

# Chapter 1

## A Link-Based Ranking Algorithm for Semantic Web Resources: A Class-Oriented Approach Independent of Link Direction

**Hyunjung Park**

*Seoul National University, Korea*

**Sangkyu Rho**

*Seoul National University, Korea*

**Jinsoo Park**

*Institute of Management Research, Seoul National University, Korea*

### ABSTRACT

*The information space of the Semantic Web has different characteristics from that of the World Wide Web (WWW). One main difference is that in the Semantic Web, the direction of Resource Description Framework (RDF) links does not have the same meaning as the direction of hyperlinks in the WWW, because the link direction is determined not by a voting process but by a specific schema in the Semantic Web. Considering this fundamental difference, the authors propose a method for ranking Semantic Web resources independent of link directions and show the convergence of the algorithm and experimental results. This method focuses on the classes rather than the properties. The property weights are assigned depending on the relative significance of the property to the resource importance of each class. It solves some problems reported in prior studies, including the Tightly Knit Community (TKC) effect, as well as having higher accuracy and validity compared to existing methods.*

DOI: 10.4018/978-1-4666-2044-5.ch001

## INTRODUCTION

The Semantic Web can be regarded as a smart web designed to understand the requests of people and machines to use Web contents (Berners-Lee, 2001), and one of the most common requests may be the ranking of Semantic Web resources. Therefore the Web had better have some understanding of the concept of ranking, and the ranking mechanism adequate for a specific domain could be defined in the domain ontology. However, there have not been many studies on ranking in the Semantic Web, while there have been extensive studies on the evaluation of the World Wide Web.

Traditionally, the importance of a particular Web page is estimated based on the number of keywords found on the page, which is subject to manipulation (Marchiori, 1997). In contrast, link analysis methods such as Google's PageRank (Brin et al., 1998; Haveliwala, 1999; Page et al., 1998) capitalize on the information that is inherent in the link structure of a Web graph. PageRank considers a page important if it is referred to by many other pages. The degree of importance also increases if the importance of the referring pages is high. Kleinberg's Hypertext-Induced Topic Selection (HITS) algorithm (Kleinberg, 1998) is another link-structure-based ranking algorithm for Web pages. The HITS algorithm differs from PageRank in that it utilizes two kinds of scores: an authority score and a 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. The link-structure-based ranking method has become an essential tool for using WWW, and its effectiveness and efficiency have been widely recognized.

On the other hand, information from the Semantic Web can be expressed using a Resource Description Framework (RDF) graph (Klyne et al., 2004; Manola et al., 2004). An RDF graph, in which a resource and a property are expressed as a node and a link, respectively, is similar to a

WWW graph in which a Web page and a hyperlink are expressed as a node and a link, respectively. Consequently, research on methods for applying the link-structure-based ranking technique of WWW to an RDF graph of the Semantic Web has great significance. The WWW graph can be thought of as an enormous class of the Web pages with only one recursive property called a 'refer to' property, so to speak. An RDF schema, in contrast, can have various classes and properties, and each link corresponding to a property can have an opposite direction depending on whether it is an active or a passive voice. As a result, RDF schemas can have many different forms because the direction of each link is changeable, even if they describe the same thing, and the direction of an RDF link does not have the same meaning as that of a WWW hyperlink. In WWW, if a page is pointed to by a directing link, we can tell that the pointed page must have some useful information. PageRank and HITS are based on this basic assumption.

Although there were some attempts made to modify the HITS algorithm to rank query results retrieved from RDF knowledge bases (Bamba & Mukherjea, 2004; Mukherjea & Bamba, 2004; Mukherjea et al., 2005), this fundamental difference has been overlooked. These attempts used the objectivity and subjectivity scores of a resource, which correspond to the authority and hub scores of a page, respectively, from Kleinberg's algorithm. The objectivity score of a certain resource is calculated by summing up all the subjectivity scores of the resources linking to it after they are multiplied by appropriate property weights. The subjectivity score of a resource is calculated similarly. If the direction of a link in a triple is changed by a schema, the objectivity and the subjectivity of the triple are reversed, causing the objectivity scores and the subjectivity scores of the linked resources to be changed. Besides this unobserved problem, there is a reported limitation of the Tightly Knit Community (TKC) effect where resources that are less important but densely connected are given

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/link-based-ranking-algorithm-semantic/74387](http://www.igi-global.com/chapter/link-based-ranking-algorithm-semantic/74387)

## Related Content

---

### Methods for the Identification of Data Outliers in Interactive SQL

Ronald Dattero, Edna M. White and Marius A. Janson (1991). *Journal of Database Administration* (pp. 7-18).

[www.irma-international.org/article/methods-identification-data-outliers-interactive/51083](http://www.irma-international.org/article/methods-identification-data-outliers-interactive/51083)

### Adaptive Modularized Recurrent Neural Networks for Electric Load Forecasting

Fangwan Huang, Shijie Zhuang, Zhiyong Yu, Yuzhong Chen and Kun Guo (2023). *Journal of Database Management* (pp. 1-18).

[www.irma-international.org/article/adaptive-modularized-recurrent-neural-networks-for-electric-load-forecasting/323436](http://www.irma-international.org/article/adaptive-modularized-recurrent-neural-networks-for-electric-load-forecasting/323436)

### An Investigation of the Impact of Organization Size on Data Quality Issues

G. Daryl Nord, Jeretta Horn Nord and Hongjiang Xu (2005). *Journal of Database Management* (pp. 58-71).

[www.irma-international.org/article/investigation-impact-organization-size-data/3337](http://www.irma-international.org/article/investigation-impact-organization-size-data/3337)

### Document Versioning in Digital Libraries

Mercedes Martínez-González (2005). *Encyclopedia of Database Technologies and Applications* (pp. 201-205).

[www.irma-international.org/chapter/document-versioning-digital-libraries/11146](http://www.irma-international.org/chapter/document-versioning-digital-libraries/11146)

### Using Temporal Versioning and Integrity Constraints for Updating Geographic Databases and Maintaining Their Consistency

Wassim Jaziri, Najla Sassi and Dhouha Damak (2015). *Journal of Database Management* (pp. 30-59).

[www.irma-international.org/article/using-temporal-versioning-and-integrity-constraints-for-updating-geographic-databases-and-maintaining-their-consistency/140545](http://www.irma-international.org/article/using-temporal-versioning-and-integrity-constraints-for-updating-geographic-databases-and-maintaining-their-consistency/140545)