Skeleton Generation for Presentation Slides Based on Expression Styles

Yuanyuan Wang and Kazutoshi Sumiya

Abstract. With the advent of PowerPoint and Keynote that can effectively create attractive presentation slides, people can use them to exchange and discuss ideas together. However, because it is necessary to prepare many slides to enable audiences to understand the content, authors need to prepare the best possible slides. Our skeleton generation method is designed to help authors to prepare slides with ease by constructing slide layouts based on the expression styles that the level positions of words expressing their role in slides from the text in the textbooks they use. By analyzing the role of the words in the slides, our method can then extract the differences between the important elements in both the texts and slides. To generate skeletons for slides from target texts in a textbook, our method derives the expression styles of the words from pre-existing texts and their slides. Finally, it generates slide skeletons by using the same expression styles of the corresponding words from the target texts arranged in slides, which are the same as the layouts of pre-existing slides. We also present the results of an evaluation of the method's effectiveness.

1 Introduction

Presentations now play a socially important role in many fields, including business and education, among others. Many university teachers have used Web services such as SlideShare [1] and CiteSeerX [2] to store the slides they use in lectures. However, because teachers prepare many slides to enable students to understand their content, the teachers should prepare the best possible slides. In fact, when authors plan their slides often refer to texts (e.g., lectures in a textbook) to determine the information

Yuanyuan Wang
University of Hyogo, Japan
e-mail: ne11u001@stshse.u-hyogo.ac.jp

Kazutoshi Sumiya
University of Hyogo, Japan
e-mail: sumiya@shse.u-hyogo.ac.jp

T. Watanabe et al. (Eds.): Intelligent Interactive Multimedia: Systems & Services, SIST 14, pp. 551–560.
springerlink.com

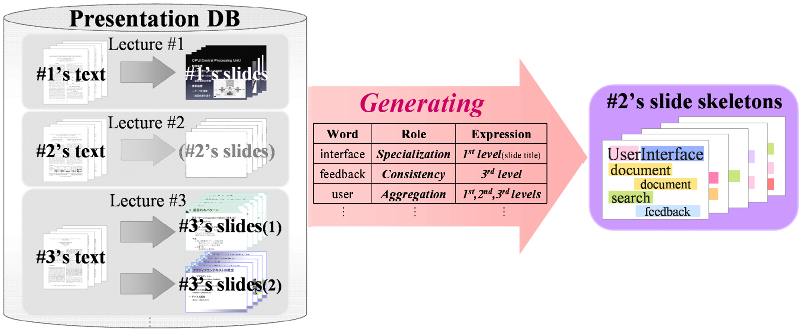

Fig. 1 Conceptual diagram of skeleton generation from textbook and its slides

should be conveyed. It is important to focus on how to express the information that will appear in slides from texts. We can generate skeletons serve as slide layouts that express typical words from the texts based on their role in slides by considering how to convey the words to create the layout of the slide. For example, a word "vegetable" appears in all the chapters of a textbook, but appears in only one slide. We consider that the role of "vegetable" is *Aggregation*, which is a concentrated summary of the information regarding "vegetable" in one slide for that textbook.

Our approach creates an editable slide skeleton that is able to produce a slide layout based on specific words to help authors prepare slides easily and efficiently. In this paper, we define the expression styles that the level positions of the words are arranged in slides for the expression of presentation, based on the role of the words in the slides by considering how each word represented in a slide differs from how it appears in text. We derived the document structure from texts by focusing on their logical units, and the document structure of slides by focusing the level of indentation in slide text that are often used to help users better organize their slide contents. As depicted in Fig. 1, when a textbook contains a number of lectures, authors can take a target text as ♯2's text to prepare slides. When the logical units that constitute ♯2's text are the same as in the pre-existing ♯1's and ♯3's texts, we can detect the expression styles of words in the slides by analyzing the differences between the pre-existing texts and their slides as input. We can therefore generate skeletons for ♯2's slides from ♯2's text, based on the expression styles of the words.

We found that there were two main features particularly helpful for deriving expression styles, based upon the differences between important elements by analyzing the document structure of texts and their slides: (1) When a word appears within the body of a text, it is an important word in the text; and when a word appears in the slide title or in lines that are less indented, it is an important word in slides [3]. (2) When a word occurs with high density in a certain passage of a segment text, which is an important description of the word in the text [4]. Also, when a number of sentences appear in lines that are deeply indented, they are an important description of a word in a slide. Therefore, we can generate skeletons for slides from a target text based on the expression styles of words by extracting the differences between the important elements of pre-existing texts and their slides that are in such a textbook.

