IEICE TRANS. INF. & SYST., VOL.E99–D, NO.4 APRIL 2016

PAPER Special Section on Data Engineering and Information Management

# Named Entity Oriented Difference Analysis of News Articles and Its Application

**Keisuke KIRITOSHI**[†a)], *Nonmember and* **Qiang MA**[†b)], *Member*

**SUMMARY** To support the efficient gathering of diverse information about a news event, we focus on descriptions of named entities (persons, organizations, locations) in news articles. We extend the stakeholder mining proposed by Ogawa et al. and extract descriptions of named entities in articles. We propose three measures (difference in opinion, difference in details, and difference in factor coverage) to rank news articles on the basis of analyzing differences in descriptions of named entities. On the basis of these three measurements, we develop a news app on mobile devices to help users to acquire diverse reports for improving their understanding of the news. For the current article a user is reading, the proposed news app will rank and provide its related articles from different perspectives by the three ranking measurements. One of the notable features of our system is to consider the access history to provide the related news articles. In other words, we propose a context-aware re-ranking method for enhancing the diversity of news reports presented to users. We evaluate our three measurements and the re-ranking method with a crowdsourcing experiment and a user study, respectively.

***key words:*** *news app, named entity, difference analysis, context aware re-ranking, crowdsourcing experiment*

## 1. Introduction

In some sense, news is never free from bias due to the intentions of editors and sponsors [1]. If users read only one article about a news event, they may be left with a biased impression of it. To understand news events, reading a diverse range of news articles is important and useful. However, efficiently seeking such diverse viewpoints from a mass of news articles is very difficult.

Many methods and systems have been proposed to analyze differences between news articles [2]–[13]. For instance, NewsCube [8] is a system that classifies news articles by some aspects of a news event to help users understand it. However, this system does not show differences among news articles, so users cannot know which articles they should read to understand the news from various viewpoints efficiently. Ogawa et al. propose a method that compares news articles by analyzing the descriptions of stakeholders, such as persons, organizations, and locations [11]. However they do not provide a scoring mechanism to support searching for or ranking news articles.

In this paper, to help users to read diverse news articles efficiently, we propose a novel named entity oriented difference analysis method. On the basis of it, we develop a news app named NewsSalad on mobile devices to help users to seek diverse information on a news event.

First, by a user survey, we reveal that showing the kinds of differences between news articles is important to help users to understand news events. We propose three measures to rank news articles on the basis of the survey results: *DC (Difference in Factor Coverage)*, *DO (Difference in Opinion)*, and *DD (Difference in Details)*. We calculate them by analyzing the descriptions of named entities. Descriptions of named entities (persons, organization, and locations) are extracted from news articles by extending the stakeholder mining method [11]. We introduce a notion of core entities to denote the most important entities in a reported news event.

The three measures are calculated by analyzing named entities and their descriptions. We summarized them as follows.

- *DC (Difference in Factor Coverage)* is a measurement of how many different things are described in two news articles reporting the same event. *DC* is calculated by differences in named entities between two articles and the number of mentioned core entities.
- *DO (Difference in Opinion)* is a measurement of differences in subjective descriptions in two news articles reporting the same event. *DO* is calculated by differences in description polarities of named entities between two articles and the number of named entities mentioned in both articles.
- *DD (Difference in Details)* is a measurement of differences in details of two articles reporting the same event. We extract topics regarding named entities by using Latent Dirichlet Allocation (LDA) [17] and calculate *DD* by comparing topic coverage and text lengths between news articles.

By utilizing these measures, we develop a news system that effectively provides different reports on the same events to support users' understanding of news on smart mobile devices. If users are interested in a news article, by clicking the links dynamically generated by the proposed system on the basis of analyzing the currently read news article, users can access diverse reports providing different information. To enhance the effect of the system's diversity-seeking, we introduce a context-aware re-ranking method on the basis of user histories.

Manuscript received July 1, 2015.
Manuscript revised November 9, 2015.
Manuscript publicized January 14, 2016.

[†]The authors are with the Graduate School of Informatics, Kyoto University, Kyoto-shi, 606–8501 Japan.
a) E-mail: kiritoshi@db.soc.i.kyoto-u.ac.jp
b) E-mail: qiang@i.kyoto-u.ac.jp

DOI: 10.1587/transinf.2015DAP0003

The major contributions of this paper can be summarized as follows:

- We have carried out a user survey to reveal the important criteria for helping users understand news articles (see also Sect. 3). On the basis of the user survey, we propose three entity-oriented measures to quantify differences between related news articles (Sect. 4.3). We also propose a named-entity mining method (Sects. 4.1 and 4.2) that extends stakeholder mining proposed by Ogawa et al. [11] to enable the computation of these three measures.
- We have performed a crowdsourcing experiment to compare and validate the ranking methods on the basis of the three measures (Sect. 4.4).
- We propose a diversity-seeking news app (Sect. 5). One of the notable features of our system is to thoroughly consider a user's news reading history to provide related reports for further reading. In other words, we propose a context-aware re-ranking method for enhancing the diversity of news reports provided to users (Sect. 5.3). We carried out simulation experiments to **evaluate** our context-aware re-ranking method (Sect. 5.4).

## 2. Related Work

Google News[†] provides news articles labeled as "Opinion", "In Depth", "From Japan", etc. to support users' understanding of news events. These labels are assigned on the basis of the meta-data of articles, such as categories specified by authors, countries of news bureaus, and the number of words in the article. Currently, this kind of service does not show the semantics of news articles, and a user cannot effectively find reports different from the articles s/he has read. On the other hand, mobile news apps, such as SmartNews[††], only provide one article per topic (event), so they cannot make a user aware of different opinions on the same event.

NewsStand [2] is a system to display news stories in map interface. The system clusters crawled news articles by news stories and analyze where the stories happened by entity feature in articles and geographic dictionary. Every time users zoom up map, the system update additional news stories in real time.

To help users to better understand news articles, news-browsing systems have been proposed that visualize and highlight the differences between news articles. The Comparative Web Browser (CWB) [3] searches for news articles that include descriptions similar to those in the article being read by the user. This enables the user to read news articles while comparing them with articles. The Bilingual Comparative Web Browser (B-CWB) [4] extends CWB to compare news articles in different languages.

Ma et al. propose a complementary information retrieval mechanism to support users obtaining supplementary information [5], [6]. They represent and structure news articles hierarchically and then generate structured queries to search for additional information for a given news article.

Opinion mining is also important to help users' understand news event. Systems are proposed to help users to understand opinion of things, person, and so on. Nikolaos et al. propose a system for opinion retrieval and mining on the web, including news articles [7]. They classify gathered web pages into three classes (articles, comment, and multiple) to analyze sentiments of subject expressions as opinions on the given query. The system shows the up-to-date evaluation of the given query in the three classes.

News systems used to help users to understand news events have also been proposed. NewsCube [8] presents various aspects of a news event by using an aspect viewer to facilitate understanding of the news. TVBanc [9] compares news articles on the basis of a notion of topic structure. It gathers related news from various media and extracts pairs of topics and viewpoints to reveal the diversity and bias of news reports on a certain news event.

Balahur and Steinberger redefine opinion mining in news articles on the basis of analysis of news articles from the viewpoints of authors' intentions and readers' backgrounds. They point out that analyzing the descriptions on named entities plays an important role in comparing news articles [15]. Ishida et al. [13] propose a system that reveals differences between news agencies to enhance users' news understanding. They analyze the subject-verb-object (SVO) triples in the descriptions of entities and extract characteristic descriptions of each news agency.

