Linking Topics of News and Blogs with Wikipedia for Complementary Navigation

Yuki Sato[1], Daisuke Yokomoto[2], Hiroyuki Nakasaki[1], Mariko Kawaba[3], Takehito Utsuro[1], Tomohiro Fukuhara[4]

[1] Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, 305-8573, Japan
[2] College of Engineering Systems, Third Cluster of Colleges, University of Tsukuba, Tsukuba, 305-8573, Japan
[3] NTT Cyber Space Laboratories, NTT Corporation, Yokosuka, Kanagawa, 239-0847, JAPAN
[4] Research into Artifacts, Center for Engineering, University of Tokyo, Kashiwa 277-8568, Japan

Abstract. We study complementary navigation of news and blog, where *Wikipedia* entries are utilized as fundamental knowledge source for linking news articles and blog feeds/posts. In the proposed framework, given a topic as the title of a Wikipedia entry, its Wikipedia entry body text is analyzed as fundamental knowledge source for the given topic, and terms strongly related to the given topic are extracted. Those terms are then used for ranking news articles and blog posts. In the scenario of complementary navigation from a news article to closely related blog posts, Japanese Wikipedia entries are ranked according to the number of strongly related terms shared by the given news article and each Wikipedia entry. Then, top ranked 10 entries are regarded as indices for further retrieving closely related blog posts. The retrieved blog posts are finally ranked all together. The retrieved blog posts are then shown to users as blogs of personal opinions and experiences that are closely related to the given news article. In our preliminary evaluation, through an interface for manually selecting relevant Wikipedia entries, the rate of successfully retrieving relevant blog posts improved.

Keywords: IR, Wikipedia, news, blog, topic analysis

1 Introduction

We study complementary navigation of news and blog, where Wikipedia entries are utilized as fundamental knowledge source for linking news articles and blog feeds/posts. In previous works, Wikipedia, news, and blogs are intensively studied in a wide variety of research activities. In the area of IR, Wikipedia has been studied as rich knowledge source for improving the performance of text classification [1, 2] as well as text clustering [3–5]. In the area of NLP, it has been studied as language resource for improving the performance of named

entity recognition [6, 7], translation knowledge acquisition [8], word sense disambiguation [9], and lexical knowledge acquisition [10]. In previous works on news aggregation such as Newsblaster [11], NewsInEssence[1] [12], and *Google News*[2], techniques on linking closely related news articles were intensively studied. In addition to those previous works on use and analysis of Wikipedia and news, blog analysis services have also become popular. Blogs are considered to be one of personal journals, market or product commentaries. While traditional search engines continue to discover and index blogs, the blogosphere has produced custom blog search and analysis engines, systems that employ specialized information retrieval techniques. With respect to blog analysis services on the Internet, there are several commercial and non-commercial services such as *Technorati*[3], *BlogPulse*[4] [13], *kizasi.jp*[5], and *blogWatcher*[6] [14]. With respect to multilingual blog services, *Globe of Blogs*[7] provides a retrieval function of blog articles across languages. *Best Blogs in Asia Directory*[8] also provides a retrieval function for Asian language blogs. *Blogwise*[9] also analyzes multilingual blog articles.

Compared to those previous studies, the fundamental idea of our complementary navigation can be roughly illustrated in Figure 1. In our framework of complementary navigation of news and blog, Wikipedia entries are retrieved when seeking fundamental background information, while news articles are retrieved when seeking precise news reports on facts, and blog feeds/posts are retrieved when seeking subjective information such as personal opinions and experiences.

In the proposed framework, we regard Wikipedia as a large scale encyclopedic knowledge base which includes well known facts and relatively neutral opinions. In its Japanese version, about 627,000 entries are included (checked at October, 2009). Given a topic as the title of a Wikipedia entry, its Wikipedia entry body text is analyzed as fundamental knowledge source for the given topic, and terms strongly related to the given topic are extracted. Those terms are then used for ranking news articles and blog feeds/posts. This fundamental technique was published in [15, 16] and was evaluated in the task of blog feed retrieval from a Wikipedia entry. [15, 16] reported that this technique outperformed the original ranking returned by "Yahoo! Japan" API.

