We are IntechOpen,
the world's leading publisher of Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

An Information Filter
for Intuitive and Simple Search

Sayaka Akioka[1], Hideo Fukumori[2] and Yoichi Muraoka[2]
[1] The University of Electro-Communications
Japan
[2] Waseda University
Japan

1.Introduction

Recent quick popularization of World Wide Web (WWW) has changed people's life; whenever people move into action, they probably go to the web first, and try to collect useful information. However, success and failure of this first step heavily relies on the queries they choose. That is, people will never run into any useful information without a set of expected keywords, which is a technical limitation of current information retrieval systems. In reality, many users find it difficult to form an appropriate set of queries to describe what exactly they are looking for.

In order to address this problem, this chapter proposes an information filter with the idea of kansei engineering. The approach expands the flexibility of an information filter, and has possibility to enable users to search items without specific queries, or helps users put their requirements into a concrete set of queries. The information filtering method described in this chapter does not necessarily provide "the very best one" result, however, promise to offer a user a certain set of "something like this" items. Our goal is to let users search information visually and sensuously, considering as many aspects as they need until they finally reach at the answer they were seeking, instead of helping users form a better set of queries.

"Kansei" is a Japanese vocabulary, and means psychological feeling or image of a product. Nagamachi founded kansei engineering as an activity to develop a new product by translating a customer's psychological needs and feeling (kansei) concerning it into design specifications (Nagamachi, 2006). Kansei engineering is widely used centering on the areas such as user interface (Berry et al., 2006; Klauser et al., 2007), music analysis (Kazama et al., 2006; Kamata et al., 2007), and virtual reality (Kaino et al., 2003).

Here, psychological feeling is often ambiguous, and hard to decide one solid expression for corresponding feeling. Therefore, kansei engineering often requires a questionnaire process in order to position a certain expression between two extreme impressions (for example, *large* and *small*), which is not a very appropriate way from the view of computational engineering. However, kansei engineering still has the potential to put flexibilities onto information retrieval or recommenders, if there exists a way to relate and position kansei expressions each other automatically, and map the relationships onto information on the

web. Therefore, the rest of this chapter describes information filter as an application of kansei engineering including the methodology to replace the questionnaire process with more computational engineering way.

This chapter is organized as follows. Section 2 introduces kansei filter, and describes the methodology to build and apply kansei filter as information filter to WWW contents. In Section 3, we compared kansei filter with the conventional collaborative filter, and then we discuss the effect of kansei information filter in Section 4. Section 5 addresses related contributions regarding recommenders, and we conclude in Section 6.

2. Kansei Information Filter

A general methodology on application of kansei engineering to design activities is following.

1. Pick up kansei words regarding the vocabularies possibly related to the features or images of a target product. For example, if a designer of the product expects to give an impression such as *airly light*, one should pick up *heavy* and *light*.

2. Conduct questionnaires and let the examinees to grade the products in the measure of two extreme kansei words extracted in the previous step. For example, examinees are requested to locate the product somewhere between *light* and *heavy*, according to their impressions and images. Semantic differential is often used for the evaluation.

3. Based on the results through the questionnaires, a certain set of features and impressions is derived. The derivertives are fed back to the product designer, and useful to be stocked as a part of the knowledge data base.

4. Repeat step 1 and step 3 as many as required, changing the set of kansei words and/or the design of the product if necessary. The set of steps repeats until the speculation of the product designer meets the results of the questionnaries.

The idea for kansei information filter, which is proposed in this chapter, is to utilize the knowledge database of kansei words as a part of information filter. More precisely, we employ the kansei database as an index of similarity of web documents or product descriptions. This section describes the details of kansei information filter as well as how to build up kansei database independent from questionnaires.

2.1 Kansei Map

Figure 1 is an example of kansei maps utilized in this chapter, and each document is scored according to these kansei maps for similarity measurements described in Section 2.2. In Figure 1, *wide* and *narrow* are the two extreme kansei words, and we call this map *wide-narrow* map. In order to generate kansei maps like the example shown in Figure 1, we execute a sequence of processes as follows.

1. Collect the web documents to process, and perform morphological analysis over the documents.

2. Extract adjectives, and find lexical and semantic synonyms, antonyms, and negatives utilizing dictionaries and/or other resources.

3. Relate extracted vocabularies and vocabularies from dictionaries each other, and build up kansei map candidates. Each set of related vocabularies becomes a kansei map candidate.

4.   Match the kansei map candidates and kansei word definitions. The survival maps of the matching are adopted as information scoring base.

In the process above, we focus only on adjectives. As our purpose here is to build up kansei map, and kansei words express feeling or impression, so we simply rely only on adjectives and without loss of generality.

