



Exploiting Social Annotations for Personalizing Retrieval

Eman El Sayed^{1*} Abeer El Korany² Akram Salah² Hesham Hefny¹

¹ *Institute of Statistical Studies and Research,
Department of Computer and Information Science, Cairo University, Egypt*

² *Faculty of Computer and Information,
Department of Computer Science, Cairo University, Egypt*

* Corresponding author's Email: emanissr@yahoo.com

Abstract: Continually, search engines improve their capabilities toward facilitating search and the retrieval enhancement. Despite the great efforts in the information retrieval field, the retrieved results may be out of user's expectation. This may be due to the huge number of web resources, and unidentified user's interests and domain. This paper proposes exploiting social annotations for improving retrieval based on personalization. The personalization focuses on web resources and retrieval process. In this context, new layer of knowledge is added to the web resource analysis and retrieval. Then, the additional knowledge leads to improve the retrieved results to be close to user's interests. So, it retrieved different results for the same query based on the user's interests. By applying the system, the experiments realize 36% precision improvement compared to non-personalized search engine. Moreover, the user satisfaction measured by evaluating search results versus user's priorities, where it was in between 92% up to 100%.

Keywords: Social annotation, Personalized information retrieval, Semantic web.

1. Introduction

Nowadays, Information Retrieval (IR) researchers' efforts are directed to improve retrieval to be close to the user's expectations. In the context of Web 2.0, "user became a part of the web not only recipient and several applications support that" [1, 2]. So, "Web 2.0 benefits provide a level of knowledge which may support retrieval improvement" [3, 4]. This has to be supported by discovering user's interests and domain.

Last few years, the "Personalized Information Retrieval (Personalized Search) became one of basic research fields which concerns to discover web user's interests" [5-7]. Personalized Search focuses on "analyzing the web users' behaviours and participations" [8]. Social annotations are one of the users' participation forms that "reflect the user's preferences as well as resource content" [2, 9].

The proposed approach combines web 2.0 and web 3.0 features for enhancing retrieval. It

emphasizes that social annotations are a good review of web resources. So, it uses social annotation for improving IR by adding personalization level. In this context, the proposed approach concerns with personalization in two levels; resource level and retrieval level, where it adds a new level of knowledge to web resources based on human perception. Further, it improves the retrieval process by keeping in touch the user's domain and priorities. In this context, the proposed system avoids the heterogeneity problem, and it retrieves different results list for the same query's keyword. So, the negative navigation is eliminated and the user's query results became close to user expectation. the proposed approach and experiments are detailed in the following sections, where section 2 looks over the previous work in this area of research, section 3 touches on the proposed approach in detail, and section 4 shows an implementation and experiments.

2. Related work

Nowadays, researchers' efforts are directed to personalized search. The researchers concern with improving the IR techniques by adding annotations as a part of the retrieval process. Moreover, some of them focus on semantic sense discovery and interest mining [10-12]. "The semantic sense discovery may be based on formal ontology or lexicon (e.g. WordNet, Wiki)" [13, 14].

Ph. Mylonas concerned with enhancing information retrieval by adding a level of personalization, where it needs to extract the users' preferences [15]. So, this framework proposed to use the user's query keywords to extract the implicit semantics and interests. These keywords are mapped to a corresponding concept in the ontology for avoiding ambiguity and add a level of standardization. Also, these interests are related to content of resources, this through using fuzzy logic and vector space. So, the relationships between user's preferences and resources contents may be constructed and represented using ontology, this supports the idea of improving personalization.

A. Dridi proposed a framework for enhancing information retrieval by using both users and document profiles [16]. It proposed to develop a social document profile on a tripartite graph (Content, Tags, and Clicks) that represents documents using content and social profile given by tags and clicks. It used social annotation for mining user's interest into topic. Then, the user's query results retrieved based on user communities and reactive action (e.g. clicks, tags, etc.). In this context, the results are re-ranked and represented as a combination of marched resources' keywords and the social relevance scores. Finally, the contribution of this framework is creating new technique which includes the users' reaction in addition to personalized search for retrieving community based results.

D. Yong proposed to exploit topic tag mining for enhancing information retrieval [17]. It improved language model based on three components; topic structures of documents, semantic structures of tags, and user interests. It calculated relation between three main parts social tags, resource, and web user. Furthermore, the web resources are clustered based on tag mining. So, it proposed to estimate the document model and rank results based on the query generative likelihood. It decomposed the model into four sub-models which combined together to develop query terms. The sub models were language annotation model, document model, user model, and query model.

