Collaborative Summarization:
When Collaborative Filtering Meets Document Summarization[*]

Yang Qu and Qunxiu Chen

Department of Computer Science, Tsinghua University,
FIT 4-506 Tsinghua University, Haidian District, Beijing 100084, China
dcatcher.qu@gmail.com, cqx@s1000e.cs.tsinghua.edu.cn

Abstract. We propose a new way of generating personalized single document summary by combining two complementary methods: *collaborative filtering for tag recommendation* and *graph-based affinity propagation*. The proposed method, named by Collaborative Summarization, consists of two steps iteratively repeated until convergence. In the first step, the possible tags of one user on a new document are predicted using collaborative filtering which bases on tagging histories of all users. The predicted tags of the new document are supposed to represent both the key idea of the document itself and the special content of interest to that specific user. In the second step, the predicted tags are used to guide graph-based affinity propagation algorithm to generate personalized summarization. The generated summary is in turn used to fine tune the prediction of tags in the first step. The most intriguing advantage of collaborative summarization is that it harvests human intelligence which is in the form of existing tag annotations of webpages, such as delicious.com bookmark tags, to tackle a complex NLP task which is very difficult for artificial intelligence alone. Experiment on summarization of wikipedia documents based on delicious.com bookmark tags shows the potential of this method.

Keywords: Collaborative Filtering, Single Document Summarization, Personalized Summarization, Tag Recommendation

1 Introduction

In this paper, we confine personalized document summarization to the extraction from a given document a few sentence which can best represent the whole content of the document in the point of view of a given user. While generic document summarization without personalization has been extensively researched, the trend of personalization is attracting more and more focus since the outgrowth of the technology of Web 2.0. For example, personalized search engine and personalized e-commence(movies, commodities) recommendation are two successful applications.

Collaborative filtering is a widely used technology for personalized recommendation. Suppose some users have rated a set of books, the key idea of collaborative filtering is that similar users should share similar rating behaviours in the future, where the similarity between users itself is defined by the similarity of their past rating behaviours. The advantage of collaborative filtering is that it harvests the computing power of individuals in a population to process the large amount of complex information. Taking the example of book recommendation, there is no model competitive with a human brain with regards to understanding a book. We can treat each individual reader as a super computational power which is reluctant to give feedback. Then the task of collaborative filtering is to harvest these human power with the least workload on each individual.

Fortunately, due to the growth of internet and web 2.0 technology, a large amount of tags data on webpages have been generated by human users. For example, people can tag their favourite

---

[*] Thanks for all the people who help collecting the dataset.

webpages using the bookmark service of delicious.com and many alike. However, current usage of tags data is limited to classification of webpages or recommendation of new pages to users. As far as we know, no attempt has been made to automatically generate a *personalized* short summary for a tagged webpage. Instead, users are optionally requested to provide a personal description of the tagged webpage by hands, which is too much a burden and often leads to copy-and-paste a whole paragraph in the webpage. In this paper, we focus on automatically generating personalized summary of a single document for a given user, which has a nature application aforementioned. Note that multi-document summarization can also benefit from our method with little modification.

Our method, called Collaborative Summarization, seamlessly combines the state-of-art affinity propagation based document summarization (Wan and Xiao, 2007) with collaborative filtering based tag recommendation (Hotho *et al.*, 2006). The combined result is three-fold. First, by collaborative filtering, we can utilize existing tag information to predict the possible tags of a new document for a given user, even if the user has not tagged the document. This eliminates the necessity to annotate a document before summarization, which is a requisite for most existing summarization methods using side information. Secondly, the predicted tags are a compromise on the document between community's view and personal view. Thus it is suitable as a guidance for generating personalized summarization which is supposed to capture both the main content of the document and the special point of interest of given user. Finally, due to the iterative nature of collaborative summarization, the generated summary is used to fine tune the predicted tags. This complements collaborative filtering by incorporating content information into tag prediction, as collaborative filtering alone only consider the user-tag-document tannery relationship without considering the actual content of the document.