We also expand the level of classification from two levels to four levels. The original kansei word definition consists of pairs of two extreme vocabularies. Consequently, we match only the two extreme vocabulary pairs in step 4. However, in order to allow higher level of classification and flexibility, we fill the gap between the two extreme vocabularies with the help of lexical definitions.
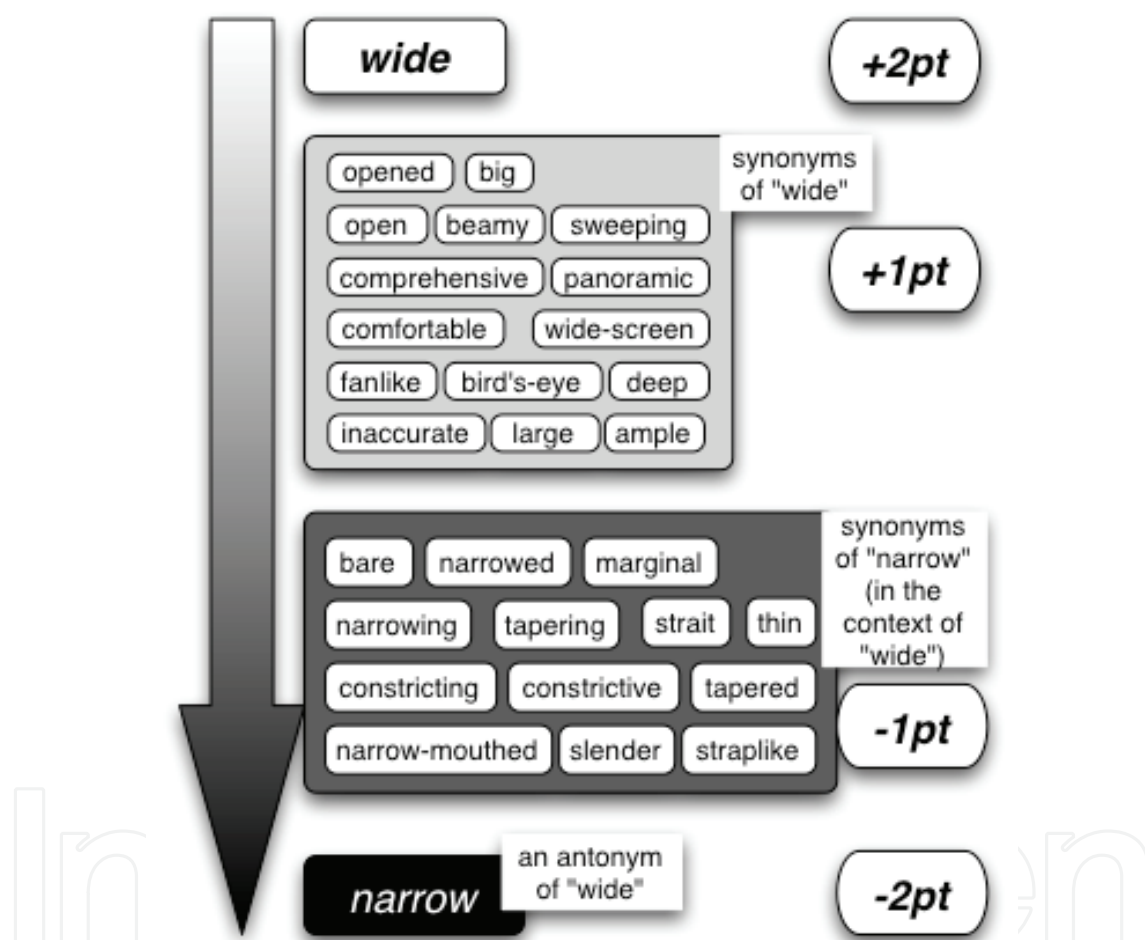
Fig. 1. An example of kansei map, and scoring according as similarity.

As shown in Figure 1, each kansei map classifies vocabularies into four levels according to direction and degree of expression on the kansei map. For example, in case of Figure 1, one opposite adjective "wide" scores two points, while the other opposite adjective "narrow" looses two points. Synonyms of "wide" scores one point, as those adjectives are semantically close to "wide". Similarly, synonyms of "narrow" loose one point, as these adjectives are semantically close to "narrow", but far from "wide".

