Personalized Image Search Optimization

Surekha R. Janrao [#1], Dr. Lata Ragha [#2]

[#] Dept. Computer Engineering, Terna Engineering College
Nerul, Navi Mumbai-400706, MS, India.

[1] surekh_99@yahoo.co.in
[2] lata.ragha@gmail.com

Abstract— **Existing work for personalized image search has been defined by using novel framework which provides us more specific and more relevant result for image search as compared to non-personalized image search. But still we are not able to achieve more optimized and quick search by using the mechanism provided in that framework. Because only ranking based multi- correlation tensor factorization and user specific topic modeling is not that much sufficient mechanism for personalized image search. To improve the performance for personalized image search a semantic search engine can be used for more personalized image search. In semantic image search PSSE (Personalized Semantic Search Engine) architecture has been used to achieve the potential of social networks. It uses the user profiles with the ranking score and ontology for calculation of personalized factor which will help us to get more personalized search result. The architecture of PSSE has two parts, Offline and Online. The Offline part consists of crawling and preprocessing processes. The Online phase includes query processing and result ranking**

Keywords— **Social Media**: A term that defines activities by which users create and publish content on the Web. Examples include Flickr, del.icio.us, Digg and many others.
Personalization: Algorithms and techniques that tailor content to individual users.
Image search: A type of Web search that returns images matching a given (text) query.
Metadata: Data about data.
Tag: A freely-chosen keyword or term associated with content by the user.
PSSE: **Personalized semantic search engine**

I. INTRODUCTION

Social sharing websites provide rich resources that can be exploited for personalized search evaluation. User's social activities, such as rating, tagging and commenting, indicate the user's interest and preference in a specific document. Recently, two types of such user feedback are utilized for personalized search evaluation. The first approach is to use social annotations. The main assumption behind is that the documents tagged by user with tag will be considered relevant for the personalized query.

The rise of the Social Web underscores a fundamental transformation of the Web. Rather than simply searching for, and passively consuming information, users of blogs, wikis and social media sites like del.icio.us, Flickr, are creating, evaluating, and distributing information. In the process of using these sites, users are generating not only content that could be of interest to other users, but also a large quantity of metadata in the form of tags and ratings, which can be used to improve Web search and personalization. As search engines became popular amongst Netizens, a need of semantic search has become a necessity. As the context and semantics of the information in the web pages indexed depends on multiple factor, semantic search has become a complex task. Google Research Lab has worked out many generic algorithms however they are successful in certain conditions only. A focus on personalization of semantic search was given where in one can restrict the domain set and search parameters depending on the personal information [4].

II. Related Work for Personalized Image Search

A. Flow Diagram Figure 1 defines the flow of novel framework of the existing system which is the collective mechanism of the online and offline stage systems approach. The novel framework has been designed by Jitao Sang, Changsheng Xu Dongyuan Lu by considering the user preference and query-related search intent into user-specific topic spaces. These approaches are used to achieve the users' social activities for personalized image search, such as annotations and the participation of interest groups. The query relevance and user preference are simultaneously integrated into the final rank list [1], [2]. Experiments on a large-scale Flickr dataset show that the novel framework greatly outperforms the baseline. To investigate on user preference and perform user modelling, the popular social activity of tagging is considered. Collaborative tagging has become an increasingly popular means for sharing and organizing resources, leading to a huge amount of user-generated annotations [3].Online photo sharing websites, such as Flickr, Picasa, Zooomr and Pinterest allow users as owners, taggers, or commenters for their contributed contents to interact and collaborate with each other in a social media dialogue. Various researchers have investigated the applicability of social annotations to improve web search [2]. Recently, social annotations are employed for automatic evaluation of personalized search.