LocalSavvy [10] finds and aggregates local news articles published by official and unofficial news sources associated with the stakeholders. They extract opinions of local social groups from the retrieved results and highlight them in the news web pages.

Ogawa et al. [11] study on analysis and visualization of differences between news articles by focusing on named entities and propose a stakeholder mining mechanism. They extract stakeholders mentioned in news articles. They then present a graph constructed on the basis of description polarity and interests of stakeholders. Such a relation graph helps users to analyze the differences between two news articles. Ogawa et al. target news articles, while Xu et al. study stakeholder mining in multimedia news, especially video news with closed captions [12].

In contrast, we are studying how to search and rank related news articles by estimating the differences between articles. Our method pays more attention to how to rank and provide users diverse reports different from the ones users have already read. In addition, to the best of our knowledge, our system is the first to attempt to effectively provide diverse news reports on smart mobile devices.

[†]https://news.google.co.jp
[††]https://www.smartnews.com/

## 3. User Survey

### 3.1 Summary

We carried out a user survey to investigate which kinds of difference among news articles are important for supporting users' news understanding. The subjects were ten students from Kyoto University. We selected five controversial news events for this survey. For each topic (event), we used one article in Japanese as the original, meaning it was the one first read by a user, and then five Japanese and five English-language articles as related ones. We asked the ten subjects to rank these related articles in accordance with their usefulness and their differences from the original article. We also asked the subjects to describe their criteria used in this ranking. The events for our user survey are shown in Table 1.

### 3.2 Analysis

Table 2 shows the subjects' top three criteria for each of their rankings. We summarized these differences into four categories: *relatedness*, *viewpoint*, *polarity*, and *detailedness*. We hypothesize these four kinds of difference are important for supporting users' news understanding.

Concerning how users read news articles, Balahur and Steinberger point out that analyzing the descriptions on entities, especially named entities, plays an important role in news comparison [15]. Consequently, we try to evaluate these four factors by analyzing the descriptions of named entities.

(1) *Relatedness* is a criterion for estimating whether the news articles are reporting on same news event and the same entities. Intuitively, there are two kinds of *relatedness*.

a) *Relatedness* at the event level: We can evaluate this kind of *relatedness* by comparing the named entities mentioned in the event with those in a news article. For example, if the article is strongly related to Event 1 in Table 1 and has named entities such as "Obama", "America" and "NSA", these entities may be strongly related to this event.

b) *Relatedness* at the level of two articles: We can evaluate this kind of *relatedness* by comparing the named entities mentioned in two articles. For example, article *a* and *b* report named entities "Obama", "NSA", "Bush", "Republican Party" and "Obama", "NSA", "America", "Bush", "Republican Party", and "FILA", respectively. In this example, articles *a* and *b* report about criticism towards Bush concerning NSA surveillance. Article *a* is therefore strongly related to article *b*.

**Table 1** News events for user survey

| | |
|---|---|
| Event1 | The Obama administration's remarks about NSA surveillance |
| Event2 | Wrestling selected as a candidate to be an Olympic event |
| Event3 | Osaka mayor Hashimoto's remarks about comfort women |
| Event4 | Mayor Hashimoto's news conference on his reflection concerning the above remarks |
| Event5 | Demotion to the minor leagues of a major league player, Munenori Kawasaki |

**Table 2** Major differences for each news event

| | |
|---|---|
| Event1 | Critical, much detailed information, high relatedness, positive opinion |
| Event2 | Positive opinion, much detailed information, polarity of descriptions, viewpoints |
| Event3 | Viewpoints |
| Event4 | Positive opinion |
| Event5 | High relatedness, viewpoints |

*Relatedness* is seemingly unimportant for obtaining diverse information. However, in observation of subjects' descriptions, we find that subjects paid attention to this category because they no longer needed barely related articles. For example, if a user reads articles about "The Obama administration's remarks about NSA surveillance", the user does not need articles mainly describing "Obama administration's remarks about the countermeasure of Ebola". Therefore, we fuse *relatedness* with other measures to ensure relevance of news events and other articles.

(2) We found that similarities of *viewpoint* and *detailedness* were highly ranked in users' questionnaires. In some cases, information from more viewpoints means more detailed information. For an example of differences in viewpoints, when we compare articles regarding Event 2 in Table 1, named entities in article *a* include "Yoshida Saori", "IOC", "FILA", and "Saint Petersburg". On the other hand, named entities in article *b* include "IOC", "FILA", "MLB", "IBAF", and "Hideki Matsui". Article *b* reports this event from a different viewpoint more focused on baseball. Therefore, we can analyze viewpoint differences between articles from the different coverage of named entities.

(3) Regarding the *polarities* of descriptions of named entities, for example, a description could be "There are negative effects on Japan". The phrase "negative effects" modifies the named entity "Japan". In this case, the polarity of descriptions of the named entity "Japan" tilts towards a minus. Therefore, we can judge the polarities of named entities from syntax trees and the positive or negative degrees of words in descriptions of named entities.

(4) We analyze the *detailedness* from the lengths of descriptions of named entities and the number of topics about named entities. For example, reporting on a person's remarks, articles *a* and *b* might quote one sentence and all sentences, respectively. In this case, we consider article *b* to be more detailed than article *a*. Because named entities often appear before and after quotes, we consider quotes about named entities along with the description of each named entity. Article *b* in this example has longer descriptions and more additional topics than *a*.

In short, we represent differences between news articles by using *relatedness*, *viewpoint*, *polarity*, and *detailedness*. The proposed measures of *Difference in factor cover-*

*age*, *Difference in opinion*, and *Difference in details* correspond to *viewpoint*, *polarity*, and *detailedness*, respectively. In the following sections, we propose a named entity mining method and describe how these three measures are calculated to rank news articles.

## 4. Named Entity Mining

We extend the stakeholder mining method [11] to extract descriptions of named entities for ranking news. The named entity mining and ranking methods consist of the following three steps.

1. Extracting named entities and descriptions of named entities
2. Extracting core entities
3. Calculating ranking measures (*DC*, *DO*, *DD*)

### 4.1 Extracting Named Entities and Descriptions of Named Entities

We use a language tool called StanfordCoreNLP [14] to analyze articles. We extract words with a NamedEntityTag for PERSON, ORGANIZATION, and LOCATION as named entities. We extract descriptions of named entities on the basis of a relationship structure constructed by StanfordCoreNLP. StanfordCoreNLP provides the following grammatical relationships between words:

$$type(governor, dependent)$$

*type* is a relationship between two words: governor and dependent. We obtain a tree structure by considering the governor as the parent and the dependent as the child. We use the conversion operations proposed by Ogawa et al. [11] to generate a tree structure suitable for computing the description polarity of named entities. Table 3 shows these conversion operations. Figure 1 shows the tree structure of the following sentence:

*Kerry said that WTO and rapidly growing China talked with Japan which concluded an good alliance with India.*

