Automated Generation of Fill-in-the-Blanks-Type Quizzes Using Wikipedia

Naoshi Sakamoto

***Abstract*—Consider generating quizzes for a knowledge of a category in order to make use of learning. Particularly, if we realized generating quizzes automatically by using Wikipedia, which is the vast treasure trove of knowledge, it might greatly contribute to learners.**

In this study, we propose a system that generates fill-in-the-blanks-type quizzes about a category consisting of the elements of the fixed number. To achieve this, the system calculates the tf-idf values for each word in retrieved documents from Wikipedia, and the degree of importance of sentences, then, outputs sentences in order of the degree of importance. In order to seek the most appropriate function for the degree of importance of a sentence, for some sample documents, we compare the correlation coefficient between the ranking obtained by human cooperators and the ranking generated by each of fifteen candidate functions.

***Index Terms*—correlated coefficient, quiz generation, rank aggregation, tf-idf, Wikipedia.**

I. Introduction

There exist various levels in learning. It is basic and important to organize the knowledge of terms. Actually, organizing the knowledge of terms is usually regarded as a fundamental learning in many areas. Let us consider a learner model. In any areas, when a learner is a beginner, they would require fundamental textbooks such as primers, books of drills and so on. In the past, they used to use textbooks and books of drills written by professionals as such educational materials. Thus, these books are usually expensive. While writing such beginners' textbook may also require some specialist knowledge, professionals do not have to write such books necessarily, since the required knowledge to write beginners' textbooks seems to form only small part of the whole knowledge that professionals acquires. If those who are not professionals could write such books, beginners could reduce the educational costs.

Wikipedia is the vast treasure trove of knowledge. It owes its great of knowledge to many people who contribute many articles. In the author's opinion, the Japanese edition of Wikipedia has a lot of articles for animations and games. Moreover, it also has a good knowledge of medical science and railways. Connecting this with the above, while medical doctors used to write beginners' textbooks in the past, if we achieved to generate books of drills about medical science, learners could reduce the educational costs. On the other hand, in new areas, that are progressive, and terms of which increase, such as network communication systems, it usually happens that primers are merely published, because of rapid progress and lack of professionals. If we could generate books of drills automatically for such areas, this would make the learning easier so that it would be expected that participants for the areas increase.

In this study, we propose a system that generates quizzes for a knowledge about an area automatically with Wikipedia. The system accepts a list of terms that form a category of the area, seeks articles in Wikipedia, then generates fill-in-the-blanks-type quizzes. However, since articles of Wikipedia are written in natural language, moreover, the notion for optimum about fill-in-the-blanks-type quizzes is not clear, it outputs not complete questions, but plural candidate sentences.

This paper is organized as follows: Section II introduces the related works. Section III prepares notions. Section IV proposes our system that generates fill-in-the-blanks-type quizzes. In Section V, we discuss the measure for sentences, do experiment, then decide a suitable measure. Section VI shows examples of candidate sentences for fill-in-the-blanks-type quizzes. Finally, we conclude in Section VII.

II. Related Works

Most of studies for Wikipedia can be divided into the studies of the data structure over articles of Wikipedia, and the studies that utilize Wikipedia as a knowledge base.

Articles of Wikipedia are written in an easy markup language called "wikitext." Thus, one can easily generate links between articles. Moreover, besides articles written in a natural language, Wikipedia has also articles called "categories" that only consist of links to other articles so that these give Wikipedia richer structure. Thus, we can say that studies for the data structure of Wikipedia focus on links.

Yazdani and Popescu-Belis calculate the relationship between texts by using articles and links of Wikipedia [1].

Zesch, Gurevych, and Mühlhäuser consider Wikipedia as a resource of a semantic dictionary, then, compare it with the former dictionaries. They find that Wikipedia could bring effective knowledge for massive natural language processing, if they could make use of Wikipedia as a resource of a semantic dictionary. However, by using present access method, the effect is limited. Thus, they develop a new API [2].

Hachey, Radford, Nothman, Honnibal, and Curran evaluate reimplement three seminal Named Entry Linking systems and present a detailed evaluation of search strategies [3]. Their experiments find that coreference and acronym handling lead to substantial improvement, and search strategies account for much of the variation between systems.

Manuscript received March 5, 2017; revised May 12, 2017.

