

SERP-level Disambiguation from Search Results

Mostafa Alli

Dept. of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Keywords: Search Engine Result Page Generation, Search Snippet Generation, Visual Search Snippet, Ambiguous Search Query, Visuality, Thumbnails, Disambiguation.

Abstract: Fast growth of search engines' popularity shows the users attraction to the Web engines. However there is a chance of misinterpretation for ambiguous queries. At this point, we propose a more adherence user interface which consist of a relevant visual content as well as generating new search snippet and title. Recent researches for meeting this aim are focused on a whole page thumbnail for assisting users to remember a recently visited Web page. Withal, this is not discussed yet that how a specific visual content of a page can allow users to distinguish between a useful and worthless page in the result page especially in an ambiguous search task. Our study^a shows that the improvement in both textual search snippet and title as well as the additional thumbnail were helpful for users to clarify the *Search Engine Result Page (SERP)* in an ambiguous search task.

^awhich consists of two parts

1 INTRODUCTION

Dramatic changes in search engines' usage (Battelle, 2005; 2012,) show the need of improvement for users' experience during a search session especially for an ambiguous query. Ambiguity in a search query is the case that search keyword has more than one underlying meaning. This is known (Song et al., 2007) that around 16% or (Mihalkova and Mooney, 2008) 7% to 23% of all search queries have ambiguity. That means 1 out of every 5 to 6 search tasks are ambiguous. Also studies (Sanderson, 2008; Beitzel et al., 2004; Jansen et al., 2005; Fang et al., 2011) show that the most ambiguous queries have the length of one, yet there is a huge quantity of ambiguity for multiple keywords as well as the average length for queries is less than 2.3. These results show that first of all, users commonly perform search with short keywords. This is proven by (Agrawal et al., 2009a) that users mostly underspecify their search intensions, i.e, they usually do not search queries with all needed information to clarify its actual intent. While it is expected that adding more keywords to the search query would make it more clear, there is a considerable number of multiple-word ambiguous queries (Sanderson, 2008).

These actual facts encouraged us to propose a better representation for the *Search Engine Result Page(SERP)* in order to lower the effort of users for

refining the intended ambiguous search query, e.g, for the keyword *China Times*. The search results are including the Web site of English newspaper, *The China Times*, the Web site of newspaper, *China Times*, current time of China, The Web site of English newspaper, *China Post*, Web site of newspaper, *Chinese Times* and the Taiwanese news agency, *Want China Times*. By adding the word *Magazine* or *Newspaper* to the search query, the results are still quite varied and even more ambiguous and hence, difficult to make a right decision.

To tackle this, there are different approaches. For instance, *Incremental Keyword Extension* is a way to increase users' satisfaction. Regardless of effectiveness of such techniques, this can be argued that it still needs users' data engagement which brings privacy issues and as we described previously, users' query length is short and there are multiple-word ambiguous queries. At this point, we introduce a novel presentation for SERP to help to clarify the result page, which gives users more insight of each result's page content by extracting the most relevant visual content of the page as well as improving the textual part of the search snippet and improving its title. The remaining of this paper is as followed, in section 2, we demonstrate three motivating examples, in section 3, we describe the related work in this area, followed by our method expansion in section 4. In

section 5, we present our experiment and finally in section 6, we give a brief conclusion to the current work and discuss possible future work.

2 MOTIVATING EXAMPLES

Example 1. Google and Bing are using a *Knowledge graph* and a *Knowledge base* called *Satori*¹ respectively for entity-based queries as an interface in SERP. A disambiguation box is a part of these graph and base which tries to distinguish between different possible taxonomies that a query may carry. Surprisingly, these boxes are insufficient and will not present even all the different topics that the search engine already returned in the first page. To illustrate this, by requesting the ambiguous query, *Kingfisher* in both Google and Bing, we will notice that the disambiguation boxes, which are illustrated in Figure 1 are not comprehensive at all. Disambiguation boxes of Bing and Google suggested 3 taxonomies to the users for disambiguation while there are more topics returned in the first page. In case of Google, the concealed topics are 1) a restaurant 2) a beverage company 3) an IT-specialized company 4) a seafood restaurant and 5) a theme park while in case of Bing, concealed topics are: 1) a fly shop 2) a restaurant and 3) a boat seller.

Example 2. To enrich the search snippet, the *SearchMonkey* was launched by Yahoo! by May 2008 (Haas et al., 2011). The aim of this service was to assist the site owners and developers to make Yahoo! Search results look more visually appealing and useful, by **sharing structured data** with Yahoo!. The main pitiful for this service is that this is a site-specific way for enriching the SERP and requires the page designers/owners to get involved to the process of the enriching the textual search snippet and- unlike our method- is only helping Yahoo! users to benefit from it, if there is any. However, the service was shut down by October 2010.