In the following sections, we will first introduce the key ideas and notations underlying collaborative filtering based tag recommendation and affinity propagation based document summarization. Then we will discuss the detailed algorithms used in our collaborative summarization method. After that, an experiment on summarize wikipedia articles using delicious.com bookmarks tags is conducted to show the potential of our method. Finally, we will compare some related work regarding document summarization with side information and conclude our work.

2 Affinity Propagation for Document Summarization

In the work of (Wan and Xiao, 2007) , An affinity graph is built for single or multi-document summarization, and has shown significant improvement over other graph-based document summarization methods. We will briefly review the key ideas and notations while focusing on single document summarization.

Suppose we have a document consisting of n sentences $\{S_1, S_2, \cdots, S_n\}$. Each sentence is presented by a vector $\vec{s}$ of $tf_t * isf_t$ value of each term t, where tf_t is the number of term t in that sentence and isf_t is inverse sentence frequency calculated by

$$isf_t = 1 + \log \frac{N}{n_t} \tag{1}$$

where N is the total number of sentences in the background corpus and n_t is the total number of sentence containing term t.

We can construct a directed graph on all the sentences by defining a similarity measurement of each pair of sentences, namely sentence affinity:

$$\text{Aff}(s_i, s_j) = \frac{\vec{s}_i \cdot \vec{s}_j}{\|\vec{s}_i\|} \tag{2}$$

Sentence affinity is a measurement of semantic similarity between sentences and it's not symmetric as opposed to cosine similarity. The affinity graph can be constructed by linking two

sentence nodes whenever the affinity value between them is not zero. Usually the affinity scores associated with each sentence are normalized, however, it's not always a necessity.

After the affinity graphed is built, a page-rank like algorithm is applied to the graph to calculate a score for each sentence node, which is called information richness of that sentence. The justification is intuitive: The more neighbours a sentence has, the more informative it is; and the more informative a sentence's neighbours are, the more informative it is. Thus a iterative update is straightforward:

$$\text{info}(s_i) = \text{dampfactor} \cdot \sum_{\forall\ j \neq i \in S} \text{info}(s_j) \cdot \text{aff}(s_j, s_i) + \frac{1 - \text{dampfactor}}{N} \qquad (3)$$

where a damp factor is used to speed up the convergence.

Finally, when converged, the information richness scores of all sentences are sorted, and sentences with the largest scores are picked as summarization for the document. Some redundancy removal methods can be used for post-processing.

3 Collaborative Filtering based Tag Recommendation

Recently, tag recommendation in online social network has attracted increasing attention, and even a new word named folksonomy is dedicated to the concept. In this section, we will focus on collaborative filtering based tag recommendation since it has a natural relationship with affinity propagation based document summarization, as will be discussed below.

Suppose we have a set of users U assigning various tags T onto a set of documents D. All the useful information can be described by a ternary relationship (U, T, D). For example, an instantiation from the relationship (u, t, d) means that user u assigned tag t to document d. We denote the set of tag assignments to be Y.

A widely referenced collaborative filtering based tag recommendation, called FolkRank (Hotho *et al.*, 2006), makes recommendations based on a tripartite graph built from such ternary relationship. We denote the tripartite graph by $G = (V, E)$. The graph nodes V are combination of all users, tags and documents,i.e. $V = U \cup T \cup D$. The edge in the tripartite graph connects two entities(user, tag or document) if both entities occur in a tag assignment and the weight of the edge amounts to the count of their co-occurrences.

More specifically, the weight of an edge connecting tag t and document d is defined by

$$w(t, r) = |\{u \in U : (u, t, d) \in Y\}| \qquad (4)$$

And similarly,

$$w(t, u) = |\{d \in D : (u, t, d) \in Y\}| \qquad (5)$$

$$w(u, d) = |\{t \in T : (u, t, d) \in Y\}| \qquad (6)$$

FolkRank applies personalized pagerank on the constructed graph G twice:

$$\vec{w} = \lambda G \vec{w} + (1 - \lambda)\vec{p} \qquad (7)$$

where $\vec{w}$ is a score vector over all graph nodes, corresponding to the information richness vector in affinity graph. $\vec{p}$ is the personalization vector over all graph nodes(usually only the tag nodes) per user. The value of $\vec{p}$ can be chosen at will as long as higher value at a node indicates higher importance. We will formulate this vector latter in equation 10. λ is a tradeoff parameter determining the influence of personalization vector.