In the first scenario of complementary navigation, given a news article of a certain topic, the system retrieves blog feeds/posts of closely related topics and show them to users. In the case of the example shown in Figure 1, suppose that a user found a news article reporting that "a long queue appeared in front of a game shop on the day a popular game Dragon Quest 9 was published".

[1] http://www.newsinessence.com/nie.cgi
[2] http://news.google.com/
[3] http://technorati.com/
[4] http://www.blogpulse.com/
[5] http://kizasi.jp/ (in Japanese)
[6] http://blogwatcher.pi.titech.ac.jp/ (in Japanese)
[7] http://www.globeofblogs.com/
[8] http://www.misohoni.com/bba/
[9] http://www.blogwise.com/

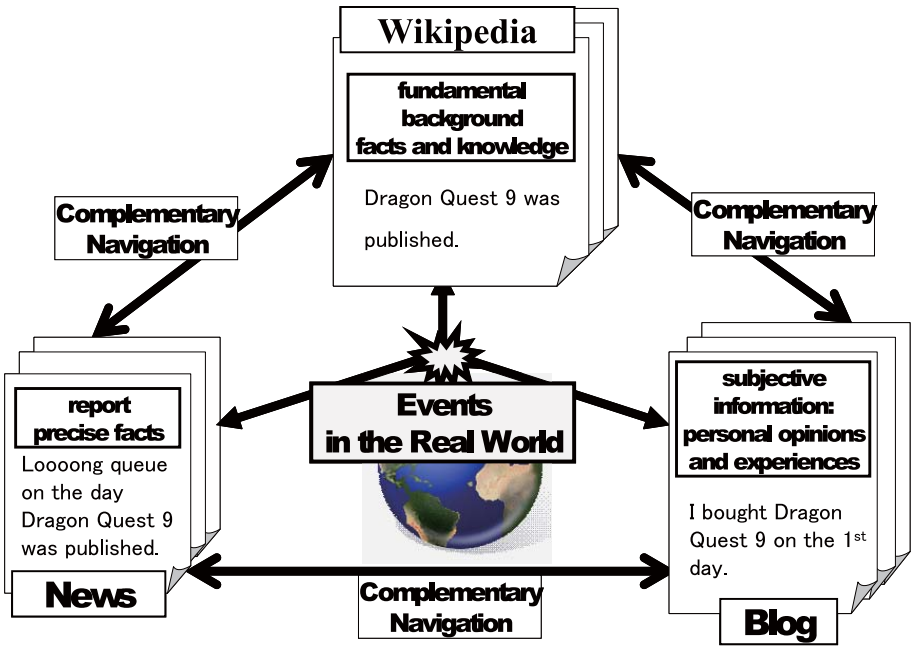


Fig. 1. Framework of Complementary Navigation among Wikipedia, News, and Blogs

Then, through the function of the complementary navigation of our framework, a closely related blog post, such as the one posted by a person who bought the game on the day it was published, is quickly retrieved and shown to the user. In the scenario of this direction, first, about 600,000 Japanese Wikipedia entries are ranked according to the number of strongly related terms shared by the given news article and each Wikipedia entry. Then, top ranked 10 entries are regarded as indices for further retrieving closely related blog feeds/posts. The retrieved blog feeds/posts are finally ranked all together. The retrieved blog feeds/posts are then shown to users as blogs of personal opinions and experiences that are closely related to the given news article.

In the second scenario of complementary navigation, which is the opposite direction from the first one, given a blog feed/post of a certain topic, the system retrieves news articles of closely related topics and show them to users. This scenario is primarily intended that, given a blog feed/post which refers to a certain news article and includes some personal opinions regarding the news, the system retrieves the news article referred to by the blog feed/post and show it to users.

Finally, in the third scenario of complementary navigation, given a news article or a blog feed/post of a certain topic, the system retrieves one or more closely related Wikipedia entries and show them to users. In the case of the example shown in Figure 1, suppose that a user found either a news article reporting the publication of Dragon Quest 9 or a blog post by a person who

bought the game on the day it was published. Then, through the function of the complementary navigation of our framework, the most relevant Wikipedia entry, namely, that of Dragon Quest 9, is quickly retrieved and shown to the user. This scenario is intended to show users background knowledge found in Wikipedia, given a news article or a blog feed/post of a certain topic.

