



City Research Online

City St George's, University of London

Citation: Loumakis, F., Stumpf, S. & Grayson, D. (2011). This image smells good: Effects of image information scent in search engine results pages. In: Proceedings of the 20th ACM international conference on Information and knowledge management. (pp. 475-484). New York: ACM. ISBN 9781450307178 doi: 10.1145/2063576.2063649

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://city-ref-test.eprints-hosting.org/id/eprint/12416/>

Link to published version: <https://doi.org/10.1145/2063576.2063649>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

This Image Smells Good: Effects of Image Information Scent in Search Engine Results Pages

Faidon Loumakis
Fluent Interaction Ltd.
17g Clerkenwell Road
London EC1M 5RD

Simone Stumpf
Centre for HCI Design
City University London
London EC1V 0HB

David Grayson
Fluent Interaction Ltd.
17g Clerkenwell Road
London EC1M 5RD

faidon@fluent-interaction.co.uk Simone.Stumpf.1@city.ac.uk dave@fluent-interaction.co.uk

ABSTRACT

Users are confronted with an overwhelming amount of web pages when they look for information on the Internet. Current search engines already aid the user in their information seeking tasks by providing textual results but adding images to results pages could further help the user in judging the relevance of a result. We investigated this problem from an Information Foraging perspective and we report on two empirical studies that focused on the information scent of images. Our results show that images have their own distinct "smell" which is not as strong as that of text. We also found that combining images and text cues leads to a stronger overall scent. Surprisingly, when images were added to search engine results pages, this did not lead our participants to behave significantly differently in terms of effectiveness or efficiency. Even when we added images that could confuse the participants' scent, this had no significantly detrimental impact on their behaviour. However, participants expressed a preference for results pages which included images. We discuss potential challenges and point to future research to ensure the success of adding images to textual results in search engine results pages.

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: On-line Information Services – *Web-based services*, H.5.2 [Information Interfaces and Presentation]: User Interfaces – *screen design*, H.1.2 User/Machine Systems – *Human Information Processing*.

General Terms

Human Factors, Theory, Design.

Keywords

Information foraging, information scent, search engine results pages, images, visual search, usability.

1. INTRODUCTION

There are currently billions of web pages, presenting the user with an overwhelming amount of information to be sifted through during their information seeking tasks. To aid users in finding the information that they are looking for, modern search engines retrieve results based on keyword queries and present them in a ranked list which usually gives details of the page title, URL and a short text snippet. In order to further help the user judge the results' relevance in search engine results pages (SERPs), some search engines now also add a visual representation of the target

page, for example, a page preview (Figure 1) gives a thumbnail visual representation of the result page. In addition to page thumbnails, there are also other approaches that aim to visually summarise the content of the target page by providing representative images [11].

Text results and visual representations in SERPs can be viewed as components within Information Foraging Theory [15, 16, 19]. Information Foraging posits that users will seek out information patches with the strongest scent, where the strength of scent is determined by textual and visual cues from the environment, reflecting the cue's relevance to the search task. In SERPs, scent can come from text cues – the text in the title, the URL and the text snippets – and also visual cues, such as styling of text or associated images. While textual cues have been explored extensively from an Information Foraging perspective, visual cues on the other hand, especially the role of images in providing scent, have received scant attention.

This paper describes two empirical studies which investigated scent of images and the effects of adding visual cues in the form of images to SERPs on user behaviour. Our first study aimed to *determine information scent for images and investigate the relationship between the presence of text and images in information scent*. Our second study built on these findings to further investigate adding images to SERPs. This second study explored *the effect of adding images to SERPs in terms of user effectiveness, efficiency and satisfaction*.

Previous work has attended to retrieved objects' relevance for a topical area within a work task and users' assessment of intellectual topicality [2]. Our approach has a different focus. We explore the usefulness of images as a visual representation of the results page when added to textual results. In our search tasks we assume the user is looking for textual information, not searching for images specifically.

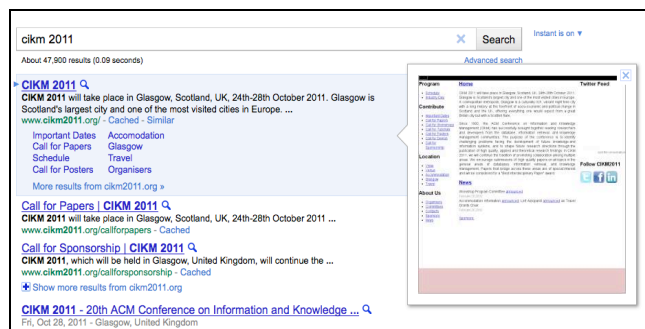


Figure 1. Search results in Google. This interface makes use of a number of scent cues: Textual cues (e.g. text snippets, URL, etc.) and visual cues (e.g. text styling, page preview image).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.