Example 3. Search engines usually suggest a set of *Synsets* for disambiguation. Nonetheless, this can be argued that this method will not be helpful all the time. This is mentioned previously that study (Sanderson, 2008) shows the most ambiguous queries have the length of one, yet there is a huge quantity of ambiguity for multiple-word queries as well. Furthermore, a search log analysis (Jansen et al., 2005) indicated that around 50% of query's length is either 1 or 2. Yi Feng et al (Fang et al., 2011) measured that the query length for Web search is

¹<https://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing/>

2.35. A large-scaled search log analysis (Beitzel et al., 2004) showed that the average length for popular queries is 1.7 and for all queries is 2.2. These results show that the offered synset of search engines were not completely successful. It shows also that expecting users to add more words to the query for narrowing the scope of a search query is not very optimistic. Besides, even multiple-word queries have ambiguity. This hypothesis is supported by (Agrawal et al., 2009a) that most users do not narrow down they search intention.

3 RELATED WORK

Attempts for clarifying a search result page in an ambiguous search task falls into 3 main categories:

3.1 Diversifying Search Results

One way to tackle the ambiguity in a Web search is to diversify the results. Consequently it can show more possible pages from different taxonomies to the user. As a result, the user have a higher chance to find the intended information.

D.Yin et.al (Yin et al., 2009) examined a new diversify method for finding out the subtopics of each ambiguous query to re-rank the result page by predicting the user intention. To do this, they discover the subtopics from other similar issued queries and then estimating users' intent based on a probabilistic ranking algorithm. R.Agrawal et al (Agrawal et al., 2009b) introduced a greedy algorithm to diversify the search result in order to overcome the *underspecified* search query problem. Their objective is to maximize the chance for users to at least find one result, related to their query within the first k results that a search engine returns. A ranking method (Zhang et al., 2005) is proposed for enhancing diversity and information richness, called, *Affinity Ranking(AR)*. This algorithm's job is to create an Affinity graph which is made from the link structure of each document, measuring each document diversity and information richness and then combining both together to get a diversity penalty score. Afterward, based on this score, they re-rank the whole documents to meet a richer and more diversified top search result.

Although diversifying the results will make the result page more probable to contain users' true intention, yet, lack of visuality for each results itself, makes it hard to figure out right one(s) from faulty one(s).

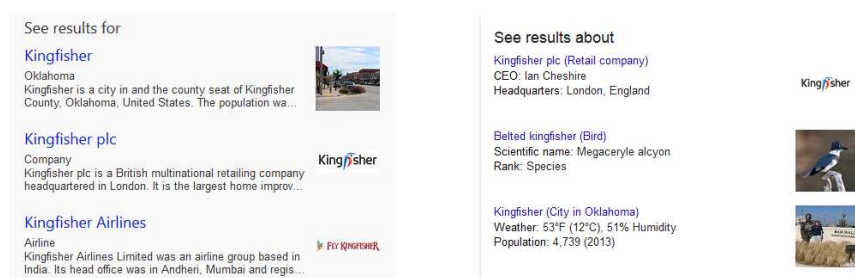


Figure 1: Disambiguation boxes of Bing (on the left) and Google (on the right) for the ambiguous query Kingfisher.

3.2 Web Search Personalizing

Personalizing web search tasks, reacquires to log users' web activity to predict one's intention when an ambiguous search query happens. This method usually require a long log of users' search history.

L.Mihalkova et al (Mihalkova and Mooney, 2008) proposed a short session logging for personalizing the search task. J.T Sun et al (Sun et al., 2005) proposed a *Clickthrough* system called *CubeSVD* which tries to use the data collected from a real-world data set from MSN search engine. The system is proposed to use a Higher-Order Singular Value Decomposition (HOSVD) which its input is the clickthrough data and output is the value that shows the association between users, queries and Web pages.

Besides the throughput of personalization of Web searches, even if a user requesting an ambiguous keyword, e.g, Jaguar, multiple of times, this does not necessarily mean that his/her "intension" was the same all the time. In addition, it brings the privacy issues for logging users' search activity.

3.3 Word Sense Disambiguation (WSD)

Word Sense Disambiguation (WSD) is the ability to distinguish between different meaning of a word in a document when it has different concepts, based on its context. This is knows that WSD is a Natural Language Processing(NLP)problem together with Artificial Intelligence(AI) and Ontology (Mallery, 1988)(Mallery, 1994).

C.Stokoe et al (Stokoe et al., 2003) introduced a sense based disambiguation system for information retrieval to compete with the term based system over ambiguous search query tasks. Their technique is taking advantage of combining high precision technique and sense frequency statics to build a more accurate retrieval system. Hence, they used the co-occurrence and collocation principles to implement an algorithm which produces 3 different implementation of their (WSD) system, namely, Sense Based(T)-Sense Query , Sense Based(S)-Sense Query and Stem Based which

is the Traditional $TF*IDF$ technique using stem words.

There are various difficulties with a WSD system , e.g. Deciding what senses are belonging to a word is usually a difficult task. In addition, the word sensing will be so dependent to the word directory that a system may use.

4 SEARCH ENGINE RESULT PAGE GENERATION

In This section, we look at our proposal for SERP generation based on a more coherence and richer textual search snippet and page title, as well as a visual feature. Hence, we first explain how we do the *search snippet generation* and then we discuss the process of relevant visual detection.