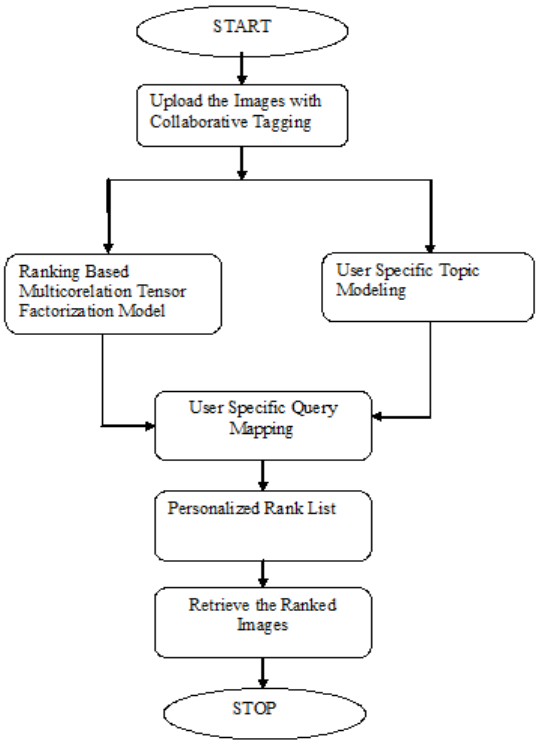

Fig. 1 Flow Diagram for personalized image search

B Detailed Description of Flow Diagram for personalized image search

*1) Uploading the Images with Tagging Data Collection:*In this section we focus on Tagging which includes different types of metadata, including social networks. Tags are freely-chosen keywords users associate with content. Tagging was introduced as a means for users to organize their own content in order to facilitate searching and browsing for relevant information. The distinguishing feature of tagging systems is that they use an uncontrolled vocabulary, and that the user is free to highlight any one of the object's properties. From an algorithmic point of view, tagging systems offer many challenges that arise when users try to attach semantics to objects through keywords. These challenges are homonymy (the same tag may have different meanings), polysemy (tag has multiple related meanings), synonymy (multiple tags have the same meaning), and "basic level" variation (users describe an item by terms at different levels of specificity, e.g., "beagle" vs. "dog"). Despite these challenges, tagging is a light weight, flexible categorization system. The growing amount of tagged content provides evidence that users are adopting tagging on Flickr, Del.icio.us and other collaborative tagging systems [4]. Following is the example of Tagging with tiger.

