International Journal of Innovative
Computing, Information and Control
Volume **10**, Number **5**, October **2014**

 ISSN 1349-4198
pp. 1827–1841

COMBINATIONAL METHOD OF DIFFERENT WRITTEN STYLES AND ITS APPLICATION TO SIGHTSEEING INFORMATION GENERATION

Naho Ito and Masafumi Hagiwara

Faculty of Science and Technology
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan
{ naho; hagiwara }@soft.ics.keio.ac.jp

Received October 2013; revised March 2014

Abstract. *In this paper, we propose a combinational method of different written styles and apply it to a sightseeing information generation system. There are various kinds of information sources on the Web and their expressions differ largely from formal to casual. For example, literary expression is used in Wikipedia, whereas colloquial expression is used in many SNSs such as twitter. Since much useful and valuable information might be contained in these different sources, research on combining different expression styles to reduce unnaturalness and to implement a user-friendly application is indispensable and very important. In the proposed method, first, N-gram is constructed from Wikipedia as examples of literary style. Then it is used to convert colloquial style to literary style. In the application to the sightseeing information generation system, we use user's location as an input. It retrieves basic information of the user's location from Wikipedia and word-of-mouth information from social media. The colloquial style in the social media data is converted to literary style. Then, selection of sentences is carried out to extract valuable and useful information for sightseeing. As for the output, the proposed system combines information from two types of sources and generates an informative sightseeing information document. We evaluated the method of conversion of written style and the proposed sightseeing information generation system. The experimental results show that the conversion method enabled colloquial style to change to literary style. As for the implementation for sightseeing, the results show that combining sightseeing information from more than one source is considered to be effective.*
Keywords: Natural language processing, Sentence generation, N-gram, Case frames

1. **Introduction.** There are a variety of information sources around us. Search engines accept not only ordinary Web pages, but also news articles, blog articles, images, movies, and real-time information like twitters. In order to provide rich information, there are several studies combining information from more than one source. Matsushita et al. [1] proposed a method to acquire and integrate information from multiple sources in order to improve the reliability of information. Kabutoya et al. [2] proposed a method to recommend an effective question using multiple Q&A sites. Liu et al. [3] proposed a probabilistic model in order to effectively blend different kinds of information obtained by Web search. However, when we want to create a document by combining several Web sources, difference of written style would degrade the understanding of the document. For example, combining information from a newspaper that is written in literary style and that from social data written in colloquial style would generate unnaturalness in the output.

In order to fill the gap between different written styles, there are several researches on paraphrasing in the natural language processing area. Paraphrasing techniques can be

used in two ways [4]. One is to use it as a preprocess step in machine translation and the other is a natural language processing technique. Zhao et al. [5] exploited multiple resources to generate various paraphrases and applied it to sentence compression, sentence simplification, and sentence similarity computation. Wubben et al. [6] used monolingual corpus of news headlines acquired automatically from Google News and a standard Phrase-Based Machine Translation framework. Ikeda et al. [7] proposed an algorithm for reducing the number of unknown words on blog documents by replacing peculiar expressions with formal expressions. First, they automatically registered peculiar expressions and formal expression pattern rules from many formally written documents. Then, the peculiar expressions are changed so that the user can read better. Carroll et al. [8] applied paraphrasing technique to assisting aphasic readers. Robin and McKeown [9] generated compact summary sentences of basketball game using on-line news.

As the application of the proposed method, we focus on the topic of "sightseeing". There are many measures taken in sightseeing area to develop the tourism industry. For example, in Japan, "Tourism Nation Promotion Basic Law" was introduced in 2007, and "Japan Tourism Agency" was inaugurated in 2008 [10]. From these backgrounds, there are many researches using information technology in the tourism industry. Ishino et al. [11] extracted useful tourism information from the blog. Kenteris et al. [12] represented a guide application using mobile devices. Mobile devices present many unique characteristics that make tourist guide attractive and convenient to user. They can provide information about the user's location using GPS, and users can receive information specific to their location [13]. Handheld devices are typically operated by a single user, so it is also able to provide personalized services [14]. By using these merits, Takeuchi and Sugimoto [15] proposed a guide system using user's location information, and Wang et al. [16] created a tourism map from pictures with location information. However, most researches use only single information resource, and as far as we know, there is no research on combining multiple information in a tourism system.

This paper proposes a novel combinational method of different written styles and is applied to a sightseeing information generation system.

The proposed combinational method of different written styles is explained in Section 2. Section 3 explains the application to the sightseeing information generation system. Evaluation experiments are shown in Section 4 and this paper is concluded in Section 5.