N. Sakamoto is with the Department of Information and Communication Engineering, Tokyo Denki University, Tokyo, Japan (e-mail: sakamoto@c.dendai.ac.jp).

DOI: 10.7763/IJCTE.2017.V9.1168

Nevertheless, our system accepts a list of elements of a category not automatic, but by hand. The reason is as follows. Suppose that people study about prefectures in a country. Their purpose may be various. One might study about prefectures comprehensively for an examination. Another might study about them to preparing a tour. That is, the granularity of the study depends on the learner's purpose. Thus, we should be able to control the quality of quizzes manually by creating a list of elements of the category by hand.

On the other hand, the studies that make use of Wikipedia as a knowledge base are also proceeded well. Our study is also contained these straightly.

Hu, Wang, Lochovsky, Sun and Chen make use of the link structure of Wikipedia as auxiliary information to group queries for the Internet [4].

Ponzetto and Strube propose an API that provides the relationship between meanings of words in Wikipedia [5].

Hu, Zhang, Lu, Park and Zhou develop a method that makes use of Wikipedia as the external knowledge base to group documents [6]. This generates a concept feature vector and a category feature vector for documents, then measures the degree of similarity between them. This also makes use of the concept categories and the number of terms in Wikipedia.

Banerjee, Ramanathan and Gupta propose a method of improving the accuracy of clustering short text items using Wikipedia as an additional knowledge source [7]. Their clustering algorithm makes use of the term frequency vector augmented with selected Wikipedia concepts.

Pereira, Botvinick and Detre show that a corpus of a few thousand Wikipedia articles about concrete or visualizable concepts can be used to produce a low-dimensional semantic feature representation of those concepts. The purpose of such a representation is to serve as a model of the mental context of a subject during functional magnetic resonance imaging (fMRI) experiments [8]. They compile words in classical lists corresponding to concepts that were deemed concrete or imageable, and related articles which are linked to from these article titles in Wikipedia by hand. Finally, they generate the text corpus by using software tools.

Mihajlović, Blok, Hiemstra, and Apers propose a logical algebra, named score region algebra (SRA), that enables transparent specification of information retrieval (IR) models for XML databases [9]. In order to test the transparency of their approach, they do experiments for four retrieval models, language model, the Okapi model, the tf-idf model, and the Garden Point XML model.

These studies make use of Wikipedia to calculate the score of the other external text at retrieving information in spite of difference methods.

III. Preliminaries

This study requires two kind of preliminary knowledge, about the data structure of Wikipedia and how to calculate the score of a word.

A. Data Structure of Wikipedia

In order to retrieve documents from Wikipedia, we make use of the API of MediaWiki [10]. MediaWiki opens the API for processing by programs to public (Fig. 1). Once a program asks a query via the API, it can obtain the answer in XML. While the meta data such as the title and the revised date can be obtained in elements or attributes of XML, the content of an article is obtained as strings in wikitext.

- Example of A Query URL

```
https://en.wikipedia.org/w/api.php?format=xml&action=query&prop=revisions&rvprop=content&titles=example1|example2
```

- Example of Retrieved XML Data

```
<?xml version="1.0" encoding="Shift_JIS" standalone="no"?>
<api batchcomplete="">
<query>
<pages>
<page _idx="9999991" ns="0" pageid="9999991" title="example1">
<revisions>
<rev contentformat="text/x-wiki" contentmodel="wikitext" xml:space="preserve">
Wiki text
...
</rev>
</revisions>
</page>
<page_idx="9999992" ns="0" pageid="9999992" title="example2">
<revisions>
<rev contentformat="text/x-wiki" contentmodel="wikitext" xml:space="preserve">
Wiki text
...
</rev>
</revisions>
</page>
</pages>
</query>
</api>
```

Fig. 1. A Behavior of the API of mediawiki.

Wikitext is a markup language which adopts many kinds of symbols. A normal sentence is recognized a sentence as it is. A single new line is ignored. On the other hand, an empty line formed by two new lines is recognized as a boundary between paragraphs. Moreover, as an important markup, "[[item name]]", an item name fenced with double square brackets, denotes a link to the other article described about the item.

By making use of these grammatical constructions, our system extracts sentences from wikitext strings, and picks up words from only the item names which are fenced with double square brackets to span a link to the article of the item.

