Serendipitous Book Explorer Using Personalized Associative Dictionaries

Takumi Oikawa and Yasuyuki Sumi(✉)

Future University Hakodate, Hakodate, Hokkaido 0418655, Japan
t-oikawa@sumilab.org, sumi@acm.org

Abstract. This paper proposes a system called AssocSearcher that utilizes associative dictionaries as assistance for exploratory search for books. Users of the system can extend their horizons to explore books by getting subtopic or related words from the other people's associative dictionaries. The associative dictionaries are created as substitutes for individual thoughts. We generated associative dictionaries of individual persons using Word2vec and GloVe from sentences read or written by the individuals. This allows each associative dictionary to respond to related words that are characteristics of each person with a particular topic or keyword. Experimental results obtained by searching books using multiple associative dictionaries showed that participants are more aware of the possibility of exploration. We also found a dictionary familiar to a particular participant improves comprehension.

Keywords: Book retrieval · Associative dictionaries · Serendipitous search · Exploratory search

1 Introduction

This paper proposes AssocSearcher, a system to support exploratory book search. AssocSearcher enables users to broaden their book search horizons by retrieving subthemes and related terms from other people's associative dictionaries. Associative dictionaries are created as a substitute for personal thoughts.

Associative dictionaries of individuals were generated from sentences read and written by individuals using Word2vec and GloVe. This allows each association dictionary to correspond to related words that are characteristic of each person with a particular topic or keyword.

Our intension is to encourage serendipitous discovery with AssocSearcher. In other words, users of the system can obtain new books, find new relevance between specific topics by using multiple diverse associative dictionaries. Our hypothesis is as follows:

- Users can extend the search space by using others' associative dictionaries.
- They encounter unexpected books through the extended exploration.

H. Mori et al. (Eds.): HCII 2023, LNCS 14056, pp. 131–150, 2023.
https://doi.org/10.1007/978-3-031-48044-7_10

This paper describes a method for generating associative dictionaries and the AssocSearcher that uses it, and discusses its effectiveness with experimental results.

The outline of this paper is as follows. First, we discuss related research on exploratory search, query expansion, serendipity, and book retrieval. Second, we present the proposed system and its expected results. Third, we describe the method for creating associative dictionaries and the exploratory search system AssocSearcher. Fourth, we describe comparative experiments and results of a search task using a single large corpus of related words and a search task using multiple associative dictionaries. Finally, we summarize the study by discussing the results, limitations, and future prospects.

2 Related Work

2.1 Exploratory Search

We aim to support exploratory search [22,32] by presenting new viewpoints to expand user's horizon and assisting their continuous exploration.

The main approach to date has been to visualize search results and topics to support exploratory search for user understanding [37]. Studies visualizing search results or topics include visualizing an overview of results (e.g., TileBars [14], Scatter/Gather [27], SenseMaker [4]); supporting broaden search space incrementally using the interactive network of data (e.g., PivotPaths [10], Apolo [8], Exploration Wall [19]); supporting topic search by visualizing text corpora enables users to find diverse perspective (e.g., ParallelTopics [12], Serendip [2], TopicPanorama [20]); and allowing users to manipulate the search result visualization partly (e.g., [1,6]).

The other approaches also aim to keep user's mind the possibility of exploration by showing data. ScentBar [35] show proportion of already viewed information of suggested queries. Also there are studies facilitate exploration using visualized timeline information [3].

The above studies facilitate users to keep extending search space by showing the possibility of finding information. In this paper, we use the association of multiple personalized associative dictionaries to facilitate it. They stimulate not only curiosity about specific topic but also stimulate curiosity about other people's thought: "What kinds of association would that person have?" We consider it makes users more aware of the possibility of exploration.

2.2 Query Expansion

We focus on enriching search results by extending or diversifying search query. There are studies to improve the quality of the search results using expansion base on semantic similarity [13,31,36]. There are several studies about personalization of expansion for search queries [5,9,21,39].

We also interested in that the effectiveness of personalization for query expansion becomes diverse for each keyword. Teevan et al. proposed predictive models to identify queries that can benefit from personalization [33]. In the specific search keyword or user, information necessity becomes diverse in complex. Therefore, we consider the flexibility for such search intents is necessary.

We consider multiple associative dictionaries enable users to fit their intent. For example, they can use their own dictionaries if they need personalized suggestion. If they want to know external words of the specific topic, they can use the dictionaries of whom have much interest in topic.

2.3 Serendipitous Search and Recommendation

We focus on facilitating serendipity, the ability to find value in unexpected information. In recommender system, the word "serendipity" often used for representing "unusualness", "unrecognized" or "surprising".

Many systems related to serendipity focus on helping users to find the information includes the above properties to facilitate serendipity. Iaquinta et al. [15] proposed method to recommend items by the system includes randomness in item choosing method. Auralist [38] attempted to change recommended items balancing accuracy, diversity, novelty and serendipity. Kamahara et al. [16] proposed method to recommend unexpected items from clusters similar to the other user.