Based on the introduction of the overall framework of complementary navigation among Wikipedia, news, and blogs above, this paper focuses on the formalization of the first scenario of complementary navigation for retrieving closely related blog posts given a news article of a certain topic. Section 2 first describes how to extract terms that are included in each Wikipedia entry and are closely related to it. According to the procedure to be presented in section 3, those terms are then used to retrieve blog posts that are closely related to each Wikipedia entry. Based on those fundamental techniques, section 4 formalizes the similarity measure between the given news article and each blog post, and then presents the procedure of ranking blog posts that are related to the given news article. Section 5 introduces a user interface for complementary navigation, to be used for manually selecting Wikipedia entries which are relevant to the given news article and are effective in retrieving closely related blog posts. Section 5 also presents results of evaluating our framework. Section 6 presents comparison with previous works related to this paper.

2 Extracting Related Terms from a Wikipedia Entry

In our framework of linking news and blogs through Wikipedia entries, we regard terms that are included in each Wikipedia entry body text and are closely related to the entry as representing conceptual indices of the entry. Those closely related terms are then used for retrieving related blog posts and news articles. More specifically, from the body text of each Wikipedia entry, we extract boldfaced terms, anchor texts of hyperlinks, and the title of a *redirect*, which is a synonymous term of the title of the target page [15–17]. We also extract all the noun phrases from the body text of each Wikipedia entry.

3 The Procedure of Retrieving Blog Posts related to a Wikipedia Entry

This section describes the procedure of retrieving blog posts that are related to a Wikipedia entry [15, 16]. In this procedure, given a Wikipedia entry title, first, closely related blog feeds are retrieved, and then, from the retrieved blog feeds, closely related blog posts are further selected.

3.1 Blog Feed Retrieval

This section briefly describes how to retrieve blog feeds given a Wikipedia entry title.

In order to collect candidates of blog feeds for a given query, in this paper, we use existing Web search engine APIs, which return a ranked list of blog posts, given a topic keyword. We use the Japanese search engine "Yahoo! Japan" API[10]. Blog hosts are limited to major 11 hosts[11]. We employ the following procedure for the blog distillation:

i) Given a topic keyword, a ranked list of blog posts are returned by a Web search engine API.
ii) A list of blog feeds is generated from the returned ranked list of blog posts by simply removing duplicated feeds.
iii) Re-rank the list of blog feeds according to the number of hits of the topic keyword in each blog feed. The number of hits for a topic keyword in each blog feed is simply measured by the search engine API used for collecting blog posts above in i), restricting the domain of the URL to each blog feed.

[15, 16] reported that the procedure above outperformed the original ranking returned by "Yahoo! Japan" API.

3.2 Blog Post Retrieval

From the retrieved blog feeds, we next select blog posts that are closely related to the given Wikipedia entry title. To do this, we use related terms extracted from the given Wikipedia entry as described in section 2. More specifically, out of the extracted related terms, we use bold-faced terms, anchor texts of hyperlinks, and the title of a *redirect*, which is a synonymous term of the title of the target page. Then, blog posts which contain the topic name or at least one of the extracted related terms are automatically selected.

4 Similarities of Wikipedia Entries, News, and Blogs

In the scenario of retrieving blog posts closely related to a given news article, the most important component is how to measure the similarity between the given news article and each blog post. This section describes how we design this similarity.

In this scenario, the fundamental component is how to measure the similarity $Sim_{w,n}(E, N)$ between a Wikipedia entry E and a news article N, and the similarity $Sim_{w,b}(E, B)$ between a Wikipedia entry E and a blog post B. The similarity measure $Sim_{w,n}(E, N)$ is used when, given a news article of a certain topic, ranking Wikipedia entries according to whether each entry is related to the given news article. The similarity measure $Sim_{w,b}(E, B)$ is used when, from the highly ranked Wikipedia entries closely related to the given news article, retrieving blog posts related to any of those entries. Then, based on those similarities

[10] http://www.yahoo.co.jp/ (in Japanese)
[11] FC2.com, yahoo.co.jp, rakuten.ne.jp, ameblo.jp, goo.ne.jp, livedoor.jp, Seesaa.net, jugem.jp, yaplog.jp, webry.info.jp, hatena.ne.jp

$Sim_{w,n}(E, N)$ and $Sim_{w,b}(E, B)$, the overall similarity measure $Sim_{n,w,b}(N, B)$ between the given news article N and each blog post B is introduced. Finally, blog posts are ranked according to this overall similarity measure.