The next section reviews related work. Section 3 describes how to determine key elements in texts and slides. Section 4 presents the generation of skeletons for slides. Experimental results and conclusions are given in Sections 5 and 6, respectively.

2 Related Work

Most of the research related to slide-making support has focused on slide generation. Mathivanan et al. [5], Beamer et al. [6] and Yasumura et al. [7] proposed a system for generating slides from academic papers. Their method extracts information from a paper by the TF-IDF method, and assigns the sentences, figures and tables in slides by identifying important phrases for bullets. Shibata et al. [8] converted Japanese documents to slides representation by parsing their discourse structure and representing the resulting tree in an outline format. However, conventional approaches that focus only on the consistency of the document structure in the text and slides, both ignore the role played by how to express words from the text to the slides. Our method focuses on the differences between the key elements of texts and their slides, and it generates skeletons for slides based on the expression styles of words.

Kan [9] proposed a system for the discovery, alignment, and presentation of such document and slide pairs. Hayama et al. [10] aligned academic papers and slides based on Jing's method, which uses a hidden Markov model. These studies are similar to ours for analyzing information that is common to texts and their slides. Our approach focuses not only on the information that is in common, but also on information that differs between texts and slides. Yokota et al. [3] can retrieve important information in slides is similar to ours. Kurohashi et al. [4] detected important descriptions of a word in a text. Their method is based on the assumption that the most important description of a word in a text is the passage where the word occurs with the highest density. We have employed the same method for detecting important descriptions of a word in a text. Therefore, our goal is to generate skeletons for slides by analyzing the differences between important elements of texts and their slides.

3 Determination of Important Elements Using Document Structures

We determine important elements by calculating the distribution of words based on the document structure in the text and then taking the document structure in slides. A chapter in a textbook is referred to as a text. We define the document structure of a text in terms of its logical units, which consist of sections, which in turn consist of a section head and paragraphs. The content of a presentation includes a number of slides that have structured text information. We define the document structure from slides, based on the indentations in the slide text. We define the slide title as the 1st level. The first item of text is considered to be on the 2nd level, and the depth of the sub-items increases with the level of indentation (3rd level, 4th level, etc.).

3.1 Determination of Important Elements in a Text

If the location in which a word b appears in the text is dispersed, b is deemed an important word in the text; it is called W_t. We explain the determination of W_t using b, and we calculate the degree of importance of b by the words dispersion.

$$W_t = \{b|min\Big(\frac{\sum_{u=1}^{n} dist(c_1, b_u)}{n}, \ldots, \frac{\sum_{u=1}^{n} dist(c_j, b_u)}{n}\Big) > \alpha\} \tag{1}$$

Where b_u is the u^{th} word b, and c_j is the j^{th} section in the text. Function *dist* calculates the distance between sections, that is a number indicates how many sections there are between two words. n is the number of times that b appears in a text. The highest degree of expectation is obtained with the lowest dispersity by using function *min*. W_t is a bag of important words in the text, if the formula is greater than a threshold α in Eq. (1), and b is determined to be the important word in W_t.

If a word m occurs in a high density in a certain range of a text segment in the text, the text segment is therefore considered an important description of m in the text, and it is called D_t. When the density of m in a text segment is high, it is determined that the text segment of m provides D_t of m in the text. We define the position i of m, and define l as a center position, and the range from l to the anteroposterior position is w as a certain range in the text. To calculate the density of m, we use the hanning window function [11] to decrease the weight of the words in the range from l to $l-w$, $l+w$. The density of m on l in the range of $|i-l| \leq w$ can be calculated as

$$D_t = \{m|\sum_{i=l-w}^{l+w} a_m(i) \cdot \frac{1}{2}(1 + cos2\pi\frac{i-l}{2w}) > \beta\} \tag{2}$$

The part of formula $\frac{1}{2}(1 + cos2\pi\frac{i-l}{2w})$ is a hanning window function. The function $a_m(i)$ indicates whether the word in l is m. If m is in l, $a_m(i)$ returns 1; otherwise, $a_m(i)$ returns 0. Here, the location starts from 0 (the head of the text segment), followed by each position as l of the hanning window, in order. For the number of sections in the text or words in each section is not consistent. Therefore, we set the range of windows ($2w$) for the average of number of words in each section.

