

## 시맨틱 웹 자원의 랭킹을 위한 알고리즘: 클래스중심 접근방법\*

노상규\*\*, 박현정\*\*\*, 박진수\*\*\*\*

### A Ranking Algorithm for Semantic Web Resources: A Class-oriented Approach

Sangkyu Rho, Hyunjung Park, Jinsoo Park

We frequently use search engines to find relevant information in the Web but still end up with too much information. In order to solve this problem of information overload, ranking algorithms have been applied to various domains. As more information will be available in the future, effectively and efficiently ranking search results will become more critical. In this paper, we propose a ranking algorithm for the Semantic Web resources, specifically RDF resources.

Traditionally, the importance of a particular Web page is estimated based on the number of key words found in the page, which is subject to manipulation. In contrast, link analysis methods such as Google's PageRank capitalize on the information which is inherent in the link structure of the Web graph. PageRank considers a certain page highly important if it is referred to by many other pages. The degree of the importance also increases if the importance of the referring pages is high.

Kleinberg's algorithm is another link-structure based ranking algorithm for Web pages. Unlike PageRank, Kleinberg's algorithm utilizes two kinds of scores: the authority score and the hub score. If a page has a high authority score, it is an authority on a given topic and many pages refer to it. A page with a high hub score links to many authoritative pages.

As mentioned above, the link-structure based ranking method has been playing an essential role in World

---

\* 본연구는 서울대학교 경영대학 경영연구소의 연구비 지원에 의해 수행되었음.

\*\* 서울대학교 경영전문대학원/경영대학

\*\*\* 서울대학교 경영대학 박사과정

\*\*\*\* 교신저자, 서울대학교 경영전문대학원/경영대학

Wide Web (WWW), and nowadays, many people recognize the effectiveness and efficiency of it. On the other hand, as Resource Description Framework (RDF) data model forms the foundation of the Semantic Web, any information in the Semantic Web can be expressed with RDF graph, making the ranking algorithm for RDF knowledge bases greatly important. The RDF graph consists of nodes and directional links similar to the Web graph. As a result, the link-structure based ranking method seems to be highly applicable to ranking the Semantic Web resources. However, the information space of the Semantic Web is more complex than that of WWW. For instance, WWW can be considered as one huge class, i.e., a collection of Web pages, which has only a recursive property, i.e., a 'refers to' property corresponding to the hyperlinks. However, the Semantic Web encompasses various kinds of classes and properties, and consequently, ranking methods used in WWW should be modified to reflect the complexity of the information space in the Semantic Web.

Previous research addressed the ranking problem of query results retrieved from RDF knowledge bases. Mukherjea and Bamba modified Kleinberg's algorithm in order to apply their algorithm to rank the Semantic Web resources. They defined the objectivity score and the subjectivity score of a resource, which correspond to the authority score and the hub score of Kleinberg's, respectively. They concentrated on the diversity of properties and introduced property weights to control the influence of a resource on another resource depending on the characteristic of the property linking the two resources. A node with a high objectivity score becomes the object of many RDF triples, and a node with a high subjectivity score becomes the subject of many RDF triples. They developed several kinds of Semantic Web systems in order to validate their technique and showed some experimental results verifying the applicability of their method to the Semantic Web. Despite their efforts, however, there remained some limitations which they reported in their paper. First, their algorithm is useful only when a Semantic Web system represents most of the knowledge pertaining to a certain domain. In other words, the ratio of links to nodes should be high, or overall resources should be described in detail, to a certain degree for their algorithm to properly work. Second, a Tightly-Knit Community (TKC) effect, the phenomenon that pages which are less important but yet densely connected have higher scores than the ones that are more important but sparsely connected, remains as problematic. Third, a resource may have a high score, not because it is actually important, but simply because it is very common and as a consequence it has many links pointing to it.

In this paper, we examine such ranking problems from a novel perspective and propose a new algorithm which can solve the problems under the previous studies. Our proposed method is based on a class-oriented approach. In contrast to the predicate-oriented approach entertained by the previous research, a user, under our approach, determines the weights of a property by comparing its relative significance to the other properties when evaluating the importance of resources in a specific class. This approach stems from the idea that most queries are supposed to find resources belonging to the same class in the Semantic Web, which consists of many heterogeneous classes in RDF Schema. This approach closely reflects the way that people, in the real world, evaluate something, and will turn out to be superior to the predicate-oriented approach for the Semantic Web. Our proposed algorithm can resolve the TKC (Tightly Knit

Community) effect, and further can shed lights on other limitations posed by the previous research. In addition, we propose two ways to incorporate data-type properties which have not been employed even in the case when they have some significance on the resource importance. We designed an experiment to show the effectiveness of our proposed algorithm and the validity of ranking results, which was not tried ever in previous research. We also conducted a comprehensive mathematical analysis, which was overlooked in previous research. The mathematical analysis enabled us to simplify the calculation procedure. Finally, we summarize our experimental results and discuss further research issues.

**Keywords :** Ranking, Semantic Web, RDF Knowledge Base, Resource Importance, Ontology

## I. 서론

정보의 홍수 속에서 살아가는 우리들은 필요한 정보를 신속하고 정확하게 찾기 위해 검색엔진을 자주 사용한다. 하지만 검색된 결과 또한 너무나 많기 때문에 원하는 정보를 얻기 위해 여전히 많은 시간과 노력을 소모하고 있다. 이러한 문제를 해결하기 위해 검색된 결과를 사용자의 의도에 맞게 정렬하는 방법에 대한 많은 연구가 진행되어 왔고 앞으로도 이러한 연구의 중요성은 더욱 커질 것으로 보인다[Ding *et al.*, 2005].

독립된 문서들의 무한한 모임을 대상으로 하는 전통적인 검색 시스템에서는 주로 검색 키워드가 문서 안에서 발견되는 횟수에 의해 문서의 중요도가 결정되었다[Boyer and Moore, 1977; Knuth *et al.*, 1977]. 이 후 문서와 문서가 하이퍼링크로 연결된 월드와이드웹(World Wide Web)에서는 거대한 웹 그래프의 문서간 링크 구조를 분석하여 객관적인 중요도 점수를 산출하는 방법이 사용되었다. 대표적인 예로 1998년에 등장하여 주목을 받아 온 구글의 페이지랭크[Brin *et al.*, 1998; Page *et al.*, 1998; Haveliwala, 1999] 알고리즘을 들 수 있다. 페이지랭크에서는 임의의 페이지를 가리키는 다른 페이지들이 많을수록, 그리고 이러한 다른 페이지들의 중요도가 높을수록 해당 페이지의 중요도가 올라간다. 월드와이드웹에서 페이지의 중요도를 결정하는 다른

방법으로 Kleinberg의 HITS(Hypertext Induced Topic Selection) 알고리즘[Kleinberg, 1998]을 들 수 있다. Kleinberg는 권위(authority)와 허브(hub)라는 개념을 도입하여 웹 페이지의 중요도를 결정하는 방법을 제안하였다. 권위는 다른 페이지들에 의해 링크되는 정도를, 허브는 다른 페이지들을 링크하는 정도를 측정하기 위한 개념이다.

위의 예에서 알 수 있는 바와 같이 링크 구조를 분석하여 랭킹 점수로 활용하는 방법은 월드와이드웹의 만족도를 높여주는 핵심적인 역할을 해오고 있으며, 오늘날 많은 사람들이 이 방법의 효과성과 효율성에 대해 상당부분 인지하고 있다. 한편, RDF(Resource Description Framework) 데이터모델은 시맨틱 웹의 근간을 이루므로[Klyne *et al.*, 2004; Manola *et al.*, 2004; Berners-Lee, 2005] 시맨틱 웹의 거의 모든 정보는 RDF 그래프로 표현 가능하다. RDF 그래프는 자원(resource)을 노드(node)로, 속성(property or predicate)을 링크(link)로 나타내어 웹 페이지를 노드로, 문서간 하이퍼링크를 링크로 나타내는 웹 그래프와 비슷하다. 그러므로 월드와이드웹의 링크 구조 분석 기술을 시맨틱 웹의 RDF 그래프에 적용하는 것에 관한 연구는 매우 중요한 의의를 가진다.

그런데 웹 그래프는 이를 테면, '참조하다'라는 하이퍼링크를 가진 하나의 재귀적(recursive) 속성을 지닌 수많은 웹 페이지들로 이루어진 복잡하면서도 단순한 하나의 거대한 클래스라고 할

수 있다. 반면에 RDF 스키마(schema)에는 다양한 종류의 속성과 이질적인(heterogeneous) 클래스들이 존재할 수 있으므로 RDF 그래프는 사이즈가 작은 경우라도 월드와이드웹 그래프 보다 훨씬 복잡하다고 할 수 있다.

이러한 시맨틱 웹 속성의 다양성에 주목하여 HITS 알고리즘을 수정한 후 RDF 지식베이스에 대한 질의 결과를 랭킹하는 것에 적용한 기존 연구가 있다[Bamba and Mukherjea, 2004; Mukherjea et al., 2005]. 이들은 Kleinberg가 정의한 웹 페이지의 권위 점수와 허브 점수에 각각 대응되는 시맨틱 웹 자원의 목적부(objectivity) 점수와 주어부(subjectivity) 점수를 정의하였다. 또, 하나의 자원이 다른 자원에 미치는 영향력을, 이 두 자원을 연결하는 속성의 특성에 따라 조절하기 위해 목적부 가중치(weight)와 주어부 가중치라는 개념을 도입하였다. 그리고 실제로 몇 가지 시맨틱 웹 시스템을 구현하여 수정된 알고리즘의 적용 가능성을 입증하였고 실험 결과와 함께 이들 연구의 한계점을 보고하였다. 이 중 대표적인 것은 강한 결합 모임 현상(Tightly-Knit Community Effect)으로, 이것은 실제로는 별로 중요하지 않은 노드들인데도 이들 사이에 링크 연결이 많으면, 더 중요하지만 링크 연결이 적은 노드들 보다 높은 점수를 받는 문제점을 일컫는 말이다. 그리고 주어진 도메인에 대해 대부분의 지식이 기술되어 있는 지식베이스에서만 타당한 결과를 보여준다는 한계점을 안고 있다. 즉, 노드 수에 대한 링크 수의 비율이 너무 낮거나, 어떤 자원에 대해서는 자세히 기술하면서 몇몇 자원에 대해서는 정보가 아주 빈약한 경우에는 의외의 결과가 나올 수 있다는 것이다.

본 논문에서는 기존 연구와 근본적으로 시각을 달리하여, 속성이 아닌 클래스를 중심으로 가중치를 설정하는 방법에 대해 다룬다. 즉, RDF 스키마의 속성에서 클래스로 시각을 전환하여 클래스를 중심으로 자원 중요도 평가 기준을 설정하는 새로운 알고리즘을 제안한다. 이 방법은

랭킹을 요하는 대부분의 질의가 결국은 하나의 클래스에 속하는 자원들을 검색하는 것이라는 점과 RDF 스키마에는 다양한 클래스가 존재한다는 것, 그리고 클래스마다 다른 기준을 적용하는 사람들의 평가 방식 등을 고려하여 착안한 것이다. 새로운 방법은 강한 결합 모임 현상을 효과적으로 해결하는 것으로 보이며 다른 한계점에 대해서도 문제 해결에 대한 기준을 제시할 수 있을 것으로 예상된다. 그리고 추가적으로 기존 연구에서 제외되었던 데이터타입속성[Schneider et al., 2004]을 자원 중요도 계산에 포함시킬 수 있는 방안도 제시한다. 본 연구에서는 기존 연구에서 시도하지 않은 방법으로 랭킹 결과의 타당성을 검증하는 실험 방법을 디자인하여 구현하였으며, 행렬 연산에 대한 수학적 분석을 전개하여 계산 과정을 더욱 단순화하였다. 본 논문에서 제안하는 랭킹 알고리즘은 RDF 그래프로 표현되는 다양한 도메인에 적용할 수 있으므로 대학 랭킹이나 쇼핑몰 순위 등 실제 생활에서 접하게 되는 여러 가지 랭킹 이슈들에도 적용 가능할 것으로 기대된다.

본 논문의 구성은 다음과 같다. 먼저 제 I장에서는 기존 연구들을 본 논문의 알고리즘과 관련되는 핵심적인 부분만 간단하게 살펴보고 제 III장에서 클래스중심 가중치 설정 알고리즘의 기본 아이디어 및 계산 과정에 대해 설명한 후 이에 대한 분석을 전개할 것이다. 그리고 제 IV장에서는 제안된 알고리즘의 평가 방법 및 결과를 제시하고 제 V장에서 결론 및 향후 연구과제에 대해 언급할 것이다.

## II. 관련 연구

이 장에서는 기존의 웹 페이지 랭킹 알고리즘과 RDF 자원 랭킹 알고리즘에 대해 살펴본다.

### 2.1 월드와이드웹(WWW)의 페이지 랭킹 알고리즘

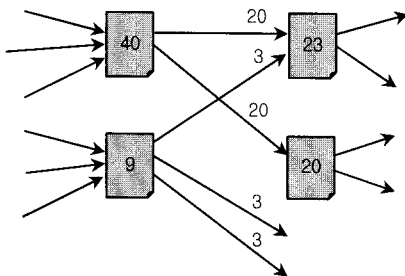
대표적인 웹 페이지 랭킹 알고리즘으로는 구

글의 페이지랭크와 Kleinberg의 HITS 알고리즘을 들 수 있다.

### 2.1.1 구글(Google)의 페이지랭크(PageRank)

앞에서 언급한 바와 같이 웹 문서를 노드(node)로, 문서들을 연결하는 하이퍼링크(hyperlink)는 노드들을 연결하는 링크로 생각하면, 월드와이드웹은 하나의 거대한 웹 그래프가 된다. 페이지랭크는 이렇게 월드와이드웹이 단순한 문서 집합과는 달리, 노드와 링크로 이루어진 그래프를 이룬다는 점에 착안하여 이러한 웹 그래프에 내재되어 있는 링크 구조로부터 객관적인 웹 문서의 중요도를 추출해내는 획기적인 방법이다. 페이지랭크의 기본 전제는 중요한 문서일수록 다른 문서에 의해 많은 추천을 받을 것이라는 것이며, 중요도가 높은 문서에 의해 많이 추천되는 문서일수록 중요도가 올라간다[Brin et al., 1998; Page et al., 1998]. 한 페이지의 중요도 점수는 이것이 가리키는 다른 페이지의 중요도 점수에 더해질 때 이 페이지에서 나가는 포워드(forward) 링크의 수로 나뉘어 균등하게 배분된다. 한 페이지의 중요도를 포워드 링크에 균등하게 배분하는 것은 페이지랭크가 웹 사용자의 서핑(surfing) 행동을 웹 그래프 상에서의 랜덤 워크(Random Walk) 프로세스로 모델링하고 있기 때문이다[Brin et al., 1998].

<그림 1>은 중요도가 전파되는 과정을 나타내



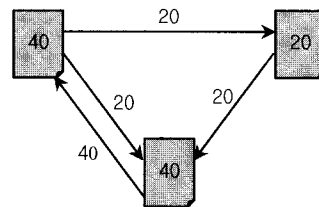
<그림 1> 단순 페이지랭크 계산

며, <그림 2>는 일단의 페이지에 대해 정상 상태 해(steady state solution)에 수렴한 모습을 보여준다.