Z. ZHOU proposed social information retrieval based on user interesting mining. It exploited the relation of user interesting, user tags, and web text context [5]. It extended the language model (LM) which is recently being part of IR researchers' interests, but LM suffers from some problems like data sparseness and term independence assumption. Then, several efforts have been made to develop it for IR based on social annotations. Z. ZHOU proposed to smooth the document query models based on the information generated by clustering and modelling the tagged web. The information includes: the topical cluster distribution of documents, and the users' interest distribution. Based on this information; the LM is expanded with user interests. The evaluation of the proposed approach is done in two levels: the first is by student judgment (test the retrieval resources through a sample of graduate student), where the number of perfect result sets over 80%, the number of bad result sets is less than 10%. So, the information retrieval performance of the proposed method is effective. A second evaluation method is by comparing the proposed model to other language model. The result of evaluation emphasized that it realized result probably greater than others.

S. Waghmare proposed approaches for developing an ontological user profile for the purpose of personalized search [18]. The profile is being updated and saved into the database automatically, when the user clicks URLs. This approach follows user behaviours through their clicks to extract user's preferences. Moreover, based on extracted preferences the retrieved query's results are re-ranked to be closer to the user interests. The performance of this application has been evaluated against the result set generated by executing the queries using re-ranking algorithm and without re-ranking algorithm. They used precision and recall measures to evaluate the query results. The evaluation showed significant improvement in retrieval time, recall and precision. It emphasized that the personalized search system helps to provide Web information that matches a user's personal interests and thus provide more effective and efficient information access. Also, a key feature in developing successful personalized Web applications is to build ontological user profiles that accurately represent user's interests.

M. Bouadjenek proposed a framework for enhancing the information retrieval based on personalization aspects [19]. It proposed to exploit annotation as a part of resource analysis in addition to the resource's content. Moreover, it applied personalization to query expansion and ranking

processes. The resource's content and tag analysis were syntactic analysis. Then, it used the vector model for measuring the similarity. Furthermore, the proposed approach is implemented in free datasets from different bookmarking systems such Delicious, Fliker, and CiteULike. The precision average is in the range 80% to 90%. In addition, the percentage of improvement compared to the non- personalized retrieval is in between 12% up to 21%.

Despite the great efforts in IR field and the researcher's successes, the search results still need improvement. "The search results still did not keep in consideration the common user's keywords and user's priorities" [20-22]. The common keywords are the popular words which used in search like txt-mining, info, and onto are unknown for the machine. The proposed approach creates new channel between web resources' content and the human mind by analyzing social annotation and keeping in consideration the web community keywords and semantic problems (detailed in section 3). Moreover, the previous efforts combined the search engine query results and the tag mining results, which leads to retrieve number of resources that may be irrelevant to user's interests. Finally, the huge number of web recourses is one of search's challenges, where the resources keyword's analysis is still part of all above researches. The proposed approach eliminates this challenge by interpreting the web resources' based on human understanding, then the results became limited through the user's scope and priorities.

3. Personalized web resource (PWR)

The search engine's capabilities need to be improved toward more satisfaction of user needs. Thus, the search engines need to be supported by metadata that reflects the user's preferences and domain. The main key for discovering user personality is his/her own words and perceptions. So, social annotation may be considered a good reflection of the user's personality. Furthermore, the annotations are a review of annotated resources where they reflect the resource content in a descriptive way. "The descriptive level of annotations provides more deep interpretation of resources than the resource summary" [23, 24].

This paper shows a proposed approach for improving IR called Personalized Web Resources (PWR). The PWR aims to improve retrieval by adding personalization aspect as well as semantic relationships. It considers annotation as a good reflection of the web resource's content. So, it exploits annotations instead of the resource's

content in analyzing and retrieval process. The PWR concerns with personalization in two levels Web Resource level, and Retrieval level. The PWR consists of two stage; pre-preparing data, and query process.

3.1 PWR pre-preparing data

The PWR pre-preparing data as represented in figure (1) aims to construct two models resource profile and user profile. Each model developed based on linguistic and semantic tag analysis. First of all, the PWR assumes that all annotated resources and annotations are written in English language. The taggers annotations are written from their perspective and understanding of resource content. So, the annotations' terms may be resource's keywords, common expressions or abbreviations for specific fields. The annotation consists of three main parts tag, tagger, and annotated resources which stored in Tagging DataBase (TDB). The tags in TDB are mapped to the tagger and annotated resource without any consideration for semantics.

The tag analysis is a part of models development, where the resource model needs it to formulate view of each resource; and the user profile uses it to represent different user's directions. The PWR uses linguistic techniques as well as semantic techniques for analyzing tags. The syntactic and semantic analysis provide more interpretation of tags and support semantic mapping.