3.2 Determination of Important Elements in Slides

If a slide has more information in terms of a word g than is contained in a prior slide in the presentation, g is thus an important word in the slides, and it is called W_s. We explain the determination of W_s using g, which is present in both slides x and y.

$$K_l(x,g) = \{k_i|k_i \in x, l(x,g) < l(x,k_i)\} \tag{3}$$

Here, $K_l(x,g)$ is a bag of words that can be considered to provide an explanation in terms of g in slide x. $l(x,g)$ is a function returns the level of g in slide x. The word k_i is included in the levels that have a hierarchical relationship with the level of g, and k_i belongs to $K_l(x,g)$ in slide x. $l(x,k_i)$ is greater than $l(x,g)$, in that k_i is a child of g in the document structure. Then, we compute the number of words in detailed information related to g for slides x and y, and compare their numbers as follows:

$$W_s = \{g||K_l(x,g)| < |K_l(y,g)|\} \tag{4}$$

where the function $|K_l(x,g)|$ extracts the total number of k_i in $K_l(x,g)$ in slide x. $K_l(y,g)$ are also bags of words in slide y, and they satisfy the same conditions as $K_l(x,g)$ in Eq. (3). W_s is a bag of important words in the slides, and if $|K_l(x,g)|$ for slide x is lower than $|K_l(y,g)|$ for slide y in Eq. (4), g is then determined to be an important word in W_s.

If a number of sentences in lines are indented deep in the level indentation of a word d, these sentences is an important description of d in the slides, and it is called D_s. When d and other words in slide x satisfy certain conditions, the lower levels of sentences $L_s(x,d)$ of d is determined to be D_s of d.

$$D_s = [d, L_s(x,d)] \tag{5}$$

$$L_s(x,d) = \{r_s|l(x,d) \leq l(x,r_s)\} \tag{6}$$

A set $L_s(x,d)$ consists of sentences from levels related to d in slide x. Sentence r_s belongs to $L_s(x,d)$ in slide x if r_s must be included in one of the indentation levels. Additionally, $l(x,r_s)$ is greater than or equal to $l(x,d)$, and words of r_s are children of d or the words of r_s and d are brothers in the document structure, $L_s(x,d)$ will also extract sentences containing d from levels from $l(x,r_s)$ to $l(x,d)$.

4 Skeleton Generation

4.1 Detecting Expression Styles

To generate skeletons, a slide layout is used, which consists of words based upon expression styles by the role of the words using the differences between the important elements in the pre-existing text and their slides. For the differences between the importance of word q in the slides and the text, which falls into 3 categories:

- tw_1: $q \in W_t \cap W_s$, q is an important word in both the text and the slides.
- tw_2: $q \in W_t$, q is an important word in the text.
- tw_3: $q \in W_s$, q is an important word in the slides.

For the differences between important descriptions of a word that appear in the text and slides, we compute the similarity of the bag of words in important descriptions of q, D_t in the text and D_s in slides. This is done using the Simpson similarity coefficient [12] as $Sim(D_t,D_s)=\frac{|D_t \cap D_s|}{min(|D_t|,|D_s|)}$. We consider that whether the text and slides contain one or multiple important descriptions of q, based upon their similarity, they