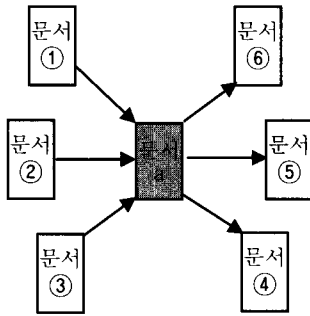
<그림 1>과 <그림 2>에서 알 수 있는 바와 같이 페이지랭크에서는 중요도가 링크의 방향을 따라 한 방향으로만 전파되며, 모든 링크에 대해 중요도의 반영 규칙이 동일하다. 즉, 모든 링크에 대해 동일하게, '참조되는' 문서가 '참조하는' 문서의 중요도를 받기만 하며, 이 때 전해 받는 중요도의 크기는 참조하는 문서의 중요도를 이 문서에서 나가는 링크 수로 나눈 만큼이다. 반면에 다음 절에서 설명할 HITS 알고리즘에서는 중요도가 양방향으로 흐르며, 중요도를 나누지 않고 더해준다.

### 2.1.2 Kleinberg의 HITS 알고리즘: 권위/허브 점수(Authority/Hub Score)

Kleinberg는 웹 페이지에 대해 권위 점수(authority score)와 허브 점수(hub score)라는 두 가지 유형의 점수를 정의하고 이를 기반으로 웹 페이지를 정렬하는 방법을 제안하였다[Kleinberg, 1998]. 웹 페이지는 많은 좋은 사이트에 의해 링크될수록 더 높은 권위 점수를 가지며, 많은 좋은 사이트를 가리킬수록 더 높은 허브 점수를 갖는다. 한 문서의 권위 점수는 그 문서를 링크하는 모든 웹 페이지의 허브 점수의 합이고, 허브 점수는 그 문서가 링크하는 모든 웹 페이지의 권위 점수의 합



<그림 2> 정상 상태 해(A Steady State Solution)



<그림 3> Kleinberg의 권위/허브 점수 계산

이다. 예를 들어, <그림 3>에서 문서 a의 권위점수는 문서 a를 가리키는 모든 문서, 즉 문서 ①, ②, ③의 허브점수를 더하여 계산하고, 문서 a의 허브점수는 문서 a가 가리키는 모든 문서, 즉 문서 ④, ⑤, ⑥의 권위점수를 그대로 더하여 계산한다.

이와 같이 Kleinberg의 방법은 링크로 연결된 쌍의 문서에 대해 추천하는 문서의 허브 점수와 추천되는 문서의 권위 점수를 서로에게 더해 주는 방식으로 모든 링크에 대해 동일하게 적용된다.

## 2.2 시맨틱 웹의 자원 랭킹 알고리즘

시맨틱 웹의 정보공간은 월드와이드웹에 비해 훨씬 복잡하고 이에 따라 시맨틱 웹에서는 랭킹의 대상도 다양하다. 이 중 링크 구조와 관련이 있는 대표적인 기존 연구들을 랭킹의 대상에 따라 정리해 보면 다음과 같다.

첫째, 본 논문에서 다루고 있는 RDF와 RDF 스키마로 이루어진 지식 베이스에 대한 질의 결과를 자원 중요도에 따라 랭킹하는 것에 관한 연구 [Bamba and Mukherjea, 2004; Mukherjea et al., 2005]와 F-Logic 지식베이스에 대한 질의 결과를 지식베이스의 유사도(similarity)에 따라 랭킹하는 것에 관한 연구 [Maedche et al., 2001; Maedche and Staab, 2002] 등이 있다.

두 번째는 의미적 관계(Semantic association)

의 랭킹에 관한 연구이다. RDF 그래프 상에 있는 임의의 두 자원은 직접적으로 연결되지 않더라도 이들 두 자원 사이에 존재하는 다른 자원과 링크를 통해 간접적으로 연결될 수 있는데, 이렇게 두 자원 사이에 형성되는 모든 패스(path)들을 의미적 관계라고 한다. 수많은 의미적 관계들을 사용자의 관심 영역이나 패스 길이에 대한 선호도 등에 따라 랭킹하는 방법에 관한 연구 [Alman-Meza et al., 2003; Halaschek et al., 2004; Anyanwu et al., 2005; Sheth et al., 2005; Alman-Meza et al., 2005]가 진행되고 있다.

마지막으로 페이지랭크의 응용으로, 문서 간 링크 유형에 따라 다른 선택 확률을 적용하여 시맨틱 웹의 문서 중요도를 결정하는 것에 관한 연구 [Ding et al., 2004; Ding et al., 2005; Finin and Ding, 2006]를 들 수 있다. 시맨틱 웹 문서(Semantic Web Document)는 RDF 그래프를 RDF 신택스 언어인 RDF/XML, N-Triple, N3 중 하나로 표현한 웹 페이지를 말한다.

위의 세 가지 연구 영역은 모두 링크 구조를 기반으로 하고 있기 때문에 서로 무관하지 않다. 예를 들어, 하나의 RDF 자원이 얼마만큼 중요한지가 결정되면 이 정보는 의미적 관계의 랭킹에도 이용될 수 있을 것이다. 하나의 의미적 관계 패스 상에 존재하는 모든 자원들의 중요도가 높을수록 그 패스의 중요도는 높다고 할 수 있기 때문이다.

### 2.2.1 RDF와 RDF 스키마, 그리고 RDF 질의 언어

RDF(Resource Description Framework)는 표현하고자 하는 개념을 자원(resource)으로 보고 이러한 자원을 서로 구별하기 위한 식별자로 URIref (Uniform Resource Identifier reference)를 사용하여 자원의 속성이나 자원과 자원 간의 관계를 기술하는 데이터 모델이다 [Klyne et al., 2004; Manola et al., 2004; 노상규와 박진수, 2007]. RDF

의 기본 단위는 '주어부(subject)-서술부(predicate or property)-목적부(object)'의 세 부분으로 이루어져 흔히 트리플(triple)이라 불리는 서술문(statement)이다. RDF 서술문은 노드와 링크로 이루어지는 RDF 그래프로도 표현된다. 노드는 서술문의 주어부와 목적부에 자원이 오는 경우에 해당되며 URIref를 포함하는 타원으로 나타난다. 목적부에 문자열 데이터가 올 때에는 직사각형 안에 그 내용을 기록한다. RDF 그래프의 링크는 서술문의 서술부에 해당되며 주어부에서 목적부로 향하는 화살표와 서술부의 개념에 대한 URIref를 화살표 옆에 명시한다.

RDF에는 공통적인 속성을 갖는 자원들을 클래스로 분류하고 이러한 클래스간의 위계구조를 설정하거나 속성의 주어부나 목적부가 될 수 있는 클래스를 제한할 수 있는 기능이 없다. 이러한 기능은 RDF를 프레임기반으로 확장한 RDF Schema에서 제공되므로 RDF Schema를 이용하면 도메인의 구성 및 클래스간 상호작용을 묘사하는데 필요한 기본 어휘와 가정들을 정의할 수 있다 [Brickley et al., 2004].

RDF자원은 SPARQL, RQL, TRIPLE, RDQL과 같은 RDF 질의어를 이용하여 검색할 수 있다. 이 중 SPARQL은 현재 W3C 권고 후보안(W3C Candidate Recommendation) [Prud'hommeaux and Seaborne, 2007]이 발표된 상태이며 2005년에 팀 버너스리가 두 번째로 발표한 시맨틱 웹 계층에도 포함되어 있다[Berners-Lee, 2005]. 그런데 아직까지, 이전의 다른 RDF 질의 언어처럼, SPAQL도 결과 데이터 셋에 대한 텍스트 위주의 정렬기능만 제공할 뿐 자원을 실질적인 중요도에 따라 랭크해서 보여주지는 않는다.

### 2.2.2 시맨틱 웹의 자원 중요도: 목적부/주어부 점수(Objectivity/Subjectivity Score)

Mukherjea and Bamba[2004]는 중요한 웹 페

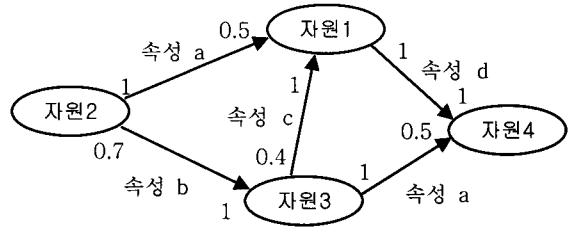
이지를 찾아내기 위해 효과적으로 사용되어온 월드와이드웹 링크 분석 기술을 수정하여, 시맨틱 웹에 대한 질의 결과를 자원 중요도에 따라 랭킹하는 데에 적용하였다. 이들은 이 랭킹 알고리즘을 생의학 특허(biomedical patent) 관리를 위한 시맨틱 웹 시스템인 BioPatentMiner에 적용하여 실용성을 입증하였다.

이들이 자원 중요도 계산에 사용한 목적부 점수(objectivity score)와 주어부 점수(subjectivity score)는 각각 Kleinberg의 권위 점수와 허브 점수에 해당한다. 이들은 RDF 그래프를 이루고 있는 트리플의 속성이 다양하며, 양방향으로 똑같이 더해지는 안 되는 경우가 있다는 사실에 주목하여 속성의 목적부와 주어부 가중치(weight) 개념을 도입하였다. 예를 들어, <그림 4>에서와 같이 (발명가, 발명하다, 특허)라는 트리플이 있을 때, '발명하다'의 주어부 가중치는 1이고 목적부 가중치는 0.4이다. 이들은 발명가의 중요도는 특허가 많을수록 올라가지만 특허의 중요도는 발명가가 많다고 하여 그대로 증가하는 것은 아니라고 생각하였다. 그래서 발명가의 주어부 점수는 특허의 목적부 점수를 그대로 더하여 계산하지만 특허의 목적부 점수는 특허를 발명한 발명가의 주어부 점수에 1보다 작은 양의 가중치인 0.4를 곱하여 계산한다. 이렇게 하면 하나의 특허를 발명한 발명가들이 여러 명 존재할 경우 단순히 발명가 수가 많아서 해당 특허의 목적부 점수가 높아지는 문제점을 완화할 수 있기 때문이다.

이와 같이 각각의 속성에 대해 속성의 특성을 고려하여 목적부와 주어부 가중치를 설정하면 전체적인 RDF 그래프를 이루고 있는 속성 링크들의 가중치가 결정된다. 처음에 가중치를 설정하는 기준은 일단 양방향으로 중요도의 흐름이 같으면 목적부와 주어부 가중치가 모두 1이고, 중요도의 흐름이 다를 때에는 큰 쪽을 1로 보고 이와 비교하여 작은 쪽의 가중치를 결정하게 된다. 이 때 가해지는 비율은 사용자나 컨텍스트에 따라 경험적으로 추정된다. 그러므로 임의의 속



<그림 4> '발명하다'의 가중치 설정 예



<그림 5> 속성중심 가중치 설정 예

성은 <그림 5>와 같이 목적부와 주어부 가중치가 모두 1이든지, 두 가지 가중치 중 하나는 1이고, 다른 하나는 1보다 작은 양수를 갖게 된다. 요약하면, 자원  $r$ 의 목적부 점수는 자원  $r$ 을 목적부로 갖는 모든 자원의 주어부 점수에 해당 링크에 대한 속성의 목적부 가중치를 곱하여 합산하고, 자원  $r$ 의 주어부 점수는 자원  $r$ 을 주어부로 갖는 모든 자원의 목적부 점수에 해당 링크의 주어부 가중치를 곱하여 합산한다.

Mukherjea와 Bamba[2005]는 UMLS(Unified Medical Language System), Biomedical Patent, TAP 시맨틱 웹에 이들의 자원 중요도 계산 알고리즘을 적용해보고 다음과 같은 한계점을 보고하였다. 첫째는 도메인에 관련된 대부분의 지식이 표현된 시맨틱 웹에 대해서만 이 알고리즘이 유용하다는 것이다. 다시 말해 한 클래스에 대해서는 자세히 기술하면서 다른 클래스에 대해서는 중요한 정보들이 빠져 있든지, 전체적으로 그래프의 링크 연결도가 너무 낮아 목적부나 주어부 자원의 수에 비해 링크의 수가 너무 적으면 적용하기 힘들다는 것이다. 둘째는 별로 중요하지 않은 노드들이, 이들 사이에 링크 연결이 많으면 현실세계에서 이들보다 더 중요한 다른 노드들 보다 높은 점수를 받게 되는 강한 결합 모임 현상(Tightly-Knit Community Effect)이다. 이러한 현상은 전체적인 그래프의 링크 연결도에 비해 강한 결합을 형성하는 노드들의 링크 연결도가 상대적으로 높을수록 더욱 강하게 나타난다. 셋째는 어떤 자원이 실제로 중요해서가 아니라 매우 흔하기 때문에 이 자원에 대한 링크 연결이 많아 높은 점수를 받게 되는 경우가 있다는

것이다.

### Ⅲ. 클래스 중심 가중치 설정 알고리즘

#### 3.1 기본 아이디어

##### 3.1.1 속성 중심에서 클래스 중심으로

Mukherjea and Bamba의 연구에서는 모든 속성의 디폴트 가중치를 1로 설정한 후, 속성에 따라 중요도의 흐름이 양방향으로 비슷하면 목적부와 주어부 가중치를 모두 1로 유지하고, 같지 않으면 무조건 큰 쪽을 1로 고정하고 이에 대한 상대적인 비율을 가늠하여 다른 쪽의 가중치를 1보다 작은 양수로 변경하였다. 이렇게 중요도의 흐름 방향을 우선적으로 고려하여 속성 중심으로 가중치를 설정했을 때 몇 가지 한계점이 발생함을 앞에서 살펴 보았다. 본 논문에서는 이러한 문제점을 해결하기 위하여 속성이 아닌 클래스 중심으로 가중치를 설정하는 방식을 제안하며 이에 대한 근거는 다음과 같다.

첫째, RDF 지식베이스에 대한 대부분의 질의는 그 결과물로 같은 클래스에 속하는 수많은 자원들을 보여주게 되며, 랭킹은 결국 이들을 대상으로 수행된다. 그런데 월드와이드웹에서는 모든 웹 페이지가 하나의 클래스에 속한다고 할 수 있으므로 언제나 한가지 기준으로 랭킹하는 것이 타당했지만, 시맨틱 웹에서는 이질적인 수많은 클래스가 존재하므로 각 클래스마다 다른 평가 기준을 적용해야 한다는 것이다.





<그림 6> 클래스에 따라 달라지는 속성 가중치 예

둘째, 클래스 중심으로 가중치를 설정하면 앞에서 제시한 Mukherjea and Bamba 연구의 첫 번째 한계점인 정보 표현의 완전성 여부를 비교적 쉽게 평가할 수 있을 것이다. 어떤 클래스에 대한 정보가 완벽하게 기술되어 있다는 것은 이 클래스에 속하는 자원의 중요도를 평가하는데 필요한 속성들이 모두 기록되어 있다는 말과 같은 의미일 것이기 때문이다. 속성 중심 방법으로는 대략 어떤 클래스에 연결된 링크 수가 많으니 완전할 가능성이 높다고 짐작하겠지만 이중에는 중요도와는 거리가 먼 속성 링크들이 많을 수도 있다.