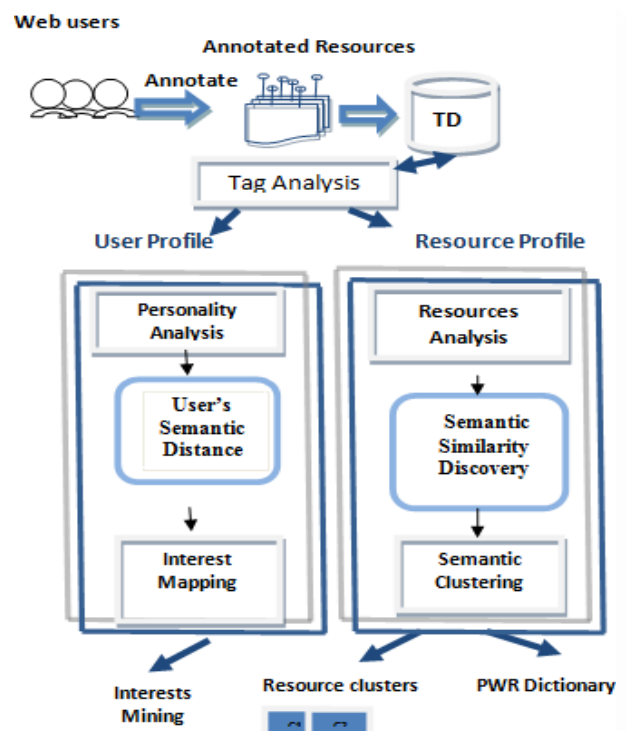


Figure.1 System model

3.1.1. Resource profile development

Each resource annotated by number of tags that contain terms reflects user's perceptions, domain terms, and common keywords. The Resource profile is developed by analyzing web resource using social annotations instead of the resource's content. The variations in users' perceptions and understanding of resources add a descriptive level of resources, which provides metadata, cannot be discovered using resources content and keywords.

The PWR resource profile is constructed based on personalization aspects and semantic sense. It is considered as a review of resource's content based on users' perceptions and domain expressions. The PWR resource profile development consists of three main stages; resource analysis, semantic similarity, and semantic clustering.

3.1.1.1. The resources analysis & semantic similarity

The resources analysis stage aims to extract the main resources' keywords through analyzing resource's tags. By applying the linguistic and syntactic analysis for each annotated resource, the weight of each tag's term will be identified based on Term Frequency Inverse Document Frequency (TFIDF).

After identifying resources' terms weight, the semantic analysis is applied using WordNet for extracting the valuable tags' terms and removing noise. It fixes the free annotations semantic problems and constructs semantic relationships that cannot be existed based on syntax matching. The popular free annotations semantic problems are heterogeneity, spelling errors, special signs. Thus, the PWR semantic analysis consists of three steps; tag filtering, semantic measurement, and PWR Dictionary development.

The Tag filtering is an essential part of semantic analysis, where the noisy terms have to be removed. The noisy term is the expression which has no meaning and rarely used, where sometimes the tagger signs resources or uses expression doesn't relate to resource. In addition, the spelling errors are considered as noisy terms that have to be filtered. So, the filtering process eliminates the useless tags, but takes into consideration the common words. Then, the remained tags will be ready to semantic relationship discovery.

The Semantic Similarity measure integrates the syntactic and semantic features for improving the relationships between resources. Then, the resources are represented as vectors in the vector space model. In general, "the vector model suffers from some

challenges like assumption of term indecency (e.g. ignore synonymy), and missing semantic information" [25]. Thus, the PWR adds a semantic level to improve retrieval and avoids the free text annotation challenges.

The PWR categories the semantic similarity into four semantic relations: equal, semi-equal, partially equal, and non-related. The equal relation represents terms that have the same syntax, semi-equal relation for synonyms, partially-equal relation for hyponymy, and non-related relation is the different terms. Each category has a pre-specified *Semantic Relation (SR)* degree which represents a threshold for identifying them. As in equation (1), the *Semantic Similarity Discovery (SSD)* identifies the web resources related degree using *cosine similarity* and *SR*. The Euclidean lengths between web resources R_n is calculated by representing them as vectors through the vector space model. The *resources' term weights* (w_n) are calculated for each vector, and then the semantic relations are discovered. The *Semantic Relation (SR)* is added to vectors and mostly improves the similarity between resources.

$$SSD(R_j, R_i) = \frac{R_j \bullet R_i}{|R_j| \bullet |R_i|} = \frac{\sum_{i=1}^t (w_j \times SR \bullet w_i)}{\sqrt{\sum_{i=1}^t (w_i \times SR)^2 \bullet \sum_{i=1}^t w_i^2}} \quad (1)$$

Finally, the PWR dictionary is constructed to support semantic resource mapping and personalized retrieval. It saves synonyms, hyponymy tag's terms, and popular expressions. The taggers may annotate resources by different expressions that have the same meaning, or the expressions may have partially semantic relationships. So, thorough applying WordNet in tag analysis different semantic relationships are discovered and added to the PWR dictionary. In addition to that, the web users use popular and agreed expressions which commonly used but have no real meaning. These expressions are human understandable terms, but cannot be matched in a retrieval process like txt mining, info, onto as a sequentially abbreviation of text mining, information, ontology. Therefore, the PWR dictionary is one of resource model contributions that will support query expansion and retrieval process.