The term "serendipity" was first used in the ancient fairy tale "The Three Princes of Serendip" by Horace Walpole [29]. In this story, there is the sentence "*as their Highnesses travelled, they were always making discoveries, by accidents and sagacity, of things which they were not in quest of*". That means making discovery needs accidents and sagacity: the wise to comprehend situation. According to Silvia, the emotion of interest, has two appraisals: "*an appraisal of something as new, unexpected, or complex*" and "*an appraisal of one's ability to comprehend the new, complex thing*" [30].

From the original story of Serendipity and Silvia's mention, we consider that the users would discard the unexpected information without ability for comprehending it. We aim not only providing new information but also facilitating comprehension of new information using human factor of other peoples' personalized associative dictionaries.

2.4 Serendipity for Book Exploration

It is said that a library itself has interest to rise exploration and serendipity [17]. There are also studies facilitate them using interface imitates actual bookshelf [11,18]. WebBook and WebForager uses 3D virtual bookshelf to represent exploration space of Web pages [7]. The other studies also use interactive visualization of information such as book cover, author and keyword [25,34].

The inspiration for this study came from thinking about Japanese-language book search sites such as SO-IMAGINE[1] and WebCat-Plus[2]. These websites have the ability to search for books using associative terms based on large corpora such as Wikipedia and newspaper article collections. These sites allow users to explore information from a variety of perspectives. Our system aims to extend the possibilities of exploration by stimulating curiosity about the thoughts of others.

3 Proposed System

3.1 Searching for Books by Distinctive Use of Multiple Personalized Associative Dictionaries

We propose a method to support exploratory search by using multiple personalized associative dictionaries. Associative dictionaries are created with information about which words are related to a given word. This is created using each individual's text collection. This collection consists of Web browsing history, bookmarks, articles written by users, sentences posted on Twitter[3], Facebook[4], Evernote[5] and WorkFlowy[6]. Each associative dictionary is different because each is composed of different reading and writing experiences and thoughts (Fig. 1).

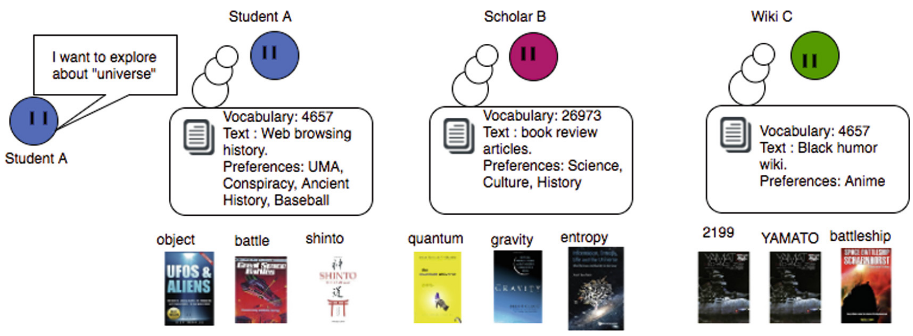

Fig. 1. Usage of the proposed system

When searching for books, users select a dictionary to use from multiple associative dictionaries, including their own dictionary. By performing an AND search or topic change using the associative terms in that associative dictionary,

[1] http://imagine.bookmap.info/index.jsp.
[2] http://webcatplus.nii.ac.jp/.
[3] https://twitter.com.
[4] https://www.facebook.com.
[5] https://www.evernote.com.
[6] https://workflowy.com/.

the user obtains search results that are different from those obtained with a single keyword.

Users are free to change the associative dictionary they use for their search depending on their search intent and interests. For example, if they wanted to look up a mystery novel, they would use the dictionary of someone who is familiar with detective novels. Or, if they want a book recommendation from someone they respect, they would use the dictionary of someone they admire. In this way, the users can flexibly use different associative dictionaries depending on their search purpose and mood.

3.2 Facilitation of Deep Understanding of Search Target Areas

By using the multiple associative dictionaries, the user can easily reach deeper into the information they want to look up. Let us take the case of "Artificial Intelligence" as an example. Word associations by a large single dictionary can only obtain words that have the same level of abstraction as the search keywords, e.g., "AI", "computer", and "robot". With the proposed system, users can obtain associative words such as "Minsky", "connectionism", and "frame problem" through the dictionaries of people who are knowledgeable in the specific field of artificial intelligence. Thus, by using multiple associative dictionaries, it is expected that information specific to the field of interest can be obtained more quickly and easily.

3.3 Discovery of New Books and Topics by Searching for Associative Terms that Users Themselves May Not Have Thought of

Multiple associative dictionaries can expand the search space of books that users can discover. The use of multiple associative dictionaries allows users to search using words that they would not have thought of on their own. Search results that are completely different from the user's expectations, but meaningful to the owner of the associative dictionaries, allow the user to obtain new search results from a different perspective. Thus, this method allows users to acquire unknown books and concepts.

4 Implementation

This section explains how to create a personalized associative dictionary and AssocSearcher which is a book explorer using them (Fig. 2).