In the first step, λ is set to 1; while in the second step, λ is set to a value in $[0, 1]$. The final score vector is

$$\vec{w} = \vec{w}_{\lambda < 1} - \vec{w}_{\lambda = 1} \qquad (8)$$

Tags with the highest scores are selected for recommendations.

Note that many improvement and variations (Abel *et al.*, 2008; Abel *et al.*, 2009a; Abel *et al.*, 2009b) have been proposed since the publication of FolkRank. However, we'll focus on the naive FolkRank for the sake of simplicity since our main concern is the fusion of the two methods. After all, various improvement can also be applied to our method to achieve even better result.

4 Collaborative Summarization

As we can see from section 2 and section 3, the key ideas underlying *Affinity Propagation based document summarization* and *Collaborative Filtering based tag recommendation* are quite similar. This motivates our method to seamlessly fuse them together.

4.1 A Real Motivating Example

Before diving into the details of our method, let's first show an example on why we need to fuse the two methods mentioned above.

Here is a random bookmark[1] selected from www.delicious.com. It is a wikipedia article about Satisficing, which is decision-making strategy taking account of human limitations to pursue a near-optimal result. Three different users who have tagged at least three tags are listed below[2].

User A *economics* , theory , *psychology* , wikipedia , *language* , reference , work

User B *psychology* , strategy , *usability* , decisionmaking , satisficing , herbertsimon

User C article , *economics* , markets , theory

As we can see, user A makes a fairly comprehensive tagging of the article with three different emphases such as *economics*, *psychology* and *language*. Meanwhile, user B seems more concerned with *psychology* issue of this article while user C cares more about its *economics* implication.

Clearly, a summary of this article taking account of tags made by all users, as in the case of naive tag oriented document summarization, totally ignores these different interest of users. On the other hands, collaborative filtering based tag recommendation can provide a personalized tag prediction. That is to say, even if a user makes no tag on this article at all, we can still predict his/her interest point on this article by using the tagging history of that user and all others users. However, collaborative filtering totally ignores the actual content of this article(only uses tags information), which is the very concern of document summarization.

Thus, it's a natural thought that we can use collaborative filtering based tag recommendation to introduce personalized side information into document summarization. In retrospect, we can also improve collaborative filtering based tag recommendation by utilizing the summary information of an article.

4.2 Fusion of the Two Methods

Fortunately, the similar principle underlying both collaborative filtering based tag recommendation and affinity propagation based document summarization provides a natural basis to build a hyper algorithm. The idea is to build a hyper graph combining four factors: users, tags, documents and sentences.

With affinity propagation, we already have an affinity graph on all sentences of an article. With collaborative filtering, we already have a tripartite graph between users, tags and documents. The missing part is a graph connecting tags and sentences in a given article. As seen in Figure 1.

Here we first define a similarity between a tag t and a sentence s by mutual information:

$$\text{sim}(t, s) = \sum_{\forall \text{ term } w \in s} P(t, w) \log \frac{P(t, w)}{P(t)P(w)} \qquad (9)$$