2. Conversion of Written Style.

This section explains the proposed method to convert sentence's written style. We treat Japanese language in this paper.

Figure 1 shows flow of the proposed conversion method. The proposed method converts sentence word and its ending. It uses N-gram data constructed from Wikipedia. By using this Wikipedia N-gram, the proposed method judges each word and ending of its colloquial style.

2.1. Construction of Wikipedia N-gram.

The proposed conversion method uses N-gram constructed from Wikipedia. N-gram is a computational model that analyzes words connection. It holds the data of N morphemes frequency. As an official N-gram data, there is Japanese Google N-gram [17], which is constructed from 20 billion sentences.

We constructed the following two types of N-gram data from Wikipedia.

- Wikipedia word N-gram
- Wikipedia ending N-gram

"Wikipedia word N-gram" holds the frequency of continuing N morphemes in each sentence. "Wikipedia ending N-gram" holds the frequency of continuing N morphemes in ending phrase. Table 1 shows the total items of each N-gram.

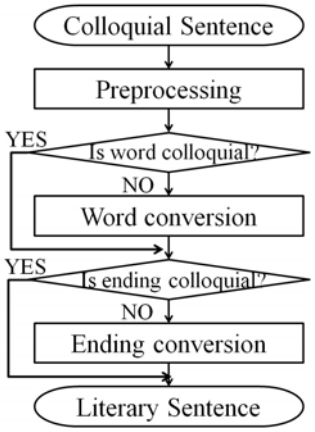

FIGURE 1. Flow of the proposed conversion method

TABLE 1. Number of items in Wikipedia N-gram

Word 1gram	1,288,819
Word 2gram	16,835,822
Word 3gram	71,858,718
Word 4gram	133,644,040
Ending 1gram	156,704
Ending 2gram	970,244
Ending 3gram	2,975,250
Ending 4gram	5,900,056
Ending 5gram	9,106,409
Ending 6gram	12,017,036
Ending 7gram	14,001,410

TABLE 2. Examples of corrupted phrases in Japanese

Type	Corrupted Example	Standard phrase
Addition of prolonged sound	もしもーーし *moshimoooshi* (helloo)	もしもし *moshimoshi* (hello)
Addition of small script	見たぁぁい *mitaaai* (Wannna seee)	見たい *mitai* (want to see)
Replacement with prolonged sound	ねーさん *neesann* (siiister)	お姉さん *oneesann* (sister)
Replacement with small script	おいしぃ *oishii* (yummy)	おいしい *oishii* (delicious)

2.2. **Preprocessing.** Preprocessing step converts corrupted phrases to standard phrases and retrieves morphemes' part of speech information.

2.2.1. *Standardization of corrupted phrase.* Table 2 shows some examples of corrupted patterns appearing in web text [18]. These patterns can be detected using Japanese language morphological analyzer JUMAN 7.0 [19].

Table 3 shows some examples of conversion of corrupted phrase using JUMAN.

2.2.2. *Acquisition of part of speech information.* The proposed method uses morphological analyzer MeCab [20] to retrieve part of speech information. In order to handle new words,

TABLE 3. Some examples of preprocessing result

Before Process	After Process
チョコばっかりでいいわねww *Choko bakkari de ii wa ne* (Lucky it's full of chocolate:))	チョコばかりでいいわね *Choko bakari de ii wa ne* (Lucky it is full of chocolate.)
2Fおもちゃ工房なども密かに人気でーす(^ー)^ノ') *2F omocha koubou nado mo hisoka ni ninnki de-su* Toy factory on the second floor is seeeecretly popular ☺.	2Fおもちゃ工房なども密かに人気です *2F omocha koubou nado mo hisoka ni ninnki de-su* Toy factory on the second floor is secretly popular.

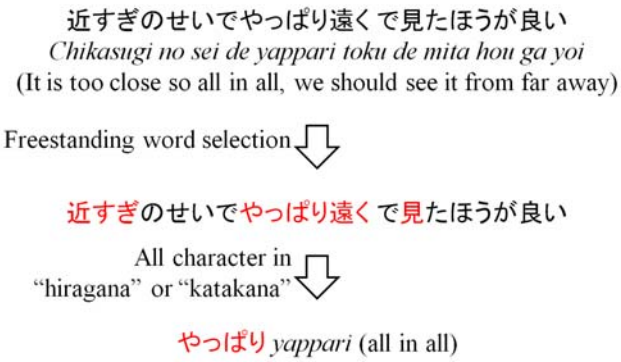

FIGURE 2. An example of word's colloquial decision

we added Wikipedia's title list to MeCab dictionary. We also considered complex words, by handling continuing nouns as one morpheme.