4.1 Similarity of a Wikipedia Entry and a News Article / a Blog Post

The similarities $Sim_{w,n}(E, N)$ and $Sim_{w,b}(E, B)$ are measured in terms of the entry title and the related terms extracted from the Wikipedia entry as described in section 2. The similarity $Sim_{w,n}(E, N)$ between a Wikipedia entry E and a news article N is defined as a weighted sum of frequencies of the entry title and the related terms:

$$Sim_{w,n}(E, N) = \sum_{t} w(type(t)) \times freq(t)$$

where $weight(t)$ is defined as 1 when t is the entry title, the title of a *redirect*, a bold-faced term, the title of a paragraph, or a noun phrase extracted from the body text of the entry. The similarity $Sim_{w,b}(E, B)$ between a Wikipedia entry E and a blog post B is defined as a weighted sum of frequencies of the entry title and the related terms:

$$Sim_{w,b}(E, B) = \sum_{t} w(type(t)) \times freq(t)$$

where $weight(t)$ is defined as 3 when t is the entry title or the title of a *redirect*, as 2 when t is a bold-faced term, and as 0.5 when t is an anchor text of hyperlinks[12].

4.2 Similarity of a News Article and a Blog Post through Wikipedia Entries

In the design of the overall similarity measure $Sim_{n,w,b}(N, B)$ between a news article N and a blog post B through Wikipedia entries, we consider two factors. One of them is to measure the similarity between a news article and a blog post indirectly through Wikipedia entries which are closely related to both of the news article and the blog post. The other is, on the other hand, to directly measure their similarity simply based on their text contents. In this paper, the first factor is represented as the sum of the similarity $Sim_{w,n}(E, N)$ between a news article N and a Wikipedia entry E and the similarity $Sim_{w,b}(E, B)$ between a blog post B and a Wikipedia entry E. The second factor is denoted as the direct document

[12] In [17], we applied machine learning technique to the task of judging whether a Wikipedia entry and a blog feed are closely related, where we incorporated features other than the frequencies of related terms in a blog feed and achieved improvement. Following the discussion in [15, 16], the technique proposed by [17] outperforms the original ranking returned by "Yahoo! Japan" API. As a future work, we are planning to apply the technique of [17] to the task of complementary navigation studied in this paper.