---

[1] http://en.wikipedia.org/wiki/Satisficing

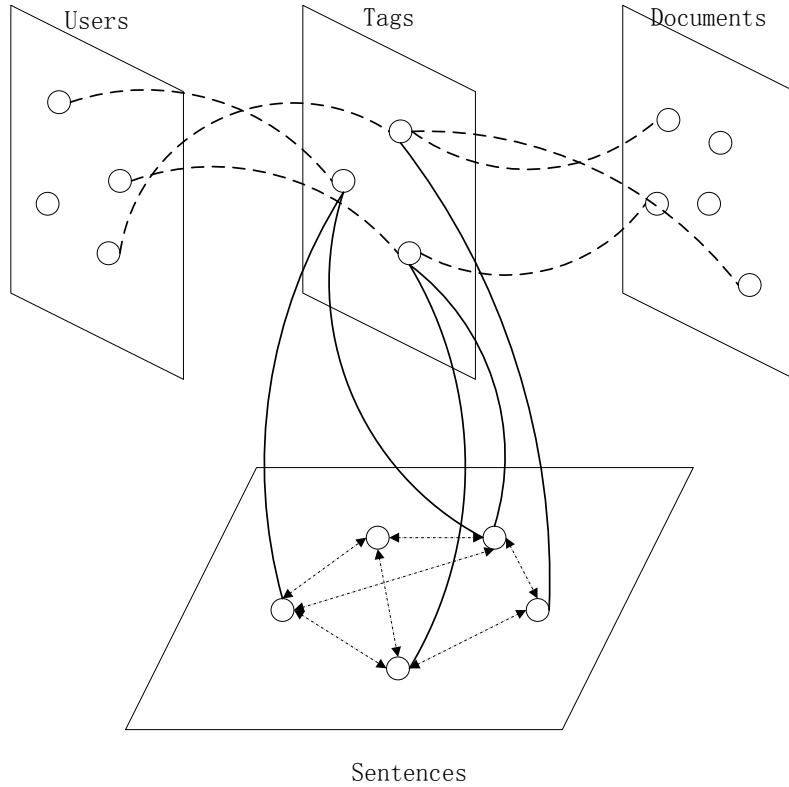[2] The emphasized tags are the most frequent tags on this article

Figure 1: Hyper graph combining users, tags , documents and sentences. The tripartite sub-graph in the upper half is built by collaborative filtering based tag recommendation model, the inter-sentence graph in the lower half is built using affinity propagation model. The edges connecting tags and sentences bridge the upper half and lower half. (Edges between users and documents are not drawn for clarification)

where $P(t, w)$ is the probability that tag t and term w occurs together in one article in the background corpus. $P(t)$ is the probability of observing tag t, which can be calculated by dividing the count of tag t by the total count of all tags. $P(w)$ is the probability of observing term w in the background corpus, which can be calculated by dividing the count of sentence containing term w by the total count of sentences in the background corpus. A similar definition can be found in the work of (Markines *et al.*, 2009).

Now we can draw an undirected edge between each pair of tag and sentence. Note that we can normalize the similarity per tag or per sentence, which will make the similarity not symmetric. Moreover, we can threshold the similarity and only connect a pair of tag and sentence only if the similarity between them is above a certain value.

The remaining problem is how to incorporate the new edge into two original graphs. As for the tripartite graph of collaborative filtering, we can achieve this by adjusting the personalization vector in equation 7. Suppose the information richness scores of sentences are already known, the personalization vector can be constructed by

$$\vec{p}(t) = d * \sum_{s \in S} \text{info}(s) * \text{sim}(t, s) + (1 - d) * \vec{p^*}(t) \qquad (10)$$

where $\vec{p^*}(t)$ is the initial value for $\vec{p}(t)$. d is damp-factor controlling the influence of initial values. It has no influence on the final result except the convergence rate. We will use the same damp-factor throughout the paper. From the equation, we can see that, the more informative sentences a

tag connects to, the more informative the tag is.

An importance issue with equation 10 is how to decide the values of $\vec{p^*}(t)$. As our candidate set of tags for a pair of user and document is the set of all tags, if the user has not assigned any tag to the document, we assign an equally large weight to the most possible tags and a small weight to all other tags. Following (Hotho *et al.*, 2006), we assign a weight of $|T|$ to all the tags used by the user and all the tags associated to the document, and assign 1 to all other tags, then normalize the weights to sum to 1. If the user has already tagged the document, we further multiply a factor of $|T|$ to user's tags then normalize. Generally, this will lead to different initial values of $\vec{p}$ for different users, in the hope to represent the users' taste.