2.3. **Examples of word's colloquial.** This step judges which word to convert to literary style. In order to define the conditions to check the morpheme's informalness, we use Kashino's et al. [21,22] analytic result. The following condition is defined to check whether each morpheme is colloquial or not.

- It is freestanding word.
- All characters are in "*hiragana*" or "*katakana*".

If all conditions are relevant, it is judged to convert to literary style.

Figure 2 shows an example of colloquial decision. In this example, word "*yappari*" corresponds to the conditions above.

2.4. **Word conversion.** The selected colloquial word is converted to literary style. Figure 3 shows the outline of this process.

1. For each colloquial word w_n, search for conversion candidates $s_n^1, \cdots, s_n^m, \cdots, s_n^M$.
2. Calculate score S_n^m (shown in Equation (1)) for each conversion candidate s_n^m.
3. Select one candidate with maximum score, and replace it with w_n.

2.4.1. *Search of conversion candidate.* The literary word to convert the colloquial word is searched using Japanese thesauruses. We use two thesauruses, Japanese WordNet [23] and Japanese Goi-taikei [24]. Synonyms defined in these thesauruses are used as conversion candidates $s_n^1, \cdots, s_n^m, \cdots, s_n^M$. We removed words that match colloquial condition from the candidates.

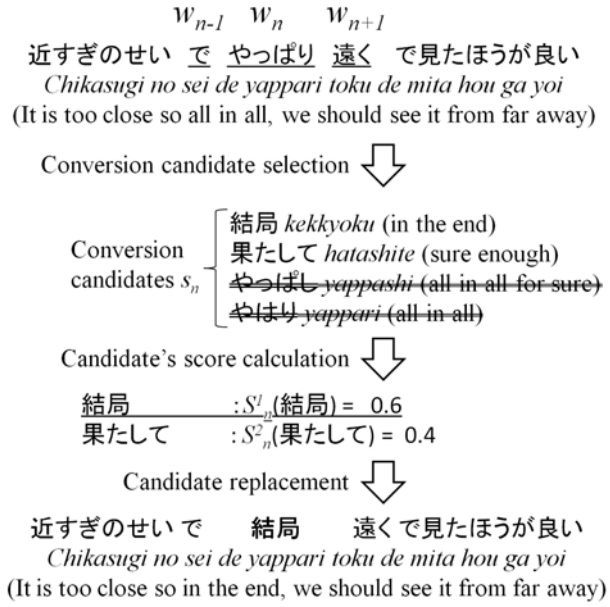

FIGURE 3. Word conversion example

2.4.2. *Calculation of candidate.* Equation (1) is the equation to calculate the score for the candidate word to be converted.

$$S_n^m = \log(F_{w_{n-1}^m, s_n^m}) + \log(F_{s_n^m}) + \log(F_{s_n^m, w_{n+1}}) \tag{1}$$

$F_{a,b}$ is frequency of phrase ab in Wikipedia word N-gram constructed in 2.1. F_a is also the frequency of phrase a in Wikipedia word N-gram.

2.4.3. *Word conversion.* The candidate word with highest score is selected as the word to convert with. However, colloquial conditions are defined in Section 2.3, so we removed words that match those conditions from the candidate. In the example in Figure 3, colloquial phrase "*yappari* (all in all)" is converted to the literal phrase "*kekkyoku* (in the end)".

2.5. **Colloquial decision for ending.** The method uses the following conditions to judge whether or not the ending is colloquial style.

- Ending is in "*desu, masu*" style
- Have sentence-ending particle ("*yo, ne, kashi*, etc.")
- Last morpheme is freestanding word

2.6. **Ending conversion.** Figure 4 shows the overview of ending conversion process. The process is in the following steps.

1. Select freestanding word z closest to the ending.
2. Search conversion candidate phrase p_m from Wikipedia ending N-gram.
3. Calculate score T_m for each conversion candidate p_m and select one ending phrase to replace.

2.6.1. *Freestanding word selection.* Search and select a freestanding word z that part of speech is "noun" or "adjective" or "adverb" or "adnominal adjective". In Figure 4, adjective "*sekaiichi*" is selected as freestanding word z.