셋째, 속성 중심으로 가중치를 설정하면 중요도가 흘러가는 방향만을 비교하기 때문에 링크로 연결된 두 자원 간의 상호작용이 똑같이 강하거나 약한 경우의 차이를 표현할 수 없다. 그리고 같은 속성이라도 연결되는 클래스에 따라 가중치를 다르게 설정해야 하는 경우가 있다. 다음 <그림 6>은 똑같은 속성 '졸업하다'가 차지하는 비중이 교수 클래스에 속하는 자원과 사업가 클래스에 속하는 자원의 중요도를 계산할 때 서로 다를 수 있음을 보여준다. 대부분의 경우 학력은 사업가 보다는 교수에게 더욱 중요한 영향을 미칠 수 있기 때문이다.

이러한 생각을 확장하면 한 클래스에 속하는 자원의 중요도에 영향을 미치는 속성들의 상대적인 비중을 고려하여 클래스를 중심으로 해당 클래스와 연결된 속성들의 상대적인 가중치를 설정할 수 있을 것이다. 그리고 어떤 클래스에 연결된 속성이 그 클래스에 속하는 자원의 중요도와는 무관하다면 해당 속성에 '0'의 가중치를 설정하여, 링크가 있더라도 관계없는 영향력을 제거할 수 있어 강한 결합 모임 현상은 거의 해결

될 것으로 예상된다.

넷째, 이렇게 클래스 중심으로 가중치를 설정하는 방식은 각 항목, 즉 속성에 대해 다른 가중치를 부과하여 합산하는 사람들의 평가 방식과 유사하다<sup>1)</sup>. 물론 그 동안 각종 평가에 사용되어 온 대부분의 속성은 링크연결과 무관한 데이터 타입속성이라 할 수 있지만, 항목에 따라서는 양적인 것 보다는 질적인 평가도 반영하는 좀더 객관적인 방식이 요구되고 있다[Ren and Taylor, 2007]. 예를 들어, 연구원 평가 시에 고려되는 발표 논문 항목에 대해 논문이 실린 저널에 따라 일률적으로 같은 점수를 부과하는 것 보다는, 논문들 사이의 링크 연결 구조를 분석하여 다른 논문들에 의해 인용되는 정도를 반영하는 실질적인 중요도를 산출하는 방식이 좀더 객관적이라 할 수 있다. 그러므로 자원들 사이의 링크 연결 구조를 파악할 수 있는 시맨틱 웹에서는 기본적으로 기존 평가 방식을 유지하되 각 항목에 대한

1) 이러한 주장에 대한 근거는 인터넷을 통해 쉽게 확인이 가능하다(Academic Ranking of World Universtie: <http://ed.sjtu.edu.cn/rank/2006/ARWU2006Methodology.ht>, U.S. News and World Reportrankin:[http://colleges.usnews.rankingsandreviews.com/usnews/edu/college/rankings/about/weight\\_brief.php](http://colleges.usnews.rankingsandreviews.com/usnews/edu/college/rankings/about/weight_brief.php), National Research Council ranking of U.S. doctoral program:[http://www7.nationalacademies.org/resdoc/ranking\\_and\\_ratings.htm](http://www7.nationalacademies.org/resdoc/ranking_and_ratings.htm), FIFA's Ranking Criteri:<http://forums.soccerfansnetwork.com/showthread.php?t=42680>, 2008학년도 서울대학 신입생 입학전형안: [http://admission.snu.ac.kr/info/gong/1206737\\_3221.htm](http://admission.snu.ac.kr/info/gong/1206737_3221.htm), 2008학년도 서울과학기술대학교 입학 전형 요강: <http://www.sshs.hs.kr/doum/index.js>, 2008 학년도 대원외국어고등학교 신입생 전형 요강:[http://daewon.seoul.kr/dflhs/daefh\\_m20/daefh\\_ms21.as](http://daewon.seoul.kr/dflhs/daefh_m20/daefh_ms21.as)).

질적인 평가를 고려하는 방식이 이상적이라 생각된다.

다섯째, HITS 알고리즘에서는 웹 문서들의 권위 점수나 허브 점수를 따로따로 비교했지만 시맨틱 웹에서는 RDF Schema의 다양성으로 인해 두 점수를 합한 최종 점수로 자원들의 중요도를 비교하는 것이 합리적이다. Mukherjea와 Bamba의 연구에서도 이 점에 대해 언급했지만 두 점수를 임의적인 비율로 더하거나[Bamba and Mukherjea, 2004], 두 점수를 따로 비교하였다[Mukherjea et al., 2005]. 클래스 중심으로 가중치를 설정하면 속성들의 상대적인 비중에 따라 목적부와 주어부 가중치가 결정되고 결과적으로 두 점수를 합하기만 하면 합리적으로 정해진 두 점수의 합산 비율이 자동적으로 반영된다.

클래스 중심으로 가중치를 설정하는 것에 대한 예는 4.1절의 <그림 12>와 <그림 13> 및 [부록]의 <표 1>과 <표 2>를 참조하기 바란다.

### 3.1.2 데이터타입속성의 반영

기존 연구에서는 데이터타입속성(owl:Datatype Property)[Schneider et al., 2004]에 해당되는 링크는 RDF 그래프에서 제거하고 객체속성(owl:ObjectProperty)만으로 중요도를 계산하였다. 이것은 데이터타입속성의 목적부에 오는 값이 자원이 아닌 단순 데이터이고, 이 값에 연결되는 다른 링크가 없기 때문일 것이다. 그런데 어떤 데이터타입속성은 해당 클래스에 속하는 자원의 중요도에 많은 영향력을 미칠 수 있다. <그림 7>에서와 같이 책의 중요도에는 판매부수가, 출판사의 중요도에는 매출액이 큰 비중을 차지할 수 있

는 바와 같다.

이러한 데이터타입속성을 고려할 수 있는 방법에는 크게 두 가지가 있을 것이다. 첫째는 객체속성만을 고려하여 계산된 자원의 중요도를 정규화한 점수와 유의미한 데이터타입속성 값을 정규화하여 얻은 점수를 스키마에서 정해진 비율에 따라 더해주는 방법이다. <그림 7>의 여러 책에 대한 판매부수 값을 0에서 1사이의 값을 갖도록 정규화할 수 있고, 링크 분석 방법으로 구해진 자원 중요도도 0에서 1사이의 값으로 정규화하여 책 클래스의 속성 가중치대로 합산할 수 있을 것이다. 이 방법은 링크 분석이 끝난 후 사후적으로 데이터타입속성 점수를 더해줌으로써 데이터타입속성이 이와 직접적으로 연결된 자원의 중요도에만 영향을 주고, 해당 자원과 연결된 다른 자원들과의 상호작용에서는 제외된다는 단점이 있다.

둘째는 데이터타입속성 값에 해당하는 더미(dummy) 자원을 만들어 링크 분석 계산에 처음부터 포함시키는 것이다. 이것은 데이터타입속성 값을 0에서 1사이의 값으로 정규화하여 이 값에 비례하도록 해당 링크의 가중치를 설정하면 가능해진다. 예를 들어, <그림 7>에서 책a의 판매부수에 대한 속성값 1000을 하나의 더미 자원으로 간주하고, 1000이란 값을 책 클래스에 있는 다른 책들의 판매부수 값과 비교하여 얻은 정규화 점수에 비례하도록 책a에 연결된 판매부수 링크의 주어부 가중치를 설정할 수 있을 것이다.

### 3.2 자원 중요도 계산 알고리즘과 행렬 연산

이 절에서는 앞에서 설명한 기본 아이디어들



<그림 7> 자원 중요도에 반영되어야 할 데이터타입 속성 예

을 반영하는 자원 중요도 계산 알고리즘을 제시한다. RDF 스키마에 의해 설계된 시맨틱 웹의 정보공간을 클래스  $C$ , 속성  $P$ , 클래스에 속하는 인스턴스 자원들  $IR$ , 그리고 자원이 아닌 문자열 데이터들  $SD$ 로 이루어진  $(C, P, IR, SD)$ 로 정의해보자. 여기에서 두 가지의 인스턴스 그래프를 추출해 낼 수 있다. 하나는 객체속성들 중에서 주어부와 목적부에 모두  $IR$ 에 속하는 자원들이 오는 속성 링크들만 포함시킨 그래프인  $instance\_Graph$ 이고, 다른 하나는  $instance\_Graph$ 에 주어부에는  $IR$ 에 속하는 자원이 오고 목적부에는  $SD$ 에 속하는 데이터가 오는 속성 링크까지 포함시킨  $instance\_data\_Graph$ 이다. 즉,  $instance\_Graph$ 에서는  $IR$ 에 속하는 자원들이 노드가 되고, 이 자원들 사이에 형성되는 관계들이 링크로 표현된다. 또,  $instance\_data\_Graph$ 는 유의미한 데이터타입속성에 연결된 데이터를 더미(dummy) 노드로 보아 이들과 이들로 향하는 링크들도 포함되도록  $instance\_Graph$ 를 확장한 것이다.

### 3.2.1 이론적인 자원 중요도 계산 알고리즘

#### (1) 클래스별 속성 가중치의 설정

먼저  $instance\_Graph$ 를 위한 가중치 설정식은 다음 식 (1)과 같다. RDF Schema 상에서 클래스별로 이 클래스와 연결된 속성들의 상대적인 비중을 고려하여 목적부와 주어부 가중치를 설정한다. 식 (1)은 클래스  $C$ 의 가중치 설정 조건을 나타내며,  $objWt_{(D,C)}$ 는 도메인(domain)이 클래스  $D$ 이고 레인지(range)가 클래스  $C$ 인 속성의 목적부 가중치이고,  $subjWt_{(C,D)}$ 는 도메인이 클래스  $C$ 이고 레인지가  $D$ 인 속성의 주어부 가중치이다. 뒤에 나올 인스턴스와 구별하기 위하여 클래스는 대문자로 표기하였다.

$$\sum_D objWt_{(D,C)} + \sum_D subjWt_{(C,D)} = 1 \quad (1)$$

다음으로  $instance\_data\_Graph$ 를 위한 가중치 설정식은 다음 식 (2)와 같다.  $dpWt_q$ 는  $C$ 에 연결된 데이터타입 속성  $q$ 에 대한 주어부 가중치이다. 모든  $q$ 에 대해  $dpWt_q = 0$ 이면 식 (2)는 식 (1)과 같아진다.

$$\sum_D objWt_{(D,C)} + \sum_D subjWt_{(C,D)} + \sum_q dpWt_q = 1 \quad (2)$$

#### (2) 반복적인 중요도 계산 단계

본 절에서는  $instance\_Graph$ 에 대한 알고리즘 위주로 설명할 것이며  $instance\_data\_Graph$ 에 대한 알고리즘은 이와 유사하므로 생략한다.  $instance\_Graph$   $G=(V, E)$ 에서  $V$ 는  $N$ 개의 자원으로 이루어진 자원 집합  $V = \{1, 2, \dots, N\}$ 이고,  $E$ 는  $V$ 안에 존재하는 임의의 자원  $r$  ( $1 \leq r \leq N$ )과  $k$  ( $1 \leq k \leq N$ )를 연결하는 방향성 있는 링크의 집합이라 하자.

스키마 상에서 설정된 속성들의 가중치에 따라 링크 집합  $E$ 에 포함되어 있는 각 속성의 인스턴스 링크에 대해 양방향 가중치를 설정한다. 이 후 수행될 계산과정에서는 이론적으로 다음과 같은 알고리즘을 가정한다. 모든 자원  $r$  ( $1 \leq r \leq N$ )에 대해  $o^r$ 은 자원  $r$ 의 목적부 점수이고  $s^r$ 은 자원  $r$ 의 주어부 점수이며, 위 첨자가 없는  $o$ 와  $s$ 는 각각 모든 자원의 목적부와 주어부 점수를 나타내는  $(N \times 1)$  벡터이다.

① 초기화 :  $o^r = s^r = 1, (1 \leq r \leq N)$ .

② 이터레이션(iteration) : 목적부와 주어부 점수 벡터  $o$ 와  $s$ 가 수렴할 때까지 다음을 반복한다. 다음은  $i$ 번째 이터레이션을 나타낸다.

a. 모든 자원  $r$ 에 대해 다음 식 (3)을 계산한다.  $objWt_{(k,r)}$ 는 자원  $k$ 에서 자원  $r$ 로 향하는 링크  $(k, r)$ 에 대한 속성의 목적부 가중치이다.

$$o_i^{*r} = \sum_{(k,r) \in E} s_{i-1}^k \times objWt_{(k,r)} \quad (3)$$

- b. 모든 자원  $r$ 에 대해 다음 식 (4)를 계산한다.  $subjWt_{(r,k)}$ 는 자원  $r$ 에서 자원  $k$ 로 향하는 링크  $(r, k)$ 에 대한 속성의 주어부 가중치이다.

$$s_i^{*r} = \sum_{(r,k) \in E} o_i^{*k} \times subjWt_{(r,k)} \quad (4)$$

- c.  $o_i^*$ 와  $s_i^*$ 를 정규화하여,  $o_i$ 와  $s_i$ 를 얻는다. 정규화 조건은 다음과 같다.

$$\sum_r (o_i^r)^2 = 1, \sum_r (s_i^r)^2 = 1$$

- ③ 앞에서 구한 목적부와 주어부 점수 벡터  $o$ 와  $s$ 를 더하여 최종 랭킹 벡터를 구한다.

위의 계산 알고리즘은 Kleinberg[1998] 및 Mukherjea et al.[2005]가 사용한 반복적인 알고리즘을 클래스중심으로 설정된 링크 가중치의 변화를 반영하여 수정한 것이다. Kleinberg[1998]의 경우에는 가중치 개념이 아예 없으므로 식 (3)과 식 (4)에서 가중치가 모두 1인 셈이고, Mukherjea et al.[2005]의 연구에서는 목적부나 주어부 가중치 중 적어도 하나는 1이고, 다른 한 쪽은 1이거나 1보다 작은 양수가 된다. 그리고 본 연구에서처럼 클래스 중심 접근 방법을 사용하면 식 (1)의 제한 조건 하에 두 가중치 모두 0에서 1사이의 값을 자유롭게 갖게 된다.

### 3.2.2 중요도 계산 알고리즘의 수렴(convergence)성 분석

앞 절에서 제시한 바와 같은 반복적인 알고리즘은 각 단계에서 구해지는 벡터들이 일정한 방향으로 수렴하는 성질에 기초하고 있다. 벡터의 방향이 정해지면 각 자원의 점수를 나타내는 벡

터 성분들의 순위가 더 이상 바뀌지 않으므로 최종 벡터가 자원의 랭킹 용도로 사용될 수 있는 것이다.