2.1 Similarity Measurement

As a first step of the similarity measurement, we calculate information entropy for each document item. Let l be the score level on kansei map, and $l \in L = \{-2, -1, 1, 2\}$. We define information entropy of item x regarding scoring level l of map m (denoted as $H_{m,l}(x)$) as follows.

$$P_{m,l}(x, w) = \frac{n_x(m, l, w)}{n_x(m, l)} \tag{1}$$

$$H_{m,l}(x) = -\sum_w P_{m,l}(x, w) \log_2 P_{m,l}(x, w) \tag{2}$$

In equation (1), $n_x(m, l)$ represents total number of adjectives of level l on map m in the descriptions of item x, and $n_x(m, l, w)$ represents whether vocabulary w of level l on map m appears in the document item x or not. That is, $n_x(m, l, w) = 1$ if vocabulary w appears, otherwise, $n_x(m, l, w) = 0$.

If none of the vocabularies of level l on map m appears in the document x, $H_{m,l}(x)$ is defined as 0, as $P_{m,l}(x, w) = 0$ for any vocabulary w. Although the definition is for descriptive purposes, this definition contains an intuitive sense. Under $P_{m,l}(x, w) = 0$ for any vocabulary w at level l of map m, the item x has nothing to do with level l of map m. Therefore, the information entropy of the item x for level l of map m is also 0.

Once we define information entropy of items, mutual information is regarded as the measure of similarity. We denote mutual information between item x and y at level l on map m as $I_{m,l}(x; y)$. $I_{m,l}(x; y)$ is given as

$$I_{m,l}(x; y) = H_{m,l}(x) + H_{m,l}(y) - H_{m,l}(x, y) \tag{3}$$

Here, with adequately huge number of document items, the event of appearance of a vocabulary of a certain level on a certain map is consider to be random, and each event is independent each other. Therefore, joint entropy $H_{m,l}(x, y)$ is given as follows.

$$H_{m,l}(x, y) = H_{m,l}(x) \times H_{m,l}(y) \tag{4}$$

Now, note that $I_{m,l}(x; y)$ represents the similarity of item x and item y regarding the connectivity between items and map m at level l, however, does not represent the similarity between item x and item y itself. Larger $I_{m,l}(x; y)$ simply means item x and y relate to map m at level l with more similar level of connectivity. So, as a second stage, we consider similarity of items on the same map. We define the similarity of item x and item y on map m, denoted as $W_m(x, y)$, as

$$W_m(x, y) = \sum_{l \in L} s_l I_{m,l}(x; y) \tag{5}$$

where s_l represents the score at level l, shown in Figure 1. The situations $W_m(x, y)$ expresses are following.

■ $W_m(x, y) = 0$

   The two items have no relationship at any level of map m, or any meaningful relationship.

■ $W_m(x, y) > 0$

   The two items are similar and located in the positive area (around the original vocabulary or synonyms of the original vocabulary) in a large sense. If $W_m(x, y)$ is larger, the two items are plotted closer on the map.

■ $W_m(x, y) < 0$

The two items are similar and located in the negative area (around the antonyms or synonyms of the antonyms) in a large sense. If $W_m(x, y)$ is smaller, the two items are plotted closer on the map.

Finally, we consider the similarity between items across maps. We define the similarity of item x and item y, denoted as $W(x, y)$, as follows.

$$W(x, y) = \sum_{m \in M} |W_m(x, y)| \qquad (6)$$

where M represents a set of kansei maps to consider. The larger $W(x, y)$ represents better similarity.

The contributions of the proposed algorithm are summarized as follows.

■   Kansei   map

The vocabularies contained in the same kansei map comprehend the same context, semantic, and kansei. With kansei map, vocabularies are categorized into groups according to the mood the vocabulary has. For example, adjectives related to breadth are categorized into the same group.

■   Scoring   system

The four-level scoring system direct vocabularies on a map according to the meaning the vocabulary has. For example, "wide" and "narrow" are recognized as a pair of opposite adjectives.

■   Map   level   similarity

The combination of mutual information and the scoring system vectorlizes documents ($W_m(x, y)$). The map level similarity not only measures the similarity between items, but also represents the mood (position on a kansei map) of the items.

■   Map   across   similarity

The map across similarity measure $W(x, y)$ allows cumulating as many maps as requested, but measures the similarity between items with only one measure.

3. Empirical Experiments

In this section, we examine kansei information filter through experiments with the actual data sets crawled from Yahoo! Shopping web services (Yahoo! Inc, 2009). We crawled the product data focusing on shoes, and extracted <summary> tags. A <summary> tag is defined at Yahoo! Product Search web service, and expected to contain a short description of the product. The XML document shown in Figure 2 is an example of the information that can be retrieved for a product item via Yahoo! Product Search web service.

3.1 Kansei filter construction

First, we processed morphological analysis over the 79,812 shoes products with the morphological analysis engine provided via LingPipe(Alas-I, Inc., 2009), for kansei map construction described in Section 2. In this experiment, we utilized Brown University Standard Corpus of Present-Day American English (Kucera et al., 1967) in Natural Language Toolkit (Loper et al., 2002; Natural Language Toolkit Projects, 2009), which is a well-utilized general-purpose corpus. After this first stage is completed, we obtained 91,093 kinds of adjectives.