B. The Score of Words

As mentioned before, on the basis of strings of wikitext retrieved from Wikipedia, our system calculates the degree of importance for each sentence, then outputs sentences in order of the degree. To calculate the degree of importance for sentences, it is also required to calculate the score of words in advance.

Our system picks up words only which are fenced with double square brackets. Note that there might exist the same word as picked up words that are not fenced with double square brackets. Thus, in order to count the number of appearances for each word, the strings of wikitext must be sought again after retrieving words that are fenced with double square brackets. Note that, for simplicity, our system analyzes strings by applying only regular expressions so that it cannot exclude duplicate counting for sub strings.

Let W be a set of words, and D be a set of documents. Let $wc(w, d)$ denote the number of appearances of $w \in W$ in $d \in D$. Then, tf-idf, which is known as the degree of importance of a word, is defined as follows:

$$tf(w, d) = \frac{wc(w,d)}{\sum_{x \in W} wc(x,d)}, \quad (1)$$

$$idf(w) = \log \frac{|D|}{|\{d \in D \mid wc(w,d) > 0\}|}, \quad (2)$$

$$tfidf(w, d) = tf(w, d) idf(w). \quad (3)$$

IV. Quiz Generation

A. Consideration for Quizzes

Our system basically accepts a category that consists of the elements of the fixed number. A famous example of such a category in Japan is about a knowledge of prefectures of Japan, which is carried out in the fourth grade of elementary schools. Moreover, the other examples are the followings:

1) OSI network model,
2) Four organic nutrients,
3) Trias politica,
4) Köppen climate classification, and
5) Normal forms of database.

On the other hand, with respect to a category which consists of many elements, our system can be available to only main part of elements for learning the category roughly, when the number of elements that form the category is so finite that we can enumerate elements. Examples of such categories are the followings:

1) Chemical elements, and
2) Taxonomic rank in biological classification.

On the other hand, our system can simply be applied to neither categories that consist of countless elements nor categories elements of which we can hardly enumerate. Examples of such categories are the followings:

1) Words for a language learning,
2) Historical people, and
3) Algorithms.

B. Summary of the Algorithm

Now, we explain the algorithm of the proposed system.

The proposed system accepts a list that contains elements that form a category, and the name of an element of them as the inputs. For example, in order to learn about counties of Ireland, then input the list of counties of Ireland, and the name of an element e.g. "County Limerick." Note that the name of every element must be registered as the name of an article in Wikipedia.

The proposed system requires a function that yields the degree of importance of a sentence. We call such a function a measure. Since the degree of importance depends on the human sense, the exact value can be calculated neither mathematically nor statistically. We consider measures in Section V concretely.

The proposed system generally works as follows:

1) Suppose that a measure is provided in advance;
2) For each element in the given list, the system retrieves an XML document from Wikipedia;
3) It analyzes the XML documents, then generates a ContentMap object that consists of the name and a Content object for each element, where each of Content objects contains a wikitext;
4) It extracts words by analyzing wikitexts in the ContentMap object;
5) It calculates the value of tf-idf for each words;
6) It generates a SentenceList object, which consists of Sentence objects corresponding to each sentence in the element designated by the input name;
7) It calculates the value yielded by the measure for each sentence, then;
8) It outputs the sentences important words of which are marked in order of the importance.

Recall that the degree of importance depends on the human sense. Thus, we cannot produce the system that outputs the complete quizzes deterministically. This yields that in order to complete quizzes, any artificial checks are required. Then, it is better that the system outputs several candidates of quizzes, rather than it outputs quizzes simply.

The system generates a Sentence object for each sentence in a document, then gathers them into a list. In this process, sentences are extracted from wikitext where they are normal sentences without grammatically special markups, moreover, end with a terminator of a natural language, such as period. Next, it extracts words, calculates the value of tf-idf of each word, then, calculates the value yielded by the provided measure.

V. The Degree of Importance of Sentences and Experiences

A. The Degree of Importance of Sentences

Fill-in-the-blanks-type quizzes help learners by letting them aware important words to fill the blanks of the sentence described about important matters. Thus, the system is required to extract not only important words, but also important sentences. However, since the degree of importance depends on human sense, it cannot be achieved by simple statistical calculation. Thus, we consider plural candidate measures, do a sensory test, then select a suitable measure.

Let us consider how to decide a suitable measure. If a sentence were important, it might have important words, and high amount of information. However, according to manners of composition, a sentence should be concise, and a single matter should be described in a single sentence or a single paragraph. Thus, a sentence with a lot of important words might be rambling. Thus, let us consider choices whether the measure counts the number of important words or not.