3.1.1.2. Semantic clustering

Semantic Clustering is a basic step that leads to constructing a resource profile. The resource profile develops depends on mining resources according to the semantic tag analysis. The PWR resource profile is the most important step toward facilitating and improving retrieval.

The PWR semantic clustering exploits the K-means algorithm and extends it. K-means is non hierarchical clustering which randomly identify seeds, and number of clusters k . The additional two issues which added through the PWR are automatic seed identification, and semantic clustering. The automatic seed identification is automatically calculation of the k seeds by extracting the number of topics that completely different. Moreover, the semantic clustering integrates the topics which have semantic relationships (e.g. part of relation, synonyms).

The output of the resource profile phase is represented in PWR dictionary and a semantic based cluster provides personalized web resources. The resource profile is a combination of descriptive resource analysis and users' common knowledge. In this context, the retrieval process is improved and became close to user expectations.

3.1.2. User profile

The user profile identifies the user interests and search scope by analyzing user's personality through his/her annotation, and annotated resources. The main purpose of constructing a user profile is to support the search engine to retrieve results close to user interests and expectations. Furthermore, one of PWR goals is constructing different levels of friendships between users for improving the user knowledge to support retrieval. The development of the PWR user profile consists of two steps; virtual user's profile, and friendship.

3.1.2.1. Virtual user's profile

The virtual user's profile purpose concerns by analyzing user's tags for the purpose of mining interests. This is one of the essential factors to follow the user's requirements. Basically, the analysis of user's tags is done syntactically semantically, where the user may use different expressions have the same meaning. So, for calculating the Tag Term Frequency (TTF), the semantic relationships have to be mapped. The TTF reflect the common user's expressions, synonyms, and hyponymy terms which used in some domain or community. When the user may have more than one

scope of interest, then the PWR sorted the user's interested topics into levels. These levels identify the user's search personality and his research priorities for supporting later the retrieval (detailed in section 2.3).

In general, users in PWR are classified into general, specified, and unknown users. This classification is based on the clarity of users' interests, and the user participation. The general user has many interests and has no domain; the specified user has a specific domain of interest or interested in one or more topics in any field; and the unknown user is new user which has no history or has not enough annotation to reflect his personality. This classification is used later in the query process to support the retrieval process (detailed in section 3.2).

3.1.2.2. Friendship

After developing the user profiles, the relationships among users can be discovered. The PWR discovers the user's relationships using the semantic distance measure. The semantic distance is the measure which determines how close two users are. The range of semantic distance is between 0 and 1, where 0 represents the close users and 1 means the user's scope is completely different. According to the calculated semantic distance, the friendship between users can be constructed.

By applying the semantic distance for each pair of users, the friendship among users are classified into three levels. The PWR friendship levels are close friends, friends, and nearby. The closed friends have close scope in an average 70% to 100%; the friends in between 40% and 69%; and the nearby less than 40%. Based on this classification the user knowledge about specific domain can be extended, where each friend in different levels has knowledge in one or more topics and domains. So, the user personality is not discovered only depended on user's annotations, but also extended by friends' knowledge. This supports both the user's search as well as retrieval process.

3.2 Query process

Usually, the different IR techniques cannot achieve the user expectation, or the results may be partly related to his/her requirements and need more navigation. It analyzes the web resources through the occurrence and frequency measure. In this context, the language model depends on the keyword matching techniques.

The PWR query process extends the language model by semantic sense and personalized retrieval. The semantic sense represents in the constructed

semantic relationships between web resources and user’s friendship. On the other hand, personalized retrieval represents in users’ profiles and semantic relations between profiles. So, the retrieval process will be improved toward retrieving query results more related to user’s domains and according to user’s interests’ priorities.

Figure 2 shows the PWR query process steps and components. First of all, the user enters his query into the PWR system, and then the query will be expanded based on semantic sense by using PWR dictionary and friendships. The PWR dictionary provides terms used as keywords in specific domains, their synonyms and hyponymy, so the query terms are expanded through a number of iterations to form multiple queries. Moreover, the user’s friendships support query expansion, especially in case of new user or general user. The new and general users have vague personality, so they need especial support.

The PWR query result is a combination of similarity based results and personality based results. The similarity based results are the query results by applying the extended queries into resource clusters. Further, the virtual users’ profiles and friendships are used to retrieve and rank results (personality based results). The two results are integrated and retrieved to the user in a suitable form. The query results are retrieved depends on the user’s interests’ priority, and specifically his/her domain. In this context, different users may use the same query keywords, but the query’s results are different. Also, the results may have different ranking based on the interests’ priority for each user.