유일한 제일 고유값(dominant eigenvalue)이 존재하고 대각화 가능한(diagonalizable) 행렬을  $M$ 이라 하고,  $M$ 의 제일 고유벡터(dominant eigenvector)와 수직(orthogonal)이 아닌 임의의 벡터를  $z$ 라 할 때,  $M^i z$ 는 차수  $i$ 가 증가함에 따라  $M$ 의 제일 고유벡터 방향으로 수렴하는 성질이 있다[Burden and Faires, 2001; Ehrlich, 1969]. 본 논문에서는 설명의 편의를 위해 이러한 성질을 '행렬수렴성 1'이라고 명명하도록 하자.

그리고 유일한 제일 고유값이 존재하고 대각화 가능하지 않은 행렬을  $M$ 이라 하고,  $M$ 의 제일 고유값에 대한 고유벡터와 일반화된 고유벡터(generalized eigenvector)들의 부공간(subspace)에 수직(orthogonal)이 아닌 임의의 벡터를  $z$ 라 할 때,  $M^i z$ 는 차수  $i$ 가 증가함에 따라  $M$ 의 제일 고유벡터 방향으로 수렴하는 성질이 있다[Ehrlich, 1969]. 본 논문에서는 이러한 성질을 '행렬수렴성 2'라고 명명하도록 하자.

또, 비음(nonnegative)이고 프리머티브(primitive)한 행렬을  $A$ 라 하면  $A$ 에 대해 유일한 양(positive)의 제일 고유값이 존재한다는 Perron-Frobenius 정리가 있다[Perron and Frobenius, 2007].

앞 절의 알고리즘에서  $N$ 개의 자원에 대한 식 (3)과 식 (4)의 계산은 행렬을 사용하여 나타낼 수 있다. 먼저 링크의 목적부 가중치를 나타내는 행렬은  $B$ 로, 주어부 가중치를 나타내는 행렬은  $C$ 로 표기하면 각각 다음과 같이 정의된다.

$$B_{rk} = objWt_{(k,r)}, C_{rk} = subjWt_{(r,k)}$$

이제 식 (3)과 식 (4)를 행렬식으로 나타내면,  $o_i^* = B s_{i-1}^*$ ,  $s_i^* = C o_i^*$ 이다.  $i = 1$ 일 때에는  $o_1^* = B s_0^*$ ,  $s_1^* = C o_1^* = C B s_0^*$ 이고, 정규화 과정에서 곱해지는 상수를  $n_1$ 이라 하면  $s_1 = n_1 C B s_0$ 이다. 다시  $i = 2$ 일

때에는  $o_i^* = Bs_1 = n_1 BCBs_0$ ,  $s_2^* = Co_2^* = n_1 CCBs_0 = n_1 (CB)^2 s_0$  이고, 정규화 과정에서 곱해지는 상수를  $n_2$  라 하면  $s_2 = n_1 n_2 (CB)^2 s_0$ 이다. 이런 식으로 처음부터  $i$  번째까지의 이터레이션을 고려하면 결국 목적부 점수 벡터  $o_i$ 는  $(BC)^{i-1} Bs_0$  방향, 주어부 점수 벡터  $s_i$ 는  $(CB)^i s_0$  방향으로의 단위벡터가 된다.  $B$ 와  $C$ 는 비음 행렬이므로  $BC$ 와  $CB$ 도 비음 행렬이며 이 행렬들은 대부분의 그래프 응용문제에서처럼 프리머티브(primitive)하다고 가정할 수 있다. 그러면 Perron-Frobenius 정리에 의해  $BC$ 와  $CB$ 는 유일한 양의 제일 고유값을 가진다. 그러므로 앞의  $(BC)^{i-1} Bs_0$ 와  $(CB)^i s_0$ 에 행렬 수렴성 1과 2를 적용하면 초기해  $s_0$ 를 각각의 조건에 맞도록 설정했을 때 목적부 점수 벡터는  $BC$ 의 제일 단위 고유벡터에, 주어부 점수 벡터는  $CB$ 의 제일 단위 고유벡터에 해당됨을 알 수 있다.

### 3.2.3 클래스 중심 속성 가중치 행렬 예

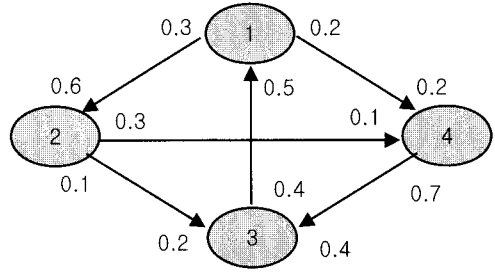
행렬 구성에 대한 아주 간단한 예로 클래스 구성과 속성 가중치가 <그림 8>과 같은 도메인이 있고, 각 클래스에 속하는 인스턴스가 하나씩만 있다고 가정하자. 식 (3)의 목적부 점수 계산은 <그림 9>와 같이, 식 (4)의 주어부 점수 계산은 <그림 10>과 같이 이루어진다.  $o_i^*$ 와  $s_i^*$ 는 각각 정규화하기 전의 목적부 점수 벡터와 주어부 점수 벡터이다. 정규화는 점수 벡터의 크기가 1이 되도록 크기만 조정하는 것이므로 벡터의 방향

$$o_i^* = Bs_{i-1}$$

$$\begin{bmatrix} o_i^{*1} \\ o_i^{*2} \\ o_i^{*3} \\ o_i^{*4} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.5 & 0 \\ 0.6 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0.4 \\ 0.2 & 0.1 & 0 & 0 \end{bmatrix} \begin{bmatrix} s_{i-1}^1 \\ s_{i-1}^2 \\ s_{i-1}^3 \\ s_{i-1}^4 \end{bmatrix}$$

<그림 9> 목적부 점수의 행렬 계산

성에는 영향을 주지 않는다.



<그림 8> 클래스 중심 가중치 설정 예

## 3.3 자원 중요도 계산 알고리즘의 구현

본 논문에서는 3.2.1절에서 제시한 반복적인 알고리즘을 사용하지 않고 3.2.2절의 분석 결과를 바탕으로 다음과 같은 자원 중요도 계산 알고리즘을 구현하였다.

### 3.3.1 실제적인 구현 알고리즘

- ① 온톨로지 스키마 상에서 클래스별로 속성에 대한 목적부 가중치와 주어부 가중치를 설정한다.
- ② 각 클래스의 모든 인스턴스(자원)를 노드로 하는 RDF 그래프에 대해 목적부 가중치 행렬  $B$ 와 주어부 가중치 행렬  $C$ 를 만든다.
- ③ 행렬  $BC$ 의 제일 고유벡터를 계산하여 목적부 점수 벡터를 구하고, 행렬  $CB$ 의 제일 고유벡터를 계산하여 주어부 점수 벡터를

$$s_i^* = Co_i^*$$

$$\begin{bmatrix} s_i^{*1} \\ s_i^{*2} \\ s_i^{*3} \\ s_i^{*4} \end{bmatrix} = \begin{bmatrix} 0 & 0.3 & 0 & 0.2 \\ 0 & 0 & 0.1 & 0.3 \\ 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \end{bmatrix} \begin{bmatrix} o_i^{*1} \\ o_i^{*2} \\ o_i^{*3} \\ o_i^{*4} \end{bmatrix}$$

<그림 10> 주어부 점수의 행렬 계산

구한다.

- ④ 목적부 점수 벡터와 주어부 점수 벡터를 합하여 최종 랭크 벡터를 구한다.

위의 첫 번째 단계는 식 (1)이나 식 (2)의 조건 하에 이루어지고 속성의 가중치는 주어진 컨텍스트나 사용자의 관점에 따라 달라질 수 있다. 만일, 공적인 기준을 적용해야 하는 객관성이 중요시 되는 경우라면 클래스별 중요 속성과 이에 대한 가중치의 결정 작업을 심도 있게 다루어야 할 것이다. National Research Council[2003]의 방법을 응용하여 도메인 전문가들의 클래스 인스턴스에 대한 질적인 평가 점수를 종속변수로, 요인 분석(factor analysis)을 통해 추출한 측정 가능한 항목들을 독립변수로 하는 회귀분석의 계수들(regression coefficients)을 가중치로 사용하는 방법을 생각해 볼 수 있지만 이에 대한 내용은 본 연구의 범위를 벗어나므로 생략하고, 일단 개념성의 원칙에 따라 설정하는 것으로 한다.

### 3.3.2 데이터타입속성의 정규화 및 가중치 변환

클래스 수준에서 설정된 데이터타입 속성  $q$ 에 대한 가중치를  $dpWt_q$ 라 하자.  $dpWt_q$ 는 RDF 그래프 상에서 항상 주어부 가중치에 해당된다.  $dpWt_q > 0$ 인 데이터타입 속성  $q$ 에 대한 자원  $r$ 의 정규화 점수  $g_{qr}$  ( $0 \leq g_{qr} \leq 1$ )은 다음 식 (5)와 같이 계산한다.  $val_{qr}$ 은 데이터타입 속성  $q$ 에 대한 자원  $r$ 의 속성값이고,  $\min_c$ 는 데이터타입 속성  $q$ 에 대한 클래스  $C$ 의 최소 속성값이며,  $\max_c$ 는 데이터타입 속성  $q$ 에 대한 클래스  $C$ 의 최대 속성값이다.

$$g_{qr} = \frac{val_{qr} - \min_c}{\max_c - \min_c} \quad (5)$$

다음으로 자원  $r$ 을 주어부로,  $r$ 의 데이터타입

속성  $q$ 에 대한 속성값인 더미(dummy) 자원을 목적부로 하는 링크의 목적부 가중치는 1로, 주어부 가중치  $dl\_sbWt_{qr}$ 은 다음 식 (6)과 같이 설정한다.

$$dl\_sbWt_{qr} = \text{조정계수} \times g_{qr} \times dpWt_q \quad (6)$$

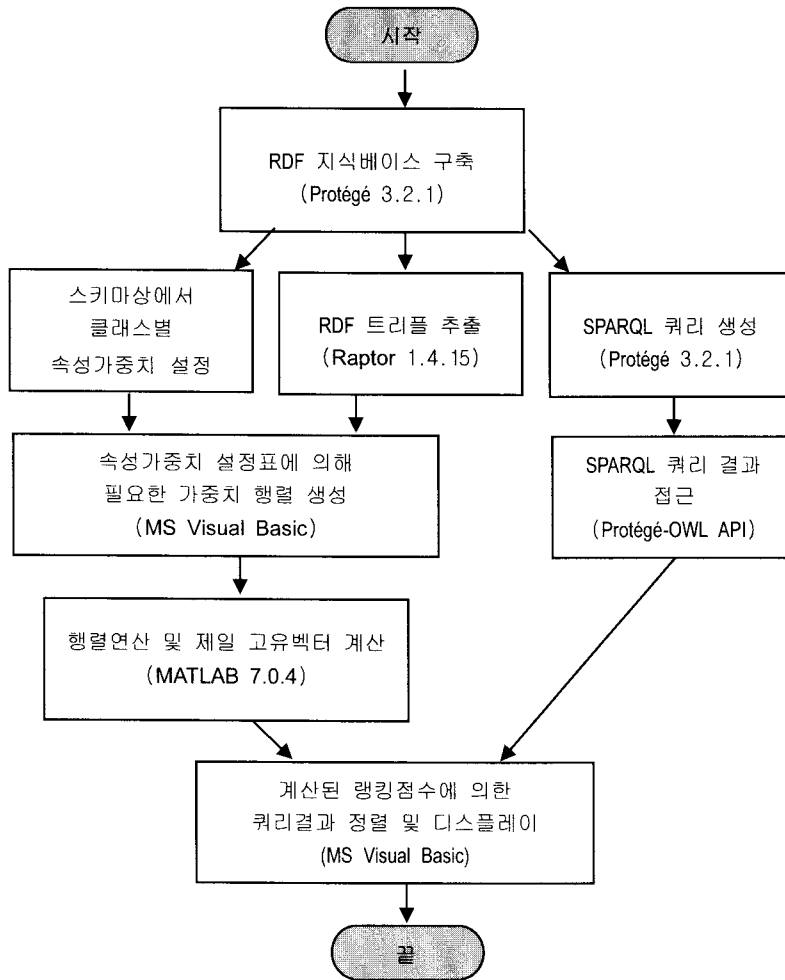
## 3.4 구현

RDF 지식베이스의 구축과 쿼리 생성 및 랭킹 점수에 의한 쿼리 결과의 정렬에 이르는 프로세스는 <그림 11>과 같다.

본 논문에서는 효율적인 트리플 정보의 구성을 위해 해당 RDF 지식베이스 파일에 대한 RDF 트리플들이 추출되어 있다고 가정하고, MS 엑셀 2003으로 트리플들을 실험 목적에 맞게 구성하여 사용하였다. 가중치 행렬은 MS 비주얼 베이직으로 프로그래밍하여 만들었다. 그리고 매트랩(MatLab) 7.0.4로 가중치 행렬관련 연산을 수행하고, 매트랩의 결과물(output)을 엑셀 파일로 받아 최종 랭킹 점수를 계산하였다.

## IV. 알고리즘 적용

이 장에서는 앞에서 제시한 알고리즘을 2개의 시나리오에 적용함으로써 알고리즘의 효과성 및 적용 가능성을 보여주고자 한다. 시나리오 1에서는 객체속성만을 반영하고 시나리오 2에서는 객체속성과 데이터타입속성을 모두 고려한다. 시나리오 1에서는 속성중심으로 가중치를 설정하는 기존 방법(PreRI : Predicate-oriented Resource Importance)과 클래스 중심으로 가중치를 설정하는 방법(ClaRI : Class-oriented Resource Importance)을 비교 분석할 것이다. 그리고 시나리오 2에서는 ClaRI 방식으로 링크 구조를 분석하여 얻은 점수와 데이터타입속성 값을 정규화하여 미리 설정된 가중치대로 합산하는 방법 (A)과, 데이터



<그림 11> 알고리즘의 구현

<표 1> 전체 시나리오

|        | 케이스 내용                                  | 가중치 설정 표                                | 도메인 구성도 | 고려 속성             |
|--------|---|---|---------|-------------------|
| 시나리오 1 | PreRI: 속성중심 가중치 설정                      | <표 2> (속성 뷰)                            | <그림 12> | 객체속성              |
|        | ClaRI: 클래스중심 가중치 설정                     | <표 3> (속성 뷰)<br>[부 록]의 <표 1><br>(클래스 뷰) |         |                   |
| 시나리오 2 | A: 데이터타입속성 값을 따로 구해<br>가중평균             | [부 록]의 <표 2><br>(클래스 뷰)                 | <그림 13> | 객체속성 +<br>데이터타입속성 |
|        | B: 데이터타입속성 값을 링크 가중치로<br>변환하여 링크 분석에 포함 |   |         |                   |

다입속성 값을 인스턴스별 링크 가중치로 변환하여 링크 분석에 처음부터 포함시켜 계산하는 방법 (B)을 살펴보기로 하겠다. 전체적인 시나리오 1은 <표 1>에 요약되어 있다.