On the other hand, the system does not count all words. That is, the system counts only words where they are fenced with double square brackets once or more. Thus, there exist many words not to count. Only fenced words are registered.

Now, suppose the degree of importance of a sentence that contains words with small score. If a measure decreased the degree of a sentence when the sentence contained a registered word, it would also happen to decrease the degree when the word would not be registered originally. Thus, measures should not decrease the degree of importance of a sentence whether the sentence contains words with small score. We consider such candidate measures as the maximum function and the summation function. Suppose that a sentence s in a document d contains words $w_1, \dots, w_k$, where the number of words s contains is $wc(w_i, s) > 0$ for each w_i. Then, we consider the three candidate measures as follows:

1) $sw_1(d,s) = \max_{i=1}^{k} tfidf(w_i, d)$,
2) $sw_2(d,s) = \sum_{i=1}^{k} tfidf(w_i, d)$, and
3) $sw_3(d,s) = \sum_{i=1}^{k} tfidf(w_i, d) wc(w_i, s)$.

Next, let us consider the length of a sentence. Let $length(s)$ denote the length of a sentence s. If a word with a score of zero added to a sentence, the degree of importance might not increase at least. Then, we can consider that when the length increases, the degree of importance of a sentence is unchanging or decreases. On the other hand, if a sentence contains an important word and is long, we can also consider that the sentence describes an important matter. That is, a long sentence might increase the degree of importance. Therefore, we consider the following five candidate measures for the length of a sentence:

1) $lw_1(d,s,sw) = sw(d,s)/length(s)$,
2) $lw_2(d,s,sw) = sw(d,s)/\log length(s)$,
3) $lw_3(d,s,sw) = sw(d,s)$,
4) $lw_4(d,s,sw) = sw(d,s) \log length(s)$, and
5) $lw_5(d,s,sw) = sw(d,s) length(s)$.

Moreover, we consider composite functions between *lw* and *sw*. We denote the functions of *sw* as max, *sum*, and *wsum*, and the functions of *lw* as 1/*length*, 1/log*length*, 1, log*length*, and *length*. Then, we denote the fifteen composite functions as max/*length*, *sum*, *wsum*log*length*, and so on as candidate measures. We consider which measure is most suitable in the following sections.

B. Correlation Coefficient

In order to decide which measure is most suitable, we do an experiment for sensory tests. This is executed by the following:

1) Select some elements of a category in Wikipedia;
2) Calculate the value of each candidate measure for all sentences in the document of each selected element, then generate the ranking of sentences;
3) Have the human cooperators select several sentences for fill-in-the-blanks-type quizzes;
4) Generate the suitable ranking from the sentences selected by the cooperators, where the selected sentences occupy the first place tie for each element;
5) Finally, compare the correlation coefficient between the ranking yielded by each measure and the suitable ranking for each element.

Let S be a set of sentences in a document, and n be its size. Let an order function induced by a function f calculating the degree of importance over S denote as $rank_f : S \to [1, \dots, n]$.

On the other hand, let R be a set of sentences selected by the human cooperators, and m be its size. We define the order function $rank_R$ induced by R as follows:

$$rank_R(x) = \begin{cases} 1 & \text{if } x \in R \\ m+1 & \text{o.w.} \end{cases} \quad (4)$$

We calculate the correlation coefficient between these ranking, then find the most suitable measure to the human sense.

For ranking of R and S, let $\bar{r}$ and $\bar{s}$ be the averages, and σ_R and σ_S be the standard deviations, respectively. The following is the definition of Pearson's correlation coefficient ρ.

$$\rho = \frac{\frac{1}{|S|}\sum_{x \in S}(rank_f(x) - \bar{s})(rank_R(x) - \bar{r})}{\sigma_R \sigma_S} \quad (5)$$

For R, we have the following:

$$\bar{r} = \frac{m + (n-m)(m+1)}{n} = \frac{mn - m + n}{n}, \quad (6)$$

$$\sigma_R^2 = \frac{1}{n}\left(\sum_{x \in R}(1-\bar{s})^2 + \sum_{x \in S \setminus R}(m+1-\bar{s})^2\right) = \frac{m^3}{n^2}(n-m). \quad (7)$$