4.1 An Improvement to the Textual Property of Returned Listing

Background. A current study (Marcos and Pérez-Montoro, 2009) indicates that the textual search snippet has a higher weight compare to the title or the page URL to the users for deciding wether or not to click on a link. We used this finding as a motivation to generate new search snippet for a higher chance of correct page selection in an ambiguous search session. However we believe that the current search snippet and page title that search engines are providing, is not comprehensive enough. Presently, search engines create the search snippet and page title using *meta data* provided by page authors or the *DMOZ directories*. This makes the snippet usually discrete and vague for the fact that author of web pages usually try to attract the search engines, using *Keyword stuffing* (Chandra et al., 2014), which is done by repeating some common keywords in the HTML meta tags and by providing a long title for the page.

4.2 Search Snippet Generation

For improvement of the textual property of each search results, we think that the snippet should always comes from the content of the page and hence, it will be consistent. Our idea is to use a segmentation method to divide page into logically different paragraphs². To make segmentation, we take advantage of building a DOM tree over the HTML tags of the page and then using a Postorder traversing method to go through the DOM nodes. By considering each group of leaf sibling node as a segment, we measure its relevancy degree with the query.

To do so, we practice the Naive Baye's classifier. The classifier is based on the Bayesian Theorem which is a probabilistic theory and hence, the classifier becomes a probabilistic classifier. We can demonstrate the simple statement of the Bayesian theorem in Equation 1:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

By re-writing the Equation 1, we can then obtain equation 2:

$$P(Seg_i|Keyword) = \frac{P(Keyword|Seg_i)P(Seg_i)}{P(Keyword)} \quad (2)$$

where Seg_i ($1 \leq i \leq n$) has a prior probability $P(Seg_i)$, $P(Seg_i|Keyword)$ is Seg_i 's posterior probability given Keyword $Keyword$, and $P(Keyword|Seg_i)$ is the conditional probability of $Keyword$ being seen in Seg_i . This can be said that the *Posterior* for $Keyword$ is equal to a fraction of *likelihood* multiplied by *prior* divided by *evidence*. Since in practice the evidence is a constant. This is same as if we say in Equation 2, as the denominator $P(Keyword)$ is independent of Seg_i , and $P(Seg_i)$ remains the same for all keywords, the likelihood that a search $Keyword$ appears in Seg_i , $P(Keyword|Seg_i)$, dominates the posterior probability $P(Seg_i|Keyword)$. so we just need to calculate the prior and likelihood to find the probability of Seg_i being chosen regarded to the $Keyword$. This can be done by using the Naive Bayesian Classifier. With $mean_i = \mu_i$ and $variant_i = \sigma_i$ of all the words' $tf*idf$ values in Seg_i , $P(Keyword|Seg_i)$ in Equation 2 can thus be approximately measured through a Normal Distribution $\mathcal{N}(\mu_i, \sigma_i^2)$:

$$P(Keyword|Seg_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \times e^{-\frac{(tf(Seg_i, Keyword) + idf(Seg_i, Keyword) - \mu_i)^2}{2\sigma_i^2}} \quad (3)$$

²A paragraph here is refereing to a group of sibling leaf nodes of Data Object Model(DOM) tree of the corresponding page

Here, we apply the stemmed $tf*idf$ principle. One way to calculate $tf(Seg_i, Keyword) * idf(Seg_i, Keyword)$, is illustrated in the Equation 4:

$$\log\left(1 + \frac{N_{Keyword}}{N_{total}}\right) * \log \frac{N_D}{N_{KD}} \quad (4)$$

Where $N_{Keyword}$ is equal to the frequency of the keyword and N_{total} is equal to the total frequency of all words, including the corresponding keyword, N_D is equal to the total number of page segments and N_{KD} is equal to the number of segments that the keyword occurrence happens there. On the other hand, $idf(Seg_i, Keyword)$ is computed as the inverse of the total number of $Keyword$ appearing in all parts. In this way, we can obtain the *mean_i* and *variance_i* of keywords' $tf*idf$ values in Seg_i by averaging all the $tf*idf$ values of $Keywords$ in $part_i$ and their deviation from the mean value. In Equation 2, as the denominator $P(Keyword)$ is independent of Seg_i , and $P(Seg_i)$ remains the same for all keywords, the likelihood that a search $Keyword$ appears in Seg_i , $P(Keyword|Seg_i)$, dominates the posterior probability $P(Seg_i|Keyword)$.

Given a user's search request containing m keywords $\{s_1, \dots, s_m\}$, for each $part_i$ of the page, we calculate the average likelihood $\sum_{i=0}^n L(part_i, s_1, \dots, s_m) = \sum_{i=0}^n P(part_i|s_1, \dots, s_m)/m$, and pick up $part_k$ with the highest likelihood value ($arg \max_k L(part_k, s_1, \dots, s_m)$).