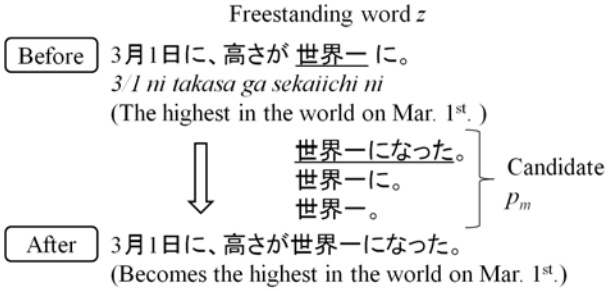

FIGURE 4. An example of ending conversion

TABLE 4. Score for each conversion candidate

Candidate p_m	Morph Num N_m	Frequency G_m
世界一になった Becomes first in the world	4	33
世界一に Is first in the world	2	4
世界一 First in the world	1	49

TABLE 5. Two examples of written style conversion

Before conversion	After conversion
入ってすぐ右側のテラス席がなかなかいい。夜はしまっちゃうけど The terrace seat, right from the entrance, is pretty nice. Though it's closed at night :(	入って近く右側のテラス席が相当いい。夜は閉まっている The terrace seat which is on the right from the entrance is nice. It is closed at night.
やっぱり近くで見るより遠くで見たほうが見やすいかも Well all in all, it's pretty better to check from far. It's tooo close.	結局近くで見るより遠くで見たほうが見やすい Since it is too close, it is better see from far in the end.

2.6.2. *Search of ending conversion candidate.* In order to search the ending conversion candidates, we use Wikipedia ending N-gram constructed in Section 2.1. It searches for phrases that start with w in Wikipedia ending N-gram and selects phrases that do not include any other freestanding word.

In Figure 4, phrases in Table 4 are searched as ending conversion candidates.

2.6.3. *Calculation of ending candidate.* By using Equation (2), score T_m for each conversion candidate p_m is calculated.

$$T_m = G_m \times N_m \tag{2}$$

G_m is frequency of Wikipedia ending N-gram, and N_m is the number of morphemes in the phrase. In the example in Figure 4, "*sekaiiichi-ni*" is converted to "*sekaiichi-ni-natta*".

2.7. **Example of written style conversion.** Table 5 shows some examples of written style conversion.

3. **Sightseeing Information Generation.** Written style conversion method is applied to a sightseeing information generation application. Figure 5 shows the flow of sightseeing information generation system.

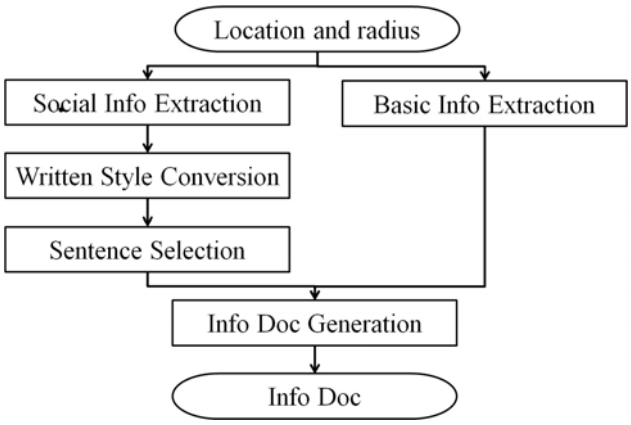

FIGURE 5. Generation flow of sightseeing information

TABLE 6. Examples of extracted word-of-mouth texts

Place's Name	Word-of-mouth Text
Tokyo Skytree's viewing deck	エレベーターを降りると360度のパノラマを楽しめます When you get off the elevator, you can enjoy panoramic view.
Restaurant	食べ物はパニーニ、ドリンクはスムージー系 Food is panino and drink is smoothie.
Tokyo Skytree	やっぱり近くで見るより遠くで見たほうが見やすいかも。 All in all, it is better to check from far. It is too close
Tokyo Skytree	最寄り駅の出口はB3だけど、A3出た時の眺めがオススメ。 The closest exit is B3 but the sight from A3 is great.

The proposed system uses user's location and radius as inputs. It uses Wikipedia to retrieve basic information and social data to retrieve word-of-mouth information. It obtains information around the input location, and as for word-of-mouth information, it converts its written style from colloquial style to literal style. After the conversion, it selects valuable information by scoring each information sentence. Finally, it combines basic information and word-of-mouth information, and generates sightseeing information document.

3.1. **Social info extraction.** The proposed system retrieves word-of-mouth information around the location from social data. It uses foursquare's API [25]. "foursquare" is a social network using location information. More than 25 million users are posting word-of-mouth text based on location. The proposed system retrieves the place's name and its word-of-mouth text. Table 6 shows the retrieved place's name and its texts around "Tokyo-Skytree" (latitude: 35.71005, longitude: 139.8107).