Moreover, if the ranking of S contains no ties, the followings are obtained regardless of a measure:

$$\bar{s} = \frac{1}{|S|}\frac{|S|(|S|+1)}{2} = \frac{n+1}{2}, \quad (8)$$

$$\sigma_S^2 = \frac{1}{|S|}\sum_{i=1}^{|S|}(i-\bar{s})^2 = \frac{n^2-1}{12} \quad (9)$$

There are several evaluation methods of ranking similarity in ordinal ranking problems [11], [12]. Now, we refer Spearman's rank correlation coefficient ρ. Under the assumption that both rankings contain no tie, this is simply yielded by substituting (8) and (9) for Pearson's correlation coefficient ρ (5). On the other hand, assuming that one ranking contains no tie, but another ranking consists of m first ranks and $(n-m)$ $(m+1)$-th ranks, we substitute (6), (7), (8), and (9) for (5), then we obtain (10). This formula requires no transformation of rankings.

$$\rho = \sqrt{3}\frac{m(n+1) - \sum_{x \in R} rank_f(x)}{\sqrt{m(n^2-1)(n-m)}}. \quad (10)$$

Unfortunately, since there exist some measures that cause ties in the ranking, we make use of not (10) but the original definition (5).

C. Experiment for Seeking a Function

We seek the most suitable measure by an experiment.

In order to do the experiment with Japanese cooperators, we select prefectures of Japan written in Japanese language as a category. Since the ordinal Japanese people study these in their fourth year of elementary school, the knowledge of prefectures in Japan is common for Japanese people. In order to select statistically normal elements, we avoid to select both Tokyo Metropolitan and Osaka Prefecture, which are huge and significant. We select the following as normal but internationally famous prefectures:

1) Kanagawa Prefecture
   This locates at the south of Tokyo Metropolitan, and includes the large port city of Yokohama.
2) Mie Prefecture
   In 2016, the city of Shima hosted the G7 summit.
3) Nagano Prefecture
   Nagano was host to the 1998 Winter Olympics.

Each of our cooperators selects five sentences from each document of these prefectures retrieved from Wikipedia, and then we gather them into a list of sentences corresponding to each of these documents. Our cooperators consist of nine Japanese students. Finally, Table I shows the number of sentences and the number of selected sentences by the cooperators for each of prefectures.

TABLE I: THE NUMBER OF SENTENCES FOR PREFECTURES

Prefecture	# of sentences	# of sentences selected by the cooperators
Kanagawa	1411	23
Mie	1473	26
Nagano	1499	20

Then, we convert the list of selected sentences into a ranking. On the other hand, we generate a ranking of the degree of importance of sentences for each of candidate measures. Then, Table II, Table III, and Table IV show the correlation coefficients between the ranking induced by the selected list and the ranking yielded by each candidate measure. According to these tables, there is no perfect function. However, max function is seemed to be better for processing the value of tf-idf. That is, a sentence that contains important words is simply worthy of a quiz.

Next, we focus on the processing of the length. The results are separated into elements where long sentences are better, and elements where short sentences are better. This might be caused by the writers' habit or special properties of the selected elements.

Nevertheless, we observe the outputs for another category to decide the most suitable measure. For "County Limerick" in Ireland, we generate and observe the top-five outputted sentences for each function of max/log*length*, max, and max*length*. They are Table V, Table VI, and Table VII respectively.

At first, we consider max. According to Table VI, there are many sentences the scores of which are the same. On the other hand, we can see that there exist sentences that simply enumerate many words in articles in Wikipedia. These sentences tend to be long and contain words with high score. In this point, Table VII indicates the similar result. On reflection, we think that sentences suitable to fill-in-the-blanks-type quizzes should be concise, and describes the property of a single important word. Thus, according to Table V, this is only a private view, but we employ max/log*length* as a suitable measure.