Secondly, we referred the dictionary definitions to build up basic relationships among adjectives, and then matched those groups of adjectives with kansei word definitions by Nagamachi (Nagamachi, 1995). In the appendix of (Nagamachi, 1995), Nagamachi defined about 400 adjective pairs as kansei expressions. As lexical definitions, we utilized WordNet. WordNet defines words in its original fashion, different from general lexical definitions, but shares the basic idea with kansei engineering. WordNet categorizes words into sets of cognitive synonyms, called synsets, and each synset expresses a distinct concept. Synsets are connected each other based on conceptual semantic or lexical relations. We employed the definitions by WordNet in order to fill between the two extreme expressions with kansei engineering. Finally, we obtained a set of kansei maps with four levels of categorization, described in Section 2.

```
<Product type="Offer">
 <Offer ID="11085491">
  <Url>...</Url>
  <ProductName>...<ProductName>
  <Price>...</Price>
  <Thumbnail>...</Thumbnail>
  <ListImage>...</ListImage>
  <GridImage>...</GridImage>
  <Summary>
    These &lt;b&gt;Adidas&lt;/b&gt; AdiStar Trail 3
shoes have everything an athlete needs to stay on
their toes and in the game Stylish women&apos;s
running shoes are also durable This supportive
footwear is made to last
  </Summary>
  <Merchant id="1000264">...</Merchant>
  <Category>...</Category>
 </Offer>
</Product>
```

Fig. 2. An example for the product information available at Yahoo! Product Search web service.

4. Discussions

In this section, we take a close look on the representative four maps, which are hard-soft map, light-dark map, round-square map, and plain-fancy map. The expected effect for each map is as follows.

- Hard-soft   map expressions for hardness of the materials, or taste of design.
- Light-dark   map expressions for the colors of shoe products.

■   Round-square   map
     descriptions for the shape of toes.
■   Plain-fancy   map
     expressions for the overall impression, design, and concept of the shoe products.

Table 1 is the details of the four obtained maps. Each row represents each map, and each column represents the expression for each level respectively. The number in parentheses appears next to an adjective is the frequency of the appearance in the whole set of the product descriptions.

As shown in Table 1, both the synonyms of the original adjectives and the synonyms of the antonymous frequently appear in the descriptions as many times as, or even more frequently than the extreme expressions. This observation indicates that the information provider utilizes fine expressions to express the details or atmosphere of the products. That is, the simple matching method only with the extreme expressions causes the poor results such as missing the products with rich expressions, or failing to reflect the fine requirements by the users.

Even with the advantage of the maps shown in Table 1, the process of map production has a point to be improved. More specifically, some misplaced adjectives are observed from the view of the context expressed with the two extreme expressions, even though these adjectives are synonyms of the original adjectives or antonyms lexically. Better placement of adjectives will plot products more precisely on a map. However, the automatic way to eliminate those unsuitable expressions requires another evolution. Therefore, we reserve this improvement for the future work.

Original Adjective	Synonyms of the Original Adjective	Synonyms of the Antonymous Adjectives	Antonymous Adjective
hard(532)	ambitious(7), arduous(1), catchy(3), delicate(490), nasty(7), rocky(91), rugged(982), serious(339), tall(1087), troublesome(2), case-hardened(7), firm(529), unyielding(1), strong(347),indulgent(53), bad(48)	compressible(7), cushioned(12063), spongy(13), velvet(51), mellow(3), dull(9), gentle(77), little(494), low(2930), tender(22), mild(153), light(1289), easy(4908)	soft(7986)
light(1289)	lightweight(8490), airy(142), buoyant(30), pale(29), powdery(1), ablaze(11), bright(115), incandescent(3), luminescent(3), white(5800), loose(143), pure(396), easy(4908), ill(1), thin(921), temperate(4), frivolous(2), shallow(10)	black(7881), dusky(1), darkish(7)	dark(735)
round(4101)	full(7458)	squared(20),	square(431)

		aboveboard(2), wholesome(3), direct(205), conventional(43)	
plain(215)	obvious(9), bare(124), dry(1852), simple(835), solid-colored(1), pure(396), direct(205), unadorned(4), unattractive(1)	baroque(9), busy(108), dressy(402), elaborate(24), fanciful(11), fantastic(150), lacy(1), puff(1)	fancy(55)

Table 1. The four representative kansei maps. Each row represents each kansei map, and the number in the parentheses is the frequency of appearance across the product descriptions.