3.2. **Basic info extraction.** Basic information of the location is retrieved from Wikipedia. Figure 6 shows the flow of information extraction.

3.2.1. *Search Wikipedia location database.* Wikipedia holds data of property for each article. Figure 7 is an example of the property. Some property includes data of location, so the database of article's name and its location is constructed. Table 7 shows some examples of Wikipedia article title and location information in the database. The constructed database includes 36,827 sets of location information.

3.2.2. *Search place's name.* The Wikipedia location database constructed in Section 3.2.1 may not include some major location's information. If there was no article around the

FIGURE 6. Extraction flow of basic information

FIGURE 7. Property information in Wikipedia

TABLE 7. Examples of Wikipedia location DB

Article Name	Latitude	Longitude
Rainbow Bridge	35.6363888889	139.763611111
Tokyo Skytree	35.7101388889	139.810833333
Printemps Ginza	35.6737500000	139.765111111

input location, the proposed system uses place's name retrieved in Section 3.1 and searches the database for articles with that name.

3.3. **Written style conversion.** The proposed system uses conversion method of written style explained in Section 2. Word-of-mouth information from social data is converted to literal style like Wikipedia.

3.4. **Sentence selection.** This step selects valuable information from obtained social data sentences. This selection is carried out in the following steps.

1. Tagging of social data sentences.
2. Scoring of each sentence based on its benefit.

3.4.1. *Tagging of social data sentences.* Table 8 shows 10 types of tags we defined in the proposed system.

Among these 10 tags, 7 tags ("Affect, Review, Advantage, Judgment, Event, Obvious, Wish") are tagged by using opinion extraction tool [26]. This tool defines 7 tags as opinion types and they judge each sentence with one opinion type.

As for the other 3 tags, the system uses keyword matching. The keyword listing for each tag was defined in the following process.

Table 9 shows the keywords defined for each tag. If any keyword appears in the sentence, the corresponding tag is applied.

TABLE 8. 10 types of tags defined

Tag	Outline	Example
Emotion	Subjective and emotional opinions	I love Kyoto.
Evaluation	Subjective but not emotional opinions	Kyoto is clean
Merit	Objective opinions about advantages or disadvantages	This card is everyday available
Adopt	Acceptance or refusal of some act	This company accepts the summer time
Event	Good/bad events or experiences	I won the first prize.
Deontic	Duties and proposals	Electric money should be introduced.
Demand	Requirements and demands	I hope this shop can be paid by electric money.
Souvenir	Written about souvenir	Original coffee is sold.
Estimation	Written about estimation	Nice view of the tower can be seen.
Time	Written about the specific time/day	It is illuminated in sky color today.

TABLE 9. Keywords for keyword matching

Souvenir	Estimation	Time
melon	too luxury	today
brandy	bright	spring
spa	novelty	morning
cherry	cool	soon

FIGURE 8. An example of tagging

Figure 8 shows an example of tagging process. In this example, 3 tags "time, merit, estimation" are tagged to the sentence "In a cloudy day, we cannot see over the second observation tower because of cloud, and we can realize how tall it is."

3.4.2. *Scoring based on benefit.* This process scores each sentence by weighting each tag based on its benefit to sightseeing information. In order to judge how beneficial each tag is, we analyzed Japanese Kyoto Sightseeing Blog Corpus [27]. Table 10 shows each tag weight u_t. This is defined by ranking the tags frequency in the sightseeing blog. Equation (3) is the equation to calculate each sentence's score U.

$$U = \sum_t u_t \tag{3}$$

Figure 9 shows an example of scoring.

TABLE 10. Tag's frequency in Kyoto Sightseeing Blog Corpus and its weight

Tag t	Frequency	Benefit weight u_t
Emotion	356	5
Evaluation	2296	7
Merit	606	6
Adopt	106	3
Event	68	2
Deontic	228	4
Demand	37	1
Souvenir	2641	8
Estimation	7962	10
Time	4600	9

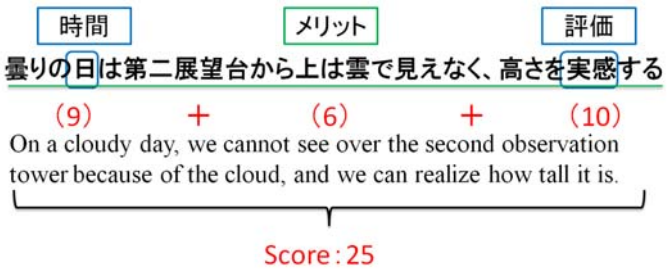

FIGURE 9. An example of score calculation

3.5. **Info doc generation.** This step combines information from social data and Wikipedia, and generates sightseeing information document. The number of sentences to generate is defined by the parameter α, and the percentage of how weigh the user is wishing to know the basic information is represented as the parameter β $(0 < \beta < 1)$.