4.2.1 Discussion

Question: Why does the keyword conditional probability follows a Normal (Gaussian) distribution?

Answer: For the weighting factor for reflecting the importance of words in a page, we used the TFIDF principle which is fundamentally equal to $tf * idf$. To calculate tf , we can use the Equation 4. On the other hand, this is shown (Bruls et al., 1999; Baayen, 1991) that the word frequency follows the *Log-normal distribution*. By considering the Equation 4, we can see that for calculating the value of $tf(Seg_i, Keyword)$, we apply the logarithm of the term frequency and hence, the conditional probability, $P(Keyword|Seg_i)$, follows a normal distribution.

4.3 Title Generation

Based on the users feedback in the first part of our user study, we decided to make the textual search snippet more coherence with its title, by generating a new title. To do so, when we extract the text as we described in previous section, we get the parent HTML tag of the extracted text fragment. Accordingly, we

look into its siblings³ textual content, one by one. By applying FIFO principle, the first sibling that has a textual content will be regarded as the new title for the improved textual search snippet. In case of no textual content in the tags, we will select the parent's parent's tag and then, we check its sibling in same manner until we get a result. In other words, we generate the title of the result by using subsection headers near the occurrence of newly generated search snippet⁴.

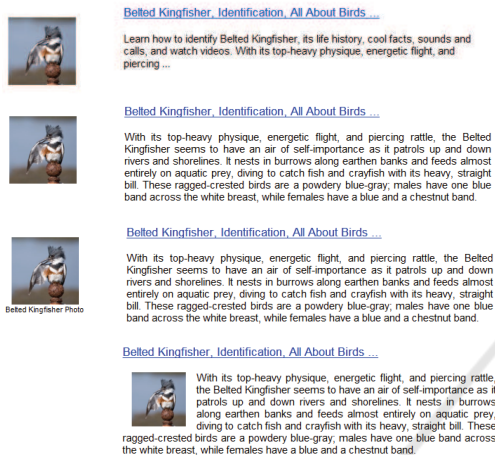


Figure 2: The four different variations of our proposed method for the SERP generation for an ambiguous search query, used in first part of the user study. from top to bottom: "Thumbnail with default snippet"; "Thumbnail with improved snippet"; "Thumbnail with caption" and "Embedded thumbnail"

4.4 Visual Relevance Detection

In addition to the previous discussion on importance of search snippet generation, here we show that how a visual factor can play a vital role for relevant page detection and then, we explain our methodology for relevant visual extraction for assisting disambiguation task.

4.4.1 Importance of Visual Features for Relevant Page Detection

Recent researches, are focused on refining previously visited pages based on a combination of textual and visual factors, or a visual factor alone. The visual factor is usually a thumbnail which is based on a snapshot of overall page lay out and the text is mostly the current default snippets provided by search engines. Withal, this is not discussed that how a specific visual content of a page can allow users to

³Siblings are adjacent tags in HTML.

⁴in the DOM tree

differentiate between a relevant and irrelevant page in the result page, especially for a searched ambiguous keyword. We believe that we are first to suggest using a visual factor for clarification from ambiguity of a query. Previous studies (Czerwinski et al., 1999; Dziadosz and Chandrasekar, 2002; Kaasten et al., 2002; Robertson et al., 1998; Teevan et al., 2009) show the importance of visual contents to help users for distinguishing between different Web pages. These results encouraged us to use an appropriate visual content to assist users to distinguish between relevant and irrelevant pages of their ambiguous query.

M.P Czerwinski et al (Czerwinski et al., 1999) ran a study to measure the effect of a thumbnail⁵ preview of a page with and without textual information for users to spot the previously visited Web sites. They concluded that a visual representation can improve the ability of remembering and distinguishing one item from another. Similarly, S.Susan Dziadosz et al (Dziadosz and Chandrasekar, 2002) compared three different versions for a SERP interface namely, text only, thumbnail only and a combination of text and thumbnail. Results of their work shows that the combination of both thumbnail and textual representation of a page can boost the accuracy of relevance decision-making for users compare to the other two.

A Session Highlight web workspace (Jhaveri and Raiha, 2005) was introduced to assist users during a web session. This is done by providing a drag and drop tool for any desired pages that a user may decide to bookmark and upon this action, a thumbnail of the same page would be added to the Workstation. They claimed that this tool was being used effectively by participant of their user study. S.Kaasten et al (Kaasten et al., 2002) has investigated the usefulness of page representation for users to spot the previously visited sites by comparing different size of a thumbnail together with a various size of titles and URLs. Accordingly, they suggested how to design a bookmark or history list for best fitting the users' chance of revisiting previously visited web sites. Besides the importance of the size of the URL and title for their recognition, they also concluded that a thumbnail will help users to accurately recognize the previously visited pages.