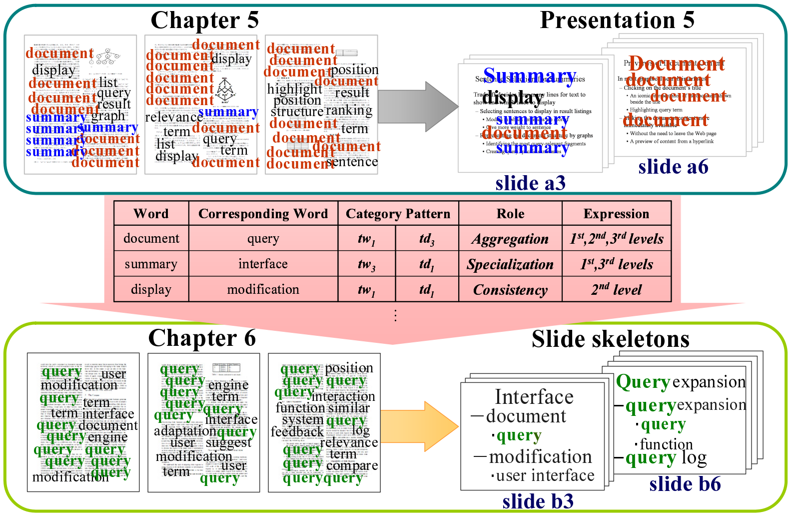

Fig. 2 Example of skeleton generation

fall into 6 categories. When $Sim(D_t, D_s) \geq 0.7$, the content of the important descriptions of q in the text and in the slides are similar, and there are 3 categories:

- td_1: one (multiple) descriptions of q in D_t corresponds to one (multiple) descriptions of q in D_s.
- td_2: one description of q in D_t corresponds to multiple descriptions of q in D_s.
- td_3: one description of q in D_s corresponds to multiple descriptions of q in D_t.

When $0.3 \leq Sim(D_t, D_s) < 0.7$, the common content of the important descriptions of q in the text and in the slides are not similar, which falls into 3 categories:

- td_4: one (multiple) descriptions of q in D_t has information in common with one (multiple) descriptions of q in D_s.
- td_5: one description of q in D_t has information in common with multiple descriptions of q in D_s.
- td_6: one description of q in D_s has information in common with multiple descriptions of q in D_t.

We can find what words are emphasized, and how the words should be described in the text and the slides, whether multiple descriptions are dispersed, or one description is centered from the differences between important elements in the text and the slides. In the example shown in Fig. 2, the word "document" is dispersed in all sections in Chapter 5, with some text segments having a high density of "document," and it also appears frequently in the body of text in slide $a6$ of Presentation 5. When "document" is an important word in both the text and slides as tw_1, multiple important descriptions in the text correspond to one important description in the slides as td_3. We consider that slide $a6$ is concentrated when it summarizes the

Table 1 Patterns in the role of words in slides

P	td_1	td_2	td_3	td_4	td_5	td_6
tw_1	*Consistency*	*Separation*	*Aggregation*	*Portion*	*Partial Separation*	*Partial Aggregation*
tw_2	*Generalization*	*Dispersion*	*Unification*	*Mention*	*Separable Mention*	*Centered Mention*
tw_3	*Specialization*	*Subdivision*	*Concentration*	*Expansion*	*Separable Expansion*	*Centered Expansion*

information in terms of "document" in Chapter 5, and the role of "document" will be *Aggregation*. On the other hand, when the word "summary" repeatedly appears in a certain text segment that has a high density of "summary", slide $a3$ is titled "summary" of Presentation 5. When "summary" is an important word in slides as tw_3, and one important description in the text corresponds to one important description in the slides as td_1. Slide $a3$ offers specialized information regarding "summary" from Chapter 5, and the role of "summary" is then *Specialization*. Therefore, we define the expression style ES that the role R of words with the expression E of presentation is represented by the level positions of the words in slides as follows:

$$ES = (R, E) \tag{7}$$

$$R = (w_i, p_{w_i})(w_i \in W, p_{w_i} \in P) \tag{8}$$

$$W = W_t \cup W_s \tag{9}$$

$$P = \{p_{w_1}(tw_1, td_1), \cdots, p_{w_6}(tw_1, td_6), \cdots, p_{w_{13}}(tw_3, td_1), \cdots, p_{w_{18}}(tw_3, td_6)\} \tag{10}$$

here, W is a bag of words that belongs to W_t or W_s that can be considered as the words that play key roles in the slides. E denotes the level positions of the words in slides by the role of the words in R, and P denotes the total of 18 patterns that intend the role of the words in R, and the words belong to W. These patterns combine 3 categories of differences in the important words and 6 categories of differences between the important descriptions of the text and slides that are shown in Table 1.

4.2 Generating Skeletons for Slides

Based upon the expression styles drawn from pre-existing texts and slides, we can generate skeletons for slides from a target text in a textbook by extracting the word in the target text that corresponds to the words in pre-existing texts. We consider texts in which the chapters in a textbook have the same document structure as the sections in each chapter. When the frequency of a word z in all sections of the pre-existing text T_a and the frequency of a word z' in all sections of the target text T_b have the same tendency, we consider that z' corresponds to z.