TABLE II: CORRELATION COEFFICIENTS FOR KANAGAWA PREFECTURE

	$1/length(s)$	$1/\log length(s)$	1	$\log(length(s))$	$length(s)$
$\max_{i=1}^{k} tfidf(w_i,d)$	0.1830	0.1830	0.1737	0.1723	0.1665
$\sum_{i=1}^{k} tfidf(w_i,d)$	0.1743	0.1644	0.1631	0.1610	0.1599
$\sum_{i=1}^{k} tfidf(w_i,d)f(w_i,d)$	0.1756	0.1707	0.1691	0.1676	0.1655

TABLE III: CORRELATION COEFFICIENTS FOR MIE PREFECTURE

	$1/length(s)$	$1/\log length(s)$	1	$\log(length(s))$	$length(s)$
$\max_{i=1}^{k} tfidf(w_i,d)$	0.2295	0.2497	0.2562	0.2681	0.2787
$\sum_{i=1}^{k} tfidf(w_i,d)$	0.2194	0.2460	0.2537	0.2600	0.2695
$\sum_{i=1}^{k} tfidf(w_i,d)f(w_i,d)$	0.2200	0.2465	0.2557	0.2609	0.2689

TABLE IV: CORRELATION COEFFICIENTS FOR NAGANO PREFECTURE

	$1/length(s)$	$1/\log length(s)$	1	$\log(length(s))$	$length(s)$
$\max_{i=1}^{k} tfidf(w_i,d)$	0.2354	0.2420	0.2446	0.2521	0.2516
$\sum_{i=1}^{k} tfidf(w_i,d)$	0.2348	0.2385	0.2381	0.2440	0.2499
$\sum_{i=1}^{k} tfidf(w_i,d)f(w_i,d)$	0.2284	0.2341	0.2342	0.2388	0.2448

TABLE V: SELECTED SENTENCES FOR MAX/LOGLENGTH MEASURE

Score	Sentence
0.00809	It is named after the (1)CITY of (2)LIMERICK.
0.00723	(1)HURLING in particular is strong in east, mid and south (2)LIMERICK.
0.00705	The (3)LIMERICK Chronicle is owned by the (1)LEADer and is primarily a (2)CITY paper.
0.00701	(5)LIMERICK (3)CITY is the (1)COUNTY TOWN and is also (2)IRELAND'S third largest (3)CITY.
0.00701	The song "(1)LIMERICK you're a lady" is traditionally associated with the county.

TABLE VI: SELECTED SENTENCES FOR MAX MEASURE

Score	Sentence
0.0306	A (9)LIMERICK is a type of humorous verse of five lines with an AABBA rhyme scheme: the poem's connection with the (4)CITY is obscure, but the name is (2)GENERALly taken to be a reference to (9)LIMERICK (4)CITY or (7)COUNTY (9)LIMERICK,3334 sometimes particularly to the (3)(6)MAIGUE POETS who were based in (5)CROOM and its environs, and may derive from an earlier form of nonsense verse parlour game that traditionally included a refrain that included "Will or won't you come (up) to (9)LIMERICK? Riverfest is an annual summer festival held in (9)LIMERICK.
0.0306	Sean South (song)—(8)SEAN SOUTH FROM GARRYOWEN is another popular (10)LIMERICK song and tells the account of the death of (10)LIMERICK (2)(3)IRISH REPUBLICAN ARMY (1922-1969)—(9)IRA member Sean South, who was (1)KILLed during an attack on a Royal (6)ULSTER Constabulary barracks in (4)COUNTY (5)FERMANAGH in 1957.
0.0306	the (12)(13)LIMERICK JUNCTION line which is the busiest line, connecting (13)LIMERICK to the (3)(11)CORK (8)RAILWAY (10)STATION—(11)CORK-(1)(2)DUBLIN HEUSTON line and to the (7)(12)(13)LIMERICK JUNCTION (8)RAILWAY (10)STATION—(12)(13)LIMERICK JUNCTION-(6)(9)CLONMEL (8)RAILWAY (10)STATION—(9)CLONMEL-(4)(5)WATERFORD (8)RAILWAY (10)STATION—(5)WATERFORD line.
0.0306	The (4)(6)N69 (1)ROAD (Ireland)—(4)(6)N69, a secondary route travels from (21)LIMERICK (15)CITY along the (9)(10)SHANNON (14)ESTUARY through (7)CLARINA, (8)KILDIMO, (11)ASKEATON (18)FOYNES & (12)(17)GLIN, (19)COUNTY (21)LIMERICK—(17)GLIN and continues towards (2)LISTOWEL in (13)COUNTY (16)KERRY.
0.0306	The (8)M18 (3)(7)MOTORWAY (Ireland)—(5)N/(8)M18 (1)ROAD links the county to (6)ENNIS and (2)GALWAY while the (10)(12)N24 (1)ROAD (Ireland)—(10)(12)N24 continues south eastwards from (14)LIMERICK towards (4)WATERFORD travelling through villages such as (13)PALLASGREEN and (11)OOLA.