Next, turning to sentence affinity graph, we can incorporate tag information by attaching a personalization value $\vec{q}(s)$ to each sentence s. This personalization vector can be updated similarly to $\vec{p}$

$$\vec{q}(s) = d * \sum_{t \in T} \vec{w}(t) * \text{sim}(t, s) + (1 - d) * \vec{q^*}(s) \tag{11}$$

where $\vec{w}$ is the score vector from equation 8.

Algorithm 1 Collaborative Summarization Algorithm

Input: A background corpus consisting of D documents with T tags assigned by U users.
  A user u and a document d on which a summarization is to be generated.
  The user's tags on the document is optional.
Output:
 1. A ranking of sentences of which the first k can be used as a summarization.
 2. A collection of tags recommended to the user.
% Initialization
build affinity graph G_A by equation 2
build tripartite graph G_{CF} by equation 4
initialize the personalization vector $\vec{q}$ of G_A to be equal to $\frac{1}{|S|}$
initialize the personalization vector $\vec{p}$ of G_{CF} following the strategy described above.
initialize $sim(t, s)$ by equation 9

while the procedure is not convergent **do**
  % Affinity Propagation
  while not convergent **do**
    update G_A using equation 12
  end while
  update $\vec{p}$ using equation 10

  % Collaborative Filtering
  while not convergent **do**
    update G_{CF} using equation 7 and equation 8
  end while
  update $\vec{q}$ using equation 11
end while

sort the information richness scores of all sentences of d and the $\vec{w}$ score vector from large to small

---

Accordingly, we need to substitute $\vec{q}$ for the original $\frac{1}{N}$ when updating information richness of sentence in equation 3.

$$\text{info}(s_i) = d \cdot \sum_{\forall\, j \neq i \in S} \text{info}(s_j) \cdot \text{aff}(s_j, s_i) + (1 - d) \cdot \vec{q}(s_i) \tag{12}$$

With modified affinity updating equation, the more important tags a sentence associated with, the more informative it becomes.

Until now, two sub-graphs induced by collaborative filtering and affinity propagation are connected together. A pagerank like algorithm can be applied to this hyper graph. To summarize, collaborative summarization following the procedure above to generate summarization.

We emphasize that this algorithm is based on naive Folkrank for tag recommendation, and various improvement over Folkrank can also be applied here to get a better $\vec{p}$ which we leave for practical implementation.

5 Experimentation

In this section, we design an experiment to demonstrate the effectiveness of collaborative summarization (CS) method. Two questions are addressed:(1) How much advantage does CS have compared with tag-based and content-based document summarization. (2) Can CS improve tag recommendation as well?

Although there are large amount of social tagging and bookmark resource on the internet, no standard summarization result for them exists, let alone personalized summary for a given document. Thus, we download a total of 100 English wikipedia articles from en.wikipedia.com as our document summarization dataset. Then we collect 1084 users' tagging data on 5000 wikipedia articles and a total of 8396 different tags from online bookmark site www.delicious.com. We select 10 active delicious.com users from the 1084 users, and let them extract the top-5 sentences from the 100 wikipedia articles dataset as the personalized summarization for the articles. Note that we don't require the selected users to tag the articles.

ROUGE-1 is used to evalute the effectiveness of the proposed method. We compare Collaborative Summarization(CS) to three other summarization methods: (1) Open Text Summarizer[3](OTS) which only scores each sentence without considering tags. (2) Affinity Propagation based summarization (AP) which is a graph-based method not considering tags. (3) EigenTag which uses all the tags assigned to a document without personalization. The result is shown in Table 1, which indicates significant improvement over all measurements.

All the dampfactor in Collaborative Summarization is set to 0.5, as it has nothing to do with final result. We set $\lambda = 0.8$ in the second step of Folkrank. The influence parameter in equation 7 is set to 0.95. The parameter d in equation 10 is set to 0.2. As for EigenTag, the parameter λ is set to 0.95 as the author advised.

For Affinity Propagation and Collaborative Summarization methods, redundancy removal is conducted in the post-process stage. Since the AP and CS share the same affinity measurement between sentences, we adopt the same redundancy removal technology as in Affinity Propagation. The idea is that we pick the most information-rich sentence out and penalize its neighbours to promote diversity, then repeat this process to generate a list of sentences. Readers are referred to (Wan and Xiao, 2007) for more details.