For each word, we rank the sections in terms of its frequency by calculating the Spearman's rank correlation coefficient $R(f(z, T_a), f(z', T_b))$ for the correlation between the section rankings $f(z, T_a)$ of z in T_a and $f(z', T_b)$ of z' in T_b. Based on the above criteria, we extract a pair C_p of z in T_a and z' in T_b as follows:

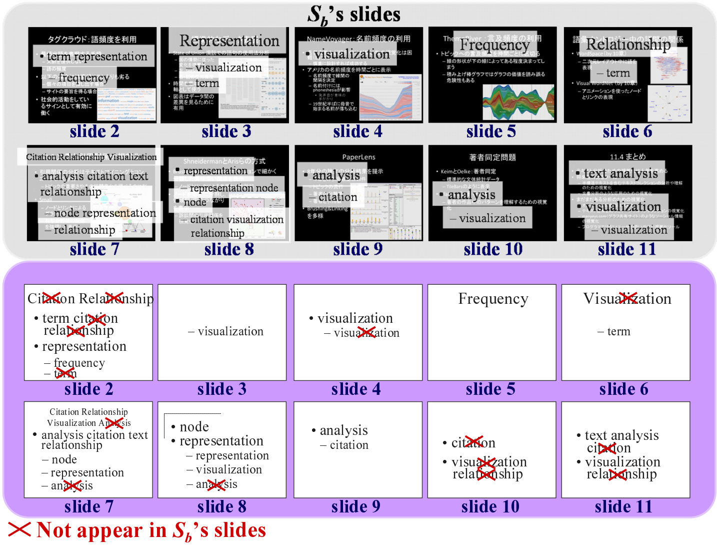

Fig. 3 Generated slide skeletons compared with slides in S_b

$$C_p = \{(z, z')|R(f(z, T_a), f(z', T_b)) > \gamma, z \in W\} \tag{11}$$

if $R(f(z,T_a), f(z',T_b))$ is greater than a threshold γ that is near to 1 in Eq. (11), z' is determined to be the corresponding word of z. Therefore, we are able to generate skeletons for layout slides by using the expression style of z' in the same expression style as z, which is performed according to Eqs. (7), (8), (9) and (10), and the number of skeletons for slides is the same as the number of pre-existing slides.

For example, an author wants to make slides for a lecture regarding Chapter 6 in a textbook. Our method generates skeletons for slides from Chapter 6, referring to Presentation 5 from Chapter 5 (see Fig. 2). In Chapter 5 the word "document" appears in all sections, and it occurs in high density in some certain ranges of the text segments. Meanwhile, if "document" appears frequently in slide $a6$ only in Presentation 5, then the role of "document" is *Aggregation*. In Chapter 6 the word "query" appears in all sections that correspond to "document" in Chapter 5. The skeleton for slide $b6$ generated from Chapter 6 shows that "query" appears frequently in slide $b6$, which explains "query expansion" in terms of "query." Next, "query" in slide $b6$ has the same role as "document" When the author makes slides referring to the skeletons for slides, such as slide $b6$, the information for "query" in slide $b6$ is constructed in the same way as it is for the level positions of "document" in slide $a6$, based upon the same expression style. The generated skeletons can be used to create layout slides that construct words according to the same roles the words play in pre-existing slides, and these skeletons then enable the author to make slides easily.

5 Evaluation: Validity of Generating Skeletons

The aim of this experiment was to verify whether our method is useful for generating skeletons for slides. We first prepared two presentation files: S_a from text T_a and S_b from text T_b were made by the same person, both from Chapter 11 in a textbook called Search User Interfaces [13]. Because of their single authorship, S_a and S_b both have the same expression styles, and T_a and T_b have the same document structure. Each presentation file contains 10 slides, not counting the cover slide. We used T_a and S_a to generate skeletons from T_b based on our method; the slides in S_b serve as correct answers regardless of whether the level positions of the words in the slides generated from skeletons from T_b are correct or not.

First, we extracted the expression styles of 14 important words in T_a and S_a and of 9 words in T_b, which correspond to 8 words of the 14 important words in T_a, based on our method. There were 40 level positions of 8 words from T_a that are in S_a. Next, we generated 10 slide skeletons from T_b with the same number of slides as in S_a, and 40 level positions of 9 words from T_b were arranged in slide skeletons based on the expression styles of the 8 corresponding words in T_a. Finally, we compared them with the correct answers as S_b's slides (see Fig. 3).

