분석 결과를 구체적으로 보여준다. 네트워크의 L은 2.65, C는 0.61로 분석되었으며, 이는 Small world의 특성을 잘 보여준다. 길이가 1~3 인 짧은 최단 경로들이 전체의 83%를 차지하고 있으며, 작은 규모의 네트워크임에도 불구하고 긴 최단 경로가 존재하는 이유는 편입, 전과 및 복학 등의 요소가 미치는 영향으로 판단된다. 그리고 구성원 중 68%가 C 값이 0.5 이상으로 높은 형태를 보인다. 특히 11명의 학생은 C 값이 1.0으로서 친구들 간의 강한 결합력을 보였다. 네트워크의 Degree distribution 분석

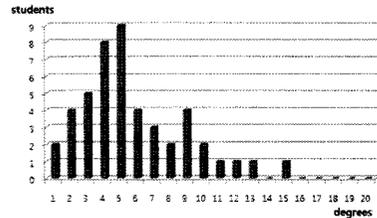
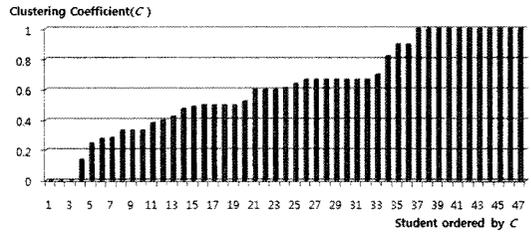
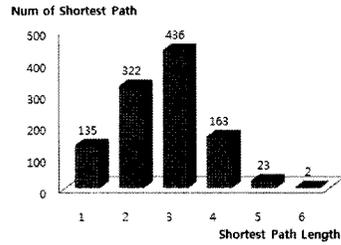


그림 6 Shortest path length, Clustering coefficient, Degree distribution 분석 결과

결과는 Power-law와 Poisson curve의 혼합된 양상을 보였다. 그래프를 보면 Degree 4를 기점으로 왼쪽은 Poisson distribution 형태에 가깝고, 오른쪽은 Power-law distribution 형태에 가깝다. 이런 분포는 1~3명의 친한 친구들과 같은 수업을 수강하는 학생들의 성향 때문에 나타난 것으로 판단된다.

6. 소셜 네트워크에서의 전문가 추천

온톨로지 기반의 소셜 네트워크에 대한 다양한 분석 결과를 통해 네트워크 내에서 중요성을 지닌 사람들을 추천할 수 있다. 본 논문에서 학생들의 각 분야에 대한 전문성은 추천도와 관심도를 통하여 추론된다. 추천도는 해당 분야에 대하여 추천 받은 수를 이용하였고, 관심도는 관심 과목과 해당 분야와의 연관성을 활용하여 이끌어냈다. 표 2는 전체 컴퓨터 분야와 10개의 세부 분야에 대한 학생들의 전문성 분석 결과를 보여준다.

다음은 이러한 정보를 활용하여 임베디드 시스템 분야에 대해서 분석한 것으로 가장 높은 점수를 보인 김동현 학생에 대한 내용이다. 그림 8은 Centrality 분석 결과를 표현한 것으로 네트워크 내에서 Degree centrality와 Closeness centrality가 상대적으로 높게 나타난다. 그리고 그림 9는 각 세부 분야에 대한 전문성을

표 1 Centrality 분석 결과 (1~5위)

	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	JungKyuKim 0.055556	SeonYongHwang 307.901681	SeonYongHwang 90.00
2	SeonYongHwang 0.048148	JungKyuKim 193.838049	JungKyuKim 91.00
3	KwangSungJun 0.044444	SungKyuKang 134.534921	SungKyuKang 93.00
4	SangWonLee 0.040741	SunWookPyo 121.540966	JungEunKim 95.00
5	SangCheolLee 0.037037	SangWonLee 95.516947	KwangSungJun 96.00

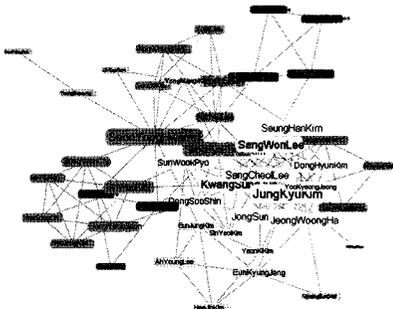


그림 7 Centrality 및 Cluster 분석 결과

표 2 컴퓨터 분야에 대한 전문성 순위 (1위~3위)