We consider descriptions on named entities as sets of sub-trees, the root of each being a verb and its descendants containing the target named entities. Suppose $e$ is a named entity in article $a$. $S_e$ is a set of sentences containing $e$. $v$ is a verb that is an ancestor of $e$. For $s_1 \in S_e$, $V_{s_1}(e)$ a sub-tree whose root is $v$. In other words, $V_{s_1}(e)$ is a description of $e$ in $s_1$. A set of descriptions of all sentences included by $S_e$ is defined as follows:

$$V(e) = \{V_{s_i} | i = 1, 2, \cdots, n\} \tag{1}$$

where $V(e)$ means the descriptions of named entity $e$. For example, we can let a sentence regarding Fig. 1 be $s_1$. When the named entity "WTO" is extracted, $v$ is "talked" and the set of descendants $V_{s_1}$ consists of descendants of "talked". Therefore, the description of "WTO' in $s_1$, $V_{s_1}(WTO)$ is expressed as follows:

$$V_{s_1}(WTO) = \{talked, WTO, Japan, China, rapidly, growing\}$$

**Table 3** Conversion operations proposed by Ogawa et al.

| *type* | Operation |
|---|---|
| *conj* | Delete this relationship and change the parent of *governor*to a parent of dependent. If both *governor* and *dependent* are verbs, change every child of *governor* except for *dependent* to children of *dependent*. |
| *appos* | Carry out the same operation as *conj*. |
| *rcmod* | Replace *governor* and *dependent*. |
| *cop* | Carry out the same operation as *rcmod*. |

### 4.2 Extracting Core Entities

Core entities in an event are named entities that have a high frequency of appearance, and we assume them to play important roles in that event. Here, we explain the processes to extract core entities. Let $e$ be a named entity in article $j$, and the appearance frequency of $e$ in $j$, $tf_{e_j}$ is calculated as follows:

$$tf_{e_j} = \frac{|e|}{n_j} \tag{2}$$

where $|e|$ is the frequency of $e$ and $n_j$ is the number of total terms in article $j$. Also, in all the articles related to an event, let $d_e$ be the number of articles including named entity $e$, so the frequency of description $df_e$ about $e$ is calculated as follows:

$$df_e = \frac{d_e}{|D|} \tag{3}$$

where $D$ is a set of the related articles and $|D|$ is its number. Named entities appearing in many articles and having a high average appearance frequency are core entities. We calculate $CoreDegree(e)$ to decide whether $e$ is a core entity:

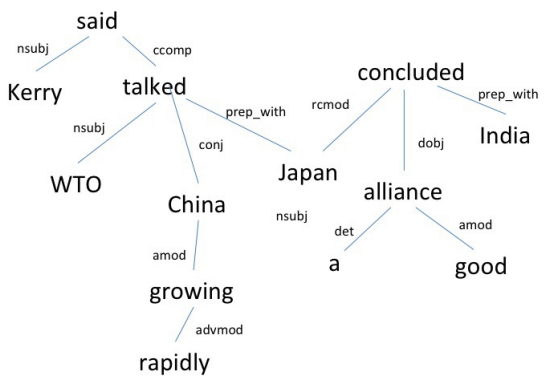


**Fig. 1** Example of tree structure

$$CoreDegree(e) = \frac{\sum_{j \in D}(tf_{e_j} \times df_e)}{|D|} \quad (4)$$

when $CoreDegree(e)$ exceeds the threshold $\theta$, $e$ is a core entity.

### 4.3 Calculating Ranking Measures

In Sect. 4.3, we explain how to calculate the three ranking criteria. Hereinafter, suppose the current article in which users are interested is $o$, the set of related articles is $A$, and a related article is $a \in A$. $E_{core}$ is a set of core entities, and $E_a$ are the named entities of article $a$.

#### 4.3.1 DC (Difference in Factor Coverage)

*DC* is the degree of how many different things are described in two news articles reporting the same event. We estimate *DC* from two aspects: 1) how many different factors are mentioned in these articles, and 2) whether these articles are related to the same event or not. For aspect 1), we can simply compare the entities mentioned in articles. The more different the entities two articles describe, the higher the difference in factor coverage between the two articles is. Aspect 2) corresponds to *relatedness*. However, *relatedness* at the level of two articles is not suitable for *DC* because two articles having different factors and the same factors are intuitively contradictory. Therefore, for aspect 2), we compare the entity mentioned in each article with a core entity set of the given event (of the related news articles).

Let $E_{core}$ be the set of core entities of a certain news event. *DC* between articles $a$ and $o$, $dc(a, o)$ is calculated as follows:

$$dc(a, o) = rel_{eve}(E_{core}, a) \times div_{dif}(a, o) \quad (5)$$

$$div_{dif}(a, o) = |E_a - E_o| \quad (6)$$

$$rel_{eve}(E_{core}, a) = \frac{|E_a \cap E_{core}|}{|E_{core}|} \quad (7)$$

where, $E_a$ and $E_o$ are sets of named entities mentioned in $a$ and $o$, respectively. *Relatedness* at the event level $rel_{eve}(E_{core}, a)$ is the extent of how related article $a$ is to the event.

#### 4.3.2 DO (Difference in Opinion)

*DO* denotes the different extent of opinions between two articles. We compare the description polarities (positive, negative, and neutral) on named entities in articles. If two news articles report the same entities while their polarities are different, we regard these two articles to contain different opinions. We extract descriptions of named entities through our named entity mining method and use an emotional word dictionary to assign a polarity score to each description. In the work described here, we used SentiWordNet [16] as the emotional word dictionary and the descriptive polarities method of Ogawa et al. [11].

The sum score of emotional words in descriptions of a named entity is the final description polarity of the named entity. After that, we compare the polarity of each entity in articles $o$ and $a$. Let $pol_a(e)$ and $pol_o(e)$ be the polarity of named entity $e \in \{E_a \cup E_o\}$, (if $e \notin E_o$, let $pol_o(e) = 0$; if $e \notin E_a$, let $pol_a(e) = 0$). In *DO* of two news articles $a$ and $o$, $do(a, o)$ is calculated as follows.

$$do(a, o) = rel_{mut}(a, o) \times pol(a, o) \quad (8)$$

$$pol(a, o) = w_{do} \times \sum_{e \in \{E_a \cup E_o\}} |pol_a(e) - pol_o(e)| \quad (9)$$

$$rel_{mut}(a, o) = \frac{|E_a \cap E_o|}{|E_a \cup E_o|} \quad (10)$$

where, $pol_a(e)$ and $pol_o(e)$ are polarities of named entity $e$ in articles $a$ and $o$, respectively. $rel_{mut}(a, o)$ is the extent of relatedness between $a$ and $o$. $w_{do}$ is introduced because core entities strongly affect whether a user's is given a positive or negative impression in articles. $w_{do}$ is the weight of core entities and is calculated as follows:

$$w_{do} = \begin{cases} w_{core,do} & (e \in E_{core}) \\ 1 - w_{core,do} & (others) \end{cases} \quad (11)$$

In *DC*, we introduce only *relatedness* at the level of two articles because weight $w_{do}$ already affects *relatedness* at the event level.

#### 4.3.3 DD (Difference in Details)

*DD* denotes the different degrees of details provided by two news articles. We compare named-entity related descriptions in articles to estimate *DD*. After extracting named-entity related descriptions, we apply Latent Dirichlet Allocation) [17] to detected topics from the description of named entities by the Mallet LDA [18]. The difference in detailedness for named entity $e$ between articles $a$ and $o$ is calculated through the following steps.