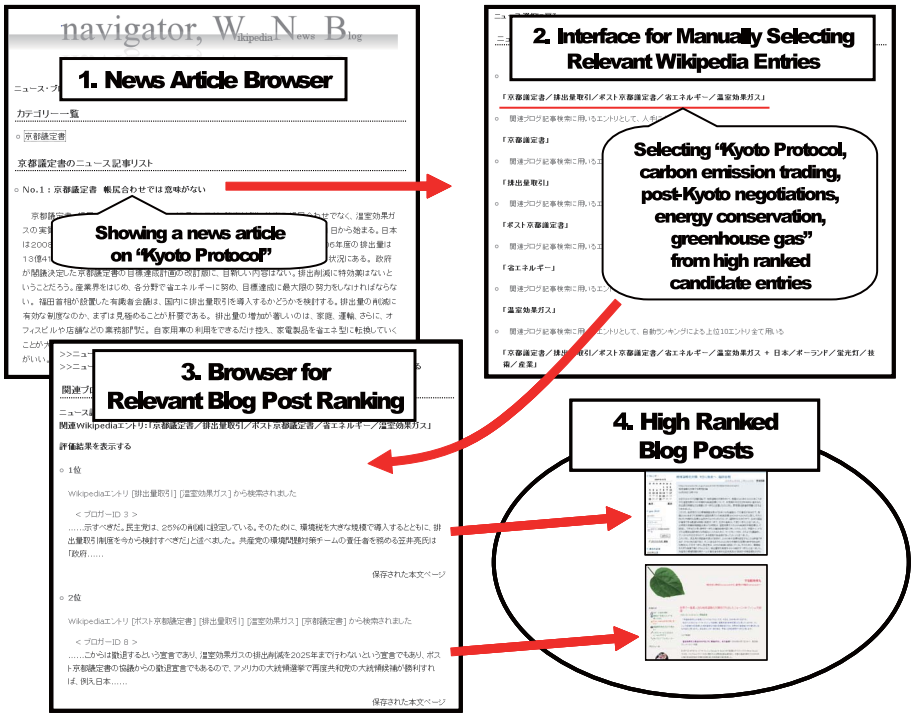


Fig. 2. Interface for Complementary Navigation from News to Blogs through Wikipedia Entries

similarity $Sim_{n,b}(N, B)$ between a news article N and a blog post B, where we simply use cosine measure as the direct document similarity. Finally, based on the argument above, we define the overall similarity measure $Sim_{n,w,b}(N, B)$ between a news article N and a blog post B through Wikipedia entries as the weighted sum of the two factors below:

$$Sim_{n,w,b}(N, B)$$
$$= (1 - K_{w,nb})Sim_{n,b}(N, B) + K_{w,nb} \sum_{E} \Big(Sim_{w,n}(E, N) + Sim_{w,b}(E, B) \Big)$$

where $K_{w,nb}$ is the coefficient for the weight. In the evaluation of section 5.2, we show results with this coefficient $K_{w,nb}$ as 1, since the results with $K_{w,nb}$ as 1 are always better than those with $K_{w,nb}$ as 0.5.

4.3 Ranking Blog Posts related to News through Wikipedia Entries

Based on the formalization in the previous two sections, given a news article N, this section presents the procedure of retrieving blog posts closely related to the given news article and then ranking them.

First, suppose that the news article N contains titles of Wikipedia entries E_1,..., E_n in its body text. Then, those entries E_1,...,E_n are ranked according to their similarities $Sim_{w,n}(E_i, N)$ $(i = 1, \ldots, n)$ against the given news article N, and top ranked 10 entries $E'_1, \ldots, E'_{10}$ are selected. Next, each E'_i $(i = 1, \ldots, 10)$ of those top ranked 10 entries are used to retrieve closely related blog posts according to the procedure presented in section 3. Finally, the retrieved blog posts $B_1, \ldots, B_m$ all together are ranked according to their similarities $Sim_{n,w,b}(N, B_j)$ $(j = 1, \ldots, m)$ against the given news article N.

5 Manually Selecting Wikipedia entries in Linking News to related Blog Posts

In this section, we introduce a user interface for complementary navigation with a facility of manually selecting Wikipedia entries which are relevant to the given news article. With this interface, a user can judge whether each candidate Wikipedia entry is effective in retrieving closely related blog posts. We then evaluate the overall framework of complementary navigation and present the evaluation results.

5.1 The Procedure

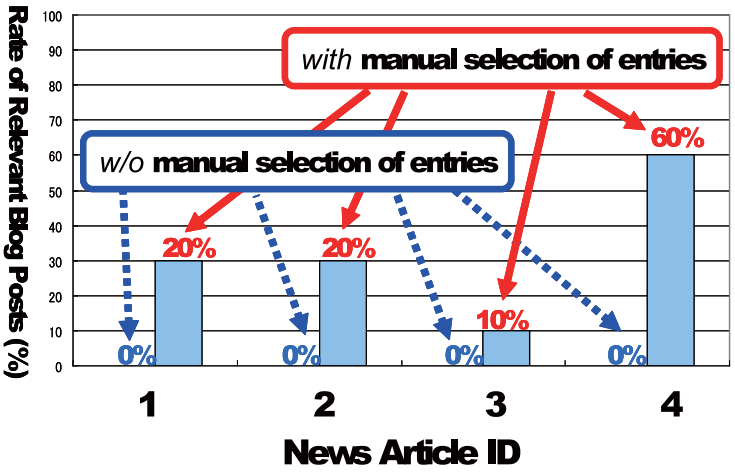
This section describes the procedure of linking a news article to closely related blog posts, where the measure for ranking related blog posts is based on the formalization presented in section 4.3. In this procedure, we also use an interface for manually selecting Wikipedia entries which are relevant to the given news article.