Figure 3 (case (a)) and Figure 4 (case (b)) are the plots of similarity measures on a particular item versus the other 49 items respectively. Figure 5 (case (a)) and Figure 6 (case (b)) are the plots of the same similarity measures on the same centering item versus the other 49 items, but in the different form. In both cases, if the absolute value is larger, the two items are more similar. At the same time, the positive similarity represents that the two items are similar in the flavor of one side of the map, while the negative similarity represents that those two items are similar in the flavor of the other side of the map, as described in Section 2.

Case (a) and case (b) are contrasting cases. For case (b), the plots are essentially located in the negative part, with a few exceptional cases. This is the result of the fact that the center item is located in the negative part of the map, and the other items are also similar in the same side of the map, or have no relationship with the center item. On the other hand, in case (a), the plots are distributed from the negative part to the positive part. As a matter of fact, the center item of case (a) is one of the exceptional in case (b). This type of similarity distribution should not appear logically, as this item has similarity equally to the other items across the map, which is not reasonable. The only reason for this observation is that the description for this center item contains expressions for all the levels of the map equally. So, if the description is correctly one-sided, the plot looks similar to case (a), and many of the 50 items actually showed this characteristic. The counter methodology for the cases such as case (b) is planned as future work.

5. Related Work

To the best of our knowledge, this is the first effort to apply kansei engineering to information filter. However, there are similar approaches utilizing WordNet applicable to information filter, as a contradictory approach to simple lexical mapping. The recent contributions are as follows. Varelas et al. propose an information retrieval model based on the semantic similarity among documents (Varelas et al., 2005). Sim developed information filtering agent utilizing the ontology provided by WordNet (Sim et al., 2004). Cao et al. built a dependency model relies on both co-occurrences of terms and definitions in WordNet (Cao et al., 2005). Zhang et al. presented an algorithm for noun phrase recognition, utilizing WordNet as well as other resources (Zhang et al., 2007).

The definitive difference between WordNet approach and kansei engineering approach is the underlying policy on map constructions. Cognitive synonyms (synsets) defined by WordNet are vocabularies related conceptually or lexically. That is, WordNet utilizes sense-to-sense and word-to-word relationships in order to break down the limitation of conventional lexical

mapping. Here, if there are two vocabularies sharing a hypernim or superordinate concept, the two expressions are recognized as sense-to-sense related vocabularies.