1. We extract all descriptions $S(e)$ of named entity $e$ from the articles describing the same event as articles $a$ and $o$. Then, we apply LDA to extract the topics $T(e)$ for entity $e$.
2. $S_o(e)$ and $S_a(e)$ are defined as sets of descriptions of named entity $e$ in articles $a$ and $o$, respectively. On the basis of results of LDA, we assign topic probabilities $p_o(s_o, t)$ and $p_a(s_a, t)$ to sentences $s_o \in S_o(e)$ and $s_a \in S_a(e)$, respectively. $(t \in T(e))$.
3. If the topic probability $P_o(s_o, t)$ is more than threshold $\gamma$, $t$ is the topic for descriptions of named entity $e$ in article $o$. We apply this operation to all sentences in $S_o(e)$ and obtain topic sets $T_o(e)$ and $T_a(e)$.
4. Let $at_{ie}, ot_{ie}$ be the numbers of words in topic $t_{ie} \in \{T_a(e) \cup T_o(e)\}$ in article $a$ and $o$, respectively (if $t_{ie} \notin T_o(e)$, let $ot_{ie} = 0$ and if $t_{ie} \notin T_a(e)$, let $at_{ie} = 0$).
   To compare topic coverage and text lengths, the *DD* for article $a$ regarding article $o$ on named entity $e$, $f_{ao}(e)$, is calculated as follows:

$$f_{ao}(e) = \sum_{t_{ie} \in \{T_a(e) \cup T_o(e)\}} (-1)^{\delta} \log\left(\frac{|at_{ie} - ot_{ie}|}{at_{ie} + ot_{ie} + 1} + 1\right) \tag{12}$$

where,

$$\delta = \begin{cases} 0 & (at_{ie} \geq ot_{ie}) \\ 1 & (at_{ie} < ot_{ie}) \end{cases} \tag{13}$$

Finally, we define $DD$ between articles $a$ and $o$, $dd(a, o)$, as the total $DD$ for all the named entities as follows.

$$dd(a, o) = \sum_{e \in \{E_a \cup E_o\}} w_{dd} \times f_{ao}(e) \tag{14}$$

where,

$$w_{dd} = \begin{cases} w_{core,dd} & (e \in E_{core}) \\ 1 - w_{core,dd} & (others) \end{cases} \tag{15}$$

*Relatedness* is not introduced in $DD$. Infrequent named entities do not have enough descriptions of named entities to learn topic probability in LDA. Therefore, calculating only highly-frequent named entities means considering *relatedness*.

### 4.4 Evaluation on the Ranking Measures

We carried out two experiments to evaluate our three ranking measures. One is for core entities, and the other is for the ranking measures.

#### 4.4.1 Experiment on the Extraction of Core Entities

We evaluated the method for extracting core entities. In this experiment, we compared core entities extracted by the proposed method with those selected manually by a user. We varied the threshold $\theta$ (also see Formula (4)) and calculated recall, precision, and the F-measure. For an event, recall $R$, precision $P$, and F-measure $F$ are calculated as follows:

$$R = \frac{A}{C} \tag{16}$$

$$P = \frac{A}{N} \tag{17}$$

$$F = \frac{2A}{N + C} \tag{18}$$

where $A$ is the number of core entities in the obtained named entities, $N$ is the number of obtained named entities and $C$ is the number of core entities in the event. Table 4 shows scores calculated depending on threshold $\theta$.

The highest average F-measure was 0.808, and in this case, $\theta$ was 0.0020. In the remaining experiments, we used $\theta = 0.0020$ to extracted core entities.

#### 4.4.2 Crowdsourcing Evaluation on Ranking

Crowdsourcing makes possible to conduct experiments extremely fast with good results at a low cost [20]. For our ranking measures to be evaluated by various people, we conducted a crowdsourcing experiment on the platform provided by CrowdFlower[†].

**Table 4** Core entity's average recall, precision, and F-measure

| threshold $\theta$ | average recall | average precision | average F-measure |
|---|---|---|---|
| 0.0010 | 1.00 | 0.406 | 0.571 |
| 0.0015 | 1.00 | 0.456 | 0.623 |
| 0.0020 | 0.950 | 0.714 | **0.808** |
| 0.0025 | 0.883 | 0.698 | 0.770 |
| 0.0030 | 0.883 | 0.698 | 0.770 |
| 0.0035 | 0.883 | 0.718 | 0.784 |
| 0.0040 | 0.833 | 0.753 | 0.787 |
| 0.0045 | 0.767 | 0.783 | 0.763 |
| 0.0050 | 0.600 | 0.833 | 0.633 |
| 0.0055 | 0.533 | 0.800 | 0.580 |
| 0.0060 | 0.483 | 0.833 | 0.574 |
| 0.0065 | 0.383 | 0.833 | 0.508 |

In the experiment, we gathered news articles with Google News US edition (Top Stories and Realtime Coverage) from June 17th to 24th, 2014. There are 20 news events (topics), and 14 news articles on average were selected randomly per topic. In each topic, we selected randomly one article as the current article and the others as its related articles.

Because the target news articles are from the US, we limited crowdsourced workers to United States residents who are familiar with US news. For each topic, we invited ten distinguished workers. We set a task's rewards as a common average hourly wage, and the required time was calculated on the basis of workers being able to read 300 words per minute [21].

In our experiment, we asked workers to compare the pair of the current article and one of its related articles. As a result, we have 13 pair-comparisons per news event (topic) on average. In each pair comparison, we asked workers to perform the following tasks.

1. Read the current article and its related article carefully.
2. Score the related article in accordance with the five-grade evaluation system from the three viewpoints: difference in factor coverage, difference in opinion, and difference in details. In the five-grade evaluation system, workers selected one of the grades to estimate the difference: "Very Strong (5)", "Slightly Strong (4)", "Neutral (3)", "Slightly Weak (2)", and "Very Weak (1)".
3. Answer questions related to the news articles. This is used for testing whether a worker has read the articles carefully or not to confirm his/her ability and sincerity. The worker who receives a low-accuracy rate in answering the questions is removed from the results.

We compared ranking by the average scores assigned by crowdsourced workers with that by our proposed ranking methods. Normalized Discounted Cumulative Gain (nDCG) [19] is the evaluation measure. The nDCG score regarding a ranking of the top $p$ is defined as follows:

[†]http://www.crowdflower.com/

**Table 5** nDCG of *DD* ($k = 3$)

| $\gamma$ — $w_{core,dd}$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.8479 | 0.8508 | 0.8513 | 0.8474 | 0.8496 | 0.8510 | 0.8517 | 0.8541 | 0.8576 |
| 0.2 | 0.8534 | 0.8522 | 0.8517 | 0.8495 | 0.8535 | 0.8532 | 0.8548 | 0.8562 | 0.8556 |
| 0.3 | 0.8448 | 0.8464 | 0.8459 | 0.8513 | 0.8520 | 0.8554 | 0.8550 | 0.8529 | 0.8541 |
| 0.4 | 0.8490 | 0.8499 | 0.8525 | 0.8553 | 0.8533 | 0.8536 | 0.8536 | 0.8536 | 0.8556 |
| 0.5 | 0.8493 | 0.8514 | 0.8524 | 0.8513 | 0.8520 | 0.8595 | 0.8602 | 0.8566 | 0.8572 |
| 0.6 | 0.8549 | 0.8556 | 0.8583 | 0.8614 | 0.8616 | 0.8604 | 0.8570 | 0.8603 | 0.8643 |
| 0.7 | 0.8619 | 0.8656 | 0.8603 | 0.8655 | 0.8680 | 0.8645 | 0.8605 | 0.8603 | 0.8630 |
| 0.8 | 0.8604 | 0.8565 | 0.8540 | 0.8592 | 0.8553 | 0.8569 | 0.8535 | 0.8534 | 0.8565 |
| 0.9 | 0.8713 | 0.8698 | **0.8747** | 0.8677 | 0.8688 | 0.8632 | 0.8630 | 0.8591 | 0.8593 |