The number of sentences to use from social data A and Wikipedia B is calculated by Equation (4).

$$\begin{aligned} A &= \alpha \times (1 - \beta) \\ B &= \alpha \times \beta \end{aligned} \tag{4}$$

Sentences in social data to use are selected from the top A sentences in the scoring result. Each sentence is added with the information of the place's name.

Wikipedia sentences to use are selected from the first B sentences in the article.

Figure 10 shows an example of information document generation.

We implemented an application using Web interface. Figure 11 shows the interface of the proposed system.

4. **Evaluation Experiment.** We executed three experiments to evaluate the following points in the proposed system.

- Accuracy of written style conversion
- Accuracy of sentence selection
- Accuracy of sightseeing information system

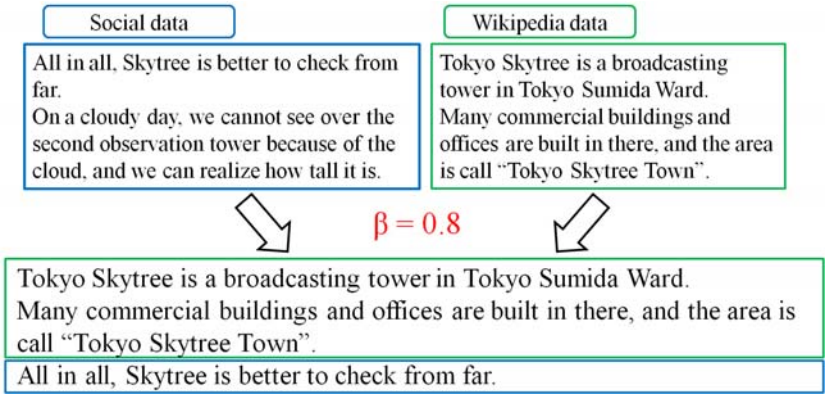

FIGURE 10. An example of information document generation

FIGURE 11. Sightseeing information generation system's web interface

TABLE 11. Articles and location information in category "Tourist spot in Tokyo"

Article Title	latitude	longitude
Tokyo Skytree	35.7101388889	139.810833333
Shinagawa Aquarium	35.5884722222	139.737500000

As for the input to the sightseeing system, we used tourist spot information extracted from Wikipedia. Wikipedia defines categories for each article, so we used articles that are under the category "Tourist spot in Tokyo". Table 11 shows some examples of extracted tourist spots and their location information.

4.1. **Evaluation of written style conversion.**

4.1.1. *Condition.* There were 80 sentences that were extracted from foursquare and those written styles were converted. 20 sentences were selected in random for evaluation. Each subject evaluated 20 sentences before conversion and 20 sentences after conversion. We also used 10 sentences from Wikipedia to measure the written style of Wikipedia. 19 subjects evaluated each sentence in the following points.

TABLE 12. Written style conversion evaluation result

	hardness [1-3]	informal [1-3]	addressing expression [1-3]
Wikipedia	2.74	1.38	1.20
Before	1.55	2.25	2.06
After	2.24	1.68	1.56

TABLE 13. Written style conversion evaluation result

difference in meaning [1-3]
1.79

TABLE 14. Some examples of sentences used in evaluation

	Higher score sentence	Lower score sentence
$(S_1, S_{n/2-1})$	何度行っても楽しい子供のお仕事天国。お菓子工場が人気過ぎてなんかいい行っても入れない。 No matter how many times you visit, it is a wonderful work place for children. Sweet factory is always filled with people.	休日だと小さい子供のいる家庭には少し厳しい。 It is not a nice place for family with small children on weekend.
$(S_{n/2}, S_n)$	HMVでは店員さんの接客が良い。 The clerk's service at HMV (CD store) is nice.	サンマルクカフェではSサイズが存在せず、Mサイズ。 There is no S size. Only M size in Saint Marc Café.