Data Mountain (Robertson et al., 1998) was suggested as a 3D alternative representation to enhance the user chance for retrieving Web pages. This Data mountain were meant to show a thumbnail of any

⁵In related work, a thumbnail is referring to a "Whole page thumbnail" while in our method, a thumbnail is an "extracted photo" from the content of the page regarded to the search query.

documents that user desires to place in an arbitrary position of a virtual 3D desktop. They mentioned that this data mountain rapidly fasciated the need of extracting pages. J.Teevan et al (Teevan et al., 2009) introduced a new way of page representation which is based on both textual and visual aspect of a page and together they made a new visual snippet that gives an overall preview of a page content. Their aim was to both improving the user ability to find a new relevant web page and re-finding of a previously seen page. They claimed that their finding shows that this method is better than textual snippet since it can fit more results at once and will be more suitable for users specially for mobile devices.

4.4.2 Extraction of Picture and Its Caption

Preface. From the discussion in last section, we can see that visual factor plays crucial role for relevancy detection. Recent researches, however, are focused on refining previously visited pages based on a combination of textual and visual factors, or a visual factor alone, the thumbnail is usually based on a snapshot of overall page lay out and the text is mostly the current default snippets provided by search engines. Withal, this is not discussed yet that how a specific visual content of a page can allow users to differentiate between a relevant and irrelevant page in the result page, specially for a prompted ambiguous keyword. We believe that we are first to suggest using a visual factor for clarification. Our idea is that if we put a relevant photo from content of each returned page, it will help users to determine between the content of different results in SERP.

Extraction Phase. A picture in a Web page is usually discriminated by tag from other elements inside the page. The main idea of how to extract the photo is the distance of the tag from the occurrence of the stemmed search query. The tag is empty, it contains attributes only, and does not have a closing tag. This tag has various attributes. One of the interesting attributes is "alt"⁶. This "alt" attribute is used by *Screen Readers*⁷ to get the content of a page that is displayed on the screen. We took advantage of the content of "alt" for our picture extraction method. Unfortunately, not all are with a useful "alt". As a consequence, we made a priority list for the attributes of a , as exhibited in Table 1, in terms of their importance for the derivation of the visual content of a page.

⁶Stands for Alternative and is used as an alternative text for an image

⁷Screen readers are usually used by blind people to identify the content of whatever is displayed on the screen

We use the textual content of these attributes, based on their priority, and compare it with the search keywords. If a highest priority attribute has at least one occurrence of the stemmed keywords, we will regard the corresponding photo as a relevant visual cue for the page. The zero distance is when the keyword is within the tag's properties, "alt", "title" and "src" in order of importance. If there was no with zero distance from the occurred keyword, we go one node further, both the parents and children of the current node that contains the to inspect the availability of stemmed keywords occurrence until we reach one occurrence. If there was more than one with same distance from the occurred keyword, we choose FIFO⁸ policy and appoint the first one that is been traversed. Moreover, the detected textual content would be regarded as the picture caption and will be placed on the bottom of it.

Picture Resolution. In consideration of the original size of a extracted photo from page and the fact that it is way too big for our purpose, we decided to resize and crop it. We decided to put a photo of 2.1*2.1cm as the thumbnail if in marginal area and for the embedded case, we reduce the size into half. To do this, we first crop the photo by 1:1 ratio and then resize it into a 2.1*2.1cm photo. If the photo has a relatively longer width or height that makes it look like a horizontal or vertical rectangle, then we resize the photo to 1.3*2.1cm size or 2.1*1.3cm respectively.

4.4.3 Comparison

In most of the previous efforts (Czerwinski et al., 1999; Dziadosz and Chandrasekar, 2002; Jhaveri and Raiha, 2005; Kaasten et al., 2002; Robertson et al., 1998) to involve visuality in SERP, the visual factor was a *whole page snapshot*, without any particular attempt on enhancing the page title and/or its description known as *search snippet*. In (Teevan et al., 2009), although the authors tried to compute a visual search snippet to "be useful for search and re-visitation", however their aim was to replace each component of a search result, i.e, title, URL and search snippet with one of the page content, i.e, first 19 characters of the page, the logo of the page and a salient image to reduce its space in order to fit better in handled devices.

A recent study (Aula et al., 2010), aiming to make a comparison of effectiveness of different combination of thumbnail/URL/title to remember previously visited Web pages, has been conducted to illustrate

⁸First In,First Out

Table 1: Priority list for useful attributes of for photo extraction.

Priority	Attribute's name	Content value type	Remarks
1	"alt"	Text	It is an alternate text for the image. Used by screen readers for vocalization.
2	"title"	Text	It usually shows a "tool tip" when hover the mouse cursor over an image. Shows an advisory information about an image.
3	"src"	URL address	Stands for "Source" and specifies the URL address of an image.

that which variation by which size of thumbnail can be the best choice for this aim. However, the study considered only a whole page snapshot as a thumbnail, without any improvement in page title and its search snippet. Although they considered various types of different pages in their user studies, however this is not yet clear that, even with their little effort for involving visual factors in the SERP, how this will help users in special condition such as an ambiguous query. Moreover, instead of validating the value of each part itself, they examined the representation style of thumbnail rather the actual value of them.