Kansei filter, proposed in this paper, stands one step ahead in the meaning of expression of concept. Our contribution aims to provide the methodology to bring more conceptual approach into information search or recommenders with kansei engineering. Kansei word definitions are pairs of opposite expressions. Especially for adjectives on valuations, kansei definitions for these adjectives are always pairs of an adjective and its negative, instead of its antonym. This feature of kansei engineering enables to construct more specific relationships, and boosts up the quality of information filtering.
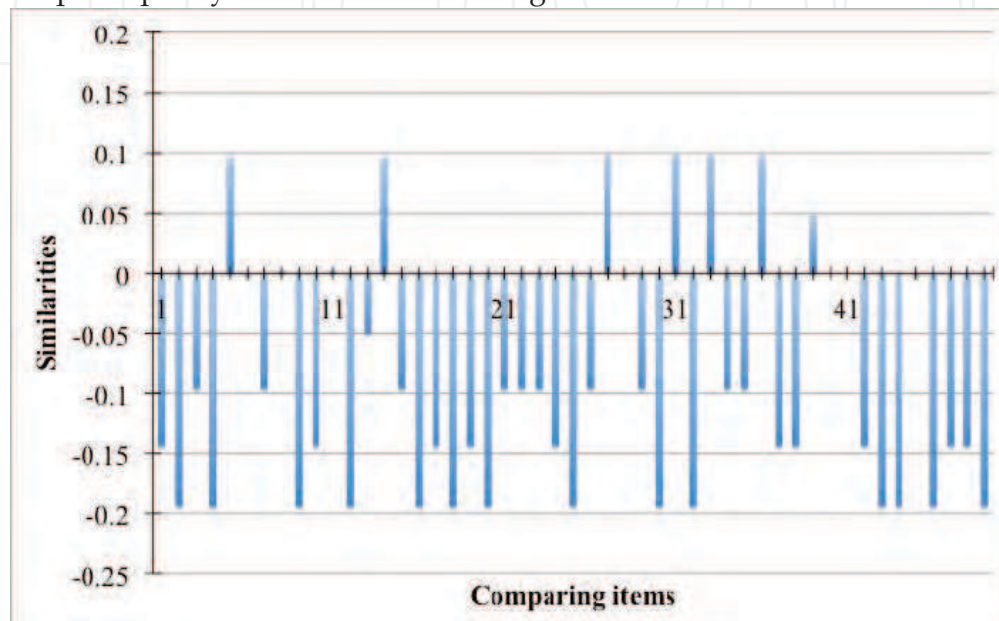
Fig. 3. Similarity measures with the other 49 items, centering on a particular item (case (a)).
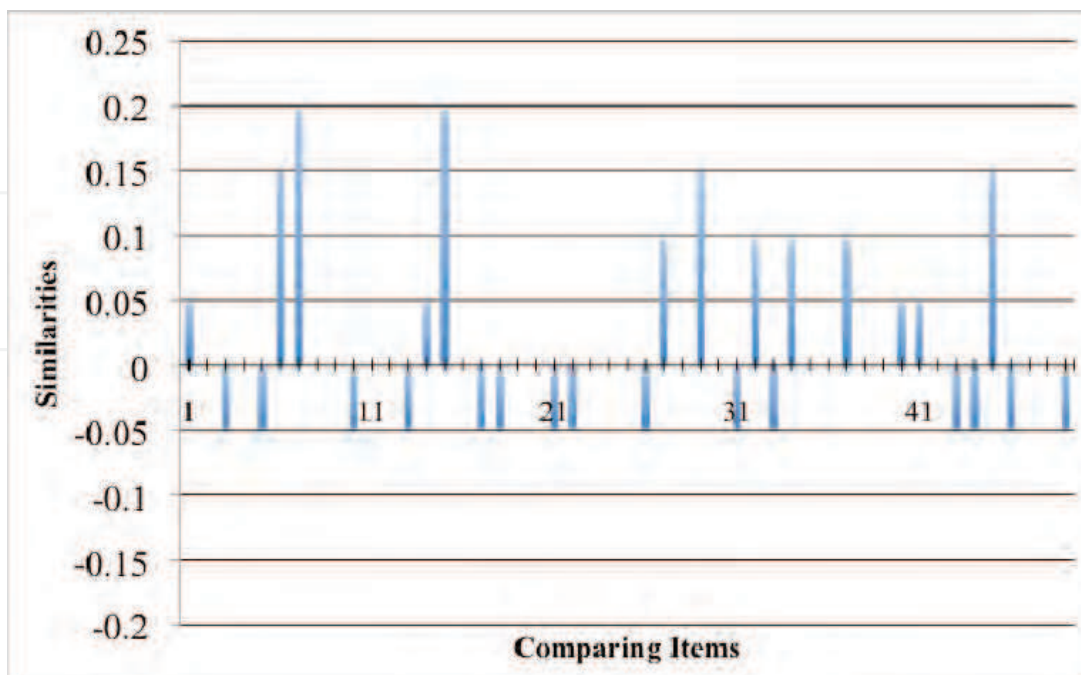
Fig. 4. Similarity measures with the other 49 items, centering on a particular item (case (b)).
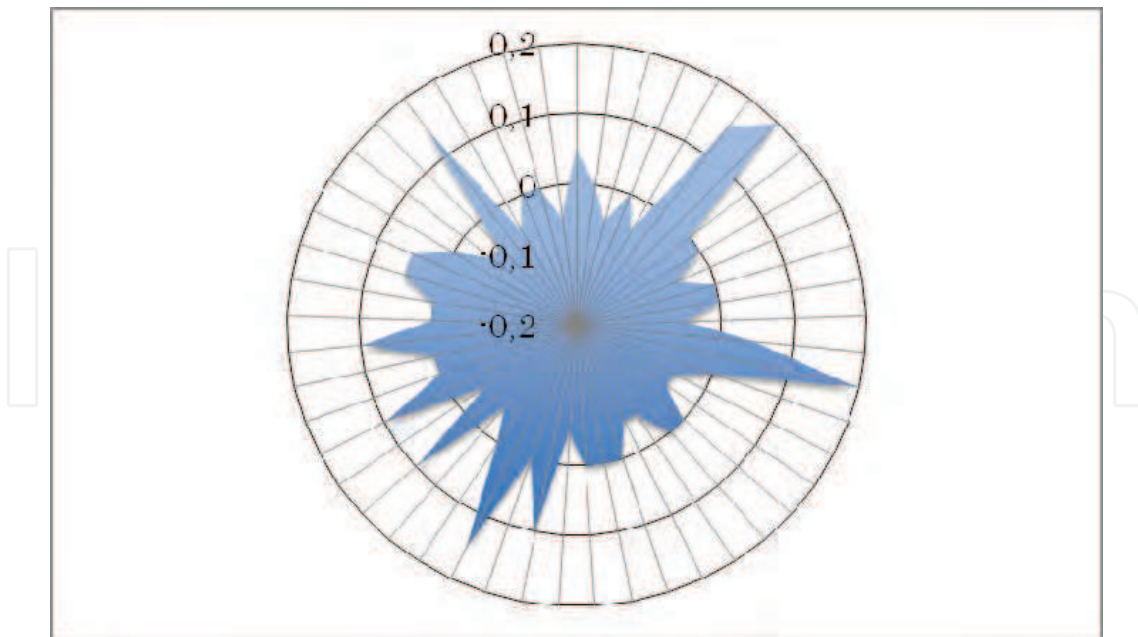
Fig. 5. Similarity measures with the other 49 items, centering on a particular item (case (a)). The same data set with Figure 3, however, in another form of plots.
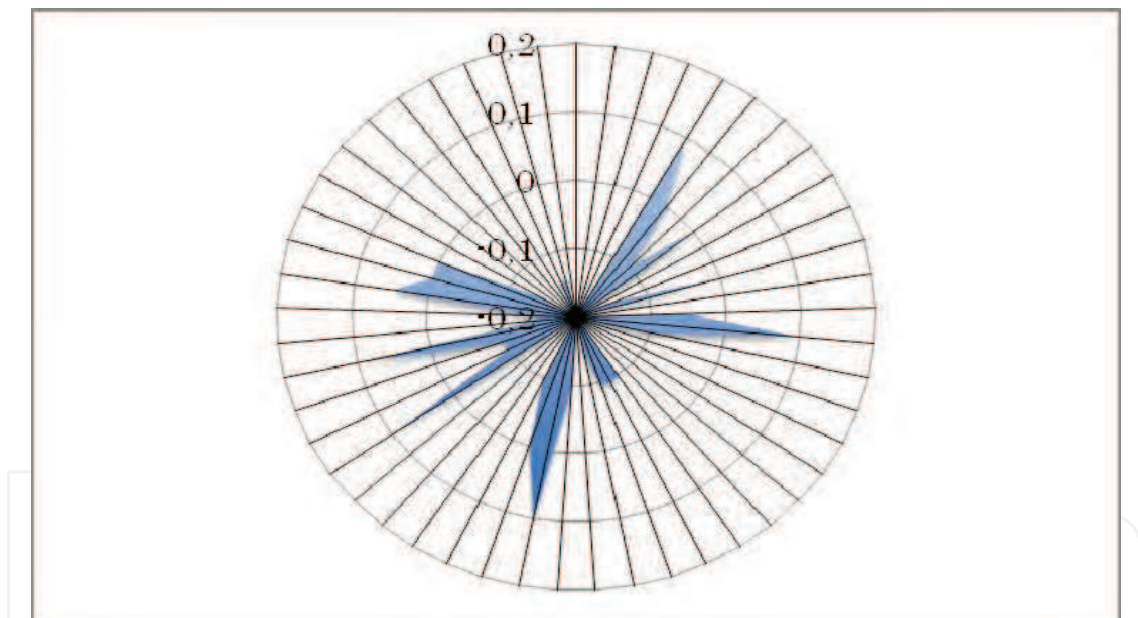

