# Chapter 8
# We Have Good Information for You:
## Cognitive Authority and Information Retrieval on the Web

**Filipe Roseiro Côgo**
*Federal University of Technology at Paraná (UTFPR), Brazil*

**Roberto Pereira**
*University of Campinas (UNICAMP), Brazil*

## ABSTRACT

*Through the concept of Cognitive Authority, information relevance and quality have been related to the expertise/skill of those who publish and share information on the Web. This chapter discusses how the concept of cognitive authority can be used in order to improve the information retrieval on folksonomy-based systems. The hypothesis is that a ranking scheme that takes into account the cognitive authority of the information sources provides results of higher relevance and quality to users. To verify this hypothesis, the Folkauthority approach is adopted; a ranking scheme called AuthorityRank is proposed; and an information retrieval system, named AuthoritySearch, is built. A real social network is used to simulate the authority relationship among users, and the AuthorityRank scheme is compared with the tf-idf scheme using the NDCG metric. The results indicate a statistically significant improvement in the quality and relevance of the information obtained through the use of the AuthorityRank scheme.*

## INTRODUCTION

The consolidation of the Web 2.0 concept favored the production and publishing of information in an easier way, without rigid control or verification of quality (Murugesan, 2007). Therefore, the production of information has taken place on a large scale that generated the so-called information overload effect (Himma 2007), demanding new ways and mechanisms for organizing and retrieving information (Plale, 2013).

Traditionally, the information retrieval and organization have been handled by using two main approaches: i) classification/categorization

schemes, which often need to be elaborated by experts; and ii) search engines, which are based on the automatic indexing and retrieval of documents. Currently, search engines represent most of the systems used for retrieving information on the Web. However, they are not yet able to address the issue of efficiently filtering its large amount of information.

In this sense, a third approach, combining both the categorization and the automatic indexing approaches, has emerged over the last decade. Named *tagging* (Trant, 2009), this approach is characterized by the use of terms (*tags*) generated by the users themselves to both describe and retrieve information in an information retrieval system (IRS). Web 2.0 applications like *Delicious*®, *Flickr*® and *Technorati*® employ the *folksonomy* approach, a tagging-based and collaborative way for manually indexing the information.

The Folksonomy-based systems (FBS) benefit from the social characteristics of Web 2.0 to improve the information organization, management and retrieval. Some of the aspects discussed by researchers as inherent benefits from folksonomies are the sense of community generated by the use of the technique, the explicit opinion from a set of users about the available information, and the possibility to reflect, almost in real time, changes on the vocabulary utilized to express about the resources (Golder and Huberman, 2006, Halpin et al., 2007, Sen et al., 2006).

In a FBS, the categorization process (Trant, 2009) — (i.e., the assignment of tags to documents for describing it or its meaning) — is held by the users themselves. This approach makes the quality of categorization directly dependent on who carried it out. By quality of categorization, we mean the use of terms that describes the information accordingly, making sense to a group of people while avoiding unnecessary ambiguities. Thus, getting better results in terms of organization and indexing of documents in FBSs depends on the

knowledge and skills of users who are performing the categorization, raising the issue of identifying the cognitive authority of the sources of information. Cognitive authority determines "who knows what about what" (Wilson, 1983), being related to the influence caused by someone in the way of thinking of an individual, because this individual judges him worthy of credit and trust.

The *Folkauthority* approach, proposed by Pereira and da Silva (2008), tries to bring to the information retrieval scenario the benefits of a social strategy people use for obtaining information in their daily activities and situations: asking for someone who may help them. *Folkauthority* considers that, as each user (an *entity*) is able to publish information (i.e., each user can be a source of information), it is also able to categorize the competence of an information source. Therefore, it suggests attributing tags to these entities in order to categorize each source of information according to their cognitive authority, generating a *meta-categorization*. The purpose of such tags, however, is not to describe what an entity "is", but to specify what it "knows", in what it is a reference, trustworthy, from the categorizer's point of view. Therefore, when retrieving information, it is possible to identify and prioritize the information produced/shared by those who are experts in a given subject for the person who is retrieving the information.

In this chapter, we consider that a ranking scheme that takes into account the cognitive authority of the information sources provides results of higher relevance and quality to users. By adopting *Folkauthority* approach, it is possible to improve the relevance and quality of the results of a query by given more importance to certain sources of information when calculating the ranking of the retrieved information. To verify this hypothesis, the *Folkauthority* approach was adopted; a ranking scheme called *AuthorityRank* was proposed; and an information retrieval sys-

## Related Content

Pattern Based Feature Construction in Semantic Data Mining

Agnieszka awrynowiczand Jdrzej Potoniec (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications  (pp. 2084-2122).*

www.irma-international.org/chapter/pattern-based-feature-construction-in-semantic-data-mining/142719

A Taxonomy of Data Mining Problems

Nayem Rahman (2018). *International Journal of Business Analytics (pp. 73-86).*

www.irma-international.org/article/a-taxonomy-of-data-mining-problems/201454

Stock Market's Reactions to Industrial Accidents: Evidence from Chinese Listed Companies

Jiuchang Wei, Han Wangand Xiumei Guo (2014). *International Journal of Business Analytics (pp. 18-33).*

www.irma-international.org/article/stock-markets-reactions-to-industrial-accidents/115518

Game Theory for Cost Allocation in Healthcare

Alexander Kolker (2014). *Encyclopedia of Business Analytics and Optimization (pp. 1067-1079).*

www.irma-international.org/chapter/game-theory-cost-allocation-healthcare/107305

Analyzing Business Intelligence Systems Based on 7s Model of McKinsey

Amin A. Shaqrah (2018). *International Journal of Business Intelligence Research (pp. 53-63).*

www.irma-international.org/article/analyzing-business-intelligence-systems-based-on-7s-model-of-mckinsey/203657