



HAL
open science

Information Retrieval enhancement using Social Networks

Bougchiche Lilia

► **To cite this version:**

Bougchiche Lilia. Information Retrieval enhancement using Social Networks. International Symposium on Web Algorithms, Jun 2015, Deauville, France. hal-01171327

HAL Id: hal-01171327

<https://hal.science/hal-01171327>

Submitted on 3 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Information Retrieval enhancement using Social Networks

Boughchiche Lilia
University of Caen,
Low Normandy

Lilia.boughchiche@unicaen.fr

Abstract

The aim of this paper is to present a preliminary work about the enhance of information retrieval systems by exploring tags coming from social networks. We discuss how to obtain folksonomies (social indexing) and their use on a search approach that models tags assigned to documents. For this, a document is represented by formal concept then we construct a bipartite graph. From this later, we retrieve emergent tags that describe better a document.

I. Introduction

Since the arise of numerical social networks (such as Facebook¹, Flickr², Twitter³, etc.), users got the possibility to create their own content (photos, movies,...) and share it. These new interactions allow users to be more actives. Cognitive intelligence emerges and social information with added valued is produced such as tags (labels, keywords), comments, judgments, bookmarks...[6]. Figure 1, shows an example of these interactions on a social tagging system Twitter.



Figure 1 : An example of social interaction

We can see a message (a tweet⁴) with URL «*capdigital.com/strategies/big...*» described with a tag #BigDataLive shared between 5 users (@Cap Digital, @dataiku, @opendatasoft, @fdouetteau et @jmlazard) on the date of Novembre, 18th.

Other situations of tagging task may be considered in different fields (political, e-trade, scientific domain,...) where comments on events, persons, products, organizations,... can be annotated by many hundred of tags but some of them do not reflect really the described elements.

Our purpose is to find which tags represent better a resource. Are they the most recent ones? Or the most popular ones?

Representing a document is one of the most important steps of the Information retrieval process. This leads us

to exploit social tags, in our work, to give a new model of document based on concepts that could enhance an information retrieval system.

II. Our approach description

Our study aims to find tags that represent well a document. For this purpose, we propose an approach based on the steps presented as follow:.

First, we collect information on a social network (ex. Flickr) about tags, documents and users. These components constitute a tag cloud or a folksonomy [11] (see Figure 2 bellow).

U/R-T	r ₁					r ₂					r ₃				
	t ₁	t ₂	t ₃	t ₄	t ₅	t ₁	t ₂	t ₃	t ₄	t ₅	t ₁	t ₂	t ₃	t ₄	t ₅
u ₁		x	x	x			x	x	x				x	x	x
u ₂		x	x	x		x	x	x	x		x	x	x	x	
u ₃		x	x	x		x	x	x	x		x	x	x	x	
u ₄						x			x		x			x	
u ₅		x	x	x	x		x	x	x	x		x	x	x	x
u ₆				x	x				x	x					
u ₇	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Figure 2: A folksonomy

In a folksonomy, a document r_1 is describes by the tags t_1, t_2, t_3, t_4 and t_5 . These tags do not exist alone, they depend on the users (u_1, u_2, u_3, u_5, u_6 and u_7). We see on the figure (Figure 2) that the user u_4 does not annotate the document r_1 .

For modeling a document r_i we need to apply clustering methods on our folksonomy. Formal Concepts Analysis (FCA⁵) [4] is a good mean to do this. A formal concept c_i is a set of all documents shared by a set of users and sharing a set of tags $c_i = (\{d_i\}, \{t_j\}, \{u_k\})$.

Let's define a formal context $K = (D, T, U, R)$ with $R \subseteq D \times T \times U$ a relation (boolean) defined on a set of documents D (the extent), a set of properties (les tags) T the intent of a concept and a set of users U . A formal context is generally represented by an adjacency matrix.

A concept formal $c_i = (\{d_i\}, \{t_j\}, \{u_k\})$ is a tri-concept that represents all documents $\{d_i\}$ shared between users $\{u_k\}$ using tags $\{t_j\}$.

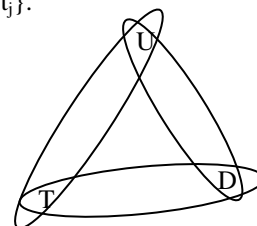


Figure 3 : An example of tripartite graph

1 <https://www.facebook.com>

2 <http://flickr.com>

3 <http://twitter.com>

4 Messages of maximum 140 caracters.

⁵ AFC : Analyse des Concepts Formels.

The next step is to represent the formal context by graphs that we know manipulating. So, to the three dimensional formal context K constituted by tri-concepts, can correspond a tripartite graph $G=(V,E)$ (as shown in Figure 3), where $V= (U,T,D)$ is a set of vertices and $E \subseteq D \times T \times U$ a set of edges between the different nodes sets.

