Chapter 10
Hidden Markov Models for Context–Aware Tag Query Prediction in Folksonomies

Chiraz Trabelsi
University Tunis El-Manar, Tunisia

Bilel Moulahi
University Tunis El-Manar, Tunisia

Sadok Ben Yahia
University Tunis El-Manar, Tunisia

ABSTRACT

Recently, social bookmarking systems have received surging attention in academic and industrial communities. In fact, social bookmarking systems share with the Semantic Web vision the idea of facilitating the collaborative organization and sharing of knowledge on the web. The reason for the apparent success of the upcoming tools for resource sharing (social bookmarking systems, photo sharing systems, etc.) lies mainly in the fact that no specific skills are needed for publishing and editing, and an immediate benefit is yielded to each individual user, e.g., organizing one's bookmarks in a browser-independent, persistent fashion, without too much overhead. As these systems grow larger, however, the users address the need of enhanced search facilities. Today, full-text search is supported, but the results are usually simply listed decreasingly by their upload date. The challenging research issue is, therefore, the development of a suitable prediction framework to support users in effectively retrieving the resources matching their real search intents. The primary focus of this chapter is to propose a new, context aware tag query prediction approach. Specifically, the authors adopted Hidden Markov Models and formal concept analysis to predict users' search intentions based on a real folksonomy. Carried out experiments emphasize the relevance of the proposal and open many issues.

INTRODUCTION

Complementing the Semantic Web effort, a new breed of so-called Web2.0 applications recently emerged on the Web. Indeed, social bookmarking systems, such as *e.g.*, *Del.icio.us*[1], *Bibsonomy*[2] or *Flickr*[3] have become the predominant form of content categorization of the Web2.0 age. The main thrust of these Web2.0 systems is their easy use that relies on simple, straightforward structures (folksonomies) by allowing their users to label diverse resources with freely chosen keywords *aka* tags. Social bookmarking systems share with the Semantic Web vision, the idea of facilitating the collaborative organization and sharing of knowledge on the web.

However, a main difference lies in the fundamentally opposite approach: the Semantic Web aims at a formal knowledge representation in form of ontologies (written in XML, RDF, or OWL), whereas social bookmark tools follow a grass-root approach: there are no limitations on the kind of tags users may select. In contrast to ontologies (Gruber, 1993), the resulting structures are called folksonomies, that is, *"taxonomies"* created by the *"folk."*

Considered as a tripartite hyper-graph (Mika, 2005) of tags, users and resources, the new data of folksonomy systems provides a rich resource for data analysis, information retrieval, and knowledge discovery applications. In fact, the success of folksonomies originated from members' ability to centrally collect and manage content collections on the web, overcoming local storage policies. Users of folksonomies are granted ubiquitous access to their collections of photos, web sites, or publications regardless their current location or the currently used device. These personal advantages came in conjunction with a social component. This derives from the fact that most folksonomies allow the sharing of content with other users, the discovery of content other users considered interesting, or the communication with other users through various channels.

Hence, aggregating the interests of up to millions of users, folksonomies reflect the dynamics of the underlying domains. High traffic websites will, then, likely reappear among the popular resources in a community such like *Del.icio.us* (Heymann *et al.*, 2008). Furthermore, external trends, *e.g.*, the advent of a new web site or a new influential blog entry, were found to reappear in folksonomies almost without any delay (Heymann *et al.*, 2008). Combining these characteristics with the existence of user-generated annotations, folksonomies have become an invaluable source for information retrieval (IR) (Pan et *al.*, 2009). These benefits come at a low cost, since all information is centrally available, and no continuous, distributed crawling process as for web indexing is required. Indeed, one of the main services provided by social tagging systems is searching. Searching occurs when the user enters a tag as a query and a, ranked by relevance, list of related resources are yielded to the user. Even though collaborative tagging applications have many benefits, they also present some thriving challenges for Information Retrieval (IR). Actually, the core of many search engines is the ranking algorithm. However, the most currently used ranking algorithms are not straightforwardly adaptable to folksonomies. Furthermore, these traditional tools for web information retrieval constitute a hindrance, since they do not consider neither social nor behavioral facts into account in the retrieval task of resources nor help understanding user's information needs.

In addition, one significant problem arising from such free-formedness stands in the tag ambiguity: tags that have several meanings, *e.g.*, *"Java"* as coffee or a programming language or an island in Indonesia. Moreover, folksonomy tags are unstructured as assigned tags to a given resource are simply enumerated in a list and finally, no special organization or categorization of the tags is made (by the folksonomy site). As per Golder and Huberman (2005), the main problems of social tagging systems include ambiguity, lack of synonymy and discrepancies in granularity. Specifically, when a

Related Content

Analysis Method of App Software User Experience Based on Multisource Information Fusion
Yongquan Chen, Ying Jiangand Haiyi Liu (2023). *International Journal on Semantic Web and Information Systems (pp. 1-22).*
www.irma-international.org/article/analysis-method-of-app-software-user-experience-based-on-multisource-information-fusion/325216

A Pattern-Based Method for Re-Engineering Non-Ontological Resources into Ontologies
Boris Carmen Villazón-Terrazas, Mari Suárez-Figueroaand Asunción Gómez-Pérez (2012). *Semantic-Enabled Advancements on the Web: Applications Across Industries (pp. 17-54).*
www.irma-international.org/chapter/pattern-based-method-engineering-non/64016

XAR: An Integrated Framework for Semantic Extraction and Annotation
Naveen Ashishand Sharad Mehrotra (2010). *Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications (pp. 235-254).*
www.irma-international.org/chapter/xar-integrated-framework-semantic-extraction/38046

Semantic-Based Optimization of Deep Learning for Efficient Real-Time Medical Image Segmentation
Zhenkun Wei, Jia Liuand Yu Yao (2024). *International Journal on Semantic Web and Information Systems (pp. 1-16).*
www.irma-international.org/article/semantic-based-optimization-of-deep-learning-for-efficient-real-time-medical-image-segmentation/340938

In Defense of Ambiguity
Patrick J. Hayesand Harry Halpin (2008). *International Journal on Semantic Web and Information Systems (pp. 1-18).*
www.irma-international.org/article/defense-ambiguity/2847