4.1 Creation of Personalized Associative Dictionaries

The associative dictionaries are created as substitutes for individual thoughts. They should respond to related words that are characteristics of each person with an input word. We generated associative dictionaries of individual persons using Word2vec and GloVe from collection of sentences read or written by the

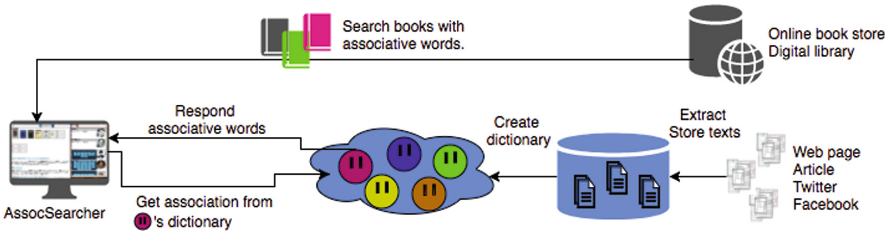

Fig. 2. System Overview

individuals. This collection consist of Web browsing history, bookmarks, articles written by users, sentences posted on twitter, Facebook, Evernote and WorkFlowy.

Word2vec is a tool to represent a language model on a neural network proposed by Mikolov et al. [23,24]. In this method, one word is represented as a vector around 200 dimensions, and the association between words can be obtained by calculating the cosine distance between vectors. Word2vec uses a model called skip-gram model. In the skip-gram model, it is assumed that neighbor of the certain word is related. Parameters of each vector are adjusted for maximizing the occurrence probability of neighbor words. Concretely, the parameters are adjusted to maximize the (1).

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq ,j\neq 0}\log p(w_{t+j}|w_t) \tag{1}$$

where T is the number of words represented in the vector space, c is the width of the neighbor words to examine (e.g. If $c = 1$, only the words before and after are evaluated). wt is the current word and $wt + j$ is the word that is located j word out of there. p is represented by the following (2).

$$p(w_{t+j}|w_t) = \frac{\exp(v'^{T}_{w_{t+j}}\ v_{w_t})}{\sum_{w=1}^{W}\exp(v'^{T}_{w}\ v_{wt})} \tag{2}$$

where v is a vector around 200 dimensions represents one word. By adjusting the parameters of v, (1) is maximized. In order to get a group of words that are related to a certain word, it can be realized by listing top rank word vectors in order of cosine distance from the word vector. Associative dictionaries created in this method was learned with 5 windows and 500 dimensions.

We also used the method called GloVe [26]. In the calculation of the similarity degree, the above method was performed on a probability, but this is a method which creating an overall co-occurrence matrix then converting it to a word vector. In this method, the dimension number is 100, the window is 10, the learning rate is 0.05, and the epoch number is 10.

We used associative words generated by both methods in the system describes later at the same time. We compared which one is used as an associative word for searching, but no difference was found. Therefore, in this paper, we define associative words that there are 10–14 words which combine with the top seven word vectors of cosine distance using each method, removing word duplications.

4.2 Example Association of Associative Dictionaries

The following describes examples of a word associative dictionaries actually created. Each table shows an example of associative words for the keyword "language".

The dictionary shown in Table 1 is a "Student A dictionary" that created from collection of the contents of the web page browsed by him (between two and three months). It can be observed this dictionary responds associative words related to information technology such as "Processing", "programming" and "function". The web browsing history includes Web page about Japanese professional baseball, conspiracy, prehistoric heritage, and programming. Another association example is "Mu" (the continent of legend which was said to exist in ancient times) for "continent". Also, when we enter a place name we get a baseball team name that is base at the place. It is understood that word association related to the information which he inputs.

Table 1. Example of association of Student A dictionary (keyword: language)

Method	Word2vec
Learning data size	11.4 MB
Vocabulary	12201

Word	Similarity
processing	0.8229
natural	0.7438
application	0.7268
speech	0.6903
factoid	0.6876
symposium	0.6564
conventionally	0.6536
word	0.6535

Method	GloVe
Learning data size	11.4 MB
Vocabulary	58120

Word	Similarity
natural	0.9622
type	0.8600
side effect	0.8458
processing	0.8350
functional	0.8336
programming	0.8330
speech	0.8074
Yiddish	0.7704

The dictionary shown in Table 2 is called "Seigo Matsuoka dictionary". Seigo Matsuoka is writer and researcher of Japanese culture. We created this dictionary from Web articles called "Thousand Nights, Thousand Books"[7]. It can be observed Seigo Matsuoka dictionary responds associative words related to

[7] http://1000ya.isis.ne.jp/.

Table 2. Example of association of Seigo Matsuoka dictionary (keyword: language)

Method	Word2vec
Learning data size	19.9MB
Vocabulary	26973

Word	Similarity
mathematical	0.6981
communication	0.6801
recognition	0.6503
eugenic	0.6439
ecology	0.6409
geography	0.6353
physiology	0.6322
symbol	0.6312

Method	GloVe
Learning data size	19.9MB
Vocabulary	77644

Word	Similarity
religion	0.9896
life	0.9772
logic	0.9762
mathematics	0.9722
environment	0.9712
spirit	0.9702
behavior	0.9701
system	0.9687

Table 3. Example of association of Professor A dictionary (keyword: language)

Method	Word2vec
Learning data size	717 KB
Vocabulary	8470

Word	Similarity
occurrence	0.6541
revelation	0.6452
pattern	0.6173
focus	0.5873
co-occurrence	0.5729
total	0.5673
protocol	0.5535
grammar	0.5527

Method	GloVe
Learning data size	717 KB
Vocabulary	14121

Word	Similarity
society	0.9887
target	0.9884
expression	0.9878
realization	0.9875
meaning	0.9874
experience	0.9867
information	0.9850
street	0.9846

communication of human despite previous dictionary. Also, this dictionary has abundant vocabulary about history and cultural history.