The snapshots of the interface are shown in Figure 2. First, in "News Article Browser", a user can browse through a list of news articles and can select one for which he/she wants to retrieve related blog posts. Next, for the selected news article, "Interface for Manually Selecting Relevant Wikipedia Entries" appears. In this interface, following the formalization of section 4.3, top ranked 10 Wikipedia entry titles are shown as candidates for retrieving blog posts that are related to the given news article. Then, the user can select any subset of the 10 candidate Wikipedia entry titles to be used for retrieving related blog posts. With the subset of the selected Wikipedia entry titles, "Browser for Relevant Blog Post Ranking" is called, where the retrieved blog posts are ranked according to the formalization of section 4.3. Finally, the user can browse through "High Ranked Blog Posts" by simply clicking the links to those blog posts.

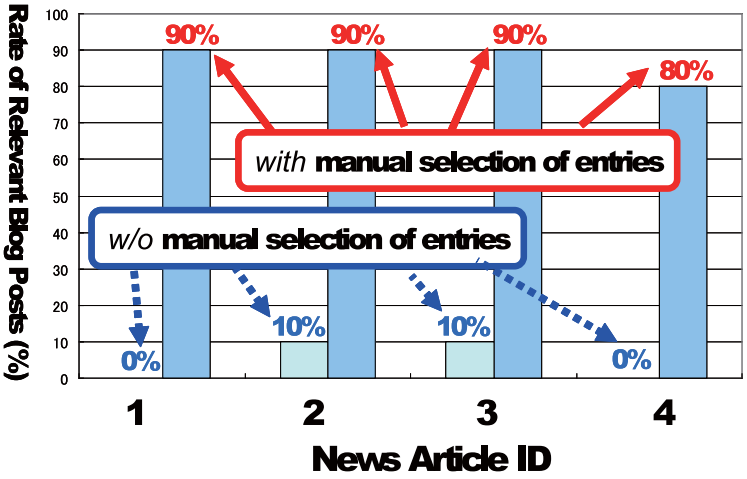
Table 1 shows a list of four news articles on "Kyoto Protocol" to be used in the evaluation of next section. For each news article, the table shows its summary and top ranked 10 Wikipedia entry titles, where entry titles judged as relevant to the news article are in squares. The table also shows the summary of an example of relevant blog posts.

Table 1. Summaries of News Articles for Evaluation, Candidates for Relevant Wikipedia Entries, and Summaries of Relevant Blog Posts

news article ID	summary of news article	top ranked 10 entries as candidates for relevant Wikipedia entries (manually selected entries are [in a square])	summary of relevant blog posts
1	Reports on Japan's activities on *"carbon offset"*, reduction of electric power consumption, and preventing global warming. (date: Jan. 25, 2008)	environmental issues, [Kyoto Protocol], Japan, automobile, [carbon offset], transport, United States, hotel, [carbon dioxide], contribution	"I understand the significance of Kyoto protocol, but I think it also has problems." (blogger A)
2	Reports on a meeting for *"carbon offset"*. (date: Mar. 31, 2008)	[Kyoto Protocol], [carbon emissions trading], Japan, [post-Kyoto negotiations], [energy conservation], Poland, fluorescent lamp, technology, [greenhouse gases], industry	"Japan has to rely on economic approaches such as *carbon offset*." (blogger A)
3	Reports on issues towards post-Kyoto negotiations. (date: Aug. 28, 2008)	[post-Kyoto negotiations], United Nations, protocol, [carbon dioxide], United States, debate, Kyoto, [greenhouse gases], minister, Poland	Referring to a news article on World Economic Forum. (blogger B)
4	Discussion on global warming such as issues regarding developing countries and technologies for energy conservation in Japan. (date: Jun. 29, 2008)	Japan, [global warming], environmental issues, United States, politics, resource, [34th G8 summit], India, [fossil fuels], society	Engineers of Japanese electric power companies make progress in research and development. (blogger C)



(a) *relevant* blog posts = closely related blog posts only



(b) *relevant* blog posts = closely related blog posts + partially related blog posts