**Table 6** nDCG of *DD* ($k = 6$)

| $\gamma$ — $w_{core,dd}$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.8852 | 0.8858 | 0.8881 | 0.8855 | 0.8846 | 0.8856 | 0.8830 | 0.8838 | 0.8862 |
| 0.2 | 0.8914 | 0.8914 | 0.8897 | 0.8867 | 0.8872 | 0.8868 | 0.8876 | 0.8881 | 0.8894 |
| 0.3 | 0.8923 | 0.8879 | 0.8851 | 0.8861 | 0.8886 | 0.8875 | 0.8849 | 0.8845 | 0.8870 |
| 0.4 | 0.8896 | 0.8879 | 0.8868 | 0.8861 | 0.8869 | 0.8880 | 0.8898 | 0.8890 | 0.8914 |
| 0.5 | 0.8867 | 0.8871 | 0.8854 | 0.8855 | 0.8881 | 0.8894 | 0.8927 | 0.8927 | 0.8929 |
| 0.6 | 0.8910 | 0.8882 | 0.8883 | 0.8906 | 0.8904 | 0.8907 | 0.8918 | 0.8944 | 0.8948 |
| 0.7 | 0.8920 | 0.8919 | 0.8909 | 0.8948 | 0.8963 | 0.8938 | 0.8938 | 0.8937 | **0.8964** |
| 0.8 | 0.8899 | 0.8907 | 0.8923 | 0.8952 | 0.8903 | 0.8905 | 0.8898 | 0.8902 | 0.8928 |
| 0.9 | 0.8934 | 0.8949 | 0.8953 | 0.8952 | 0.8945 | 0.8905 | 0.8918 | 0.8919 | 0.8927 |

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{19}$$

where $DCG_p$ is a weighted score. Let the $i$-th score be $rel_i$, so $DCG_p$ is expressed as follows:

$$DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i} \tag{20}$$

where $IDCG_p$ is a value that applies the ideal ranking arranged in descending order of scores to expression (19). We assume that the ranking by the average scores assigned by crowdsourced workers is the ideal one.

We calculated nDCG for the top $k$ ranking results ($k = 3, 6$). We varied each parameter $\boldsymbol{w_{core,do}}$ (Formula 11), $\gamma$ (for *DD*; Sect. 4.3) and $\boldsymbol{w_{core,dd}}$ (Formula 15). We calculated the average nDCG of the 20 topics (events). The nDCG results with different parameters of *DD* are shown in Tables 5 and 6.

In the nDCG results of *DD*, when $k = 3$, the highest evaluation value was 0.8747 and $w_{core,dd} = 0.3$ and $\gamma = 0.9$ in this case, and when $k = 6$, the highest evaluation value was 0.8964 $w_{core,dd} = 0.9$ and $\gamma = 0.7$. These values are sufficiently high.

We find that when $w_{core,dd}$ becomes a high value, *DD* will achieve higher values. Usually, users focus more on the main named entities in news events. The change of results with $w_{core,dd}$ confirmed this feature.

Table 7 shows the average nDCG of the ranking by *DC* with different parameters. Here, *DCB* denotes the comparative method, which calculates difference in factor coverage without considering relatedness between the target articles. The score of *DCB* is calculated by Formula (5) as follows.

$$DCB = div_{dif}(a, o) \tag{21}$$

Similarly, for the difference in opinion, we define comparative method *DOB* and calculate its score by Formula (8) as follows.

**Table 7** nDCG of *DC*

| | $k = 3$ | $k = 6$ |
|---|---|---|
| DCB | 0.9109 | 0.9332 |
| DC | 0.9047 | 0.9242 |

**Table 8** nDCG of *DO*($k = 3$)

| $\boldsymbol{w_{core,do}}$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| DOB | 0.8943 | 0.9024 | 0.9119 | **0.9191** | 0.9134 | 0.9095 | 0.9042 | 0.9061 | 0.9105 |
| DO | 0.8867 | 0.8841 | 0.8828 | 0.8862 | 0.8817 | 0.8880 | 0.8881 | **0.8938** | 0.8902 |

**Table 9** nDCG of *DO*($k = 6$)

| $w_{core,do}$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| DOB | 0.9186 | 0.9204 | 0.9251 | **0.9273** | 0.9225 | 0.9229 | 0.9206 | 0.9222 | 0.9236 |
| DO | 0.9086 | 0.9076 | 0.9086 | 0.9088 | 0.9087 | 0.9130 | 0.9142 | **0.9159** | 0.9142 |

**Table 10** Ranking of *DC* and *DO* with noisy articles

| Article | DCB | DC | DOB | DO |
|---|---|---|---|---|
| noisy article1 | 3 | 5 | 2 | 18 |
| noisy article2 | 2 | 1 | 1 | 18 |
| noisy article3 | 1 | 8 | 6 | 18 |

$$DOB = pol(a, o) \tag{22}$$

The results are shown in Tables 8 and 9.

From the experiment as results, it is hard to say which method (considering or not considering *relatedness*) is better for calculating difference in opinion and difference in factor coverage. One considerable reason is that the news articles used in the experiment are strongly related because we gathered them from Google Realtime Coverage. No noisy article needed to be detected from the candidates in our data set. To confirm this assumption, we conducted a small additional experiment. We manually inserted some noisy articles of a news event into our data set to see how they were ranked. The news event was "Hobby Lobby and Obamacare", and the noisy articles mentioned other events related to Obama. The ranking results with noisy articles are shown in Table 10. As we expected, the noisy articles were assigned low ranks by the method in consideration of relatedness. From these results, we can say the proposed method works well and is independent of the accuracy of gathering related news articles.

## 5. Diversity-Seeking Mobile News App

By utilizing the proposed ranking measures, we developed a system to help users to acquire diverse reports of news events on a smart mobile device.