On the contrast, our algorithm will extract a specific picture from the Web page based on the technique which we discussed in previously. In addition, we generated new search snippet and title for each result in the SERP. In our user study⁹ we will make a content-determined comparison of each part rather rely solely on the value of each part's way representation.

5 EVALUATION

5.1 Participants, Settings and Baseline

Our evaluation includes total number of 43 participants, including 13 Females and 30 Males with age varied from 18 to 50. Most of participants had no educational background related to computer major. Hence the results of this experiment reveals the point of view of common users. The study itself consists of two parts. In first part, We pre-selected 2 ambiguous queries, one for Bing and the other for Yahoo!. Based on a current study (Agrawal et al., 2015), there is a significant equality between the results of two common search engines, Google and Bing. Consequently, by applying this user study on either, the result can be applicable on the other. In this case, we choose the baseline interface as Yahoo! and Bing. We presented 5 different interfaces based on each query, the baseline interfaces together with

⁹Which consists of two parts

4 variations of our proposed methods. The four variations are illustrated in Figure 2.

In second part, we used a *within-subject* design. One drawback for this case is the concept of *learning* which is also called *Carryover effect*. This is caused for the sake of treatments' order. At this point, the experiment design for a within-subject design should be *Counterbalanced*. This will make the experiment sure to give all the participant different order of representation and hence, the effect of learning will be avoided. A *Balanced Latin Square* is one mean to do such. If the number of conditions are even, the first row of a Latin square will follow 1, 2, n , 3, $n - 1$, 4, $n - 2$... where n indicates the number of conditions and for the rest of rows, the number would be add up by one and it will return to 1 for the n . More details of both parts is provided in corresponding sections.

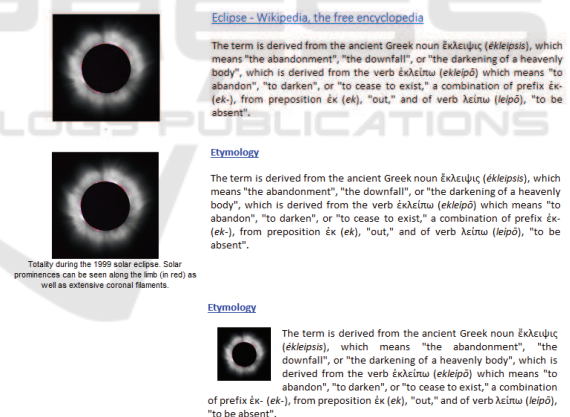


Figure 3: The three different variations of our proposed method for the SERP generation for an ambiguous search query, used in second part of the user study. We combined the first two interface from previous part to form "Improved default" interface as well as enlarged the thumbnails with 3:2 ratio. From top to bottom are : "Improved default interface", "Thumbnail with caption and improved title" and "Embedded thumbnail."

5.2 Procedure and Results

First Part. In first part of our experiment, we asked users to review the baseline interfaces as well as 4 variations of our method, which is illustrated

in Figure 2, and then, voting for each interface. Accordingly, we asked each participant to express reason(s) for such selection. In addition, we gave participants the opportunity to give any suggestions, if necessary. 53.8% of all participants selected the marginal thumbnail with caption as their favorable choice. On the other side, 33.3% privileged the embedded method and finally, 92.3% of all parties found the generated search snippet to be useful, while only 7.7% preferred to be with classical page description. From the user study results we can make this closure that both visual and textual aspects are vital to the users for disambiguation the search result page if the visual feature of the page content surrounds with pertinent textual description. From some of the participants feedback, we realized that the current title that search engines are providing are not always helpful and meaningful, regarding to the users intention and not cohesive with the textual search snippet. Consequently, for second part of our study, we decided to generate a new title. Moreover, according to some of the participants' comments, pointing out that a slightly bigger thumbnail would make it more helpful, we made search results' thumbnails larger with 3:2 ratio. Moreover, according to the results which indicates that 87.1% of all votes are for two interfaces, we combined the other less desired interface into one and make a new experiment together with previous stated changes.

Second Part. For second part of evaluation, participants were asked¹⁰ to initiate an ambiguous query, then, by using the presented techniques, we provided each ambiguous keyword with 3 variations of our method and together with the default interface, we ordered these interfaces as explained in previous section in a balanced Latin square. Consequently, we asked users to review these different interfaces as are illustrated in Figure 3 based on the presented order and rating each interface according to following criteria from 1(as the lowest) to 5(as the highest) for their effectiveness:

- **Helpfulness:** Upon each model, we asked users to score each **part** of each interface based on its helpfulness for disambiguation. By helpfulness, we mean the necessity of a part that its absence cause some critical information loss. Helpfulness itself is divided into smaller sections, regarding to different part of each interface, i.e, **Title, Search snippet and Knowledge base** for Default interface, **Title, Improved search snippet and Marginal thumbnail** for Improved default interface, **Improved title, Improved search snippet, Caption and Marginal thumbnail** for Marginal

¹⁰Unlike first part

	Type of user interface	N	Mean Rank
User Satisfaction	Default	8	12.69
	Improved default	8	14.50
	Thumbnail with caption	8	20.31
	Embedded thumbnail	8	18.50
	Total	32	
Transparency	Default	8	12.81
	Improved default	8	16.44
	Thumbnail with caption	8	18.38
	Embedded thumbnail	8	18.38
	Total	32	
Helpfulness	Default	8	12.94
	Improved default	8	18.12
	Thumbnail with caption	8	18.50
	Embedded thumbnail	8	16.44
	Total	32	

Figure 4: Mean rank results.

thumbnail with caption interface and finally, **Improved title, Improved search snippet and Embedded thumbnail** for the Embedded thumbnail interface. The total score for Helpfulness is computed accordingly based on the mean of these subelement scores.