### 4.1 시나리오 1

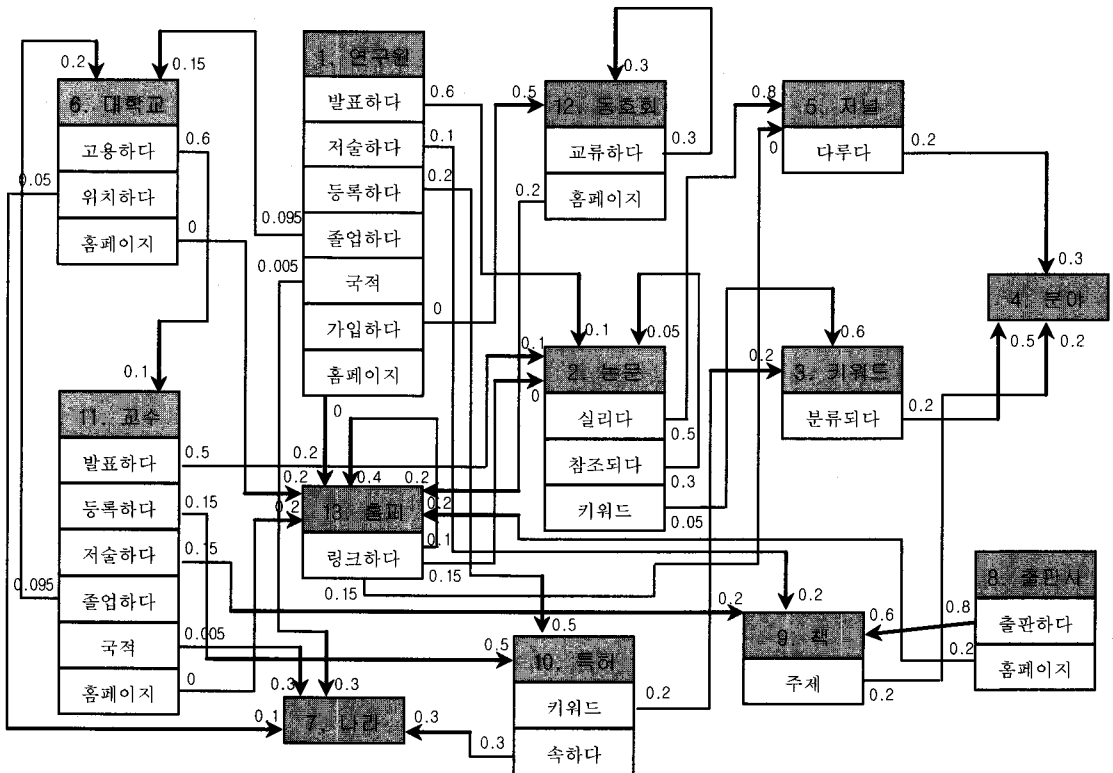
#### 4.1.1 데이터 집합

시나리오 1은 <그림 12>와 같은 스키마를 가진 도메인을 대상으로 하고 있으며 온톨로지 구성에 있어 RDF 스키마 이상에서 제공되는 클래스 간 위계구조는 단순화하여 한 계층만 있는 것으로 가정하였다. 속성들에 대한 가중치는 각각의 경우에 적합하게 <표 2>와 <표 3>과 같이 설정하였으며, 컨텍스트에 따라 다른 가중치를 사

용할 수도 있을 것이다. 설정된 가중치 값에 따라 각 방법의 결과가 약간씩 달라질 수는 있겠지만 전반적인 효과성의 비교에는 큰 영향을 주지 않을 것으로 판단된다.

<표 4>는 시나리오 1에서 사용된 클래스들의 인스턴스 수와 이들에 대한 정보를 기술하는 트리플 수를 보여준다.

시나리오 1에서는 ClaRI에 의한 강한 결합 모임 현상의 해소를 관찰할 수 있는 대상으로 연구원 클래스를 선정하였다. 시나리오 1의 두 가지 방식은 모두 같은 트리플 집합을 사용하였고, 이 안에 포함된 연구원 인스턴스들의 속성 값을 분석해 보면 대략 <표 5>와 같다. 모든 트리플 정보를 구성함에 있어 인스턴스와 속성의 이름은 간결성을 위해 URL과 '#'이 없는 단편 식별자 (fragment identifier) 형태를 사용하였고, 인스턴



<그림 12> 시나리오 1의 클래스 구성



<표 2> 시나리오1 - PreRI 속성 가중치(속성 부)

|    | domain | predicate | range | subwt | objwt |
|----|--------|-----------|-------|-------|-------|
| 1  | 교수     | 국적        | 나라    | 0.3   | 1     |
| 2  | 교수     | 등록하다      | 특허    | 1     | 0.2   |
| 3  | 교수     | 발표하다      | 논문    | 1     | 0.3   |
| 4  | 교수     | 저술하다      | 책     | 1     | 0.3   |
| 5  | 교수     | 졸업하다      | 대학교   | 1     | 0.8   |
| 6  | 교수     | 홈페이지      | 홈피    | 0.8   | 1     |
| 7  | 논문     | 실리다       | 저널    | 1     | 1     |
| 8  | 논문     | 참조되다      | 논문    | 1     | 0.4   |
| 9  | 논문     | 키워드       | 키워드   | 0.6   | 1     |
| 10 | 대학교    | 고용하다      | 교수    | 1     | 0.8   |
| 11 | 대학교    | 위치하다      | 나라    | 0.4   | 1     |
| 12 | 대학교    | 홈페이지      | 홈피    | 0.8   | 1     |
| 13 | 동호회    | 홈페이지      | 홈피    | 0.8   | 1     |
| 14 | 동호회    | 교류하다      | 동호회   | 1     | 1     |
| 15 | 연구원    | 가입하다      | 동호회   | 0.8   | 1     |
| 16 | 연구원    | 국적        | 나라    | 0.3   | 1     |
| 17 | 연구원    | 등록하다      | 특허    | 1     | 0.2   |
| 18 | 연구원    | 발표하다      | 논문    | 1     | 0.3   |
| 19 | 연구원    | 저술하다      | 책     | 1     | 0.3   |
| 20 | 연구원    | 졸업하다      | 대학교   | 1     | 0.8   |
| 21 | 연구원    | 홈페이지      | 홈피    | 0.8   | 1     |
| 22 | 저널     | 다루다       | 분야    | 0.4   | 1     |
| 23 | 책      | 주제        | 분야    | 0.5   | 1     |
| 24 | 출판사    | 출판하다      | 책     | 1     | 0.7   |
| 25 | 출판사    | 홈페이지      | 홈피    | 0.8   | 1     |
| 26 | 키워드    | 분류되다      | 분야    | 0.4   | 1     |
| 27 | 특허     | 속하다       | 나라    | 0.2   | 1     |
| 28 | 특허     | 키워드       | 키워드   | 0.6   | 1     |
| 29 | 홈피     | 링크하다      | 논문    | 1     | 0.8   |
| 30 | 홈피     | 링크하다      | 저널    | 1     | 0.8   |
| 31 | 홈피     | 링크하다      | 홈피    | 1     | 0.8   |

<표 3> 시나리오1 - ClaRI 속성 가중치(속성 부)

|    | domain | predicate | range | subwt | objwt |
|----|--------|-----------|-------|-------|-------|
| 1  | 교수     | 국적        | 나라    | 0.005 | 0.3   |
| 2  | 교수     | 등록하다      | 특허    | 0.15  | 0.5   |
| 3  | 교수     | 발표하다      | 논문    | 0.5   | 0.1   |
| 4  | 교수     | 저술하다      | 책     | 0.15  | 0.2   |
| 5  | 교수     | 졸업하다      | 대학교   | 0.095 | 0.2   |
| 6  | 교수     | 홈페이지      | 홈피    | 0     | 0.2   |
| 7  | 논문     | 실리다       | 저널    | 0.5   | 0.8   |
| 8  | 논문     | 참조되다      | 논문    | 0.3   | 0.05  |
| 9  | 논문     | 키워드       | 키워드   | 0.05  | 0.6   |
| 10 | 대학교    | 고용하다      | 교수    | 0.6   | 0.1   |
| 11 | 대학교    | 위치하다      | 나라    | 0.05  | 0.1   |
| 12 | 대학교    | 홈페이지      | 홈피    | 0     | 0.2   |
| 13 | 동호회    | 교류하다      | 동호회   | 0.3   | 0.3   |
| 14 | 동호회    | 홈페이지      | 홈피    | 0.2   | 0.2   |
| 15 | 연구원    | 가입하다      | 동호회   | 0     | 0.5   |
| 16 | 연구원    | 국적        | 나라    | 0.005 | 0.3   |
| 17 | 연구원    | 등록하다      | 특허    | 0.2   | 0.5   |
| 18 | 연구원    | 발표하다      | 논문    | 0.6   | 0.1   |
| 19 | 연구원    | 저술하다      | 책     | 0.1   | 0.2   |
| 20 | 연구원    | 졸업하다      | 대학교   | 0.095 | 0.15  |
| 21 | 연구원    | 홈페이지      | 홈피    | 0     | 0.2   |
| 22 | 저널     | 다루다       | 분야    | 0.2   | 0.3   |
| 23 | 책      | 주제        | 분야    | 0.2   | 0.2   |
| 24 | 출판사    | 출판하다      | 책     | 0.8   | 0.6   |
| 25 | 출판사    | 홈페이지      | 홈피    | 0.2   | 0.2   |
| 26 | 키워드    | 분류되다      | 분야    | 0.2   | 0.5   |
| 27 | 특허     | 속하다       | 나라    | 0.3   | 0.3   |
| 28 | 특허     | 키워드       | 키워드   | 0.2   | 0.2   |
| 29 | 홈피     | 링크하다      | 논문    | 0.15  | 0     |
| 30 | 홈피     | 링크하다      | 저널    | 0.15  | 0     |
| 31 | 홈피     | 링크하다      | 홈피    | 0.1   | 0.4   |

스 이름은 '클래스이름 클래스번호 - 인스턴스번호' 형식으로 부여하였다. 인스턴스 번호가 작을 수록 [부록]의 <표 1> 기준에 의해 대략 높은 점수를 가지도록 속성값을 설정하여 예상한대로 랭킹 점수 결과가 나오는지 살펴볼 것이다. 단적인

예로 <표 5>에서 살펴볼 수 있는 바와 같이 '연구원 1-1'은 논문을 10편 발표한 반면에 '연구원 1-25'는 발표한 논문이 하나도 없다. 강한 결합을 형성하기 위해 '연구원 21-25'는 동호회에, '연구원 17-25'는 홈피에 연결하였고 동호회와 홈피,

<표 4> 시나리오 1의 인스턴스와 트리플 수

|          | 인스턴스 수 |
|----------|--------|
| 1. 연구원   | 25     |
| 2. 논문    | 100    |
| 3. 키워드   | 15     |
| 4. 분야    | 5      |
| 5. 저널    | 5      |
| 6. 대학교   | 3      |
| 7. 나라    | 3      |
| 8. 출판사   | 3      |
| 9. 책     | 15     |
| 10. 특허   | 10     |
| 11. 교수   | 9      |
| 12. 동호회  | 5      |
| 13. 홈페이지 | 30     |
| 인스턴스 총 수 | 228    |
| 트리플 총 수  | 1160   |

홈피와 홈페이지, 홈페이지와 다른 클래스 간에도 링크를 만들어 주었다. '연구원 1-25'는 연구원 중요도 평가에 반영되지 않는 동호회에는 5개나 가입되어 있다.

이 외에 클래스중심 가중치 설정 방법이 다른 클래스에 대해서도 주어진 트리플 정보에 부합하는 랭킹 순위를 보여주는지와, 특정한 한 자원의 중요도에 영향을 주는 링크 정보를 추가하거나 삭제했을 때 실제로 해당 자원의 랭킹 점수에 영향이 있는지 등의 기본적인 평가도 수행하였다.

#### 4.1.2 적용 결과

시나리오 1의 PreRI에 의한 연구원 클래스의 랭킹 결과는 <표 6>과 같다. 목적부 점수(objectivity)

<표 5> 시나리오 1-연구원 클래스 인스턴스별 속성 값

|          | 논문 수<br>(논문번호) | 책 수<br>(책번호) | 특허 수<br>(특허번호) | 학교      | 국적     | 동호회 수<br>(동호회번호) | 홈페이지<br>(홈페이지번호) |
|----------|----------------|--------------|----------------|---------|--------|------------------|------------------|
| 연구원 1-1  | 10편(1~10)      | 4권(1~4)      | 3개(1~3)        | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-2  | 8편(11~18)      | 3권(5~7)      | 2개(4~5)        | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-3  | 7편(19~25)      | 1권(8)        | 0              | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-4  | 6편(26~31)      | 0            | 0              | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-5  | 5편(32~36)      | 0            | 0              | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-6  | 5편(37~41)      | 0            | 0              | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-7  | 5편(42~46)      | 0            | 0              | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-8  | 5편(47~51)      | 0            | 0              | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-9  | 5편(52~56)      | 0            | 0              | 대학교 6-1 | 나라 7-1 | 0                | no               |
| 연구원 1-10 | 4편(57~60)      | 0            | 0              | 대학교 6-1 | 나라 7-2 | 0                | no               |
| 연구원 1-11 | 4편(61~64)      | 0            | 0              | 대학교 6-2 | 나라 7-2 | 0                | no               |
| 연구원 1-12 | 4편(65~68)      | 0            | 0              | 대학교 6-2 | 나라 7-2 | 0                | no               |
| 연구원 1-13 | 4편(69~72)      | 0            | 0              | 대학교 6-2 | 나라 7-2 | 0                | no               |
| 연구원 1-14 | 4편(73~76)      | 0            | 0              | 대학교 6-2 | 나라 7-2 | 0                | no               |
| 연구원 1-15 | 4편(77~80)      | 0            | 0              | 대학교 6-2 | 나라 7-2 | 0                | no               |
| 연구원 1-16 | 3편(81~83)      | 0            | 0              | 대학교 6-2 | 나라 7-2 | 0                | no               |
| 연구원 1-17 | 3편(84~86)      | 0            | 0              | 대학교 6-2 | 나라 7-2 | 0                | yes(10)          |
| 연구원 1-18 | 3편(87~89)      | 0            | 0              | 대학교 6-2 | 나라 7-3 | 0                | yes(9)           |
| 연구원 1-19 | 2편(90~91)      | 0            | 0              | 대학교 6-3 | 나라 7-3 | 0                | yes(8)           |
| 연구원 1-20 | 2편(92~93)      | 0            | 0              | 대학교 6-3 | 나라 7-3 | 0                | yes(7)           |
| 연구원 1-21 | 2편(94~95)      | 0            | 0              | 대학교 6-3 | 나라 7-3 | 3(1~3)           | yes(6)           |
| 연구원 1-22 | 2편(96~97)      | 0            | 0              | 대학교 6-3 | 나라 7-3 | 3(1~3)           | yes(5)           |
| 연구원 1-23 | 1편(98)         | 0            | 0              | 대학교 6-3 | 나라 7-3 | 3(1~3)           | yes(4)           |
| 연구원 1-24 | 1편(99)         | 0            | 0              | 대학교 6-3 | 나라 7-3 | 4(1~4)           | yes(3)           |
| 연구원 1-25 | 0              | 0            | 0              | 대학교 6-3 | 나라 7-3 | 5(1~5)           | yes(1,2)         |

가 모두 0인 것은 <그림 12>의 스키마에서 알 수 있는 바와 같이 연구원 클래스에 속하는 인스턴스는 트리플의 목적부가 될 수 없고 주어부에만 올 수 있기 때문이다. 연구원 클래스에 연결되는 링크 구조를 이렇게 디자인한 것은 기존 연구에서 목적부나 주어부 점수를 따로따로 비교하던지[Mukherjea *et al.*, 2005], 아니면 임의로 설정한 비율대로 두 가지 점수를 합하여 랭킹에 사용했기[Bamba and Mukherjea, 2004] 때문에 본 논문에서 제안하고 있는 ClaRI와의 비교를 좀더 객관적으로 수행하기 위해서이다. 속성 중심으로 가중치를 설정했을 때에는 논문을 7편 발표하고 책을 1권 저술한 '연구원 1-3'이나, 논문을 6편 발표

한 '연구원 1-4'보다 논문을 한 편도 쓰지 않은 '연구원 1-25'가 훨씬 높게 랭크 되었음을 살펴볼 수 있다. 그리고 동호회나 홈페이지로 연결된 다른 연구원들의 중요도도 높게 평가되어 있음을 볼 수 있다.