There are several ways to keep up with news: newspapers, TV programs, news websites, and so on. With the spread smart phones and tablets, news applications on smart mobile devices have become widely used. To help us to keep up with news, for example, the best news app of 2014[†] SmartNews[††] allows us to read news articles easily

[†]http://www.idownloadblog.com/2014/12/16/the-best-news-apps-of-2014/

[††]https://www.smartnews.com/

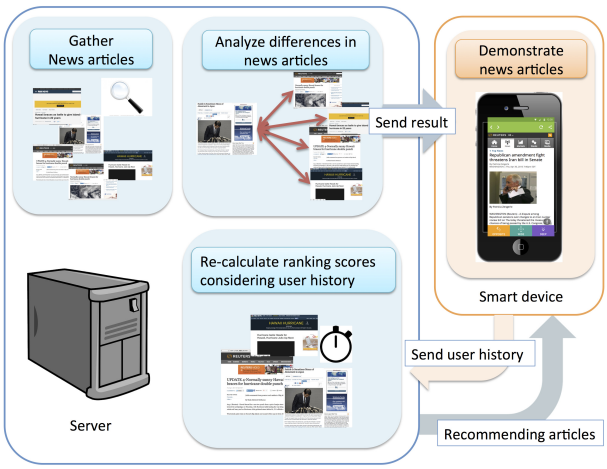


**Fig. 2** Overview of diversity-seeking news system

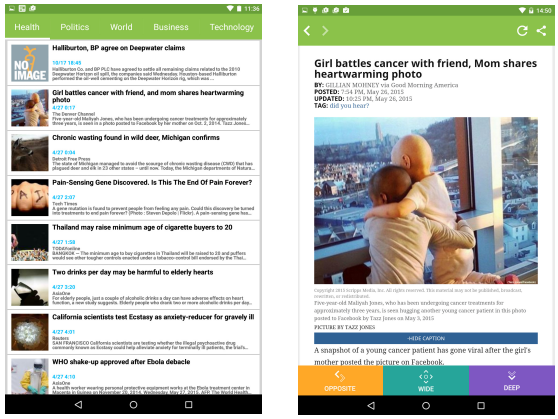


**Fig. 3** View modes of news client

and speedily, even when we are offline. However, a news app usually provides only one article per topic. In addition, mobile search is more difficult than a general web one due to the limitations of the environment and devices. As a result, a user may lose the chance to obtain information from multiple viewpoints to avoid being left with a biased impression.

When a user is interested in a news event, it is easier to search for related news articles on PCs than on smart mobile devices. News web sites, such as Google News, BBC, CNN, and so on, always have search functions to help users find articles they want to read, and each article has links to related articles. On the other hand, when users have little time during commuting or waiting, they may use mobile devices to check news. Due to the time limitations and features of mobile devices, it is not yet easy to find diverse reports on a certain news event even when the users are interested in it.

Currently, the typical news apps, such as SmartNews, specialize in providing news speedily, but they present only one article per news event. This may make mobile device users miss detailed reports and objective opinions and probably leave them with more biased impressions than PC users. In short, although a diversity-seeking news application is necessary for both PC and mobile users, the mobile one is the priority. This is why we focus on a news app on smart mobile devices rather than on PCs. However, as explained below, our system will also be available for PCs when we develop a news client on PCs.

The overview of our system is shown in Fig. 2. The system consists of a news server and news client apps on smart devices. The main functions of the news sever are as follows.

(1) Gathering news articles and analyzing their difference of news articles,
(2) Delivering news articles to smart devices, and
(3) Re-ranking news article considering user histories

The news client app will present news articles to users and record their news reading history for further ranking and recommendation.

### 5.1 News Server: Gathering and Analyzing News Articles

Our diversity-seeking news system targets English news articles. We gather news articles in RSS by using Google News Realtime Coverage[†], which is a Google News' function to present articles related to top news articles. We gather top news articles and their related articles from Realtime Coverage for each topic.

After gathering news articles, the news server analyzes and ranks news articles to top news articles for each topic with the method introduced in Sect. 4.3. Differences between news articles are quantified by *DC*, *DO*, and *DD*. The information about top news articles and the ranking of their related articles will be stored for further processing. Each top news article and its related articles with top ranks of *DC*, *DO*, and *DD* will be delivered to news clients and then presented.

### 5.2 News Client: Presenting News Articles

The news client of our diversity-seeking news system has two view modes. One is the *Top news view* to list that day's top news. The other is *Article details view*, which presents the content of a news article with links to its three top-ranked related articles. Figure 3 illustrates the running examples of these two views.

- *Top news view* shows top news articles gathered from Google News per category. Each news item is presented as a row with its thumbnail and snippets obtained from RSS. A user can click an interesting article to view its details in *article details view*.
- *Article details view* presents the details of a news article. At the bottom part, there are three link buttons named "Opposite" (different opinions), "Wide" (diverse viewpoints), and "Deep" (detailed information),

[†]https://support.google.com/news/answer/2602970

**Table 11** News events for re-ranking evaluation

| | |
|---|---|
| Event 1 | Typhoon of Philippines |
| Event 2 | Investigation of Jerry Sandusky |
| Event 3 | Result of Tiger Woods in British Open |
| Event 4 | Death of Tony Gwynn |
| Event 5 | Hobby Lobby and Obamacare |
| Event 6 | Remarks of Janet L Yellen |
| Event 7 | Tour de France |
| Event 8 | Play of Neymar in World Cup 2014 |
| Event 9 | Murder trial by Oscar Pistorius |
| Event 10 | Rupert Murdoch's bid for Time Warner |

corresponding to the top ranked articles of *DO, DC*, and *DD*, respectively. By clicking these buttons, a user can access diverse reports on the current event.

### 5.3 Context-Aware Re-Ranking

By using the mode of article details view, a user can access top-ranked articles different from the current viewing one. Because the rank is decided by comparison with the current viewing article, the presented top-ranked different articles may be similar to the ones a user has read already. For example, suppose the current article is $o$, and its top-ranked different articles are $a$, $b$, and $c$. First, a user clicks $a$ and then selects $b$. If $a$ has similar contents to $b$, $b$ is no longer useful for the user. This reduces the diversity and novelty of articles provided to users. As an effective solution, to present more diverse information from different articles, we propose a context-aware re-ranking method. That is, the compare target to ranking includes not only the current article but also the articles accessed before.

Suppose that $A$ is the set of related news articles about a certain news event and $H(H \subset A)$ is the set of articles the user has read before. We update the ranking score $d(b, H)$ (represent $dc(\ )$, $do(\ )$, and $dd(\ )$)of article $b \in \{A - H\}$ by merging all articles in $H$. In other words, $H$ is regarded as one virtual article $h_{merged}$ and then we have $d(b, H)$ update $d(b, h_{merged})$. Hereafter, we call this method merge method.

We notice that when $A$ consists of a large number of articles reporting a controversial news topic, the re-ranking may be time-consuming because we need to calculate differences between all articles every time the current article changes. This poor responsiveness will worsen the user experience, especially for a mobile application. Therefore, in reality, not articles but intermediate results, $E_a$, $E_{core}$ (Formula (5)), $pol_a(e)$ (Formula (8)), and $at_{ie}$ (Formula (12)) are merged. The system database saves these intermediate results of all articles to input the responsibility.

### 5.4 Experiment on Context-Aware Re-Ranking Method

We conducted an experiment to evaluate our context-aware re-ranking method by simulating the news reading sequences. We randomly selected ten news events that were used in our crowdsourcing experiment. Table 11 shows these events.

To evaluate the context-aware method, we compared the merge method with the average method [24], in which re-ranking score $d(b, H)$ is simply calculated by the average difference between articles a user has read before and related articles as follows:

$$d(b, H) = \frac{1}{|H|} \sum_{h_i \in H} d(b, h_i) \tag{23}$$

where $h_i \in H$ and $d(b, h_i)$ denote the ranking score of $b$ against $h_i$.

In our preliminary experiment, we found that the average method is superior to the simple way which just removes articles have been read [23], [24]. Thus, in this work, we use the average method to compare with the merge method.