TABLE VII: SELECTED SENTENCES FOR MAX LENGTH MEASURE

Score	Sentence
16.3	A (9)LIMERICK is a type of humorous verse of five lines with an AABBA rhyme scheme: the poem's connection with the (4)CITY is obscure, but the name is (2)GENERALly taken to be a reference to (9)LIMERICK (4)CITY or (7)COUNTY (9)LIMERICK,3334 sometimes particularly to the (3)(6)MAIGUE POETS who were based in (5)CROOM and its environs, and may derive from an earlier form of nonsense verse parlour game that traditionally included a refrain that included "Will or won't you come (up) to (9)LIMERICK? Riverfest is an annual summer festival held in (9)LIMERICK.
9.44	Sean South (song)—(8)SEAN SOUTH FROM GARRYOWEN is another popular (10)LIMERICK song and tells the account of the death of (10)LIMERICK (2)(3)IRISH REPUBLICAN ARMY (1922-1969)—(9)IRA member Sean South, who was (1)KILLed during an attack on a Royal (6)ULSTER Constabulary barracks in (4)COUNTY (5)FERMANAGH in 1957.
8.85	the (12)(13)LIMERICK JUNCTION line which is the busiest line, connecting (13)LIMERICK to the (3)(11)CORK (8)RAILWAY (10)STATION—(11)CORK-(1)(2)DUBLIN HEUSTON line and to the (7)(12)(13)LIMERICK JUNCTION (8)RAILWAY (10)STATION—(12)(13)LIMERICK JUNCTION-(6)(9)CLONMEL (8)RAILWAY (10)STATION—(9)CLONMEL-(4)(5)WATERFORD (8)RAILWAY (10)STATION—(5)WATERFORD line.
7.75	The (4)(6)N69 (1)ROAD (Ireland)—(4)(6)N69, a secondary route travels from (21)LIMERICK (15)CITY along the (9)(10)SHANNON (14)ESTUARY through (7)CLARINA, (8)KILDIMO, (11)ASKEATON (18)FOYNES & (12)(17)GLIN, (19)COUNTY (21)LIMERICK—(17)GLIN and continues towards (2)LISTOWEL in (13)COUNTY (16)KERRY.
7.69	The (8)M18 (3)(7)MOTORWAY (Ireland)—(5)N/(8)M18 (1)ROAD links the county to (6)ENNIS and (2)GALWAY while the (10)(12)N24 (1)ROAD (Ireland)—(10)(12)N24 continues south eastwards from (14)LIMERICK towards (4)WATERFORD travelling through villages such as (13)PALLASGREEN and (11)OOLA.

VI. SAMPLE QUIZZES

We show some example candidate sentences for quizzes. These candidate sentences are outputted by the proposed system by applying max/log*length* as the measure. Note that numbers fenced with parentheses indicates the ranking of the score of words in the sentence. A quiz can be made by removing such words.

A. Network Layer of OSI

1) score=0.004287253241947884
   Since many networks are partitioned into subnetworks and connect to other networks for wide-area communications, networks use specialized (1) HOSTS, called gateways or (5)ROUTER (computing)—(4)(5)ROUTERS, to forward (2)PACKETS between networks.
2) score=0.0040612505745062305
   The (8) NETWORK LAYER is responsible for (4) PACKET FORWARDING including (6)ROUTING through intermediate (7)(9)ROUTERS, since it knows the address of neighboring (3)NETWORK NODEs, and it also manages (5)QUALITY OF SERVICE (QoS), and recognizes and forwards local host domain messages to the (1)TRANSPORT LAYER (layer 4).
1) score=0.0037026526252733716
   (1)(2)IPV6 has a better designed solution.
2) score=0.0032586899521613003
   The (1) (2) TCP/ (3) IP MODEL describes the protocols used by the Internet.
3) score=0.003118577653436723
   The (1) TCP/ (3) IP (4) INTERNET LAYER is in fact only a subset of functionality of the (2)NETWORK LAYER.

B. Protein of Nutrition

1) score=0.009082792669625223
   (1)PROTEIN can be found in a wide range of food.
2) score=0.008800964579107166
   This value is known as the "crude (1)PROTEIN" content.