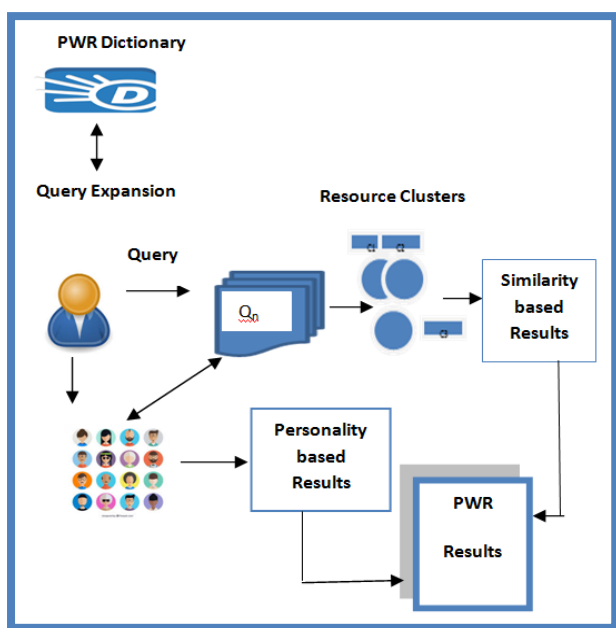


Figure.2 The PWR query system

4. Experiments and evaluation

4.1 The PWR system performance

As CiteULike is an academic bookmarking system which provides free dataset. It has a set of annotated resources (<http://www.springer.com/about+springer/citeulike>). It allows users to tag several references (e.g., academic papers or books) included in its library. The CiteULike dataset consists of directory documents that contain over 180 documents in text format, and annotated with 807 annotations.

PWR assumes the annotated resources have to be annotated at least by ten terms through more than three taggers, and ignore other cases. Continually, the PWR system connects to TDB either for getting, updating, and adding data. For each annotated resource R_i , and tagger T_i , the PWR system collects and analysis all tags’ terms syntactically and semantically. In addition, the JAWS API is used to support semantic analysis and WordNet to Java connection. The semantic challenges that faced PWR approach are spelling mistakes, non-meaning common word, stemming errors, and heterogeneity. These problems are fixed through the different WordNet semantic capabilities and PWR system.

The PWR system applied the resource model through analyzing all annotations attached to annotated resources. The semantic challenges are fixed by using WordNet and semantic cosine similarity measure. The WordNet supported resources’ mapping by discovering the synonyms and part-of relations. Then the semantic cosine similarity measure is applied and realized good result, where new resources’ semantic relations are discovered and some existing are improved.

The resource model based on social annotations improved the resources’ mapping. This supported the retrieval process by adding resources related to the user’s query, where the web resources’ retrieval became based on human understanding and semantic aspects. So, the user’s query interpretation will be close to resources’ keywords. By comparing the RWP model to non-personalized resources’ matching, the PWR improved the resources’ mapping similarity for 96.6% of resources and added average of new discovered resources’ relationships in percentage of 10%.

On the other side, the retrieval process was personalized by keeping in touch the user’s domain and his/her interests’ priorities. The PWR essential goal is retrieving right result to the suitable user and avoiding the negative search. This means two users may write the same query keywords and the

retrieved results are different to be close to the user's scope. Actually, the PWR realized this goal and it has been tested. For example, the dataset includes two users U_1 and U_2 , where U_1 interested in biomedical topics and social network and U_2 interested in cognition and statistics. Both U_1 and U_2 search for the same topic using the same keyword as "Network"; the PWR retrieved results related to each user's scope and the results retrieved to each are different. In another case, when the users have close interested the retrieved results for the same query may have different rank; this depends on the priority of the topic for each user.

The PWR performance is measured using precision and recall. The target of PWR is achieving high precision and low recall. The PWR approach has been applied into up to three hundred queries, and the number of query terms is between two to five terms. The query terms were tags' keywords, resources' keyword, and random expressions which selected from the dataset. Practically, the PWR achieved a precision average in between 85.7% and 90.77%, and recall average in between 62.2% and 65.11%.

Finally, the PWR performance proves that the PWR achieved the personalization goals in resource level as well as retrieval level. It may be a step toward semantic web goals achievement. Also, it is considered as an approach that exploits the web 2.0 benefits to realize the web 3.0 goals.

4.2 Evaluation

For evaluating the PWR accuracy and performance, set of queries applied to it and to a search engine then the results are compared. This comparison aims to answer two essential questions; the first is "Are Tags considered as a Good Reflection of Resources Content?", and the second one is "What's the Retrieval Improvement That Achieved through the Proposed Approach?" Besides, the PWR approach is compared to the previous work in the same research area.