Fig. 3. Evaluation Results of the Ratio of *Relevant* Blog Posts (%): Comparison of *with / without* Manual Selection of *Relevant* Wikipedia Entries

5.2 Evaluation

The Procedure To each of the four news articles on "Kyoto Protocol" listed in Table 1, we apply the procedure of retrieving related blog posts described in the previous section. We then manually judge the relevance of top ranked N blog posts into the following three levels, i.e., (i) closely related, (ii) partially related,

and (iii) not related. Next, we consider the following two cases in measuring the rate of *relevant* blog posts:

(a) Only closely related blog posts (judged as (i)) are regarded as *relevant*.
(b) Both closely related blog posts (judged as (i)) and partially related blog posts (judged as (ii)) are regarded as *relevant*.

For both cases, the rate of *relevant* blog posts is simply defined as below:

$$\text{rate of } \textit{relevant} \text{ blog posts} = \frac{\text{the number of } \textit{relevant} \text{ blog posts}}{N}$$

In the evaluation of this section, we set N as 10.

Evaluation Results In terms of the rate of *relevant* blog posts, Figure 3 compares the two cases of *with / without* manually selecting Wikipedia entries relevant to the given news article through the interface introduced in the previous section. In Figure 3 (a), we regard only closely related blog posts as *relevant*, where the rates of *relevant* blog posts improve from 0% to 10∼60%. In Figure 3 (b), we regard both closely and partially related blog posts as *relevant*, where the rates of *relevant* blog posts improve from 0∼10% to 80∼90%.

With this result, it is clear that, the current formalization presented in this paper has its weakness in the similarity measure for ranking related Wikipedia entries. As can be seen in the list of top ranked 10 Wikipedia entry titles in Table 1 as well as those manually selected out of the 10 entries, general terms and country names such as "automobile", "transport", "Japan", and "United States" are major causes of low rates of relevancy. Those general terms and country names mostly damage the step of retrieving related blog posts and the final ranking of those retrieved blog posts. However, it is also clearly shown that, once closely related Wikipedia entries are manually selected, the rates of *relevant* blog posts drastically improved. This result obviously indicates that the most important issue to be examined first is how to model the measure for ranking Wikipedia entries which are related to a given news article. We discuss this issue as a future work in section 7.

6 Related Works

Among several related works, [18, 19] studied linking related news and blogs, where their approaches are different from that proposed in this paper in that our proposed method conceptually links topics of news articles and blog posts based on Wikipedia entry texts. [18] focused on linking news articles and blogs based on cites from blogs to news articles. [19] studied to link news articles to blogs posted within one week after each news article is released, where a document vector space model modified by considering terms closely related to each news articles is employed.

[20] also studied mining comparative differences of concerns in news streams from multiple sources. [21] studied how to analyze sentiment distribution in news

articles across 9 languages. Those previous works mainly focus on news streams and documents other than blogs.

Techniques studied in previous works on text classification [1, 2] as well as text clustering [3–5] using Wikipedia knowledge are similar to the method proposed in this paper in that they are based on related terms extracted from Wikipedia, such as hyponyms, synonyms, and associated terms. The fundamental ideas of those previously studied techniques are also applicable to our task. Major differences between our work and those works are in that we design our framework as having the intermediate phase of ranking Wikipedia entries related to a given news article.

7 Conclusion

This paper studied complementary navigation of news and blog, where Wikipedia entries are utilized as fundamental knowledge source for linking news articles and blog posts. In this paper, we focused on the scenario of complementary navigation from a news article to closely related blog posts. In our preliminary evaluation, we showed that the rate of successfully retrieving relevant blog posts improved through an interface for manually selecting relevant Wikipedia entries. Future works include improving the measure for ranking Wikipedia entries which are related to a given news article. So far, we have examined a novel measure which incorporates clustering of Wikipedia entries in terms of the similarity of their body texts. The underlying motivation of this novel measure is to prefer a small number of entries which have quite high similarities with each other, and we have already confirmed that this approach drastically improves the ranking of Wikipedia entries. We are planning to evaluate this measure against a much larger evaluation data set and the result will be reported in the near future.