- **Transparency:** by transparency we mean the level of information that each method reveals from content of the page and makes it easier for users to understand the topic of each result page based on the page description/snippet and other elements.
- **User Satisfaction:** the score for a method based on the experience that each user had within the model during a search session

To see any significant impact on any of the 3 major factors that we explained in the last section, we ran a *Kruskal-Wallis* test against the helpfulness, transparency and user satisfaction scores. By considering the *P_{value}* for user satisfaction, transparency and helpfulness 0.282, 0.557 and 0.613 respectively, we can conclude that there is no significant change in using any of 4 different user interfaces. However, from the listed *Mean rank* for each examined category, there is a notable improvement made by our suggested interface. The value of page rank for *Thumbnail with caption* is the highest in all the 3 different categories by the mean rank value of 20.31 , 18.38 and 18.50 for user satisfaction, transparency and helpfulness, respectively. The result for mean rank is illustrated in Figure 4:

These results can be shown visually in Figure 5. *The multiple line chart* on the left and *clustered bar* on the right are visually demonstrating the positive effect of our method against current user interface of Bing for disambiguation. Please note that for thumbnail with caption interface, the highest user satisfaction is achieved by the highest transparency

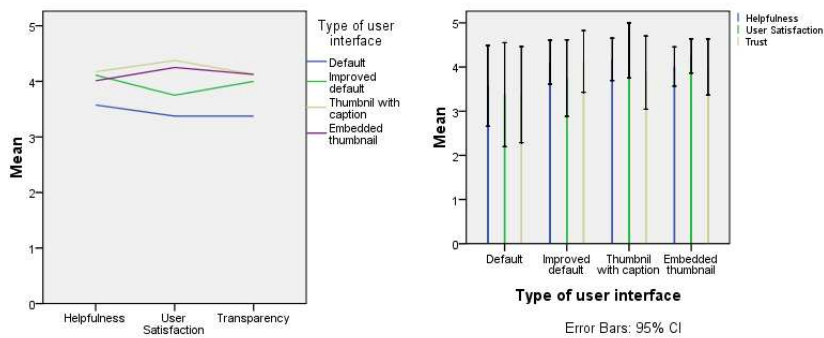


Figure 5: The comparison of the four different interfaces' mean for helpfulness, user satisfaction and transparency. The multiple line chart at left side and the clustered bar at right side show that higher user satisfaction is earned by both higher helpfulness and transparency.

and helpfulness. On the contrary, the lowest user satisfaction is earned by default user interface when it has the lowest transparency and helpfulness. This may be concluded that for achieving the higher user satisfaction, all the factors should work well together to enhance the user experience.

6 CONCLUSION AND FUTURE WORK

Conclusion. Web searching has been grown through years. Nonetheless, the ambiguity of queries makes it necessary to apply appropriate techniques to help users to find the appropriate Web page regarding to their intended ambiguous query. Despite different existing approaches that are available to deal with such cases, we believe that we are the first one to notice that visuality as well as a more appropriate textual snippet and title will help users for disambiguation process. As a result, we decided to generate new search snippet and titles as well as add an appropriate visual feature and introducing different variations of our technique which were evaluated them in our study against state-of-the-art. The results show that the visuality together with a better relevant textual snippet and page title, will boost the users' satisfaction with an interface during an ambiguous search session.

Future Work. To extend this current work, we would like to embed similar results of an ambiguous queries, forming them under one same title and search snippet and common thumbnail and caption. Since both parts of our experiments show that the interface with thumbnail and caption is the most useful interface, we will continue this work based on the interface with caption. The title and search snippet would be generated according to a summarization to produce a comprehensive snippet that carries as most as possible of all the similar results snippets.

In addition, the thumbnail will be computed, using a visualization technique. We will use a *Tree map* (Shneiderman, 1992; Bruls et al., 1999) to give space to each thumbnail according to their rank in the result page. The space for each thumbnail in a tree map will be computed according to the recent eye track study (Cutrell and Guan, 2007) which indicates how much of users look at each links according to their rank. In case of caption, we will only select the largest thumbnail's caption and place it under the tree map.