Fig. 6. Similarity measures with the other 49 items, centering on a particular item (case (b)). The same data set with Figure 4, however, in another form of plots.

6. Conclusions and Future Work

Recent quick spread of web services has change the style of information retrieving, and caused the explosive amount of information provided. Such situation requests a certain skill to obtain information effectively. Especially, ambiguous search is one of the major demands, however, such a service is known as one of the technical challenges.

In this paper, we proposed an information filter employed kansei engineering concept, in order to enable flexible filtering. Through the experiments with the actual data collected on the web, we verified that the concept of kansei filtering perform reasonably.

One of the unsolved problems is the methodology for the evaluation. Even though this paper successfully represented the flexibility and usefulness of information filter with kansei engineering, we have no clear way to measure the degree of satisfaction of users. The ideal benchmark set should be a set of data, such as product descriptions, evaluations by the users for the products, the history of purchases, and so on. We keep considering this problem for better justification.

7. References

Alias-I, Inc. (2009). LingPipe. http://www.alias-i.com/lingpipe

Berry, R.; Naemura, M.; Kobayashi, Y.; Tada, M.; Inoue, N.; Pisan, Y. & Edmonds, E. (2006). An Interface Test-bed for 'Kansei' Filters using the Touch Designer Visual Programming Environment. *Proceedings of 7thAustralian User Interface Conference (AUIC2006),* Vol. 50, pp. 173 – 176, ISSN 1-920682-32-5, Hobart, Australia, January 2006, Australian Computer Society, Darlinghurst, Australia

Cao, G.; Nie, J.-Y. &Bai, J. (2005). Integrating Word Relationships Into Language Models. *Proceedings of 28th Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR2005),* pp. 298 – 305, ISSN 1-59593-034-5, Salvador, Braszil, August 2005, ACM, New York, USA