반면에 <표 7>에서는 일련번호 순서가 랭크 순위와 거의 일치함을 확인할 수 있다. 여기에서도 <표 6>과 같은 이유로 목적부 점수는 모두 0이다. 목적부 점수와 주어부 점수가 다 양의 값을 갖는 클래스 예로 <표 8>에 특히 클래스의 랭킹 결과를 제시하였다.

이처럼 클래스 중심으로 가중치를 설정하면 아무리 강한 결합을 보이는 노드들이 있어도 중

<표 6> 시나리오 1-PreRI: 연구원 클래스 랭킹 결과

| Ranking |          | objectivity        | subjectivity       | ranking score      |
|---------|----------|--------------------|--------------------|--------------------|
| 1       | 연구원 1-1  | 0.0000000000000000 | 0.0280030507813939 | 0.0280030507813939 |
| 2       | 연구원 1-2  | 0.0000000000000000 | 0.0213430021103920 | 0.0213430021103920 |
| 3       | 연구원 1-25 | 0.0000000000000000 | 0.0180022646307396 | 0.0180022646307396 |
| 4       | 연구원 1-18 | 0.0000000000000000 | 0.0127843375684079 | 0.0127843375684079 |
| 5       | 연구원 1-17 | 0.0000000000000000 | 0.0127130956030976 | 0.0127130956030976 |
| 6       | 연구원 1-20 | 0.0000000000000000 | 0.0120720263735658 | 0.0120720263735658 |
| 7       | 연구원 1-19 | 0.0000000000000000 | 0.0117446526759738 | 0.0117446526759738 |
| 8       | 연구원 1-24 | 0.0000000000000000 | 0.0096891442421732 | 0.0096891442421732 |
| 9       | 연구원 1-21 | 0.0000000000000000 | 0.0095702531798360 | 0.0095702531798360 |
| 10      | 연구원 1-22 | 0.0000000000000000 | 0.0095702531798360 | 0.0095702531798360 |
| 11      | 연구원 1-23 | 0.0000000000000000 | 0.0095574853938614 | 0.0095574853938614 |
| 12      | 연구원 1-15 | 0.0000000000000000 | 0.0050150163038354 | 0.0050150163038354 |
| 13      | 연구원 1-16 | 0.0000000000000000 | 0.0038441292780735 | 0.0038441292780735 |
| 14      | 연구원 1-14 | 0.0000000000000000 | 0.0026835036809852 | 0.0026835036809852 |
| 15      | 연구원 1-3  | 0.0000000000000000 | 0.0006218765376622 | 0.0006218765376622 |
| 16      | 연구원 1-4  | 0.0000000000000000 | 0.0005802088373936 | 0.0005802088373936 |
| 17      | 연구원 1-5  | 0.0000000000000000 | 0.0005623722545404 | 0.0005623722545404 |
| 18      | 연구원 1-6  | 0.0000000000000000 | 0.0005620909781709 | 0.0005620909781709 |
| 19      | 연구원 1-7  | 0.0000000000000000 | 0.0005594629999432 | 0.0005594629999432 |
| 20      | 연구원 1-8  | 0.0000000000000000 | 0.0005506387457501 | 0.0005506387457501 |
| 21      | 연구원 1-9  | 0.0000000000000000 | 0.0005506387457501 | 0.0005506387457501 |
| 22      | 연구원 1-10 | 0.0000000000000000 | 0.0004709909670102 | 0.0004709909670102 |
| 23      | 연구원 1-11 | 0.0000000000000000 | 0.0003827768125259 | 0.0003827768125259 |
| 24      | 연구원 1-12 | 0.0000000000000000 | 0.0003827768125259 | 0.0003827768125259 |
| 25      | 연구원 1-13 | 0.0000000000000000 | 0.0003599564785046 | 0.0003599564785046 |

<표 7> 시나리오 1 -ClaRI: 연구원 클래스 랭킹 결과

| Ranking |          | objectivity        | subjectivity       | ranking score      |
|---------|----------|--------------------|--------------------|--------------------|
| 1       | 연구원 1-1  | 0.0000000000000000 | 0.0155959095081723 | 0.0155959095081723 |
| 2       | 연구원 1-2  | 0.0000000000000000 | 0.0051286892012097 | 0.0051286892012097 |
| 3       | 연구원 1-3  | 0.0000000000000000 | 0.0033880976031599 | 0.0033880976031599 |
| 4       | 연구원 1-4  | 0.0000000000000000 | 0.0028516085955419 | 0.0028516085955419 |
| 5       | 연구원 1-5  | 0.0000000000000000 | 0.0022749359582430 | 0.0022749359582430 |
| 6       | 연구원 1-6  | 0.0000000000000000 | 0.0022641151361006 | 0.0022641151361006 |
| 7       | 연구원 1-7  | 0.0000000000000000 | 0.0018189884972129 | 0.0018189884972129 |
| 8       | 연구원 1-8  | 0.0000000000000000 | 0.0001034070826220 | 0.0001034070826220 |
| 9       | 연구원 1-9  | 0.0000000000000000 | 0.0001034070826220 | 0.0001034070826220 |
| 10      | 연구원 1-10 | 0.0000000000000000 | 0.0000528428733103 | 0.0000528428733103 |
| 11      | 연구원 1-11 | 0.0000000000000000 | 0.0000094887230912 | 0.0000094887230912 |
| 12      | 연구원 1-12 | 0.0000000000000000 | 0.0000094887230912 | 0.0000094887230912 |
| 13      | 연구원 1-13 | 0.0000000000000000 | 0.0000088860386709 | 0.0000088860386709 |
| 14      | 연구원 1-15 | 0.0000000000000000 | 0.0000088331428200 | 0.0000088331428200 |
| 15      | 연구원 1-14 | 0.0000000000000000 | 0.0000088331428200 | 0.0000088331428200 |
| 16      | 연구원 1-17 | 0.0000000000000000 | 0.0000087888709563 | 0.0000087888709563 |
| 17      | 연구원 1-16 | 0.0000000000000000 | 0.0000087888709562 | 0.0000087888709562 |
| 18      | 연구원 1-18 | 0.0000000000000000 | 0.0000083393722875 | 0.0000083393722875 |
| 19      | 연구원 1-20 | 0.0000000000000000 | 0.0000052870768580 | 0.0000052870768580 |
| 20      | 연구원 1-19 | 0.0000000000000000 | 0.0000052870768580 | 0.0000052870768580 |
| 21      | 연구원 1-21 | 0.0000000000000000 | 0.0000052870768580 | 0.0000052870768580 |
| 22      | 연구원 1-22 | 0.0000000000000000 | 0.0000052870768580 | 0.0000052870768580 |
| 23      | 연구원 1-24 | 0.0000000000000000 | 0.0000052608409277 | 0.0000052608409277 |
| 24      | 연구원 1-23 | 0.0000000000000000 | 0.0000052608409277 | 0.0000052608409277 |
| 25      | 연구원 1-25 | 0.0000000000000000 | 0.0000052348640916 | 0.0000052348640916 |

<표 8> 시나리오 1 -ClaRI: 특허클래스 랭킹 결과

|          | objectivity        | subjectivity       | ranking score      |
|----------|--------------------|--------------------|--------------------|
| 특허 10-1  | 0.0014695855651039 | 0.1185432831844480 | 0.1200128687495520 |
| 특허 10-2  | 0.0014695855651039 | 0.1185432831844480 | 0.1200128687495520 |
| 특허 10-3  | 0.0014695855651039 | 0.1185432831844480 | 0.1200128687495520 |
| 특허 10-4  | 0.0006897810251708 | 0.0030453994039200 | 0.0037351804290908 |
| 특허 10-5  | 0.0004257485115526 | 0.0001127339341673 | 0.0005384824457199 |
| 특허 10-6  | 0.0002640325136182 | 0.0000407860780610 | 0.0003048185916792 |
| 특허 10-7  | 0.0001485768065311 | 0.0000402531124557 | 0.0001888299189869 |
| 특허 10-8  | 0.0001450083726445 | 0.0000125036375940 | 0.0001575120102385 |
| 특허 10-9  | 0.0001408984514883 | 0.0000125036375940 | 0.0001534020890824 |
| 특허 10-10 | 0.0001349454940768 | 0.0000123827349429 | 0.0001473282290198 |

요도에 영향을 미치지 않는 링크들은 제외시키는 효과가 있으므로 훨씬 안정적이라고 할 수 있다. 강한 결합 모임 말고도 기존 연구의 다른 한계점인 정보 표현의 완전성에 대해서도 효율적인 지침을 제시한다. 온톨로지 스키마 상에서 중요도에 영향을 미치는 속성들에 대해서는 누락된 정보가 없어야 정확한 랭킹 점수를 얻을 수 있다는 것은 당연한 결과인 것이다. 어떤 자원이 혼잡하기 때문에 높은 점수를 받는 현상도 결국 강한 결합 모임 현상과 맥락을 같이 한다.

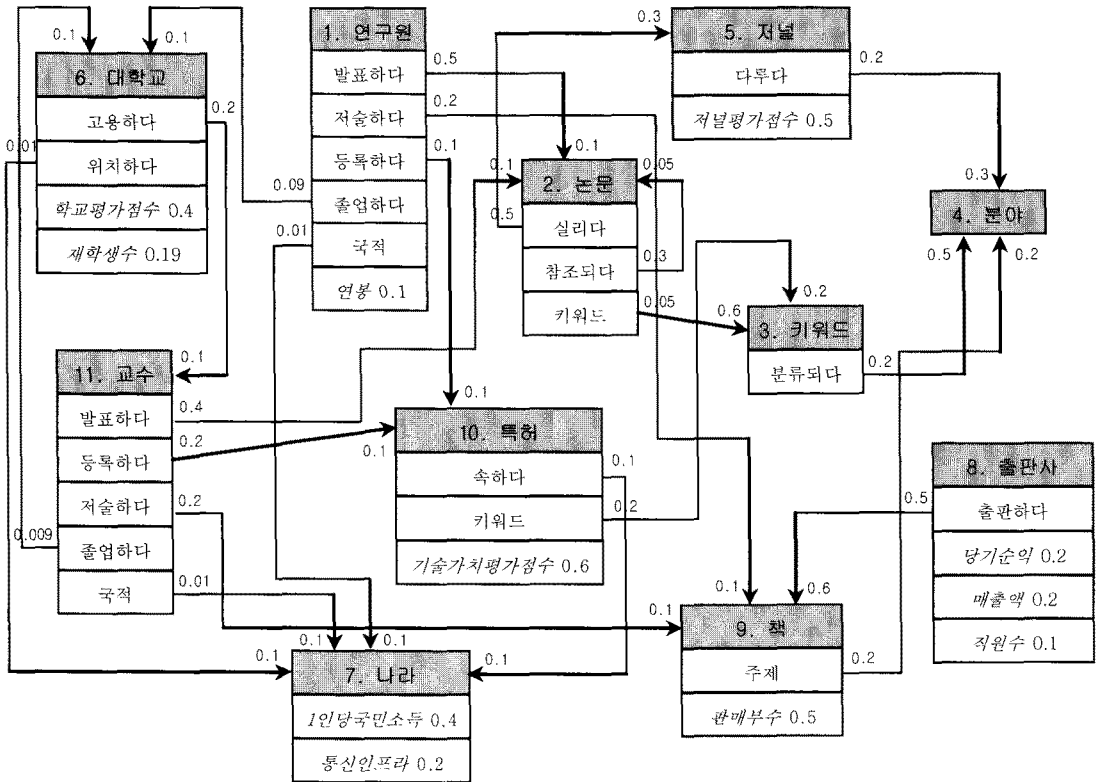
시나리오 1의 ClaRI는 다른 클래스들에 대해서도 트리플 링크 구조를 고려할 때 합리적인 결과를 보여주었고 특정 자원의 중요 속성에 대한 링크 연결을 추가하거나 삭제했을 경우에도 예상되는 바와 같이 랭킹 점수가 증가하거나 감소하였다. 이에 대한 결과는 따로 제시하지 않는다.

## 4.2 시나리오 2

### 4.2.1 데이터 집합

시나리오 2는 시나리오 1에서 '동호회'와 '홈피' 클래스가 사라지고 데이터타입속성이 추가된 <그림 13>과 같은 도메인을 바탕으로 한다. 추가된 데이터타입속성은 이탤릭체로 표현하였다.

시나리오 2에서는 [부 록]의 <표 2>에서 데이터타입 속성 반영 비율이 높으면서 클래스 인스턴스의 수가 적지 않은 '책' 클래스를 선정하여 A와 B 방법의 적용 결과를 살펴보았다. 데이터타입속성인 '판매부수'에 대한 인스턴스별 속성 값은 실험 결과를 보여주는 <표 10>와 <표 11>에 함께 제시되어 있다. <표 9>는 시나리오 2에서 사용된 클래스들의 인스턴스 수와 이러한 인



<그림 13> 시나리오 2의 클래스 구성

&lt;표 9&gt; 시나리오 2의 인스턴스와 트리플 수

|          | 인스턴스 수   |
|----------|----------|
| 1. 연구원   | 20(20×1) |
| 2. 논문    | 100      |
| 3. 키워드   | 17       |
| 4. 분야    | 5        |
| 5. 저널    | 5(5×1)   |
| 6. 대학교   | 3(3×2)   |
| 7. 나라    | 3(3×2)   |
| 8. 출판사   | 3(3×3)   |
| 9. 책     | 10(10×1) |
| 10. 특허   | 10(10×1) |
| 11. 교수   | 12       |
| 12. 동호회  | 0        |
| 13. 홈페이지 | 0        |
| 인스턴스 총 수 | 253      |
| 트리플 총 수  | 873      |

스턴스들간의 관계 및 데이터타입속성 값을 기술한 트리플의 총 수를 보여준다. 괄호 안에 있는 수자는 데이터타입속성에 대한 더미 자원의 수를 나타낸다.

#### 4.2.2 평가 결과

<표 10>은 <그림 13>에서 ClaRI에 의해 객체 속성만을 반영하여 구한 책 인스턴스들의 링크 분석 결과를 정규화한 점수와 데이터타입속성인 '판매부수' 값을 정규화한 점수를 설정된 가중치 대로 합산한 내용을 보여준다.