1. hardness [3: hard expression, ..., 1: soft expression]
2. informal [3: informal, ..., 1: formal]
3. addressing expression [3: addresses expression, ..., 1: does not address expression]
4. difference in the meaning [3: different meaning, ..., 1: same meaning]

4.1.2. *Result.* Tables 12 and 13 show the results of the evaluation. By using the conversion method, sentences from social data are changed closer to Wikipedia's written style. Also, since the meaning of the sentence is relatively small, this conversion method is able to convert colloquial written style to literary style without failing the meaning of the sentence.

4.2. **Evaluation of sentence selection.**

4.2.1. *Condition.* 40 articles were randomly selected and used to generate sightseeing information documents. In order to evaluate sentence selection process, we scored social data sentences using the method explained in Section 3.4.2. By ordering all sentences in descending order $S_1, \cdots, S_n$, we selected two datasets $((S_1, S_{n/2-1}), (S_{n/2}, S_n))$ from $S_1, \cdots, S_n$, and randomly showed it to 19 subjects. Each subject selects which sentence is worthy to sightseeing information.

Table 14 shows some examples of sentences used in evaluation.

4.2.2. *Result.* Table 15 shows the accuracy rate. From this result, the sentence selection process is able to select worthy sentences to sightseeing from social data.

TABLE 15. Accuracy rate of sentence selection

	Accuracy weight to select sentence with higher score
$(S_1, S_{n/2-1})$	71.0%
$(S_{n/2}, S_n)$	62.0%

TABLE 16. Accuracy of proposed sightseeing information system

β	worth as an overview of the spot [1-3]	worth to know before going [1-3]	worth to know when visiting the spot [1-3]
0	1.51	2.48	2.28
0.5	2.58	2.57	2.67
1.0	2.76	1.80	1.86

4.3. **Evaluation of proposed sightseeing information system.**

4.3.1. *Condition.* 10 tourist spots' location information was randomly selected from Wikipedia articles, and used to generate sightseeing information for each tourist spot. Each parameter was set up as follows:

$$\alpha = 10$$
$$\beta = (0, 0.5, 1.0)$$

We asked 19 subjects to evaluate three documents (one for each parameter $\beta = (0, 0.5, 1.0)$) for each tourist spot. The following points are the evaluation items.

1. Worth to understand the overview of the spot [3: Worthy, ..., 1: Not worthy]
2. Worth to know before going to the tourist spot [3: Worthy, ..., 1: Not worthy]
3. Worth to know when visiting the tourist spot [3: Worthy, ..., 1: Not worthy]

4.3.2. *Result.* Table 16 shows the result for each parameter $\beta = (0, 0.5, 1.0)$. $\beta = (0, 1.0)$ are results of documents generated only from social data and Wikipedia. Since they use only one source, at least one evaluation item is scored low. However, result of $\beta = 0.5$ has all evaluation items in high score. This result is the document generated from two types of sources, so we can assume that by combining information from more than one source, we can include more worthy information for sightseeing.

5. **Conclusion.**

This paper proposed a combinational method of different written styles and its application to a sightseeing information generation system. In the combinational method of written style, first, N-gram data is constructed from Wikipedia as examples of literary style. Then it is used to convert colloquial style to literary style. In the application to the sightseeing information generation system, first, basic information about the user's location is retrieved from Wikipedia. Then word-of-mouth information from social media is collected and is converted to literary styles. They are combined and useful and valuable sentences sightseeing are extracted and shown to the user.

Evaluation experiments for the method of conversion of written style and the proposed sightseeing information generation system were carried out. The results show that the proposed conversion method effectively changes colloquial style to literary style. As for the implementation for sightseeing, the results show that combining sightseeing information from more than one source is considered to be effective.

REFERENCES

[1] M. Matsushita, H. Sato and T. Kato, Information integration method for generating adequate linguistic expression, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol.14, no.5, pp.29-39, 2002.

[2] Y. Kabutoya, T. Iwata, H. Shiohara and K. Fujimura, Effective question recommendation based on multiple features for question answering communities, *IPSJ Transaction on Database (TOD)*, vol.3, no.4, pp.34-47, 2010.

[3] N. Liu, J. Yan and Z. Chen, A probabilistic model based approach for blended search, *Proc. of the 18th International World Wide Web Conference*, pp.1075-1076, 2009.

[4] K. Inui and A. Fujita, A survey on paraphrase generation and recognition, *Journal of Natural Language Processing*, vol.11, no.5, pp.151-198, 2004.

[5] S. Zhao, X. Lan, T. Liu and S. Li, Application-driven statistical paraphrase generation, *Proc. of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Stroudsburg, PA, USA, pp.834-842, 2009.