The dictionary shown in Table 3 is called "Professor A dictionary", a dictionary created from articles written by him, sentences posted on twitter. Professor A dictionary responds associative words such as "protocol (language protocol analysis)", "co-occurrence (co-occurrence reaction of language)" and "pattern (pattern language)", related to his research interests.

Thus, by using a variety of training data sets, each dictionary can present related terms that are characteristic of each person interested in a particular topic or keyword. Using the dictionaries thus created, we developed a book explorer called AssocSearcher.

4.3 AssocSearcher

We created a system called AssocSearcher that utilizes associative dictionaries as assistance for exploratory search for books. Users of the system can extend their horizons to explore books by getting subtopic or related words from the other peoples' associative dictionaries. This system is able to manage list of associative dictionaries to use, bookmarks and reading list, explore books using dictionaries.

On the exploring view (Fig. 3), search results of books and web pages are displayed on the left, the dictionary list of current user on the right. When a button at the lower left of the dictionary icon is pressed, a list of associative words for the search keyword by the dictionary is displayed. When a specific word is clicked from the list, the result on the left is updated to that search on the search keyword AND the associative word. When a button the right of current associative word, current associative word becomes current keyword, then refresh results into that on next current keyword. When a book image is pressed, a detailed view of the book is displayed, with book title, author name, detailed explanatory note, ISBN. There are "Later" button and close button at the lower right, when "Later" button is pressed, current book is registered for "reading list".

Fig. 3. Exploring view of AssocSearcher

We utilized search results of honto store[8] which is one of the largest Japanese online stores of books. Web search results assist understanding for associative words were obtained from bing[9]. This system is available on the Web, and this system collecting user's usage data.

The system detects the events such as input of a search word, browsing of book information, registration to read later, browsing of Web sites, association of a dictionary, query to get search result. For each event, date, time, search query, bibliographic information, associative word, name of dictionary associated the word.

[8] https://honto.jp/netstore.html.
[9] https://www.bing.com/.

5 Experiment

5.1 Goal

We hypothesize as follows.

– Users of multiple associative dictionaries can obtain more books that will be new discoveries for themselves (more books will be registered for "reading list").
– The exploring space become extended by using multiple associative dictionaries (nouns of detailed explanatory note of books will be diverse).

In order to verify it, a comparative experiment was conducted when using a multiple personalized associative dictionaries and when using a large single related dictionary.

5.2 Participants

6 students aged 21 to 24 (average age: 22) participated in the experiment. Table 4 shows basic information on experiment participants. Both participants are engaged in research activities, at least look through 1 to 2 books per month.

Table 4. Detail of each participant

ID	Gender	Frequency reading books	Keyword	Order
A	female	3–10/month	creature	single->multiple
B	male	1–2/month	gaze	multiple->single
C	male	3–10/month	sense	single->multiple
D	male	3–10/month	robot	multiple->single
E	male	3–10/month	language	multiple->single
F	male	1–2/month	learning	single->multiple

5.3 Task

The participant selected one keyword from 10 prepared keywords. These keywords include the words related to the topics the participants want to explore that were found by hearing in advance. The participant browsed books using the system starting from searching on selected keyword. While exploring, when the participant found a book that he wants to read, he pushed button to register the book for "reading list".

The participants were asked two tasks with AssocSearcher.

Single Dictionary Task. The participants explore books with single large dictionary utilizing Hatena associative words API[10].

Multi Dictionaries Task. The participants explore books with 10 personalized associative dictionaries (Table 5).

The participants performed each task for 15 min with the same keyword (balanced order between participants). We asked the participants to say whatever comes into their mind as much as possible during task.

Table 5. Dictionaries the participants use

dictionary	resources	notes
D1	book reviews	Scholar
D2	literatures	Historically prominent novelist
D3	blog articles	Blogger mentioning about AI
D4	twitter, research articles	University professor
D5	Web browsing history	Colleague of participants
D6	twitter	Colleague of participants
D7	Bookmarks, blog, memo	Colleague of participants
D8	twitter	Colleague of participants
D9	literatures	Historically prominent novelist
D10	official documents	Compendium of laws

5.4 Questionnaire

After completing both tasks, we asked to the participants two kinds of questionnaires. The first is asking the reason why the word was selected for exploring. For each associative word that the participant used for exploration, we asked a question with the following answer in a questionnaire "Please check all of the reasons for choosing this associative word which apply to you".

1. The word presented is relevant to my interests.
2. The word presented surprising to me.
3. The dictionary used to present the word has reliability, familiarity and expectation for me.

The second is a questionnaire on the overall impression evaluation of the task utilizing ResQue framework [28].