Intuitively, the merge method may be superior to the average method. For example, in the news event about TPP, Article A and B mention Obama's and Abe's remarks, respectively. Article C mentions both the remarks of Obama and Abe. For users who have already read Article A and B, Article C could not provide additional (different) information. The merge method will rank C in a lower place because C could not provide additional information comparing with the virtual article V consisting of A and B. However, the average method may rank C in a higher place because it compares C with A and B separately: C provides additional information to A, and C is different from B.

The simulation based experiment was carried out by Algorithm 1 for each news event. Input and output of the simulation are as follows.

- Input: News Articles of Event $e$, Re-ranking Method $m$, evaluator $u$
- Output: Reading Sequence $S_{e,m}$, Aspect Sequence $X_{e,m}$, Gain Sequence $U_{e,m}$

For a given re-ranking method, the simulation will generate a reading sequence consisting of one initial article $a$ and six related articles. These six articles are picked out by considering the access history and from the aspects of *DC*, *DO*, and *DD*. For each article, the evaluator $u$ will grade it by considering the history and the difference from four aspects[††].

The five-grade evaluation was conducted to estimate how much different information can be obtained from the new article compared with the articles read before from four aspects: factor coverage, opinion, detailed information, and relevance. There aspects are acquired by the user survey that investigated how users felt news articles differed in Sect. 3. The details are as follows:

- Factor coverage: the extent of different things (Person, organization, location, and so on) have been described

[†]In this experiment, we assumed that for a news event, a user will read six related articles

[††]Actually, only the aspect $x_i$ is necessary. To relax the personal difference and help our analysis, in this experiment, we asked each evaluator also grade the other three differences.

**Algorithm 1** The Simulation Based Experiment

1: Randomly select an article $a$ as the current news article. The rest news articles are $a$ 's related article set $A$;
2: Ask the evaluator $u$ read $a$;
3: $H = \{a\}$. $S_{e,m} = \{a\}$, $X_{e,m} = \phi$, $U_{e,m} = \phi$;
4: **for** $i = 1$ to $i = 6^{\dagger}$ **do**
5: Randomly choose one aspect $x_i \in \{DC, DO, DD\}$; {each aspect should not be selected more than twice. }
6: $X_{e,m} = X_{e,m} \cup \{x_i\}$ ;
7: Perform $m$ to create ranks of articles in $A$ from aspect $x_i$;
8: Choose the top-ranked article $t$;
9: $S_{e,m} = S_{e,m} \cup \{t\}$;
10: Ask the evaluator $u$ read article $t$ and then evaluate how much different information $t$ has compared with accessed articles $H$ on a five-grade scale. The evaluators are asked to grade difference from four aspects, which are described later. Such score is the gain score $u_{e,m,t,x_i}$;
11: $U_{e,m} = U_{e,m} \cup \{u_{e,m,t,x_i}\}$;
12: $a \leftarrow t$. $A = A - \{t\}$, $H = H \cup \{t\}$;
13: **if** $A = \phi$ **then**
14: break;
15: **else**
16: continue;
17: **end if**
18: **end for**
19: **return** $S_{e,m}, X_{e,m}, U_{e,m}$;

in the news articles, compared with previously read articles about the same event. This corresponding to aspect *DC*.

- Opinion: the extent of difference of opinions on in the news articles, compared with previously read articles about the same event. This corresponding to aspect *DO*.
- Detailed information: the extent of difference in details of two articles reporting the same event, compared with previously read articles about the same event. This corresponding to aspect *DD*.
- Relevance: the extent of news articles reporting on the same event as previously read articles[†].

Five evaluators scored articles from these four aspects on a five-grade scale, where five stands for "Very strong" and one stands for "Very weak". To conduct the simulation, we randomly assigned each evaluator four events; two for the merge method and the others for the average method. We assigned distinguished events to an evaluator. This is on the basis of the consideration that evaluators may confuse and be hard to grade difference if they read two different article sets of the same event.

To relax the score differences caused by evaluators' personal experiences, we convert score $u(e, m, t, x)$, the difference score of event $e$'s article $t$, from aspect $x$ by evaluator $u$, to linear scaling to unit variance [22]. Function $F(\ )$ converting $u(e, m, t, x_i)$ is defined as follows:

$$F(u(e, m, t, x_i)) = \frac{u(e, m, t, x_i) - \mu_{x,u}}{\rho_{x,u}} \tag{24}$$

[†]We used relevance to judge whether relevance affected users' scores.

**Table 12** Re-ranking evaluation

| Event | E 1 | E 2 | E 3 | E 4 | E 5 | E 6 | E 7 | E 8 | E 9 | E 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| M [1] | −1.409 | −0.862 | 1.931 | 1.437 | 1.757 | 2.825 | −4.116 | −0.550 | −3.075 | 5.302 |
| A [2] | −0.416 | −2.371 | −0.698 | −0.340 | −1.253 | 1.497 | 1.124 | 1.757 | −0.392 | −0.480 |

[1] *M*: merge method.
[2] *A*: average method.

where, $\mu_{x,u}$ and $\rho_{x,u}$ are average and standard deviation of scores assigned by $u$ from aspect $x$. For a given event $e$ and a re-ranking method $m$, the user-gain of evaluator $u$, $g(u, e, m)$ is calculated as follows.

$$g(u, e, m) = \sum_{U_{e,m}} F(u(e, m, t, x)) \tag{25}$$

Larger user-gains denote better user satisfaction of obtaining different information. As shown in Table 12, the merge method achieved better results in six events out of ten events. We could not find out significant differences between these two methods.

By analysis the results, we found that in most cases, the merge method is superior to the average method when re-rank articles by *DC* and *DD*. For example, for Event 2, comparing with the six articles have been read, the article entitled "*Report shows prosecutor pushing Sandusky charges*"[††] has been top-ranked from the aspect of *DC* and returned as the seventh article by both the merge and average methods. However, because it provides almost the same content as the article entitled "*Sandusky report faults police, prosecutors for long delays in charges*"[†††] which is the fourth article in the reading sequence returned by the average method, the evaluator assigned lower score to it. In contrast, there is no such near-duplicate article in the reading sequence returned by the merge method. In addition, different from the previous six articles, the seventh article returned by the merge method provides new information about the Democrat spokesman, the prosecutor's supervisor, and the attorney and so on. Thus, the evaluator assigned higher score to it in the reading sequence returned by proposed method.

On the other hand, for the results of re-ranking from the aspect of *DO*, there appear some cases in which the merge method is inferior to the average method. One of the considerable reasons is that we model the opinion by using the descriptive polarities of entities. The number of entities mentioned in articles will affect the estimation of difference in opinion. For example, in Event 1, the article entitled "*UPDATE 2-Typhoon kills at least 38 in the Philippines, heads for China*"[††††] is top-ranked as a DO article comparing with the first three articles by the merge method. However, it provides almost the same information as the third article entitled "Typhoon takes aim at China after killing 38 in Philippines". Because this event is about typhoon, there are many

[††]http://www.seattletimes.com/nation-world/report-shows-prosecutor-pushing-sandusky-charges/
[†††]http://www.foxsports.com/college-football/story/jerry-sandusky-report-details-police-prosecutor-delays-in-investigation-062314
[††††]http://www.reuters.com/article/2014/07/18/us-philippines-typhoon-idUSKBN0FM01320140718

entities (locations) mentioned in the articles. Then, when the proposed method merges the historic articles, a huge virtual article containing too many different entities is generated. As a result, comparing with such that virtual article, the article with a few entities will be higher ranked as a *DO* article. We also found that most of the entities are local regions of Philippines and China. The evaluators often pay more attention to the major entities (China and Philippines in this example). They do not care the difference in the descriptions of these local regions and aggregate the descriptions at the country level. Hence, grouping the entities on the basis of their relationships (geographical hierarchies, etc.) is a considerable approach to improve the representation of opinions.