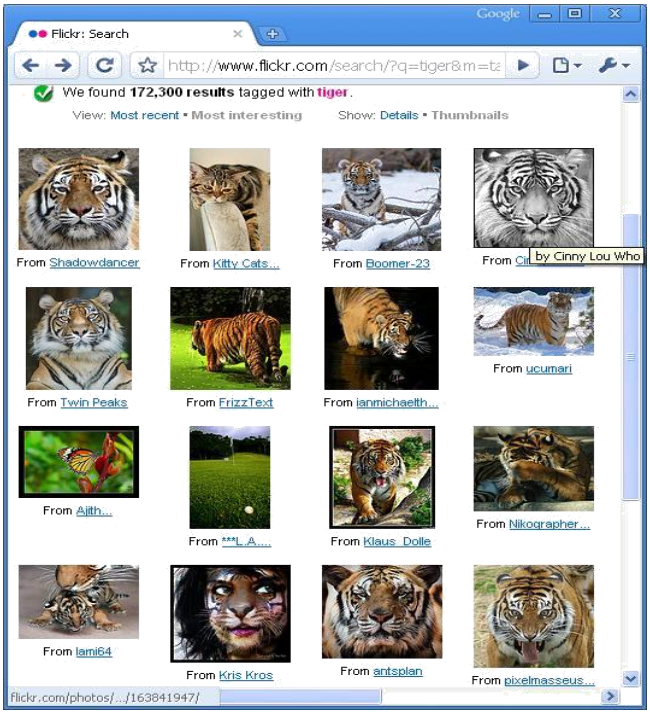

Fig. 2 Tagging of Images

*2) RMTF (Ranking Based Multi-correlation Tensor factorization Model):*To improve the sparsity and noisy problem, we present a novel method named Ranking based Multi-correlation Tensor Factorization (RMTF) to better leverage the observed tagging data for users' annotation prediction. Zhu ET. al. [2] has demonstrated that the semantic space spanned by image tags can be approximated by a smaller subset of salient words from the original space. Illustrated by this, we employ low rank approximation to extract the compact representation for image, tag and user, and at the same time reconstruct the user-image-tag ternary relations for annotation prediction. With the observed user-tag-image ternary relations as input, the reconstructed ternary relations can be viewed as user's potential annotations for the images. Following the assumption we mentioned in the introduction, we can straightly utilize the predicted user annotations for Personalized image search, i.e., if a user has a high probability to assign the tag t to an image, the image should be ranked higher when the user issues query t. However, this formulation has two problems. 1) It is unreasonable to assign the query to a single tag in the tag vocabulary, e.g., when a user searches "cheer dance", he/she would like the images that he/she annotated with semantic related tag "cheerleader" are also ranked higher. 2) There are variations in individual user's tagging patterns and vocabularies, e.g., the tag "jaguar" from an animal specialist should be related to "leopard", while a car fan will consider "jaguar" more related to "autos". To address the two problems, we perform User-specific Topic Modelling to build the semantic topics for each user. The user's annotation for an image is viewed as document. The individual tag to the image is *word*. User's annotations for all the images constitute the corpus. As the original annotation is too sparse for topic modelling, we use the reconstructed ternary relations as the

document collections. The user's topic distribution per image can be considered as his/her preference over the image on the learned user-specific topic space. Therefore, after the offline stage, two outcomes are stored in the system, the user-specific topics and topic-sensitive user preferences. For the online stage, when a user *u* submits a query *q*, we first map the query *q* to user *u*-specific topics. The query distribution is then sent to the rank module and employed as the weight on topics to calculate the user *u*'s topic sensitive preferences over the images. Finally, the images are ranked according to the calculated user's preferences, which simultaneously consider the query and user information.

3) Annotation and Prediction: Image annotation and Prediction can be understood as a learning process, in which the known relations between test images and annotated words are estimated by exploring available resources. Thus, how to estimate and integrate these relations is a key issue. In this section, we will address the issue by proposing an extended RMTF algorithm for image annotation [1]. In the problem of image annotation, there are two media types, image and word. We can have three kinds of relations: word-word relation, word-image relation and image- image relation. The word-image relation in the problem of image annotation can be analogous to user-item relation in recommender system. Furthermore, the available relation of annotated words and images is usually very sparse and imbalanced. Due to the scarce of high-quality image tagged dataset, the probabilistic matrix factorization algorithm as a natural and feasible option is employed to conduct our work. However, the standard probabilistic matrix factorization model can only employ one relation

*4) User Specific Topic Modelling:*With the reconstructed user-tag-image ternary interrelations, we can directly perform the personalized image search when user *u* submits a query *q*, the rank of image *i* is inversely proportional to the probability of *u* annotating *i* with tag *q* [1].

However in practice, the queries and tags do not follow one to one relationship one query usually corresponds to several related tags in the tag vocabulary. Besides, the query-tag correspondence differs from user to user. Therefore topic spaces are used for each user to exploit this user-specific one too many relationships. Example: Following tables shown in Figure 3 which defines the user specific topics for two users.

USER-SPECIFIC TOPICS FOR TWO EXAMPLE USERS.
User A Topic 1 military, aircraft, battleship, navy, artillery, Iraq, aircraft carrier, barracks
Topic 2 apple, computer, art, girl, cellphone, cool, vintage, digital
Topic 3 athlete, basketball, baseball, actor, sports, art, film, black

User B Topic 1 Buddha, Buddhist, temple, religion, Buddhism, Thailand, Asia, ancient
Topic 2 blossoms, blooms, nature, macro, flower, bravo, butterfly, spring
Topic 3 airplane, Boeing, aircraft, airport, aviation, jet, aero plane, cockpit

Fig. 3 Topic Modelling

5) User Specific Query Mapping: User specific information is considered to distinguish the exact intentions of the user queries and rerank the list results. Given the large and growing importance of search engines, personalized search has the potential to significantly improve searching experience.

III. Proposed Work for Personalized Image Search

To improve the performance of personalized image search the following proposed framework as shown in Figure 4 has been designed in which in spite of general specific search engine semantic search engine can be used. The idea of this search engine can be taken from PSSE (Personalized Semantic Search Engine) architecture [4].