Table 1: Comparison of various methods on document summarization

	OST	AP	EigenTag	CS
Precision	0.424824	0.451688	0.484586	**0.524037**
Recall	0.4280267	0.437867	0.463866	**0.508515**
F-measure	0.426419	0.444670	0.474000	**0.516159**

Although it's not our main concern in this paper, we also compare Collaborative Summarization on its performance of tag recommendation with naive Folkrank. Not surprisingly, CS outperforms

---

[3] http://libots.sourceforge.net

Folkrank because the summary of document provides useful side information for picking tags. The result is evaluated using standard precision and recall as shown in Table 2. It indicates that the fusion of both methods benefit each other.

Table 2: Comparison of Collaborative Summarization and Folkrank on tag recommendation

	FolkRank	CS
Precision	0.362330	**0.499431**
Recall	0.181965	**0.233019**
F-measure	0.242264	**0.317774**

6 Related Work

In this section, we describe the relationship between our method and other approaches in both graph-based document summarization and tag recommendation. The contexts of these two topics are too rich to cover up here, so we only focus on graph-based approaches.

6.1 Graph based Document Summarization

Graph based document summarization methods have been proposed for both single and mult-document summarization (Erkan and Radev, 2004; Mani and Bloedorn, 2000; Mihalcea and Tarau, 2005). Many of them are based on PageRank (Brin and Page, 1998) or HITS (Kleinberg, 1999).

One advantage of graph-based document summarization methods is that it can easily incorporate various side information, such as user profile, user query (Saggion *et al.*, 2003; Ge *et al.*, 2003; Conroy and Schlesinger, 2005) and specific topic (Farzindar *et al.*, 2005). However, there are two issues with most existing user/query/topic-oriented document summarization. First, current graph-based methods usually focus on given single or multiple documents to generate summarization while ignoring all other documents in the corpus, which seems rather reasonable at the first glance. However, the information or regularity contained in the whole corpus of document can serve as a common sense for the summarization of specific documents. It has been shown in (Wan and Xiao, 2008a; Wan and Xiao, 2008b) that incorporating these information significantly improves the performance. Our method is also along this line of thinking but following a different direction, i.e. by incorporating the wisdom of the population of people rather than documents.

Secondly, most existing document summarization methods using side information post a heavy burden on users. A user either has to specify his/her profile or topic of interest (Zhang *et al.*, 2003), or has to first tag a document (Zhu *et al.*, 2009) in order to receive a personalized summarization. The issue is most obvious in the scenario that a user needs personalized summarization of a totally new document, which is often the case in real life. As far as we know, no effort has been made to passively mining user's point of interest to provide a personalized summarization. Our work, using ideas from collaborative filtering, may shed some light in this direction of research.

6.2 Tag Recommendation

Tag recommendation embraces a broad range of objects including document, image, video and so on. Since the content to be tagged varies a lot across different applications, collaborative filtering based tag recommendation has an appealing advantage that it is content independent. After Folkrank was proposed for solving the problem, a lot of refinement (Wu *et al.*, 2006) is made to incorporate more useful information, such as the concept of groups of tags (Abel *et al.*, 2008). However, when facing the problem of document summarization, we need to further incorporate the information from document content. Our method provides one seamless way to do so while some related work (Tso-Sutter *et al.*, 2008) follows other directions.

7   Conclusion

In this paper, we argue that single document summarization can be significantly improved by harvesting human computing power made available by the rapid out-growth of internet technology. There is no surprise that population of human brains can and will complement most existing language models and algorithms in complex NLP tasks for a long time. The philosophy implied here definitely has a larger range of application than document summarization. For example, active learning in machine learning literature shares the same characteristics involving an optimal collaboration between machines and users.

More specifically, by combining collaborative filtering and affinity propagation together, we show that we can make personalized summarization for a given user even on a totally new document for that user. Thus, the proposed method reduces the burden of user in a personalized summarization system. Moreover, the side information contained in tags presents both the general meaning and personal interest of the document. Incorporating such side information is shown to boost the performance significantly.