If we project this tripartite graph on two dimensional axes without loss of information between sets, we might obtain three bipartite graphs $G1=((D,T),E1)$ with $E1 \subseteq D \times T$, $G2=((D,U),E2)$ with $E2 \subseteq D \times U$ and $G3=((T,U),E3)$ with $E3 \subseteq T \times U$.

On the next steps we focus our work on the bipartite graph $G1$ describing relations between documents and tags, but the same process will be applied to the others bipartite graphs ($G2$ and $G3$).

The bipartite graph $G1$ represents all tags $\{t_j\}$ describing a document d_i . On this graph, we search emergent patterns that gather tags which represent better a document. Searching $G1$ will be more efficient using the bipartite graph Document-Concept proposed by Navarro et al. [8] based on Infomap algorithm [9] that allows fast retrieval. In his study [8], he proves that the searching results are more significant than if the document-tag graph is used.

III. Related work

In the last decades, work on Information Retrieval (IR) does not base only on traditional models [3]. Several studies enhance IR process by using concepts of social networks like folksonomies, friendship, comments, votes, judgments, and so on, to do recommendations or search personalization.

Folksonomy's components modelisation (documents, users and tags) attracted many researchers [7]. Some users characteristics like social influence, authorship were of interest [2]. Lot of studies concentrate on tags exist. Some researchers use bayesian networks (probabiliste models) to model social annotations [1] [12]. Others use conceptual models [4] matching documents concepts (represented by tags which are document topics) with users concepts which elements are tags (user's interests) [4].

In litterature, graphs [10] are used to represent folksonomies. Data mining technics are used to analyse folksonomies and to extract knowledge from great data sets.

IV. Conclusion and outlooks

In this paper, we presented a preliminary strategy to give a document's social representation based on tags assigned by users on social networks. A lot of tags may be attributed to a single document, from which we can find those are useful to describe a document, those are more representative for a document or others non important.

More representative tags for a document appear on emergent concepts given by emergent patterns.

As future work, we intend to study other social factors like trends, votes, likes. We also intend to implement a

complete information retrieval system using the discussed ideas.

References

- [1] Bao Shenghua, Xu Shengliang, Yunbo Cao et Yong Yu, "Using social annotation to improve language model of IR", CIKM'07, November 6-8, Lisboa, Portugal, 2007.
- [2] Ben Jabeur, Lamjed, Leveraging social relevance : using social networks to enhance literature access and microblog search, doctorat de l'université de toulouse 3 Paul Sabatier (UT3 Paul Sabatier) 2013.
- [3] Boughanem M. 2008.« Introduction à la recherche d'information ». In *Mohand Boughanem and Jacques Savoy, editors, recherche d'information : état des lieux et perspectives*, chapter1, pages1944. Lavoisier, Paris, France.
- [4] Chi Wang, Rajat Raina, David Fong, " Learning Relevance in a Heterogeneous Social Network and Its Application in Online Targeting" , *SIGIR'11* , Beijing, China Jul 24–28, 2011.
- [5] Ganter & Wille, Formal Concept Analysis. Mathematical Foundations, Springer Verlag, 1999.
- [6] Hammond Tony and Hannay Timo and Lund Ben and Scott Joanna. Social Bookmarking Tools (I): A General Review. D-Lib Magazine, 2005.
- [7] A. Hotho, R. Jaschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *ESWC 2006: Proc. of the 3rd European Semantic Web Conference*, 2006.
- [8] Navarro E., Prade H., Gaume B., Clustering Sets of Objects Using Concepts-Objects Bipartite Graphs", In *Scalable Uncertainty Management - 6th International Conference, SUM 2012, Marburg, Germany* (Eyke Hüllermeier, Sebastian Link, Thomas Fober, Bernhard Seeger, eds.), Springer, vol. 7520, pp. 420-432, 2012.
- [9] Rosvall, M. and Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*,105(4):1118_1123, 2008.
- [10] Schmitz C., Grahl,A.Hotho M., Stumme G., C.Cattuto,A.Baldassarri, V.Loreto, V.DPServedio, "Network properties of folksonomies", *Proceedings of the WWW2007 International World Wide Web Conference*, 2007.
- [11] Vander Wal, Thomas. 2004. You down with folksonomy? August 4, 2004. Retrieved, from <http://www.vanderwal.net/random/entrysel.php?blog=1529>
- [12] Zhou D., Bian J.,Zheng S., Zha H., C. Lee Giles, Exploring Social Annotations for Information Retrieval, *World Wide Web Conference*, Beijing, China, 2008.