[10] http://developer.hatena.ne.jp/en/documents/keyword/apis/association/.

6 Results

6.1 Objective Measures

To test our hypothesis, we analyzed usage data of the participants (Table 6). For calculating Variation of books, list of nouns is extracted from a description note of each book registered for "reading list". The value was obtained by dividing the length of lists excluding duplication by that including duplication. Also, from the timestamp for each query history, we calculated topic stay time which is the time the participant keep exploring with the same keyword. Relevance is the rate that the participant checked "The word presented is relevant to my interests" for each questionnaire about associative words. Surprising is the rate that the participant checked "The word presented surprising to me" for that.

In each item, when t-test between two tasks was performed, no significant difference was found for all items ($p > .05$). In number of Associations ($t\ (6) = -2.05$), Topic stay time ($t\ (6) = -2.05$), Relevance ($t\ (6) = -2.53$), marginal differences were observed between the two tasks ($p < .10$).

Table 6. Overall usage data of each participants. Where TM represents multi dictionaries task and TS represents single dictionaries task.

	A		B		C		D		E		F		mean	
	TM	TS	TM	TS	TM	TS	TM	TS	TM	TS	TM	TS	TM	TS
Associations	18	19	8	10	17	19	11	28	3	10	12	14	11.5	16.6
Book views	12	14	33	33	14	18	26	21	33	26	9	14	21.2	21.0
Registrations	8	8	8	6	6	5	7	6	12	13	3	8	7.33	7.66
Variation of books	0.82	0.83	0.80	0.80	0.93	0.81	0.59	0.83	0.65	0.76	0.78	0.83	0.61	0.67
Topic stay time	105.0	109.5	327.5	195.8	141.7	116.4	339.5	139.8	442.4	154.1	147.0	165.4	250.5	143.8
Relevance	0.11	0.52	0.25	0.3	0.41	0.47	0.64	0.64	0.60	0.90	0.25	0.43	0.39	0.55
Surprising	0.44	0.16	0.38	0.40	0.59	0.52	0.18	0.29	0.33	0.30	0.33	0.29	0.38	0.33

6.2 Summary of User Responses Based on Questionnaire

The results of the overall questionnaire are shown in the Table 7. In Q1, Q2, Q3, the answers to each of the questions using the 1–5 likert scales, where 1 indicates "strongly disagree" and 5 is "strongly agree". A t-test performed and we found a significant difference between the two tasks on Q1 ($p < .05$).

We also asked users "The task ended in 15 min. Please choose your feeling of the moment that it ended". The answers to the question are "It ended early", "It ended at just right timing", "It ended late". 5 out of 6 participants answered "It ended early" on multi dictionaries task. 1 out of 6 participants answered "It ended early" on single dictionary task.

Table 7. Summery of overall questionnaire. Where TM represents multi dictionaries task and TS represents single dictionaries task.

Question	TM	TS
Q1 The items recommended to me are novel and interesting	4.0	3.0
Q2 The recommender system helps me discover new products	4.3	3.7
Q3 The items recommended to me are diverse	4.3	3.7

6.3 User Studies

In below, we show user studies about differences in exploration strategies and characteristics between two tasks. We mention them describing the overall usage data (Table 6), the dictionaries list used by participants (Table 5), and the distribution of dictionaries use for each user (Table 8).

Table 8. Distribution of usage of associative dictionaries in multi dictionaries task.

Participant	Distribution(%)
A	D8(33.3),D1(27.8),D5(16.7),D3(11.1),D4(5.6),D6(5.6)
B	D4(37.5),D1(25.0),D6(25.0),D3(12.5)
C	D6(23.5),D4(23.5),D8(17.6),D3(17.6),D5(11.8),D7(5.9)
D	D4(45.5),D7(27.3),D5(18.2),D3(9.1)
E	D1(66.7),D6(33.3)
B	D5(33.3),D2(33.3),D1(25.0),D3(8.3)

Book Exploration from a New Perspective. We will describe the case the user could explore books from new perspectives.

In single dictionary task, participant A started with “creature”, explored books by changing current keyword into associative word. However, the participant recognized that new keywords were not lexically far away from start keyword such as “taxonomy”, “owls” and “humans”. After the task, the participant said, “I have the impression that I have found books related to my favorite topic”. In the evaluation on associative words, Relevance was 0.52, Surprising was 0.16, we observed the participant explored books with more relevant associative words.

In the multi dictionary task, we observed the admirations such as “oh”, “wow” and “yes” at the more than half time before she registered for “reading list”. After the task, the participant said “I found somewhat related to living things with surprising, they seemed like interesting books”. In the evaluation on associative words, Relevance was 0.11, Surprising was 0.44, the tendency became opposite to the single dictionary task.

Selection of Dictionaries Based on Known Features. We observed that users use dictionaries depending on the characteristics of dictionaries which they knew in advance.

In multi dictionaries task, we observed selection of dictionaries based on dictionary's characteristics. At the beginning of task, participant C said "Writer or Scholar would associate difficult terms", "I think I can get interesting association from people that is familiar with me". Consequently, participant C did not use D1, D2, D9, D10 which were not familiar with him.