1) score=0.008601116757364325
   (1)WHOLE (3)GRAINs and (2)(4)CEREALS are another source of (5)(6)PROTEINS.
2) score=0.008393344244670977
   (4)(6)PROTEINS are (1)ESSENTIAL (3)(5)NUTRIENTs for the (2)HUMAN BODY.
3) score=0.008361561340782567
   (2)(3)PROTEINS are also used in membranes, such as (1)GLYCOPROTEINs.

VII. Conclusion

In this study, we propose a system generates candidate sentences for fill-in-the-blanks-type quizzes to learn a knowledge of a category by making use of Wikipedia. We select the measure yielding the degree of importance of a sentence by the experiment.

Since, we have not showed how the selected measure is truly appropriate, we would like to consider the validity of measures in the future. On the other hand, since our system is available to an only category consisting of elements of the fixed number, we would also like to improve the system to deal with a category consisting of elements of a large number in the future. Moreover, we would like to develop systems that generate other kind of quizzes, such as true-false problems, multiple-choice tests, and so on.

Acknowledgment

This study expands the joint research with Yusuke Ooishi.

References

[1] M. Yazdani and A. Popescu-Belis, “Computing text semantic relatedness using the contents and links of a hypertext encyclopedia,” *Artif. Intell.*, vol. 194, pp. 176–202, 2013.

[2] T. Zesch, I. Gurevych, and M. Mühlhäuser, “Analyzing and accessing wikipedia as a lexical semantic resource,” *Data Structures for Linguistic Resources and Applications,* 2007.

[3] B. Hachey, W. Radford, J. Nothman, M. Honnibal, and J. R. Curran, “Evaluating entity linking with wikipedia,” *Artif. Intell.*, vol. 194, pp. 130–150, January 2013.

[4] J. Hu, G. Wang, F. Lochovsky, J.-t. Sun, and Z. Chen, “Understanding user’s query intent with wikipedia,” in *Proc. of the 18th International Conference on World Wide Web, ser. WWW ’09*, New York, NY, USA: ACM, 2009, pp. 471–480.

[5] S. P. Ponzetto and M. Strube, “An api for measuring the relatedness of words in wikipedia,” *Companion Volume to the Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics,* pp. 23–30, 2007.

[6] X. Hu, X. Zhang, C. Lu, E. K. Park, and X. Zhou, “Exploiting Wikipedia as external knowledge for document clustering,” in *Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’09.* New York, NY, USA: ACM, 2009, pp. 389–396.

[7] S. Banerjee, K. Ramanathan, and A. Gupta, “Clustering short texts using Wikipedia,” in *Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR ’07.* New York, NY, USA: ACM, 2007, pp. 787–788.

[8] F. Pereira, M. Botvinick, and G. Detre, “Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments,” *Artif. Intell.,* Jan. 2013, vol. 194, pp. 240–252.

[9] V. Mihajlović, H. E. Blok, D. Hiemstra, and P. M. G. Apers, “Score region algebra: Building a transparent XML-R database,” in *Proc. of the 14th ACM international conference on Information and knowledge management, ser. CIKM ’05.* New York, NY, USA: ACM, 2005, pp. 12–19.

[10] Api: Main page, mediawiki. [Online]. Available: https://www.mediawiki.org/wiki/API:Main_page

[11] J. Mazurek, “Evaluation of ranking similarity in ordinal ranking problems,” *Acta academica karviniensia,* vol. 2, pp. 119–128, 2011.

[12] R. Fagin, R. Kumar, and D. Sivakumar, “Comparing top k lists. in *Proc. of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ser. SODA ’03,* 2003, pp. 28–36.

Naoshi Sakamoto was born in Japan in 1964. He is currently a professor at Tokyo Denki University. He graduated from the University of Electro-Communication in 1987 and received his M.S. and Dr.S. degrees from the Tokyo Institute of Technology in 1989 and 2001, respectively.

From 1992 to 1997, he was a research assistant at the Computer Center of Hitotsubashi University. From 1997 to 2001, he was a research assistant at the Tokyo Institute of Technology. From 2001 to 2014, he was an associate professor at Tokyo Denki University and became a full professor in 2014.

He received the 1999 IEICE best paper award and his paper was selected as one of the best papers in the 2015 SNPD. His research interests are distributed algorithms, randomized algorithms, and complexity theory.