References

1. Gabrilovich, E., Markovitch, S.: Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge. In: Proc. 21st AAAI. (2006) 1301–1306
2. Wang, P., Domeniconi, C.: Building semantic kernels for text classification using Wikipedia. In: Proc. 14th SIGKDD. (2008) 713–721
3. Hu, J., Fang, L., Cao, Y., Zeng, H.J., Li, H., Yang, Q., Chen, Z.: Enhancing text clustering by leveraging Wikipedia semantics. In: Proc. 31st SIGIR. (2008) 179–186
4. Huang, A., Frank, E., Witten, I.H.: Clustering document using a Wikipedia-based concept representation. In: Proc. 13th PAKDD. (2009) 628–636
5. Hu, X., Zhang, X., Lu, C., Park, E.K., Zhou, X.: Exploiting Wikipedia as external knowledge for document clustering. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2009) 389–396
6. Cucerzan, S.: Large-scale named entity disambiguation based on Wikipedia data. In: Proc. EMNLP-CoNLL. (2007) 708–716

7. Kazama, J., Torisawa, K.: Exploiting Wikipedia as external knowledge for named entity recognition. In: Proc. EMNLP-CoNLL. (2007) 698–707
8. Oh, J.H., Kawahara, D., Uchimoto, K., Kazama, J., Torisawa, K.: Enriching multilingual language resources by discovering missing cross-language links in Wikipedia. In: Proc. 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. (2008) 322–328
9. Mihalcea, R., Csomai, A.: Wikify! linking documents to encyclopedic knowledge. In: Proceedings of the 16th ACM Conference on Information and Knowledge Management. (2007) 233–242
10. Sumida, A., Torisawa, K.: Hacking Wikipedia for hyponymy relation acquisition. In: Proc. 3rd IJCNLP. (2008) 883–888
11. McKeown, K.R., Barzilay, R., Evans, D., Hatzivassiloglou, V., Klavans, J.L., Nenkova, A., Sable, C., Schiffman, B., Sigelman, S.: Tracking and summarizing news on a daily basis with Columbia's Newsblaster. In: Pro. 2nd HLT. (2002) 280–285
12. Radev, D., Otterbacher, J., Winkel, A., Blair-Goldensohn, S.: NewsInEssence: Summarizing online news topics. Communications of the ACM **48** (2005) 95–98
13. Glance, N., Hurst, M., Tomokiyo, T.: Blogpulse: Automated trend discovery for Weblogs. In: WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. (2004)
14. Nanno, T., Fujiki, T., Suzuki, Y., Okumura, M.: Automatically collecting, monitoring, and mining Japanese weblogs. In: WWW Alt. '04: Proc. 13th WWW Conf. Alternate Track Papers & Posters. (2004) 320–321
15. Kawaba, M., Nakasaki, H., Utsuro, T., Fukuhara, T.: Cross-lingual blog analysis based on multilingual blog distillation from multilingual Wikipedia entries. In: Proceedings of International Conference on Weblogs and Social Media. (2008) 200–201
16. Nakasaki, H., Kawaba, M., Yamazaki, S., Utsuro, T., Fukuhara, T.: Visualizing cross-lingual/cross-cultural differences in concerns in multilingual blogs. In: Proceedings of International Conference on Weblogs and Social Media. (2009) 270–273
17. Kawaba, M., Yokomoto, D., Nakasaki, H., Utsuro, T., Fukuhara, T.: Linking Wikipedia entries to blog feeds by machine learning. In: Proc. 3rd IUCS. (2009)
18. Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, M., Konig, A.C.: Blews: Using blogs to provide context for news articles. In: Proc. ICWSM. (2008) 60–67
19. Ikeda, D., Fujiki, T., Okumura, M.: Automatically linking news articles to blog entries. In: Proc. 2006 AAAI Spring Symp. Computational Approaches to Analyzing Weblogs. (2006) 78–82
20. Yoshioka, M.: IR Interface for Contrasting Multiple News Sites. In: Prof. 4th AIRS. (2008) 516–521
21. Bautin, M., Vijayarenu, L., Skiena, S.: International Sentiment Analysis for News and Blogs. In: Proc. ICWSM. (2008) 19–26