Fellbaum, C. (1998). *WordNet: An Electric Lexical Database,*ISBN 978-0262061971, May 1998, MIT Press, Cambridge, USA

Kaino, H. & Hagiwara, M. (2003). 3-D Virtual Space Creation System Reflecting User's Emotion by Interactive Evolutionary Method. *Proceedings of the Second International Conference on Entertainment Computing (ICEC2003),* pp. 1 – 6, Pittsburgh, USA, May 2003, Carnegie Mellon University, Pittsburgh, USA

Kamata, M. & Furukawa, K. (2007). Three Types of Viewers' Favourite Music Videos. *Proceedings of the International Conference on Advances in Computer Entertainment Technology (ACE2007),* pp. 196 – 199, ISBN 978-1-59593-640-0, Salzburg, Austria, June 2007, ACM, New York, USA

Kazama, M. & Shiizuka, H. (2006). Music Analysis in Consideration of Feelings. *Proceedings of 2006 International Conference on Game Research and Development (CyberGames2006),* pp.117 – 120, ISBN 86905-901-7, Perth, Australia, December 2006, ACM, New York, USA

Klauser, K. & Walker, V. (2007). It's About Time: An Affective and Desirable Alarm Clock. *Proceedings of 2007 International Conference on Designing Pleasure Products and Interfaces (DPPI2007),* pp. 407 – 422, ISBN 978-1-59593-942-5, Helsinki, Finland, August 2007, ACM, New York, USA

Kucera, H. & Francis, W. N. (1967). *Computational Analysis of Present-Day American English,*ISBN 978-0870571053, June 1967, Brown University Press, Providence, USA

Loper, E. & Bird, S. (2002). NLTK: The Natural Language Toolkit. *Proceedings of the ACL-02Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics,* Vol. 1, pp. 62 – 69, Philadelphia, USA, July 2002, ACL, Morristown, USA

Miller, G. A.; Fellbaum, C.; Tengi, R.; Wakefield, P.; Poddar, R.; Langone, H. & Haskell, B. (2009). WordNet: A Lexical Database for the English Language. http://wordnet.princeton.edu/

Nagamachi, M. (1995). *Kanseikougaku no Ohanashi,* ISBN 978-4-542-90186-5, July 1995, Japanese Standards Association, Tokyo, Japan (in Japanese)

Nagamachi, M. (2006). Kansei Engineering and Rough Sets Model. *Proceedings of 5th International Conference on Rough Sets and Current Trends in Computing (RTCS2006),* Vol. 4259, pp. 27 – 37, ISBN 3-540-47693-8, Kobe, Japan, November 2006, Springer, Berlin, Germany

Natural Language Toolkit Projects. (2009). Natural Language Toolkit. http://nltk.sourceforge.net/index.php/Main_Page

Sim, K. M. (2004). Toward an Ontology-enhanced Information Filtering Agent. *ACM SIGMOD Record,* Vol. 33, Issue 1, pp. 95 – 100, ISBN 0163-5808, March 2004, ACM, New York, USA

Yahoo! Inc. (2009). Documentation for Yahoo! Shopping Web Services: Product Search. http://developer.yahoo.com/shopping/V3/productSearch.html

Varelas, G.; Voutsakis, E. &Raftopoulou, P. (2005). Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web. *Proceedings of 7th ACM International Workshop on Web Information and Data Management (WIDM2005),* pp. 10 – 16, ISBN 978-1595931948, Bremen, Germany, November 2005, ACM, New York, USA

Zhang, W.; Liu, S.; Yu, C.; Sun, C.; Liu, F. &Meng, W. (2007). Recognition and Classification of Noun Phrases in Queries for Effective Retrieval. *Proceedings of 16th Conference on Information and Knowledge Management (CIKM2007),* pp. 711 – 720, ISBN 978-1-59593-803-9, Lisbon, Portugal, November 2007, ACM, New York, USA

Web Intelligence and Intelligent Agents

Edited by Zeeshan-Ul-Hassan Usmani

This book presents a unique and diversified collection of research work ranging from controlling the activities in virtual world to optimization of productivity in games, from collaborative recommendations to populate an open computational environment with autonomous hypothetical reasoning, and from dynamic health portal to measuring information quality, correctness, and readability from the web.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sayaka Akioka, Hideo Fukumori and Yoichi Muraoka (2010). An Information Filter for Intuitive and Simple Search, Web Intelligence and Intelligent Agents, Zeeshan-Ul-Hassan Usmani (Ed.), ISBN: 978-953-7619-85-5, InTech, Available from: http://www.intechopen.com/books/web-intelligence-and-intelligent-agents/an-information-filter-for-intuitive-and-simple-search

INTECH
open science | open minds