References

Abel, Fabian, Matteo Baldoni, Cristina Baroglio, Nicola Henze, Daniel Krause and Viviana Patti. 2009a. Context-based ranking in folksonomies. *HT '09: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia.*

Abel, Fabian, Nicola Henze and Daniel Krause. 2008. A Novel Approach to Social Tagging: GroupMe! *4th Int. Conf. on Web Information Systems and Technologies (WEBIST).*

Abel, Fabian, Nicola Henze, Daniel Krause and Matthias Kriesell. 2009b. On the Effect of Group Structures on Ranking Strategies in Folksonomies. *Weaving Services and People on the World Wide Web.*

Brin, S. and L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30(1-7)*, 107-117.

Conroy, J.M. and J. D. Schlesinger. 2005. CLASSY query-based multi-document summarization. *Proceedings of the 2005 Document Understanding Workshop.*

Erkan, G. and D. Radev. 2004. LexPageRank: prestige in multi-document text summarization. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, Barcelona, Spain.

Farzindar, A., F. Rozon and G. Lapalme. 2005. CATS a topic-oriented multi-document summarization system at DUC 2005. *Proceedings of the 2005 Document Understanding Workshop.*

Ge, J., X. Huang and L. Wu. 2003. Approaches to event-focused summarization based on named entities and query words. *Proceedings of the 2003 Document Understanding Workshop.*

Hotho, Andreas, Robert Jäschke, Christoph Schmitz and Gerd Stumme. 2006. FolkRank: A Ranking Algorithm for Folksonomies. *Proc. FGIR 2006.*

Kleinberg, J.M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM, 46(5)*, 604-632.

Mani, I. and E. Bloedorn. 2000. Summarizing Similarities and Differences Among Related Documents. *Information Retrieval 1(1).*

Markines, Benjamin, Ciro Cattuto, Filippo Menczer, Dominik Benz, Andreas Hotho and Gerd Stumme. 2009. Evaluating Similarity Measures for Emergent Semantics of Social Tagging. *WWW '09: Proceedings of the 18th International Conference on World Wide Web.*

Mihalcea, R. and P. Tarau. 2005. A language independent algorithm for single and multiple document summarization. *IJCNLP 2005. LNCS (LNAI), vol. 3651*, pp. 19-24. Springer, Heidelberg.

Saggion, H., K. Bontcheva and H. Cunningham. 2003. Robust generic and query-based summarization. *Proceedings of EACL-2003.*

Tso-Sutter, Karen H. L., Leandro Balby Marinho and Lars Schmidt-Thieme. 2008. Tag-aware recommender systems by fusion of collaborative filtering algorithms. *SAC '08: Proceedings of the 2008 ACM Symposium on Applied Computing.*

Wan, Xiaojun and Jianguo Xiao. 2007. Towards a Unified Approach Based on Affinity Graph to Various Multi-document Summarizations. *ECDL2007*, pp. 297-308.

Wan, Xiaojun and Jianguo Xiao. 2008a. Single Document Keyphrase Extraction Using Neighborhood Knowledge. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008*, Chicago, Illinois, USA.

Wan, Xiaojun and Jianguo Xiao. 2008b. CollabRank: Towards a Collaborative Approach to Single-Document Keyphrase Extraction. *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008).*

Wu, H., M. Zubair and K. Maly. 2006. Harvesting social knowledge from folksonomies. *The 17th Conf. on Hypertext and Hypermedia (HT '06)*, pp. 111-114, New York, NY, USA. ACM Press.

Zhang, Haiqin, Zheng Chen, Wei-ying Ma and Qingsheng Cai. 2003. A study for documents summarization based on personal annotation. *Proceedings of the HLT-NAACL 03 Workshop on Text Summarization.*

Zhu, Junyan, Can Wang, Xiaofei He, Jiajun Bu, Chun Chen, Shujie Shang, Mingcheng Qu and Gang Lu. 2009. Tag-oriented document summarization. *WWW '09: Proceedings of the 18th International Conference on World Wide Web.*