Because the merge and average methods are good at different criteria in the context re-ranking, we are planning to conduct further study on choosing the re-ranking method dynamically based on the criteria (the merge method for *DC* and *DD*, and the average method for *DO*, etc.).

The evaluator's impression is also an important factor in the experiment. For example, for Event 3 "Result of Tiger Woods in British Open", one evaluator labeled this event "British Open" and assigned higher relevance scores to all the articles. In contrast, the other evaluator labeled it as "Tiger Woods" and assigned lower relevance scores to some articles, such as "*Rory McIlroy beats Friday jinx, finds 'inner peace' at British Open 2014*"[†]. Event 7 and 10 are the similar cases. Such kind of different impressions may lead different judgments of the gain scores. Further experiments, such as asking the same user to rank the same event twice with different reading sequences are planned.

## 6. Conclusion

In this paper, we proposed three entity-oriented ranking measures on the basis of a user survey to support users obtaining diverse reports on the same news event. For one application of these ranking measures, we developed a diversity-seeking news app on smart mobile devices. A context-aware re-ranking method was also proposed to provide more diverse information. We conducted a crowdsourcing experiment to validate our entity-oriented ranking measures. We compared two context-aware re-ranking methods by a simulation-based analysis and discussed their different effectiveness. Further study on the re-ranking methods are planned.

For an important future work, we need to conduct a user study of our news app. In addition, we plan to carry out experiments to investigate how using our system changes a user's media literacy.

## Acknowledgements

This work is partly supported by KAKENHI (No.25700033) and SCAT Reseach Funding.

[†]http://news.nationalpost.com/sports/golf/rory-mcilroy-beats-friday-jinx-finds-inner-peace-at-british-open-2014

### References

[1] B.H. Baker, T. Graham, and S. Kaminsky, "How to identify, expose & correct liberal media bias," Alexandria, Media Research Center, VA, 1994.
[2] B.E. Teitler, M.D. Lieberman, D. Panozzo, J. Sankaranarayanan, H. Samet, and J. Sperling, "NewsStand: a new view on news," Proc. 16th ACM SIGSPATIAL international conference on Advances in geographic information systems - GIS '08, Article no.18, 2008.
[3] A. Nadamoto and K. Tanaka, "A comparative web browser (CWB) for browsing and comparing web pages," Proc. 12th international conference on World Wide Web - WWW '03, pp.727–735, 2003.
[4] A. Nadamoto, Q. Ma, and K. Tanaka, "B-CWB: Bilingual Comparative Web Browser Based on Content-Synchronization and Viewpoint Retrieval," World Wide Web, vol.8, no.3, pp.347–367, 2005.
[5] Q. Ma and K. Tanaka, "Topic-structure-based complementary information retrieval and its application," ACM Trans. Asian Lang. Inf. Process., vol.4, no.4, pp.475–503, 2005.
[6] Q. Ma, A. Nadamoto, and K. Tanaka, "Complementary information retrieval for cross-media news content," Inf. Syst., vol.31, no.7, pp.659–678, 2006.
[7] N. Pappas, G. Katsimpras, and E. Stamatatos, "Distinguishing the Popularity between Topics: A System for Up-to-Date Opinion Retrieval and Mining in the Web," Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science, vol.7817, pp.197–209, 2013.
[8] S. Park, S. Kang, S. Chung, and J. Song, "NewsCube: delivering multiple aspects of news to mitigate media bias," Proc. SIGCHI conference on Human factors in computing systems - CHI '09, pp.443–452, 2009.
[9] Q. Ma and M. Yoshikawa, "Topic and Viewpoint Extraction for Diversity and Bias Analysis of News Contents," Advances in Data and Web Management, Lecture Notes in Computer Science, vol.5446, pp.150–161, 2009.
[10] J. Liu and L. Birnbaum, "LocalSavvy: aggregating local points of view about news issues," Proc. first international workshop on Location and the web - LOCWEB '08, pp.33–40, 2008.
[11] T. Ogawa, Q. Ma, and M. Yoshikawa, "News Bias Analysis Based on Stakeholder Mining," IEICE Transactions on Information and Systems, vol.94-D, no.3, pp.578–586, 2011.
[12] L. Xu, Q. Ma, and M. Yoshikawa, "A Cross-Media Method of Stakeholder Extraction for News Contents Analysis," Web-Age Information Management, Lecture Notes in Computer Science, vol.6184, pp.232–237, 2010.
[13] S. Ishida, Q. Ma, and M. Yoshikawa, "Analysis of News Agencies' Descriptive Features of People and Organizations," atabase and Expert Systems Applications, Lecture Notes in Computer Science, vol.5690, pp.745–752, 2009.
[14] C.D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit," Proc. 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp.55–60, 2014.
[15] A. Balahur and R. Steinberger, "Rethinking Sentiment Analysis in the News: from Theory to Practice and back," WOMSA, pp.1–12, 2009.
[16] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," LREC, vol.6, pp.417–422, 2006.
[17] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent dirichlet allocation," Journal of machine Learning research, vol.3, pp.993–1022, 2003.
[18] Andrew Kachites McCallum, "MALLET: A Machine Learning for Language Toolkit," http://mallet.cs.umass.edu. 2002.
[19] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," ACM Trans. Inf. Syst., vol.20, no.4, pp.422–446, 2002.
[20] O. Alonso, "Crowdsourcing for Information Retrieval Experimenta-

tion and Evaluation," Multilingual and Multimodal Information Access Evaluation, Lecture Notes in Computer Science, vol.6941, p.2, 2011.
[21] M. Ziefle, "Effects of Display Resolution on Visual Performance," Human Factors, vol.40, no.4, pp.554–568, 1998.
[22] A.K. Jain and R.C. Dubes, Algorithms for clustering data, Prentice-Hall, 1988.
[23] K. Kiritoshi and Q. Ma, "Named Entity Oriented Related News Ranking," Database and Expert Systems Applications, Lecture Notes in Computer Science, vol.8645, pp.82–96, 2014.
[24] K. Kiritoshi and Q. Ma, "A Diversity-Seeking Mobile News App Based on Difference Analysis of News Articles," Database and Expert Systems Applications, Lecture Notes in Computer Science, vol.9262, pp.73–81, 2015.

**Keisuke Kiritoshi** received B.E. degree from Kyoto University in 2014. He is on the Graduate School of Informatics, Kyoto University as a master student. His general research interests are in the area of information retrieval, databases and machine learning. His current interests include Web mining, Web information system and information network.

**Qiang Ma** received his Ph.D. degree from Department of Social Informatics, Graduate School of Informatics, Kyoto University in 2004. He was research fellow (DC2) of JSPS from 2003 to 2004. He joined National Institute of Information and communications Technology as a research fellow in 2004. From 2006 to 2007, he served as an assistant manager at NEC. From October 2007, he joined Kyoto University and has been an associate professor since August 2010. His general research interests are in the area of databases and information retrieval. His current interests include Web information system, Web mining, knowledge discovery, and multimedia Mining and search.