	Rank 1	Rank 2	Rank 3
Total Computing	JeongWoongHa 52.355372	DongHyunKim 43.760331	SeonYongHwang 42.900826
Artificial Intelligence	SangWonPark 9.772095	DongSooShin 9.462799	SeonYongHwang 9.302613
Advance Math	KwangSungJun 10.405128	DongHyunKim 9.059307	DongSooShin 8.233070
Basic Part	KwangSungJun 8.370923	JeongWoongHa 7.360923	SungWonCho 6.500000
Database	JungEunKim 9.733702	EunJeongChoi 9.598230	MinJoonSon 8.479110
Embedded System	DongHyunKim 12.634172	SungKyuKang 7.841736	WonGullPark 7.637397
Network	SeonYongHwang 9.366036	SangYeobJeon 8.571429	EunJungKim 7.523865
Operating System	JeongWoongHa 11.946019	HaeSungOh 9.391898	YongMarooKim 8.666667
Programming Language	JeongWoongHa 9.462502	SangCheolLee 8.496513	HaeSungOh 8.343289
Software Engineering	JungKyuKim 12.008540	DeokYongLee 10.555556	YuMILee 8.604160
Web Part	SeonYongHwang 11.471299	SunHoKim 10.587908	MinSuPark 8.128222

나타내는 그래프이다. 이러한 다양한 분석을 통해 김동현 학생이 임베디드 시스템 분야에 대해서 잘 알고 있을 가능성이 가장 높고, 그룹 내에서 임베디드 시스템 관련 활동에 대한 역할과 관할력이 높다고 판단된다. 그리고 김동현 학생이 추천한 학생들과, 같은 클러스터 내에서 높은 점수를 보이는 학생들에 대해서도 주목할 필요가 있다.

7. 결론

본 논문에서는 OWL 온톨로지와 RDFS 추론 및 SparQL 등의 시맨틱 웹 기술을 적용한 소셜 네트워크 분석 시스템을 제안하였다. 그리고 이러한 시스템을 활용하여 소셜 네트워크의 구조와 특성을 파악하고 전문가를 분석해내는 방법을 제시하였다. 소셜 네트워크 분석 과정에서 온톨로지 기반 추론을 이용하면 사람들의 함축적인 상호 관계 및 특성을 이끌어 낼 수 있다. 그리고 분석 시스템은 효과적인 시각화를 제공하기 때문에

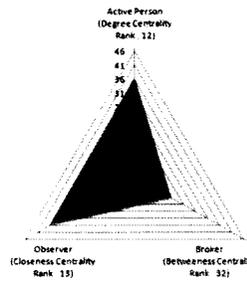


그림 8 네트워크 내 영향력

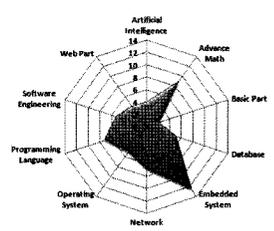


그림 9 전문 분야 분석

다양한 특징을 한 눈에 파악할 수 있도록 도와준다. 이번 연구는 학생들의 네트워크를 기반으로 이루어졌기 때문에 소셜 네트워크의 분석 방법과 활용 방법에 다소 제한이 있었다. 하지만 제한된 분석 방법은 소비자 네트워크 분석을 통한 집중적인 마케팅, 회사 조직 분석을 통한 효율적인 조직 편성, 그리고 지식 경영 시스템에의 적용처럼 다양한 분야에서 활용될 수 있을 것이다.

참고 문헌

- [1] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications," Cambridge, Cambridge University Press, 1994.
- [2] <http://www.foaf-project.org>
- [3] <http://www.w3.org/TR/owl-features/>
- [4] <http://www.w3.org/TR/rdf-schema/>
- [5] <http://www.w3.org/TR/rdf-sparql-query/>
- [6] <http://www.prefuse.org/>
- [7] P. Mika, "Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks," Journal of Web Semantics, 2005.
- [8] <http://www.openrdf.org/>
- [9] <http://jung.sourceforge.net/>
- [10] D. J. Watts, "Networks, dynamics, and the small-world phenomenon," American Journal of Sociology, 1999.
- [11] S. Milgram, "The small-world problem," Psychology Today 1, 61-67, 1967.
- [12] D. J. Watts, "Small Worlds: The Dynamics of Networks Between Order and Randomness," Princeton, Princeton University Press, 1999.
- [13] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature 393, 440-442, 1998.
- [14] A. Barabasi and R. Albert, "Emergence of scaling in random networks," Science 286, 509-512, 1999.
- [15] A. Barabasi and R. Albert, "Statistical mechanics of complex networks," Review of Modern Physics, 2002.
- [16] <http://jena.sourceforge.net/>
- [17] <http://www.w3.org/TR/rdf-nt/>