[6] S. Wubben, A. van den Bosch and E. Krahmer, Paraphrase generation as monolingual translation: Data and evaluation, *Proc. of the 6th International Natural Language Generation Conference*, Stroudsburg, PA, USA, pp.203-207, 2010.

[7] K. Ikeda, T. Yanagihara, K. Matsumoto and Y. Takishima, Automatic rule generation approach for morphological analysis of peculiar expressions on blog documents, *IPSJ Transactions on Databases*, vol.3, no.3, pp.68-77, 2010.

[8] J. Carroll, G. Minnen, Y. Canning, S. Devlin and J. Tait, Practical simplification of english newspaper text to assist aphasic readers, *Proc. of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pp.7-10, 1998.

[9] J. Robin and K. McKeown, Empirically designing and evaluating a new revision-based model for summary generation, *Artificial Intelligence*, vol.85, no.1-2, pp.135-179, 1996.

[10] Ministry of Land, Infrastructure, Transport and Tourism, *Tourism Nation Promotion Basic Law*, http://www.mlit.go.jp/kankocho/kankorikkoku/kihonhou.html.

[11] A. Ishino, H. Nanbu and T. Takezawa, Automatic compilation of travel information from automatically identified travel blog entries, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol.22, no.6, pp.667-679, 2010.

[12] M. Kenteris, D. Gavalas and D. Economou, An innovative mobile electronic tourist guide application, *Personal and Ubiquitous Computing*, vol.13, pp.103-118, 2009.

[13] U. Varshney, Issues, requirements and support for location-intensive mobile commerce applications, *International Journal of Mobile Communications*, vol.1, no.3, pp.247-263, 2003.

[14] H. S. Ying and K. S. Ho, The attraction of personalized service for users in mobile commerce: An empirical study, *SIGecom Exch.*, vol.3, no.4, pp.10-18, 2002.

[15] Y. Takeuchi and M. Sugimoto, A user-adaptive city guide system based on location data history, *IEICE Transactions on Information Systems*, vol.90, no.11, pp.2981-2988, 2007.

[16] J. Wang, M. Noda, T. Takahashi, D. Deguchi, I. Ide and H. Murase, Creation of a sight-seeing map with visual classification of photos on the web, *Jounal of Information Processing*, vol.52, no.12, pp.3588-3592, 2011.

[17] T. Kudo and H. Kazawa, *Web Japanese N-gram Version 1*, Gengo Shigen Kyokai, 2007.

[18] R. Sasano and N. Kaji, Robust nlp for real-world data: 2. processing of new words and informal spellings, *Information Processing Society of Japan*, vol.53, no.3, pp.211-216, 2012.

[19] S. Kurohashi and D. Kawahara, *User-Extensible Morphological Analyzer for Japanese Version 7.0*, Department of Intelligence Science and Technology, Kyoto University, 2012.

[20] T. Kudo, K. Yamamoto and Y. Matsumoto, Applying conditional random fields to japanese morphological analysis, *Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp.230-237, 2004.

[21] W. Kashino, S. Tachibana, S. Yasuda, T. Maruyama, M. Okumura, S. Sato, T. Tokunaga, H. Otsuka and S. Sadoshima, Analysis of textual formality and informality: In the case of the book samples in the balanced corpus of contemporary written japanese, *The 1st Workshop on Courpus Japanese Linguistics*, pp.131-138, 2012.

[22] W. Kashino, S. Tachibana, S. Yasuda, R. Iida, T. Maruyama, M. Okumura, S. Sato, T. Tokunaga, H. Otsuka, S. Sadoshima, M. Tsubakomoto and H. Numata, Annotation of writing styles of the book samples in the balanced corpus of contemporary written Japanese, *The 18th Annual Meeting of the Association for Natural Language Processing*, 2012.

[23] National Institute of Information and Communications Technology, *Japanese Wordnet*, http://nlpww.nict.go.jp/wn-ja/index.en.html.
[24] Iwanami Shoten, *Goitaikei – A Japanese Lexicon CD-ROM*, http://www.kecl.ntt.co.jp/icl/lirg/resources/GoiTaikei/index-en.html.
[25] Foursquare, *Foursquare*, http://foursquare.com/.
[26] National Institute of Information and Communications Technology, *Opinion Extraction Tool*, http://alaginrc.nict.go.jp/opinion/index_e.html.
[27] *Kyoto Tour Blog Corpus with Opinion Information*, https://alaginrc.nict.go.jp/images/documents/KyotoBlog_ALAGIN_V1_README.pdf.