4.2.1. Non personalized search vs. PWR retrieval system

For evaluating the PWR, more than five hundred queries applied into both non personalized search engine and the PWR system. The non-personalized search retrieved query results based on keyword retrieval technique, when the PWR integrates semantic and syntactic techniques. Then, the results are compared through precision and recall measures. The average of PWR precision and recall is 94.4% and 62.4%. Further, the PWR system achieved 85%

of the all non-personalized search engine results; so this approves that annotations are considered as a good reflection of resource. In addition, the average of precision improvement which achieved through the PWR compared to search engines is 36%. Table (1) shows a sample of queries, results retrieved through PWR, non-personalized search engine results (SE), and a number of additional results (AR) through PWR. It aims to represent the effect of adding knowledge to resources and the semantic role.

As shown in table (1) the query sample includes keywords extracted from the web resources' content and some selected tags. The combination of queries is exploited as an evidence to prove that social annotations are a good reflection of user's interests as well as resource content. Also, the role of personalization of resources and retrieval is appeared in the additional resources which retrieved through the PWR system. The PWR field represents the number of resources that semantically and syntactically matched. Also, the SE field represents the non-personalized results which lack of poor

Table 1. The PWR improvements based on semantic relations

Query	PWR	SE	AR
Bayesian reasoning causality	126	5	123
rat			
Rna small Arabidopsis	130	6	126
Learning psycholinguistics	106	19	92
Evolution proteins	149	42	112
Interaction domain protein	57	46	16
statistics			
Regulatory network motif	149	46	108
p2p			
Expression noise network	60	54	14
Stem cell interaction protein	54	40	29
Statistics analysis	110	43	77
Gene network inference	63	48	23
expression			
Comparative genome	93	58	47
Genetics regulation network	136	58	87
Protein expression	142	59	92
stochastic			
Blog theory image	94	59	44
ethnography			
Networks motifs	142	60	91
Genome perturbed protein	143	60	92
network metabolic			
Whole genome sequencing	98	23	84
Small world networks	130	41	96
Genomic interaction	105	46	65

Table 2. Comparison between the PWR approach and related research in the same research area

	Resource profile	User profile	Vector model	Prob model	Semantic Sense	Precision Average	Improvement
PWR	√	√	√	×	√	94.4.6%	36.1%
Wu	√	√	√	√	√	72%	-15%
Mylonase	×	√	√	×	√	92%	25%
Bouadene	√	√	√	×	×	90%	21%
Wagmar	√	√	√	×	√	75%	---
ZHOU	√	√	×	√	×	80%	---
Yong	√	√	√	×	√	56.6%	10%
Niebler	×	√	√	×	√	32%	
Alonso	×	√	√	×	√	60%	--

semantics, unknown user’s identity, and common user’s search keywords.

Moreover, the AR field shows the number of resources which retrieved by PWR and did not retrieved through non personalized search. The AR is produced from the semantic procedures and the Meta knowledge that used in the retrieval matching process. This shows the PWR capability to discover additional resources’ relationships and improve the resources’ mapping.

4.2.2. The previous work vs. PWR retrieval system

The PWR approach compared to the previous work which concerns with the same research area. Table (2) shows the comparison based on some criteria that represents the main features in the personalized search field. The first two criteria are resource profile and user profile; this means the approach constructs a profile for resources and users. Also, the two models that commonly used similarity measures in this field are vector model and probabilistic model. Moreover, the semantic criteria are ontology and semantic senses, where some researchers focused on ontology as a formal and standard semantic representation model; or the semantic sense which concerns with discovering the semantic relations through lexicons. Finally, the implementation results represented in two criteria; the average of precision and the improvement average.

The represented previous work in most cases developed resources’ profile and user’s profile. The resources’ profile is developed based on the resource’s keywords, and then the annotations’ keywords may be used as inference in the retrieval process. In other cases, the annotations are considered as a part of resources’ content and indexed as resource’s keywords. When the web resources number increases continuously, the combination of resource’s keywords and annotations may be time consuming. Thus, the PWR considered

social annotations as a good reflection of resources, and exploited annotation instead of the resource’s content. This is one of the factors for minimize the huge number of indexed terms which may be useless in the retrieval process. Also, the tagger extracts the main keywords and utilizes them to form annotations. In addition, the human understanding of resources and the search common words in specific area are also formed as annotation. All the above support the resources with descriptive knowledge and concentrated keywords without going deeply into useless keywords.

The second comparison factor is user profile. It is a joint factor in all personalized research, where it is the core factor to represent user’s directions. The purpose of exploiting user’s profile in the PWR is not only as a reflection of user’s interests, but also it constructs users’ friendships. The friendship in different categories adds layers of knowledge to the user’s personality through the system. This improves and supports the retrieval process. Moreover, the PWR user profile is decomposed into layers of user’s priorities. So, this avoids the negative search challenge.

The third and fourth factors are vector model and probabilistic model. Mostly, the vector model is used in personalized retrieval researches where the distance between resources, users’ tags, and query is measured. The PWR uses the vector model for clustering resources and users, and developing their profiles. The vector model lacks semantic considerations, so the PWR added semantic level to it by measuring the semantic distance. The semantic distance really improved and strengthened the retrieval.