Participant D thought that colleague's dictionaries would associate words related to "robot" because they were interested in such topics. The participant frequently used dictionaries of his supervisor and laboratory members (D4, D5, D7), and D3 based on articles about Artificial Intelligence. Also, we observed when he feels disappointed at D6, his colleague's dictionary that could not respond to keyword on contrary to his expectations.

We observed participant B distinguished the usage of dictionaries. For example, when keyword is the word close to his research field, he used a dictionary of his research supervisor or colleagues. When he wants to broaden the search topic, he used dictionaries of literary giants and scholars. Also, no participants used association of D9 and D10 which were unfamiliar with them.

Selection of Dictionaries Based on Dictionaries Features in the Exploring Process. We observed that users use dictionaries depending on the characteristics of dictionaries which they learned during task.

In multi dictionaries task, we observed selection of dictionaries based on dictionary's characteristic after using particular dictionaries. Participant F looked through the associations of all dictionaries, until he comprehends their characteristic. We observed speeches about tendency of dictionaries such as "D2 will not answer this keyword", "D3 will able to respond such topic infinitely" and "D2 is weak in the keyword consist of European language".

Participant A looked through the associations of all dictionaries like participant F. After she said "association of D8 is interesting", she become positive to use this dictionary. Consequently, D8 marked the highest frequency of use for this participant.

Reactions to Unknown Concepts. We observed behaviors trying to comprehend the meanings of the associative words which are unfamiliar with them.

When participant A could not understand the word associated by the D1, she searches for the word in a single query, attempting to understand the meaning by reading the Web search result. She registered the book found at that process for "reading list". In the questionnaire about that associative word, the item "The dictionary used to present the word has reliability, familiarity and expectation for me" was checked.

Also participant B got unknown word from D4, the dictionary of his supervisor. He searched for that word and opened the detailed information from one book by one book in order to comprehend meaning of it. We also observed the case unknown associative words by personalized associative dictionary were not effective.

Conversely, participant D said, "Although associative words had unexpectedness, I could not find any related books at all, I could not enjoy the search of books". In participant D, the proportion of "relevance" was 0.18, there were no books registered using these unexpected words.

7 Discussion

At beginning of evaluation, we hypothesized as follows.

- Users of multiple associative dictionaries can obtain more books that will be new discoveries for themselves (more books will be registered for "reading list").
- The exploring space become extended by using multiple associative dictionaries (nouns of detailed explanatory note of books will be diverse).

There was no difference in the number of the books registered for "reading list" between two tasks. Also, there was no difference in the variation of the books that were registered for "reading list" between two tasks. We consider that is because the participants took time selecting associative words and browsing association of dictionaries. Our hypothesis was rejected, however, we also consider positively even if the participants took much time to do this, it do not cause decrease of the number of finding books and narrow exploration space.

We also found discussions as follows.

7.1 Search Books Using Words that Do Not Recognize Relevance

In the multiple dictionaries task, there was a tendency to search using associative words with low "relevance". Also in the questionnaire, "the book that came up was surprising and interesting", there was a significant difference between task.

We also found a dictionary familiar to a particular participant improves his comprehension capability. It seems that the user was able to conduct a search using words that are not conscious of the relevance with multiple associative dictionaries.

7.2 Exploration Sustainability

We believe that AssocSearcher helps users continue their book search.

In a single dictionary, keywords changed frequently because there were few options for obtaining associative words. Therefore, the time spent per topic was short. Conversely, in the case of multiple dictionaries, participants often continued AND searches for the same keywords using associative words from various dictionaries. This allowed participants to use multiple associative dictionaries to search for books while creating different perspectives of the search space.

In the single-dictionary task, we observed several "out of ideas" moments. No such moments were observed in the multiple dictionary task. 5 out of 6 participants reported they finished quickly and a marginal difference was observed in relevance of associative words between the two tasks. That means, with multiple associative dictionaries, participants can be more aware of the possibility of exploration. Therefore, we consider AssocSearcher gives users exploration sustainability.

7.3 Intentional Selection of Dictionaries

From the participants' utterances, it was found that participants varied the frequency of dictionary use based on their interest in a particular dictionary and its associations. Some participants limited the dictionaries they used, and unfamiliar dictionaries D9 and D10 were not used by all participants. It was suggested that intentional choice of dictionaries influenced book discovery.

7.4 Recognition of Individual Differences in Associative Dictionaries by Users

We believe that users can recognize the differences and features of multiple personalized dictionaries. In this study, we did not quantitatively evaluate the associative words of the dictionaries themselves. However, we did see predictions of trends in the associative words of certain dictionaries and the selection of dictionaries to use based on dictionary preferences. One participant stated, "It was fun just to see what words were associated with each dictionary, since I could see the characteristics of each dictionary". These results suggest that the AssocSearcher can be used to create a situation in which there is an associative dictionary that matches the personality of the individual user.

8 Limitations

The difference between the two tasks was considered to be the number of books found within the specified time. However, we found that the two tasks differed in terms of the likelihood of exploration perceived by the participants at the end of the task. Therefore, for a more accurate assessment, an experiment comparing the two tasks in terms of task completion time is needed. This can be accomplished by leaving the decision of task completion to the participants in the experiment.