<표 11>은 책 인스턴스들의 판매부수 속성값을 정규화하여 인스턴스별 링크가중치로 변환한 후 ClaRI에 의한 링크 분석에 처음부터 포함시켜 계산한 결과를 보여준다. <표 11>의 랭킹 점수를 <표 10>의 링크분석 점수와 비교할 때 최대값은 더 크고 최소값은 더 작음을 볼 수 있다. 이것은 판매부수 값의 차이가 반영된 결과로 보이며, 판매부수 값도 일련번호가 낮을수록 높게 설정되어 랭킹 순위에는 변동이 없는 것을 볼 수 있다.

### 4.3 연구의 시사점

지금까지 시나리오 1에서는 링크구조를 분석하여 랭킹에 활용하는 알고리즘에서 클래스 중심으로 가중치를 설정하면 속성중심으로 가중치를 설정할 때 나타나는 강한 결합 모임 현상과 같은 문제점들이 해결됨을 관찰할 수 있었다. 그리고 본 논문에서 제안하는 알고리즘이 데이터 집합에 타당한 랭킹 결과를 보여주고 있음도 확인하였다. 시나리오 2에서는 중요도 계산에서 제외되었던 데이터타입 속성도 랭킹 점수에 반영될 수 있음을 살펴볼 수 있었다.

이러한 실험 결과로부터 다음과 같은 시사점들을 생각해 볼 수 있겠다. 첫째, 본 논문에서 제안하는 알고리즘은 보완 과정을 거쳐 시맨틱 웹의 자원에 대한 기본적인 정렬 메커니즘으로 활용될 수 있을 것으로 보인다. 시맨틱 웹은 RDF 데이터 모델을 기초로 하고 있으므로 대부분의 도메인에 대한 정보가 그래프로 표현될 수 있으며 링크구조 분석 방법을 적용하면 각 클래스에 속하는 자원들의 중요도를 쉽게 계산할 수 있다. 실제 세계에서 자주 기사화되고 있는 대학 랭킹이나 쇼핑몰 순위, 또는 저널 랭킹 등 많은 분야에 응용될 수 있을 것으로 기대된다.

둘째, 노드 간에 중요도를 전파함으로써 계산되는 링크구조 분석 방법은 이웃하는 노드의 절대적인 수뿐만 아니라 노드들의 실질적인 중요도도 고려하므로 데이터를 조작하기가 상대적으로 어려워 더욱 안전한 알고리즘이라고 할 수 있다. 페이지랭크가 웹 그래프로부터 전통적인 랭킹 알고리즘보다 더욱 객관적인 중요도 점수를 산출했던 것처럼 자원 간에 복잡한 상호 작용에 의해 계산되는 링크구조 기반 방식은 악의적인 개입을 좀더 어렵게 한다.

셋째, 사람들은 클래스마다 이에 속하는 자원에 대해 서로 다른 판단 기준을 적용하는데 클래스 중심으로 가중치를 설정하는 방법은 사람들의 정신작용과 비슷한 메커니즘을 소프트웨어

<표 10> 시나리오 2 -A: 책 클래스 랭킹 결과(판매부수 속성 가중치 = 0.5)

|        | 링크분석 점수            | 링크 정규화 점수          | 판매부수 | 판매 정규화점수           | 최종 랭킹 점수           |
|--------|--------------------|--------------------|------|--------------------|--------------------|
| 책 9-1  | 0.0211158322171767 | 1.0000000000000000 | 8000 | 1.0000000000000000 | 1.0000000000000000 |
| 책 9-2  | 0.0209980612747133 | 0.994416209938332  | 7000 | 0.8333333333333333 | 0.913874771635833  |
| 책 9-3  | 0.0209718587670499 | 0.993173889049821  | 6000 | 0.6666666666666667 | 0.829920277858244  |
| 책 9-4  | 0.0021112883643479 | 0.098951058110468  | 5000 | 0.5000000000000000 | 0.299475529055234  |
| 책 9-5  | 0.0014356776145794 | 0.066918805406992  | 4500 | 0.4166666666666667 | 0.241792736036829  |
| 책 9-6  | 0.0006028526625789 | 0.027432667376506  | 4000 | 0.3333333333333333 | 0.180383000354920  |
| 책 9-7  | 0.0005946751016884 | 0.027044950489182  | 3500 | 0.2500000000000000 | 0.138522475244591  |
| 책 9-8  | 0.0000572924804120 | 0.001566409881177  | 3000 | 0.1666666666666667 | 0.084116538273922  |
| 책 9-9  | 0.0000454220365117 | 0.001003604954160  | 2500 | 0.0833333333333333 | 0.042168469143747  |
| 책 9-10 | 0.0000242544245480 | 0.0000000000000000 | 2000 | 0.0000000000000000 | 0.0000000000000000 |

<표 11> 시나리오 2 -B: 책 클래스 랭킹 결과(조정계수 = 20, 판매부수 속성 가중치 = 0.5)

|        | 판매부수 | 정규화점수              | 인스턴스 링크가중치<br>(조정계수*정규화점수*속성가중치) | ranking score      |
|--------|------|--------------------|----------------------------------|--------------------|
| 책 9-1  | 8000 | 1.0000000000000000 | 10.00000                         | 0.0286011806011376 |
| 책 9-2  | 7000 | 0.8333333333333333 | 8.33333                          | 0.0254760658470699 |
| 책 9-3  | 6000 | 0.6666666666666667 | 6.66667                          | 0.0230398080065576 |
| 책 9-4  | 5000 | 0.5000000000000000 | 5.00000                          | 0.0027921381886903 |
| 책 9-5  | 4500 | 0.4166666666666667 | 4.16667                          | 0.0013590619006261 |
| 책 9-6  | 4000 | 0.3333333333333333 | 3.33333                          | 0.0004335665893297 |
| 책 9-7  | 3500 | 0.2500000000000000 | 2.50000                          | 0.0004093888546677 |
| 책 9-8  | 3000 | 0.1666666666666667 | 1.66667                          | 0.0000456496989496 |
| 책 9-9  | 2500 | 0.0833333333333333 | 0.83333                          | 0.0000298994011345 |
| 책 9-10 | 2000 | E-16               | 5E-16                            | 0.0000163708290858 |

에이전트들에게도 제공할 수 있는 효과적인 방식이라고 생각된다. 사람과 에이전트의 유사성을 높일수록 시맨틱 웹의 성장 속도는 빨라질 것이다.

넷째, 같은 스키마라도 컨텍스트에 따라 속성 가중치를 다르게 조정할 수 있고 데이터타입 속성도 고려할 수 있기 때문에 다양한 기준에 의해 랭킹 결과를 비교해 봄으로써 주어진 상황을 다각적으로 분석해 볼 수 있어 의사 결정의 질을 높여줄 것으로 예상된다.

### V. 결론 및 향후 연구과제

컴퓨터도 사람처럼 생각하고 추론할 수 있는 시맨틱 웹의 이상은 정밀하게 구축된 온톨로지를 통해 이루어진다. 온톨로지는 대부분 수많은 클래스와 이들간의 상호역학적인 관계를 나타내는 속성들로 묘사된다. 이 속성들 중에는 클래스에 속하는 자원 중요도에 영향을 미치는 것도 있고 단순히 정보를 기록하기 위한 것들도 있다. 한

편 클래스에 속하는 인스턴스 데이터가 축적됨에 따라 이것들을 중요도에 따라 랭킹하는 작업이 사람에게 절실하게 필요한 만큼 컴퓨터에게도 사람의 판단과 유사한 정렬 메커니즘을 만들어주는 일은 큰 의미가 있다.

시맨틱 웹의 기초를 이루고 있는 RDF 데이터 모델은 클래스와 이에 대한 인스턴스를 노드로, 속성을 링크로 나타낸다. 그리고 OWL과 같은 많은 온톨로지 언어가 이러한 RDF 데이터 모델을 기반으로 하고 있으므로 임의의 도메인에 대한 온톨로지는 웹 그래프와 비슷한 RDF 그래프로 표현될 수 있다. 그러므로 그 동안 많은 연구가 진행되어온 링크구조 기반 랭킹 알고리즘을 시맨틱 웹의 자원을 중요도에 따라 랭킹하는 정렬 메커니즘에 접목하려는 시도는 상당히 합리적이라고 할 수 있다.

이 과정에서 클래스 중심으로 속성의 가중치를 설정하는 방법은 한 클래스에 속하는 자원들을 평가함에 있어 속성의 상대적인 비중을 고려하는 사람들의 평가방식과 유사하다. 컨텍스트에 따라 다양한 조합의 가중치를 설정할 수도 있으므로 여러 상황에 이용될 수 있다. 온톨로지 설계에 있어서도 중요하게 다루어야 할 속성들에 대한 기준을 제시해주며, 양의 가중치를 갖는 속성들에 대한 데이터는 누락하지 않도록 하는 관리 지침을 세워주기도 한다. 이 방법은 한 클래스에 속하는 자원의 중요도에 무의미한 속성 링크를

제거함으로써 기존 연구가 안고 있는 강한 결합모임과 같은 한계점을 효과적으로 해결한다. 본 논문에서 제안한 알고리즘은 데이터타입 속성을 반영할 수 있는 방안에 대한 연구와 함께 RDF 그래프로 표현되는, 랭킹이 필요한 많은 도메인에 적용될 수 있을 것으로 기대되며, 시맨틱 웹의 핵심인 온톨로지를 구축할 때에 이러한 정렬 메커니즘에 대한 검토가 필수적으로 이루어져야 할 것으로 판단된다.

그런데 시맨틱 웹의 클래스 중심으로 가중치를 설정할 때 가중치 설정의 공적인 기준이나 근거에 대한 연구는 아직 많이 부족한 상태이므로 계속 진행되어야 한다. 본 논문은 컨텍스트에 따른 가중치의 변화를 유연하게 수용할 수 있는 프레임워크의 개발을 목표로 하였기 때문에 객관적인 가중치의 설정 방식에 대한 내용은 자세히 다루지 않았지만, 사회적으로 중요한 도메인에 대해서는 어느 정도 합의를 거친 가중치 설정 기준이 마련되어야 할 것이다. 이와 관련하여 요인 분석과 회귀분석을 포함하는 다양한 기법에 대한 연구가 요구된다. 그리고 본 논문에서는 클래스와 속성의 위계구조가 하나의 계층으로 이루어진 경우를 가정하였는데 좀더 복잡한 위계구조를 가지고 있는 온톨로지에 대해 생각해 볼 필요가 있다. 본 논문에서 다룬 내용을 최하위 계층의 클래스와 속성을 대상으로 한 것이라고 보면 이에 대한 연구는 그리 어렵지 않을 것으로 예상된다.

## 〈참 고 문 헌〉

- [1] 노상규와 박진수, "인터넷 진화의 열쇠 온톨로지," *가즈토이*, 2007.
- [2] Aleman-Meza, B., Halaschek-Wiener, C., Arpinar, I.B., and Sheth, A., "Context-Aware Semantic association Ranking" *Semantic Web and Database Workshop Proceedings*, Belin, September pp. 7-8, 2003.
- [3] Aleman-Meza, B., Halaschek-Wiener, C., Arpinar, I.B., Ramakrishnan, C., and Sheth, A., "Ranking Complex Relationships on the Semantic Web," *IEEE Internet Computing*, Vol. 9, No. 3, 2005, pp. 37-44.
- [4] Anyanwu, K., Maduko, A., and Sheth, A., "SemRank: Ranking Complex Relationship Search Results on the Semantic Web," *International World Wide Web Conference*



- Committee(IW3C2)*, Chiba, Japan, 2005.
- [5] Bamba, B. and Mukherjea, S., "Utilizing Resource Importance for Ranking Semantic Web Query Results," *Proc. Second Toronto International Work. Semantic Web Databases (SWDB)*, 2004, pp. 185-198.
- [6] Berners-Lee, T., "Web for real people," 2005. Available at [http://www.w3.org/2005/Talks/0511-keynote-tbl/#\[17\]](http://www.w3.org/2005/Talks/0511-keynote-tbl/#[17]).
- [7] Boyer, R.S. and Moore, J.S., "A Fast String Searching Algorithm," *Comm. ACM*, Vol. 20, No. 10, 1977, pp. 762-772.
- [8] Brickley, D. and Guha, R.V. eds., "RDF Vocabulary Description Language 1.0: RDF Schema," *W3C Recommendation*, 10 February 2004.
- [9] Brin, S. and Page, L., "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Special Issu. 7<sup>th</sup> International World Wide Web Conf. Computer Networks and ISDN Systems*, Vol. 30, Vol. 1-7, 1998, pp. 107-117.
- [10] Brin, S., Motwani, R., Page, L., and Winograd, T., "What can you do with a Web in your Pocket," *Bull. IEEE Computer Society Technical Comm. Data Engineering*, 1998.
- [11] Burden, R.L. and Faires, J.D., "Numerical Analysis," seventh edition, BROOKS/COLE, 2001.
- [12] Ding, L., Finin, T., Joshi, A., Peng, Y., Pan, R., Reddivari, P., "Search on the Semantic Web," *IEEE Computer Society*, Vol. 38, No. 10, 2005, pp. 62-69.
- [13] Ding, L., Pan, R., Finin, T., Joshi, A., Peng, Y., and Kolari, P., "Finding and Ranking Knowledge on the Semantic Web," *Proc. 4<sup>th</sup> Galway IE International Semantic Web Conf.*, 2005, pp. 156-170.
- [14] Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R.S., Peng, Y., Reddivari, P., Doshi, V.C., and Sachs, J., "Swoogle: A Semantic Web Search and Metadata Engine," *Proc. 13<sup>th</sup> ACM Conf. Info. and Knowledge Management*, 2004, pp. 652-659.
- [15] Ehrlich, L.W., "Rate of Convergence Proofs of the Method for Finding Roots of Polynomials(or Eigenvalues of Matrices) by the Power and Inverse Power Methods," National Technical Information Service AD707331, 1969.
- [16] Finin, T. and Ding, L., "Search Engines for Semantic Web Knowledge," *Proceedings of XTech 2006: Building Web 2.0*, Amsterdam, May pp. 16-19, 2006.
- [17] Friedberg, S.H., Insel, A.J., and Spence, L.E., *Linear Algebra*(4<sup>th</sup> Edition), Prentice Hall, 2003.
- [18] Halaschek, C., Aleman-Meza, B., Arpinar, I.B., and Sheth, A., "Discovering and Ranking Semantic Associations over a Large RDF Metabase," *Proceedings of the 30<sup>th</sup> VLDB Conference*, Toronto, Canada, 2004.
- [19] Haveliwala, T.H., "Efficient Computation of PageRank," Unpublished manuscript, Stanford University, 1999.
- [20] Karam, N., Benbernou, S., Debrauwer, L., and Schneider, M., "Semantic Ranking of Web Documents," Technical report RR-04-23, LIMOS, 2004.
- [21] Kleinberg, J., "Authoritative sources in a hyperlinked environment," *Proc. 9<sup>th</sup> ACM-SIAM Symp. Discrete Algorithms*, 1998, pp. 668-677. Extended version in *J. ACM*, Vol. 46, No. 5, 1999, pp. 604-632.
- [22] Klyne, G. and Carroll, J. eds., "Resource Description Framework (RDF): Concepts