The fifth factor is semantic sense, which concerns with having an inference for unifying expression or discovering semantic relations. On one side, ontology adds a level of standardization and formal representation that supports free text

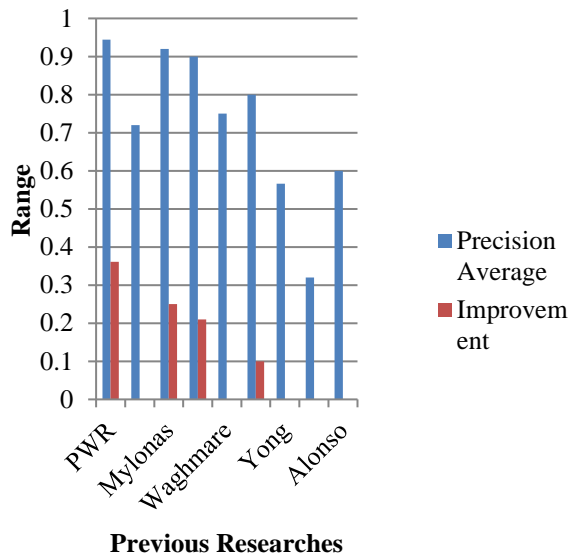


Figure.3 The precision and improvement of PWR versus related work

annotations. While on the other side, lexicons may be used for extracting semantic relations and improve the machine understanding of free text annotations. Usually, ontology uses to support specific domain (e.g. gene ontology). When, the PWR are not directed toward specific domain, and then it uses the WordNet as inference. Through WordNet the PWR avoided semantic challenges like polysemy and heterogeneity.

As represented in table (2) and figure (3) the precision and improvement average is calculated compared to non-personalized search results. Some researchers did not compare themselves to search engine capabilities, so the improvement field is empty. Also, others realized negative improvement, where their results were less than the search engine.

Figure 3 has two axes X and Y. The X axis represents the percentage between 0-1 to show the precision and improvement average. The second axis is Y which represents the previous researches. The previous researches pointed by the research's family name.

The PWR approach realized high precision and improvement, and it is approximately close to Mylonas research [15] which achieved the highest precision and improvement average. The non-improvement represented researches did not compare themselves to the search engine result like Waghmare [13], ZHOU [7], and Alonso [26]. On the other side, Wu [27] realized a negative improvement, where its results were less than the search engine in an average 15%.

Finally, the PWR approach achieved its goal through personalizing resources and retrieval

process. It concerned with decomposing users into categories based on their domains, user's interests and priorities. This enhances the search engine results to be close to user expectation. Further, the PWR proved that annotation is a good reflection of resource content. So, it improves the indexing process, where the indexed keywords became concentrated. Also, the semantic sense eliminates the free text semantic challenges. All the above realized retrieval results close to user's expectations.

5. Conclusion

Personalized information retrieval became part of IR researcher's goal; it is considered as web revolution. Personalization supports search engines to be close to user's needs and expectations. The PWR added personalization level to the resource's analysis and retrieval process. The main goal was retrieving right query results to right user. So, the retrieval process takes in consideration user's priorities and domains, thus the same query keywords retrieved different results to different users. The results became concentrated and close to user expectations.

The future work may be directed to avoid the free text annotation problems by integrating them with formal annotations. The formal annotation will be used as an inference through mapping the free text annotations to ontology. Further, the PWR may be improved by supporting the multi-word expressions.

References

- [1] A. Gangemi, "A Comparison of Knowledge Extraction Tools for the Semantic Web", In: *Proc. of 10th International Conference on the Semantic Web: Semantics and Big Data*, Montpellier, France, pp. 351-366, 2013.
- [2] X. Wang and Q. Patrick, "An Ontology Based Personalizing Search Measure with the Protection of Original Engine Integrity", In: *Proc. of the 2nd International Congress on Computer Applications and Computational Science*, New York, USA, pp.155-160, 2011.
- [3] C. Ziyu and A. Haining, "The Building of Digital Archives Personalized Service Website based on Web 2.0", In: *Proc. of International Conference on Solid State Devices and Materials Science*, Macao, China, pp.2096-2102, 2012.
- [4] J. Kampa, J. Karlgren, and R. Schenkel, "The third workshop on Exploiting semantic annotations in information retrieval", In: *Proc. of 10th International Conference on*