REFERENCES

- 2012, G. Z. Google zeitgest 2012.
- Agrawal, R., Gollapudi, S., Halverson, A., and Ieong, S. (2009a). Diversifying search results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM '09*, pages 5–14.
- Agrawal, R., Gollapudi, S., Halverson, A., and Ieong, S. (2009b). Diversifying search results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM '09*, pages 5–14, New York, NY, USA. ACM.
- Agrawal, R., Golshan, B., and Papalexakis, E. E. (2015). A study of distinctiveness in web results of two search engines. In *Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume*, pages 267–273.
- Aula, A., Khan, R. M., Guan, Z., Fontes, P., and Hong, P. (2010). A comparison of visual and textual page previews in judging the helpfulness of web pages. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pages 51–60, New York, NY, USA. ACM.
- Baayen, H. (1991). A stochastic process for word frequency distributions.
- Battelle, J. (2005). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Portfolio.

- Beitzel, S. M., Jensen, E. C., Chowdhury, A., Grossman, D., and Frieder, O. (2004). Hourly analysis of a very large topically categorized web query log. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 321–328, New York, NY, USA. ACM.
- Bruls, M., Huizing, K., and van Wijk, J. (1999). Squarified treemaps. In *In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization*, pages 33–42. Press.
- Chandra, A., Suaib, M., and Beg, R. (2014). Low cost page quality factors to detect web spam. *CoRR*, abs/1410.2085.
- Cutrell, E. and Guan, Z. (2007). What are you looking for?: An eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 407–416, New York, NY, USA. ACM.
- Czerwinski, M. P., van Dantzich, M., Robertson, G., and Hoffman, H. (1999). The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3d. In *In Proc. Human-Computer Interaction INTERACT '99*, pages 163–170.
- Dziadosz, S. and Chandrasekar, R. (2002). Do thumbnail previews help users make better relevance decisions about web search results? In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '02, pages 365–366, New York, NY, USA. ACM.
- Fang, Y., Somasundaram, N., Si, L., Ko, J., and Mathur, A. P. (2011). Analysis of an expert search query log. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 1189–1190, New York, NY, USA. ACM.
- Haas, K., Mika, P., Tarjan, P., and Blanco, R. (2011). Enhanced results for web search. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 725–734, New York, NY, USA. ACM.
- Jansen, B. J., Spink, A., and Pedersen, J. (2005). A temporal comparison of altavista web searching: Research articles. *J. Am. Soc. Inf. Sci. Technol.*, 56(6):559–570.
- Jhaveri, N. and Raiha, K.-J. (2005). The advantages of crosssession web workspace. In *In Proc. CHI 2005*, pages 1949–1952.
- Kaasten, S., Greenberg, S., and Edwards, C. (2002). How people recognize previously seen web pages from titles, urls and thumbnails. In *In X. Faulkner, J. Finlay, F. Detienne (Eds.) People and computers XVI (Proc. Human Computer Interaction)*, pages 247–265.
- Mallery, J. C. (1988). Thinking about foreign policy: Finding an appropriate role for artificial intelligence computers. Master's thesis, MIT Political Science Department.
- Mallery, J. C. (1994). Thinking about foreign policy: Finding an appropriate role for artificial intelligence computers. Technical report.
- Marcos, M.-C. and Pérez-Montoro, M. M. (2009). Comportamiento de los usuarios en la elección de resultados en los buscadores: un estudio con tecnología eye tracking. In *X Congreso Internacional de Interacción Persona-Ordenador*.
- Mihalkova, L. and Mooney, R. (2008). Search query disambiguation from short sessions. In *Beyond Search: Computational Intelligence for the Web Workshop at NIPS*.
- Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., and van Dantzich, M. (1998). Data mountain: Using spatial memory for document management. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, UIST '98, pages 153–162, New York, NY, USA. ACM.
- Sanderson, M. (2008). Ambiguous queries: Test collections need more sense. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 499–506, New York, NY, USA. ACM.
- Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.*, 11(1):92–99.
- Song, R., Luo, Z., Wen, J.-R., Yu, Y., and Hon, H.-W. (2007). Identifying ambiguous queries in web search. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 1169–1170, New York, NY, USA. ACM.
- Stokoe, C., Oakes, M. P., and Tait, J. (2003). Word sense disambiguation in information retrieval revisited. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, pages 159–166, New York, NY, USA. ACM.
- Sun, J.-T., Zeng, H.-J., Liu, H., Lu, Y., and Chen, Z. (2005). Cubesvd: A novel approach to personalized web search. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 382–390, New York, NY, USA. ACM.
- Teevan, J., Cutrell, E., Fisher, D., Drucker, S. M., Ramos, G., André, P., and Hu, C. (2009). Visual snippets: Summarizing web pages for search and revisitation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2023–2032, New York, NY, USA. ACM.
- Yin, D., Xue, Z., Qi, X., and Davison, B. D. (2009). Diversifying search results with popular subtopics. In *Proceedings of the Eighteenth Text REtrieval Conference (TREC 2009)*.
- Zhang, B., Li, H., Liu, Y., Ji, L., Xi, W., Fan, W., Chen, Z., and Ma, W.-Y. (2005). Improving web search results using affinity graph. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 504–511, New York, NY, USA. ACM.