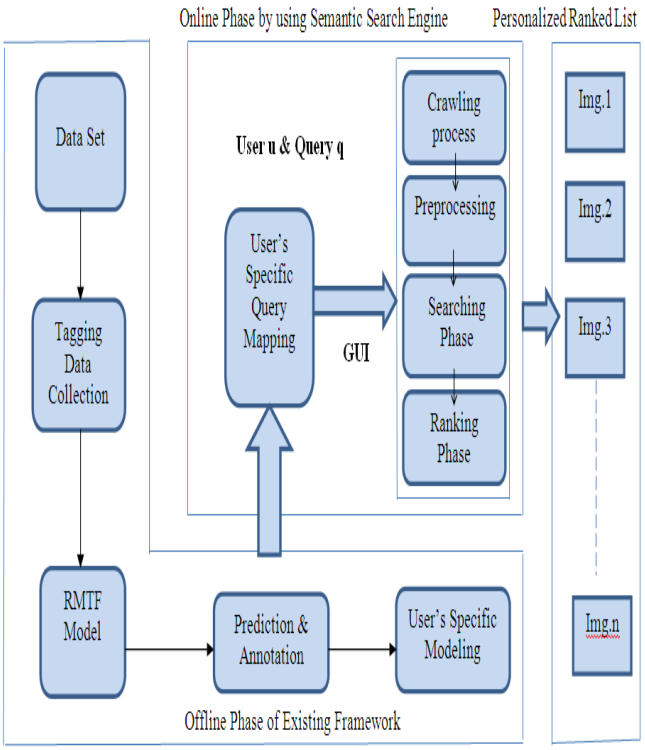

Fig. 4 Proposed framework for Personalized Image Search.

Proposed framework for personalized image search can be used to optimize the image search and which can be achieved by using semantic search engine in spite of general specific engine which has been used in existing novel framework. Offline phase of the proposed framework is same as we have seen in novel framework for existing work i.e. back end part of the system.

But in online phase semantic search engine has been used to get more optimized result as a personalized ranked list images in the output. The steps for the same are listed below and completely explained in next section by using PSSE architecture [4].

A) Crawling process
B) Pre-processing stage
C) Searching stage
D) Ranking module

IV. PSSE Personalized Semantic Search Engine

Following, Figure 5. PSSE uses the profiles with the ranking score and ontology for calculation of personalized factor which will help us to get more personalized search result. The architecture of PSSE has two parts, Offline and Online. The Offline part consists of crawling and pre-processing processes. The Online phase includes query processing and result ranking. The following figure gives the architecture of PSSE.

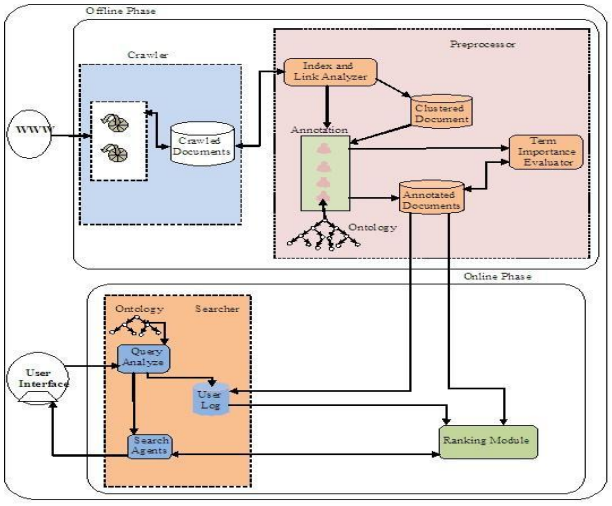

Fig. 5 PSSE (Personalized Semantic Search Engine) Architecture

A. Crawling Process

PSSE uses Multi-crawlers (web spiders) that traverse World Wide Web, collect web resources and store them in database [5]. Crawlers work with the aid of information extraction techniques to find link information in the retrieved pages.