In the experimental results, the correct rate of the level positions of words in slides by the generated skeletons based our method was 62.5%(25/40), and the correct rate of the expression styles of the words was 66.7%(6/9). The result for the skeleton generation was low, and it was dependent upon the expression styles of the words that were arrayed in the slides. For example, our method determined the expression style of a word that has one important description in S_a; however, we used the same expression style for the corresponding word has multiple important descriptions in the correct answer S_b. In addition, we need to consider the figure captions for determining the important elements in the text. S_a and S_b, which were written by the same person, contain a number of important words in slides, and they appear in figure captions in the texts. However, those words in the body of the text that appear once cannot be determined the important word by our method.

This experiment showed that our method can arrange the words in slides using generated skeletons based on their expression styles. However, our method could not extract the corresponding words by using the frequency of each word in all sections in T_a and T_b, when some words appeared frequently in one section only, or when some words appeared just once in one section. This was one of the reasons why the rate of correct responses was low. Therefore, these corresponding words in the target text that is used for generating skeletons also need to be considered.

6 Concluding Remarks

In this paper, we proposed a method of skeleton-generation that provides support for making slides based on the expression styles of words. We described in detail how to expression styles are determined by extracting the patterns that combine the differences between the important words and the important descriptions of words in texts and slides, respectively. To generate skeletons for slides from a target text, we extracted the words in the target text that correspond to the words in pre-existing text, and we then used the same expression styles of the words in the target text.

In the future, we plan to improve our algorithm for skeleton generation and to evaluate it using a large set of actual presentation data. We also plan to enhance our method for extracting corresponding words based on the document structures of texts, not only in terms of sections but also in terms of paragraphs in a section.

References

1. SlideShare, `http://www.slideshare.net/`
2. CiteSeerX, `http://citeseer.ist.psu.edu/index`
3. Yokota, H., Kobayashi, T., Okamoto, H., Nakano, W.: Unified contents retrieval from an academic repository. In: Proc. of International Symposium on Large-scale Knowledge Resources (LKR 2006), pp. 41–46 (March 2006)
4. Kurohashi, S., Shiraki, N., Nagao, M.: A Method for Detecting Important Descriptions of a Word Based on Its Density Distribution in Text. IPSJ (Information Processing Society of Japan) 38(4), 845–854 (1997)
5. Mathivanan, H., Jayaprakasam, M., Prasad, K.G., Geetha, T.V.: Document summarization and information extraction for generation of presentation slides. In: Proc. of International Conference on Advances in Recent Technologies in Communication and Computing (ARTCOM 2009), pp. 126–128 (October 2009)
6. Beamer, B., Girju, R.: Investigating automatic alignment methods for slide generation from academic papers. In: Proc. of the 13th Conference on Computational Natural Language Learning (CoNLL 2009), pp. 111–119 (June 2009)
7. Yoshiaki, Y., Masashi, T., Katsumi, N.: A support system for making presentation slides. In: Transactions of the Japanese Society for Artificial Intelligence, pp. 212–220 (2003) (in Japanese)
8. Shibata, T., Kurohashi, S.: Automatic Slide Generation Based on Discourse Structure Analysis. In: Dale, R., Wong, K.-F., Su, J., Kwong, O.Y. (eds.) IJCNLP 2005. LNCS (LNAI), vol. 3651, pp. 754–766. Springer, Heidelberg (2005)
9. Kan, M.: Slideseer: A digital library of aligned document and presentation pairs. In: Proc. of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2007, pp. 81–90 (2007)
10. Hayama, T., Nanba, H., Kunifuji, S.: Alignment between a technical paper and presentation sheets using a hidden markov model. In: Proc. of the 2005 International Conference on Active Media Technology (AMT 2005), pp. 102–106 (May 2005)
11. Blackman, R.B., Tukey, J.W.: Particular Pairs of Windows. In: The Measurement of Power Spectra, From the Point of View of Communications Engineering, pp. 95–101. Dover, New York (1959)
12. Simpson, E.H.: Measurement of diversity. Nature 163, 688 (1949)
13. Hearst, M.A.: Search user interfaces, pp. 281–296. Cambridge University Press (November 2009)