It was also observed that the use of multiple associative dictionaries was counterproductive because if the associative words were too unrelated, no valid search results were obtained. It is necessary to consider how to create associative dictionaries that reflect the individuality of the associative words and have sufficient validity for book searches.

In the experiment, all participants used the same 10 dictionaries. Because the experiment was time-limited, it was not possible to observe in depth the changes in the usage of the dictionaries used by the participants. The experiment suggests that further analysis of the choice of dictionaries by users is needed.

9 Future Directions

We consider possible behavioral changes and social interactions of dictionary owners. During the task, we observed the user speeches such as "I'm wondering what other people think about my favorite word" and "If my dictionary exist, it could respond some associative words to explore for this keyword". People are interested in the information what other people are thinking about a certain word. AssocSearcher allows sharing such information among users. We focus on not only users of dictionaries but also the people who have the dictionary which is used by other people. For example, they might attempt to read or write information that they not usually see to nurture their own dictionary if their dictionaries are not used because of low vocabulary or narrow covering topics.

We believe that AssocSearcher could also recommend the owners of dictionaries which is used by other people discovering books. In this system, the specific user could find the book using the specific user's dictionary. By showing the book to the owner of the dictionary, recommendation to the owner could be realized. In other words, we can recommend the owner of the dictionary which the user get interested in, for the books which the user get interested in. We will analyze the effectiveness of this way of recommendation.

We believe that one person's associative dictionary can be classified into several new dictionaries. A person's interests change over time, and accordingly, the words that come to mind for a given word change. In addition, we can distinguish between private and public writings that a person writes or reads. For example, we believe that a person can create a new dictionary with different characteristics depending on time, purpose of writing, and purpose of reading, such as a dictionary for the year 2014, a dictionary for writing public documents, etc.

10 Conclusions

In this paper, we proposed a method that assists exploratory search by using multiple personalized associative dictionaries. We generated associative dictionaries of individual persons using Word2vec and GloVe from sentences read or written by the individuals. We described a system called AssocSearcher that enables users to extend their horizons to explore books by getting subtopic or related words from the other people's associative dictionaries. A discussion of experimental results comparing the use of multiple personalized associative dictionaries with the use of a large single associative dictionary suggested the following.

- AssocSearcher has enabled users to sustainably search for books.
- Human factor in associative dictionaries could influence book exploration and discovery.
- AssocSearcher could provide users with a situation where they could explore books using multiple perspectives with multiple personalities.

Future directions include discussing user immersion by comparing the execution times of the two conditions; and observing the effectiveness of personalized associative dictionary recommendations to the owners themselves.

References

1. Ahn, J.W., Brusilovsky, P.: Adaptive visualization for exploratory information retrieval. Inf. Process. Manag. **49**(5), 1139–1164 (2013). https://doi.org/10.1016/j.ipm.2013.01.007
2. Alexander, E., Kohlmann, J., Valenza, R., Witmore, M., Gleicher, M.: Serendip: topic model-driven visual exploration of text corpora. In: 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 173–182 (2014). https://doi.org/10.1109/VAST.2014.7042493
3. Alonso, O., Baeza-Yates, R., Gertz, M.: Exploratory search using timelines. In: SIGCHI 2007 Workshop on Exploratory Search and HCI Workshop, No. 1, pp. 1–4 (2007)
4. Baldonado, M.Q.W., Winograd, T.: SenseMaker: an information-exploration interface supporting the contextual evolution of a user's interests. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 11–18. ACM (1997). http://dl.acm.org/citation.cfm?id=258563
5. Biancalana, C., Micarelli, A.: Social tagging in query expansion: a new way for personalized web search. In: International Conference on Computational Science and Engineering (CSE 2009), vol. 4, pp. 1060–1065. IEEE (2009)
6. Bostandjiev, S., O'Donovan, J., Höllerer, T.: TasteWeights: A Visual Interactive Hybrid Recommender System, p. 35. ACM Press (2012). https://doi.org/10.1145/2365952.2365964
7. Card, S.K., Robertson, G.G., York, W.: The webbook and the web forager: an information workspace for the World-Wide Web. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 111–117. ACM (1996)
8. Chau, D.H., Kittur, A., Hong, J.I., Faloutsos, C.: Apolo: making sense of large network data by combining rich user interaction and machine learning. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 167–176. ACM (2011). http://dl.acm.org/citation.cfm?id=1978967
9. Chirita, P.A., Firan, C.S., Nejdl, W.: Personalized query expansion for the web. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 7–14. ACM (2007). http://dl.acm.org/citation.cfm?id=1277746
10. Dörk, M., Riche, N.H., Ramos, G., Dumais, S.: PivotPaths: strolling through faceted information spaces. IEEE Trans. Visual Comput. Graph. **18**(12), 2709–2718 (2012). https://doi.org/10.1109/TVCG.2012.252
11. Detken, K., Martinez, C., Schrader, A.: The search wall: tangible information searching for children in public libraries. In: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 289–296. ACM (2009)
12. Dou, W., Wang, X., Chang, R., Ribarsky, W.: Paralleltopics: a probabilistic approach to exploring document collections. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 231–240. IEEE (2011)
13. Gong, Z., Cheang, C.W., Leong Hou, U.: Web query expansion by WordNet. In: Andersen, K.V., Debenham, J., Wagner, R. (eds.) DEXA 2005. LNCS, vol. 3588, pp. 166–175. Springer, Heidelberg (2005). https://doi.org/10.1007/11546924_17
14. Hearst, M.A.: TileBars: visualization of term distribution information in full text information access. In: 1995 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 59–66 (1995) https://doi.org/10.1145/223904.223912