B. Pre-processing Stage

This stage is used to maintain resources that are downloaded from Web sites. The main task of query Indexer and link analyser is to cluster the crawled web documents to enable parallel processing. This can be done in three steps: first indexer and link analyser builds a graph of the crawled pages. Link analysis is then performed to calculate authoritativeness of web pages. And finally the graph is clustered by identifying its connected components [6]. These clusters are then annotated by annotation agents that work in parallel to reduce processing time. Afterwards, annotations are weighted so as to determine their relevancy to web resource using term relevancy evaluator. Annotations are assigned weights that are calculated based on their relevancy to document. PSSE system uses vector space model namely TF-IDF to represent documents in weighted terms [7].

C. Searching Stage

This component is responsible for searching and retrieving relevant results. First query analyzer performs mapping of query terms as well as query expansion using ontology [8]. This component is responsible too for maintaining user log and keeping track of user search history. Afterwards, search agent retrieves relevant results from resources database. Retrieved results are then passed to ranking module to be reranked.

D. Ranking Module

This module is responsible for ranking the retrieved results. Three factors contribute to the score. The first one is the page authoritativeness which is calculated during the preprocessing phase using link analysis techniques [9]. The second is the relevancy of resource content to query terms which depends on content analysis. And finally, the third factor is the personalization factor (PF). PF is used to support tailoring results according to user's interests and preferences. Personalization factor is calculated based on the analysis of user's log file. Analyzing user's search history can result in a value that represents user's interests in a particular query term. The final ranking score is the combination of these three factors [10].

V CONCLUSIONS

Use of internet has increased in past ten years to the extent that the users are taking help of it for almost all the planning's in their day to day life. The search engines like Google and Yahoo are so famous that they are in use now and then for searching various type of information available on web. The search using Google is easy but sorting the expected data out of the search results is very difficult and keyword based search algorithms used in search engines adds more and more confusion in identifying requisite data. So some semantic approach is necessary to optimize the image search which has been explained in this paper by using PSSE architecture. Semantic Search engine which is a basic idea of semantic search web development structure and it achieves more user specific ranking result as compare to novel framework used for personalized image search. Personalized Semantic Search gives us the results which are up to the mark but within some particular domains.

REFERENCES

[1] Jitao Sang, Changsheng Xu Dongyuan Lu, "Learn to Personalized Image Search From the Photo Sharing Websites," IEEE Trans. Multimedia, vol. 14, no.4, Aug.- 12.

[2] D. Zhou, J. Bian, S. Zheng, H. Zha, and C. L. Giles, "Exploring social annotations for information retrieval," in *Proc. WWW*, 2008, pp.715–724.

[3] Y. Cai and Q. Li, "Personalized search by tag-based user profile and resource profile in collaborative tagging systems," in CIKM, 2010.

[4] S.G.Choudhary, S.R.Kalmegh, Dr. S. N. Deshmukh," Semantic Search Algorithms based on Page Rank and Ontology", 3rd International Conference on Intelligent Computational Systems (ICICS'2013) January 26-27, 2013 Hong Kong (China).

[5] C.Castillo, "effective web crawling", SIGIR Forum, ACM Press, Volume 39, Number 1, New York, NY, USA, p.55-56 (2005)".

[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein "Introduction to Algorithms, Second Edition", MIT Press, 2001.

[7] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries", First International Conference on Machine Learning, 2003.

[8] A. Sieg, B. Mobasher, R. Burke, "Learning Ontology-Based User Profiles: A SemanticApproach to Personalized Web Search". IEEE INTELLIGENT INFORMATICS BULLETIN, VOL. 8, NO. 1, 2007

[9] L. Page, S. Brin, R. Motwani, and T. Winograd , "The PageRankCitation Ranking: Bringing Order to the Web", Technical report, Stanford Digital Library Technologies project, 1998.

[10] A. M. Riad,Hamdy K. Elminir, Mohamed Abu ElSoud, Sahar. F. Sabbeh "PSSE: An Architecture For A Personalized Semantic Search Engine", International Journal on Advances in Information Sciences and Service Sciences Vol 2, No 1, March 2010.