- and Abstract Syntax," *W3C Recommendation*, 10 February 2004.
- [23] Knuth, D.E., Morris, J.H., and Pratt, V.R., "Fast Pattern Matching in Strings," *SIAM Journal on Computing*, Vol. 6, No. 2, 1977, pp. 323-350.
- [24] Maedche, A. and Staab, S., "Measuring Similarity between Ontologies," *European Conference of Knowledge Acquisition and Management(EKAW2002)*, Lectures Notes in Computer Science, Madrid, Spain, Springer, 2002.
- [25] Maedche, A., Staab, S., Stojanovic, N., Studer, R., and Sure, Y., "SEAL-A Framework for Developing SEMantic Web PortALs," *Lecture Notes in Computer Science*, 2097, 2001.
- [26] Manola, F. and Miller, E. eds., "RDF Primer," *W3C Recommendation*, 2004.
- [27] Mukherjea, S. and Bamba, B., "BioPatent Miner: An Information Retrieval System for BioMedical Patents," *Proc. 30<sup>th</sup> Toronto Conf. Very Large Databases (VLDB)*, 2004, pp. 1066-1077.
- [28] Mukherjea, S., Bamba, B., and Kankar, P., "Information Retrieval and Knowledge Discovery Utilizing a BioMedical Patent Semantic Web," *IEEE Trans. Knowledge and Data Eng.*, Vol. 17, No. 8, 2005, pp. 1099-1110.
- [29] National Research Council, "Assessing Research-Doctorate Programs: A Methodology Study," 2003.
- [30] Page, L., Brin, S., Motwani, R., and Winograd, T., "The PageRank Citation Ranking: Bringing Order to the Web," Technical Report, Stanford University, 1998.
- [31] Perron, O. and Frobenius, F.G., "Perron-Frobenius Theorem," Available at [http://en.wikipedia.org/wiki/Perron%E2%80%93Frobenius\\_theorem](http://en.wikipedia.org/wiki/Perron%E2%80%93Frobenius_theorem), 2007.
- [32] Prud'hommeaux, E. and Seaborne, A. eds., "SPARQL Query Language for RDF," *W3C Candidate Recommendation*, 14 June 2007.
- [33] Ren, J. and Taylor, R.N., "Automatic and Versatile Publications Ranking for Research Institutions and Scholars," *Communications of the ACM*, Vol. 50, No. 6, June 2007.
- [34] Schneider, P., Hayes, P., Horrocks, I. eds., "OWL Web Ontology Language Semantics and Abstract Syntax," *W3C Recommendation*, 10 February 2004.
- [35] Sheth, A., Aleman-Meza, B., Arpinar, I.B., Halaschek, C., and Ramakrishnan, C., "Semantic Association Identification and Knowledge Discovery for National Security Applications," *Special Issue. Jour. Database Tech. Enhancing National Security*, Vol. 16, No. 1, 2005, pp. 33-53.

[부록] 클래스중심 가중치 설정 예

<표 1> 시나리오 1- ClaRI: 클래스중심 가중치 설정 예(클래스 뷰)

| 1. 연구원 |       |     | 2. 논문      |      |            | 3. 키워드    |     |          | 4. 분야      |     |           | 5. 저널      |     |            |
|--------|-------|-----|------------|------|------------|-----------|-----|----------|------------|-----|-----------|------------|-----|------------|
| 발표하다   | 0.6   | 논문  | 실리다        | 0.5  | 저널         | (obj) 키워드 | 0.6 | (dom) 논문 | (obj) 분류되다 | 0.5 | (dom) 키워드 | (obj) 실리다  | 0.8 | (dom) 논문   |
| 등록하다   | 0.2   | 특허  | 참조되다       | 0.3  | 논문         | 분류되다      | 0.2 | 분야       | (obj) 다루다  | 0.3 | (dom) 저널  | 다루다        | 0.2 | 분야         |
| 저술하다   | 0.1   | 책   | (obj) 발표하다 | 0.1  | (dom) 연구원  | (obj) 키워드 | 0.2 | (dom) 특허 | (obj) 주제   | 0.2 | (dom) 책   | (obj) 링크하다 | 0   | (dom) 홈페이지 |
| 졸업하다   | 0.095 | 대학교 | (obj) 발표하다 | 0.1  | (dom) 교수   | :         | :   | :        | :          | :   | :         | 저널평가점수     | 0   |            |
| 국적     | 0.005 | 나라  | 키워드        | 0.05 | 키워드        |           |     |          |            |     |           | 저널명        | 0   |            |
| 가입하다   | 0     | 동호회 | (obj) 참조되다 | 0.05 | (dom) 논문   |           |     |          |            |     |           | :          | :   |            |
| 홈페이지   | 0     | 홈피  | (obj) 링크하다 | 0    | (dom) 홈페이지 |           |     |          |            |     |           |            |     |            |
| 연봉     | 0     |     | :          | :    |            |           |     |          |            |     |           |            |     |            |
| 생년월일   | 0     |     |            |      |            |           |     |          |            |     |           |            |     |            |
| 전화     | 0     |     |            |      |            |           |     |          |            |     |           |            |     |            |

| 6. 대학교     |      |           | 7. 나라      |     |           | 8. 출판사 |     |    | 9. 책       |     |           | 10. 특허     |     |           |
|------------|------|-----------|------------|-----|-----------|--------|-----|----|------------|-----|-----------|------------|-----|-----------|
| 고용하다       | 0.6  | 교수        | (obj) 국적   | 0.3 | (dom) 연구원 | 출판하다   | 0.8 | 책  | (obj) 출판하다 | 0.6 | (dom) 출판사 | (obj) 등록하다 | 0.5 | (dom) 연구원 |
| (obj) 졸업하다 | 0.2  | (dom) 교수  | (obj) 국적   | 0.3 | (dom) 교수  | 홈페이지   | 0.2 | 홈피 | 주제         | 0.2 | 분야        | (obj) 등록하다 | 0.5 | (dom) 교수  |
| (obj) 졸업하다 | 0.15 | (dom) 연구원 | (obj) 속하다  | 0.3 | (dom) 특허  | 당기순의   | 0   |    | (obj) 저술하다 | 0.2 | (dom) 연구원 | 속하다        | 0.3 | 나라        |
| 위치하다       | 0.05 | 나라        | (obj) 위치하다 | 0.1 | (dom) 대학교 | 대출액    | 0   |    | (obj) 저술하다 | 0.2 | (dom) 교수  | 키워드        | 0.2 | 키워드       |
| 홈페이지       | 0    | 홈피        | 1인당 국민소득   | 0   |           | 직원수    | 0   |    | 판매부수       | 0   |           | 기술가치 평가점수  | 0   |           |
| 학교평가점수     | 0    |           | 통신 인프라     | 0   |           | 주소     | 0   |    | :          | :   |           | :          | :   |           |
| 재학생수       | 0    |           | :          | :   |           | 전화번호   | 0   |    |            |     |           |            |     |           |
| :          | :    |           |            |     |           | :      | :   |    |            |     |           |            |     |           |

| 11. 교수     |       |           | 12. 동호회    |     |           | 13. 홈페이지   |      |            |
|------------|-------|-----------|------------|-----|-----------|------------|------|------------|
| 발표하다       | 0.5   | 논문        | (obj) 가입하다 | 0.5 | (dom) 연구원 | (obj) 링크하다 | 0.4  | (dom) 홈페이지 |
| 등록하다       | 0.15  | 특허        | 교류하다       | 0.3 | 동호회       | (obj) 홈페이지 | 0.2  | (dom) 출판사  |
| 저술하다       | 0.15  | 책         | 홈페이지       | 0.2 | 홈피        | (obj) 홈페이지 | 0.2  | (dom) 연구원  |
| (obj) 고용하다 | 0.1   | (dom) 대학교 | :          | :   |           | (obj) 홈페이지 | 0.2  | (dom) 대학교  |
| 졸업하다       | 0.095 | 대학교       |            |     |           | 링크하다       | 0.15 | 논문         |
| 국적         | 0.005 | 나라        |            |     |           | 링크하다       | 0.15 | 저널         |
| 홈페이지       | 0     | 홈피        |            |     |           | 링크하다       | 0.1  | 홈피         |
| 전화         | 0     |           |            |     |           | :          | :    |            |
| :          | :     |           |            |     |           | :          | :    |            |

시맨틱 웹 자원의 랭킹을 위한 알고리즘

<표 2> 시나리오 2- A, B: 클래스중심 가중치 설정 예(클래스 뷰)

| 1. 연구원 |      | 2. 논문 |            | 3. 키워드 |           | 4. 분야     |     | 5. 저널    |          |     |           |          |     |          |
|--------|------|-------|------------|--------|-----------|-----------|-----|----------|----------|-----|-----------|----------|-----|----------|
| 발표하다   | 0.5  | 논문    | 실리다        | 0.5    | 저널        | (obj) 키워드 | 0.6 | (dom) 논문 | 분류되다     | 0.5 | (dom) 키워드 | (obj)실리다 | 0.3 | (dom) 논문 |
| 등록하다   | 0.1  | 특허    | 참조되다       | 0.3    | 논문        | 분류되다      | 0.2 | 분야       | 다루다      | 0.3 | (dom) 저널  | 다루다      | 0.2 | 분야       |
| 저술하다   | 0.2  | 책     | (obj) 발표하다 | 0.1    | (dom) 연구원 | (obj) 키워드 | 0.2 | (dom) 특허 | (obj) 주제 | 0.2 | (dom) 책   | 저널평가점수   | 0.5 |          |
| 졸업하다   | 0.09 | 대학교   | (obj) 발표하다 | 0.1    | (dom) 교수  | :         | :   | :        | :        | :   | :         | 저널명      | 0   |          |
| 국적     | 0.01 | 나라    | 키워드        | 0.05   | 키워드       | :         | :   | :        | :        | :   | :         | :        | :   |          |
| 연봉     | 0.1  |       | (obj) 참조되다 | 0.05   | (dom) 논문  |           |     |          |          |     |           |          |     |          |
| 이름     | 0    |       | :          | :      |           |           |     |          |          |     |           |          |     |          |
| 생년월일   | 0    |       |            |        |           |           |     |          |          |     |           |          |     |          |
| 전화     | 0    |       |            |        |           |           |     |          |          |     |           |          |     |          |

| 6. 대학교     |      | 7. 나라     |            | 8. 출판사 |           | 9. 책 |     | 10. 특허 |            |     |           |            |     |           |
|------------|------|-----------|------------|--------|-----------|------|-----|--------|------------|-----|-----------|------------|-----|-----------|
| 고용하다       | 0.2  | 교수        | (obj) 국적   | 0.1    | (dom) 연구원 | 출판하다 | 0.5 | 책      | (obj) 출판하다 | 0.6 | (dom) 출판사 | (obj) 등록하다 | 0.1 | (dom) 연구원 |
| (obj) 졸업하다 | 0.1  | (dom) 교수  | (obj) 국적   | 0.1    | (dom) 교수  | 당기순의 | 0.2 |        | 주제         | 0.2 | 분야        | (obj) 등록하다 | 0.1 | (dom) 교수  |
| (obj) 졸업하다 | 0.1  | (dom) 연구원 | (obj) 속하다  | 0.1    | (dom) 특허  | 매출액  | 0.2 |        | (obj) 저술하다 | 0.1 | (dom) 연구원 | 속하다        | 0.1 | 나라        |
| 위치하다       | 0.01 | 나라        | (obj) 위치하다 | 0.1    | (dom) 대학교 | 직원수  | 0.1 |        | (obj) 저술하다 | 0.1 | (dom) 교수  | 키워드        | 0.2 | 키워드       |
| 학교평가점수     | 0.4  |           | 1인당 국민소득   | 0.4    |           | 주소   | 0   |        | 판매부수       | 0.5 |           | 기술가치 평가점수  | 0.6 |           |
| 재학생수       | 0.19 |           | 통신 인프라     | 0.2    |           | 전화번호 | 0   |        | :          | :   |           | :          | :   | :         |
| :          | :    |           | :          | :      |           | :    | :   |        | :          | :   |           | :          | :   | :         |

| 11. 교수     |       |           |
|------------|-------|-----------|
| 발표하다       | 0.4   | 논문        |
| 등록하다       | 0.2   | 특허        |
| 저술하다       | 0.2   | 책         |
| (obj) 고용하다 | 0.1   | (dom) 대학교 |
| 졸업하다       | 0.009 | 대학교       |
| 국적         | 0.01  | 나라        |
| 전화번호       | 0     |           |
| :          | :     |           |

◆ 저자소개 ◆



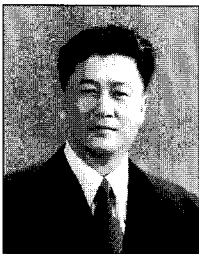
노상규 (Rho, Sangkyu)

서울대학교 경영학과를 졸업하고 미국 미네소타 대학에서 MBA 및 경영학 박사학위를 취득하였으며 현재 서울대학교 경영전문대학원/경영대학 교수로 재직 중이다. 주요 연구분야로는 인터넷 비즈니스, 온톨로지, 정보시스템 모델링, 데이터 마이닝 등이 있으며 IEEE Transactions on Knowledge and Data Engineering, Strategic Management Journal, Long Range Planning, Annals of Operations Research, Journal of Database Management 등 우수 학술지에 다수의 논문을 게재하였다. 저서로는 '인터넷 진화의 열쇠 온톨로지: 웹 2.0에서 3.0,' '한국온라인게임산업의 발전과정과 향후과제' 등이 있다.



박현정 (Park, Hyunjung)

현재 서울대학교 경영학과에서 MIS 전공으로 박사과정에 재학 중이다. KAIST 경영과학과를 졸업하고 동 대학원 경영과학과에서 정보통신시스템 최적화로 석사학위를 취득하였으며, KT 전송기술연구소와 경영연구소에서 전임연구원으로 재직하였다. 주요 관심분야는 시맨틱 웹과 온톨로지, Ranking Algorithm, Knowledge Management, U-Learning 등이다.



박진수 (Park, Jinsoo)

현재 서울대학교 경영전문대학원/경영대학 부교수로 재직 중이다. The University of Arizona에서 경영정보시스템을 전공하여 경영학 박사를 취득했으며, University of Minnesota의 Carlson School of Management에서 조교수, 고려대학교 경영대학에서 조교수를 역임했다. 주요 관심분야는 온톨로지, 정보시스템 통합, 지식 공유, 에이전트, 시맨틱 모델링, 웹 정보시스템 등이 있다. MIS Quarterly, IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Computer, ACM Transactions on Information Systems (TOIS), Information Systems Frontiers, Communications of the AIS, Journal of Global Information Technology Management (JGITM), International Journal of Electronic Business, Information Systems Review, 한국경영정보학연구, 한국전자거래학회지 등 우수 학술지에 다수의 논문을 게재하였으며, 대표 저서로는 '인터넷 진화의 열쇠 온톨로지: 웹 2.0에서 3.0,'가 있다.

◆ 이 논문은 2007년 05월 15일 접수하여 1차 수정을 거쳐 2007년 12월 02일 게재확정되었습니다.