15. Iaquinta, L., de Gemmis, M., Lops, P., Semeraro, G., Filannino, M., Molino, P.: Introducing serendipity in a content-based recommender system. In: 2008 Eighth International Conference on Hybrid Intelligent Systems (HIS 2008), pp. 168–173. IEEE (2008)
16. Kamahara, J., Asakawa, T., Shimojo, S., Miyahara, H.: A community-based recommendation system to reveal unexpected interests. In: 2005 Proceedings of the 11th International Multimedia Modelling Conference (MMM 2005), pp. 433–438. IEEE (2005)
17. Khoo, M., Rozaklis, L., Hall, C.: A survey of the use of ethnographic methods in the study of libraries and library users. Lib. Inf. Sci. Res. **34**(2), 82–91 (2012)
18. Kleiner, E., Rädle, R., and Reiterer, H.,: Blended shelf: reality-based presentation and exploration of library collections. In: CHI 2013 Extended Abstracts on Human Factors in Computing Systems, pp. 577–582 (2013). https://doi.org/10.1145/2468356.2468458
19. Klouche, K., Ruotsalo, T., Cabral, D., Andolina, S., Bellucci, A., Jacucci, G.: Designing for exploratory search on touch devices. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 4189–4198. ACM (2015)
20. Liu, S., Wang, X., Chen, J., Zhu, J., Guo, B.: TopicPanorama: a full picture of relevant topics. In: 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 183–192. IEEE (2014)
21. Liu, S., Liu, F., Yu, C., Meng, W.: An effective approach to document retrieval via utilizing WordNet and recognizing phrases. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 266–272. ACM (2004). http://dl.acm.org/citation.cfm?id=1009039
22. Marchionini, G.: Exploratory search: from finding to understanding. Commun. ACM **49**(4), 41–46 (2006). https://doi.org/10.1145/1121949.1121979
23. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
24. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013). http://papers.nips.cc/paper/5021-distributed-representations
25. Pearce, J., Chang, S.: Exploration Without Keywords: The Bookfish Case, pp. 176–179. ACM Press (2014). https://doi.org/10.1145/2686612.2686639
26. Pennington, J., Socher, R., Manning, C.: GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014). http://www.aclweb.org/anthology/D14-1162
27. Pirolli, P., Schank, P., Hearst, M., Diehl, C.: Scatter/gather browsing communicates the topic structure of a very large text collection. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 213–220. ACM (1996)
28. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys 2011), pp. 157–164. ACM (2011). https://doi.org/10.1145/2043932.2043962
29. Remer, T.G.: Serendipity and the Three Princes. University of Oklahoma Press (1965)

30. Silvia, P.J.: Emotional responses to art: from collation and arousal to cognition and emotion. Rev. Gen. Psychol. **9**(4), 342–357 (2005). https://doi.org/10.1037/1089-2680.9.4.342
31. Song, M., Song, I.Y., Hu, X., Allen, R.B.: Integration of association rules and ontologies for semantic query expansion. Data Knowl. Eng. **63**(1), 63–75 (2007)
32. Teevan, J., Alvarado, C., Ackerman, M.S., Karger, D.R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 415–422. ACM (2004)
33. Teevan, J., Dumais, S.T., Liebling, D.J.: To personalize or not to personalize: modeling queries with variation in user intent. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 163–170. ACM (2008)
34. Thudt, A., Hinrichs, U., Carpendale, S.: The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1461–1470. ACM (2012). http://dl.acm.org/citation.cfm?id=2208607
35. Umemoto, K., Yamamoto, T., Tanaka, K.: ScentBar: a query suggestion interface visualizing the amount of missed relevant information for intrinsically diverse search. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 405–414. ACM (2016)
36. Voorhees, E.M.: Query expansion using lexical-semantic relations. In: Croft, B.W., van Rijsbergen, C.J. (eds.) SIGIR 1994, pp. 61–69. Springer, London (1994). https://doi.org/10.1007/978-1-4471-2099-5_7
37. Zhang, J.: Visualization for Information Retrieval, vol. 23. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75148-9
38. Zhang, Y.C., Séaghdha, D.Ó., Quercia, D., Jambor, T.: Auralist: introducing serendipity into music recommendation. In: Proceedings of the fifth ACM International Conference on Web Search and Data Mining, pp. 13–22. ACM (2012). http://dl.acm.org/citation.cfm?id=2124300
39. Zhou, D., Lawless, S., Wade, V.: Improving search via personalized query expansion using social media. Inf. Retrieval **15**(3–4), 218–242 (2012)