- Information and Knowledge Management*, Toronto, Canada, pp. 1-38, 2010.
- [5] Z. Zhou, "Social Information Retrieval Based on User Interesting", *Journal of Computational Information Systems*, Vol. 4, No.3, pp. 130-136, 2011.
- [6] G. Metal, "Personalizing Image Search From The Photo Sharing Website", *International Journal of Research In Computer Applications and Robotics*, Vol. 2, No. 6, pp. 18-23, 2014.
- [7] D. Zhou, J. Bian, and S. Zheng, "Exploring Social Annotation For Information Retrieval", In: *Proc. of 17th International Conference on World Wide Web, Beijing, China*, pp.715-724, 2008.
- [8] G. Pasi, "Implicit Feedback Through User System Interactions For Defining User Models in Personalized Search", In: *Proc. of the 6th International Conference on Intelligent Human Computer and Interaction, Sciencedirect, Evry, France*, pp. 8-11, 2014.
- [9] Z. Mezghani, I. Amous, and C. Amel, "A User Profile Modeling Using Social Annotations", In: *Proc. of the 21st International Conference on World Wide Web, New York, USA*, pp. 969-976, 2012.
- [10] T. Ruotsalo and E. Hyvönen, "Exploiting Semantic Annotations for Domain-Specific Entity Search", In: *Proc. Of European Conference on Information Retrieval Top of Form, Vienna, Austria*, pp. 358-369, 2015.
- [11] S. Kalarani and S. Uma, "Integration of Semantic Web & Knowledge Discovery for Enhanced Information Retrieval", *International Journal of Computer Applications*, Vol. 1, No. 1, pp. 99-103, 2010.
- [12] P. Laublet, "Meaning Of A Tag: A Collaborate Approach To Bridge The Gap Between Tagging and Linked Data", In: *Proc. of the Linked Data on the Web Workshop at WWW2008, Beijing, China*, pp. 1-5, 2008.
- [13] A. Sharif, "Combining Ontology and Floksonomy an Integrated", In: *Proc. of Emerging Trends In Technology: Libraries Between Web 2.0, Semantic Web and Search Technology, Italy*, pp. 1-13, 2009.
- [14] H. Ching, "Integrating ontology technology with folksonomies for personalized social tag recommendation", *Applied Soft Computing*, Vol. 13, No. 8, pp. 3745-3750, 2013.
- [15] Ph. Mylonas, D. Vallet, P. Castells, "Personalized Information Retrieval Based on Context and Ontological Knowledge", *The Knowledge Engineering Review Journal*, Vol. 23, No. 1, pp. 73-100, 2008.
- [16] A. Dridi, "Information Retrieval Framework Based On Social Document Profile", In: *Proc. of Conference on Advanced Information Systems Engineering, Thassaloniki, Greece*, pp. 1-8, 2015.
- [17] D. Yong, "Enhanced Web Information Retrieval by Topic Tag Mining", *Journal of Convergence Information Technology*, Vol. 6, No. 4, pp.18-24, 2011.
- [18] S. Waghmare and R. Krishna, "Implementation of Personalized Search Model Using Ontology", *International Journal of Computer Science & Communication Networks*, Vol. 4, No. 3, pp. 130-136, 2013.
- [19] A. M. Reda, H. hacid, M. Bouzeghoub, and A. Vakali, "Using social annotations to enhance document representation for personalized search", In: *Proc. of the 36th international ACM conference on research and development in information retrieval*, Dublin, Ireland, pp. 1049-1052, 2013.
- [20] N. Sarkas, S. Das, and N. Koudas, "Improved Search For Socially Annotated Data", *The VLDB Endowment Journal*, Vol. 2, No. 1, pp. 778-789, 2009.
- [21] X. Han, X. Luo, and C. Miao "Folksonomy-Based Ontological User Interest Profile Modeling and Its Application In Personalized Search", In: *Proc. Of International Conference on Active Media Technology, Toronto, Canada*, pp. 34-46, 2010.
- [22] W. Ahmad and R. Ali, "Textual content based information retrieval from Twitter", In: *Proc. of International Conference on Advances in Computing Communications and Informatics, Jaipur, India*, pp. 2668-2672, 2016.
- [23] D. Vallet, J. Jose, and I. Cantador, "Personalizing Web Search with Folksonomy-Based User and Document Profiles", In: *Proc. of European Conference on Advances in Information Retrieval, Milton Keynes, UK*, pp. 420-431, 2010.
- [24] O. Medelyan, E. Frank, and H. Witten, "Human-competitive tagging using automatic key phrase extraction". In: *Proc. Of International Conference of Empirical Methods in Natural Language Processing, EMNLP, Singapore*, pp. 1318-1327, 2009.
- [25] C. Manning, P. Raghavan, and H. Schutze, "An Introduction to Information Retrieval", *Cambridge University, USA*, 2009.
- [26] O. Alonso and J. Kamps, "Seventh Workshop On Exploiting Semantic Annotations In Information Retrieval", In: *Proc. of The 23rd ACM International Conference on Conference*

on Information and Knowledge Management,
Shanghai, China, pp. 2094-2095, 2014.

- [27] X. Wu and L. Zhang, "Exploring Social Annotations for the Semantic Web", In: *Proc. of The International World Wide Web Conference Committee*, Edinburgh, Scotland, pp. 417-426, 2006