Copyright 2011 ACM 978-1-4503-0717-8/11/10...\$10.00.

In the remainder of this paper, we will explore previous research relevant to Information Foraging and the presentation of SERPs. We then present the set-up and results of our first study, followed by the methods and findings for our second study. We end with an overall discussion of the combined findings from both studies and directions for future work.

2. RELATED WORK

2.1 Information Foraging

Information Foraging Theory [15, 16, 19] aims to explain how information seekers navigate information environments such as the web, drawing on an analogy of how animals forage for food. According to this theory, users forage for information by seeking out information patches, e.g. web pages, drawn by the patch's "scent". The strength of scent is determined by textual and visual cues, and it is argued that if scent decreases or is not strong enough, users will move on to the next information patch.

Models built on this theory have been used to predict user behaviour during the information seeking process [5, 10, 13, 18] and also to predict the information scent for specific interfaces and web pages [1, 6, 7]. However, in these models involving information scent, visual cues are mostly ignored. Although information scent has been recognised as an important concept in interface design [24], no studies have been carried out that leverage Information Foraging for making design decisions involving SERPs with images.

2.2 Search Engine Results Pages

Current SERPs usually present textual cues in the form of words in the page title, URL, text snippet, etc., combined with some visual cues such as text styling and visual treatment. More recently, other visual cues such as page previews have been added. Research into the presentation of SERPs has shown the importance of the length of text snippets on a user's click behaviour [9], however, what matters to a user in the presentation of results is strongly tied to the user's task [9, 20]. Search tasks can be categorised into informational, navigational, and transactional types [3]. In this paper, we focus on informational tasks, i.e. how to find pages that contain relevant information, and will not further discuss navigational tasks, e.g. how to re-find pages using SERPs.

Visual cues are often combined with text such as colour styling on sponsored ads, text size, and so on, to draw attention to them. Styling needs to be carefully used because, whilst it can increase users' efficiency, it may also interfere if it draws attention to the wrong item [21]. However, some research has also suggested that styling does not override textual cues in users' information seeking behaviour [4].

Visual cues can also come from images associated with search results. Again, there are problems with using this type of cue in SERPs. Woodruff et al. [25, 26] compared textually enhanced thumbnails of target pages, plain thumbnails and text as elements of a SERP. Results showed that enhanced thumbnails proved to be more consistently successful than plain thumbnails. When presenting users with visual summaries in addition to text results, users have been shown to be less effective in finding relevant information [14].

Our studies in contrast focus on the scent of images, and compare them to textual cues. In addition, we were interested in the

behaviour of users when the scent of images is manipulated in SERPs. Our studies therefore contribute to deepening the understanding of image scent, the impact of images on users' information foraging behaviour and the design of SERPs extended with images.

3. STUDY 1: INFORMATION SCENT OF IMAGES

How can we determine the scent of an image, and how does it compare to the scent of text? In this first study, we aimed to measure information scent for images and investigate the relationship between cues from text and images in information scent.

3.1 Methodology

There are two main approaches to measuring information scent. The most common way is to use independent human judges to rate scent on a scale [16, 19]. The second is to use algorithms which automatically assign scent based on keywords found in the text using a model-driven approach [12, 17]. There is currently no automatic way to determine the information scent of an image, thus our study used human judges to rate the images.

Previous studies involving human judges of scent used between four and ten judges [19]. Since images can be more ambiguous than text, we decided to use a much higher number of judges by distributing a scent rating elicitation tool (SRET) over the Internet. Invitations to take part were distributed via email, Facebook and Twitter. Participants who completed all the tasks in the SRET were offered the chance to enter a competition to win a £30 voucher from Amazon.co.uk. In total, we had 81 participants.

We presented a series of questions to participants, stimulating an information need and mimicking a search task. Topic familiarity has previously been shown to influence the rating of relevance criteria [22], thus we chose questions so that they covered a wide breadth of topics and so that it would be unlikely that the participants would know the answers. In total we asked eight questions:

1. *What is the problem with the antenna of the iPhone 4 and how can it be fixed?*
2. *What is the length of a Japanese bullet train?*
3. *Where does the blobfish live?*
4. *How many stairs does the Eiffel tower have?*
5. *What is the maximum speed a cheetah can reach?*
6. *When (day, month, year) did Martin Luther King deliver the famous "I have a dream" speech?*
7. *Who invented the pen Bic and how did they come up with the idea?*
8. *What is the address of the Google offices in London?*

For each question, we provided a number of "search results" i.e. cues that may lead to a target page on which to find the right answer. These cues were presented in three cue treatments:

- *Text-only* cues, where the cue showed a link with the title of the result, a text snippet, and a URL,
- *Image-only* cues, where the cue was a visual representation of the results page information,

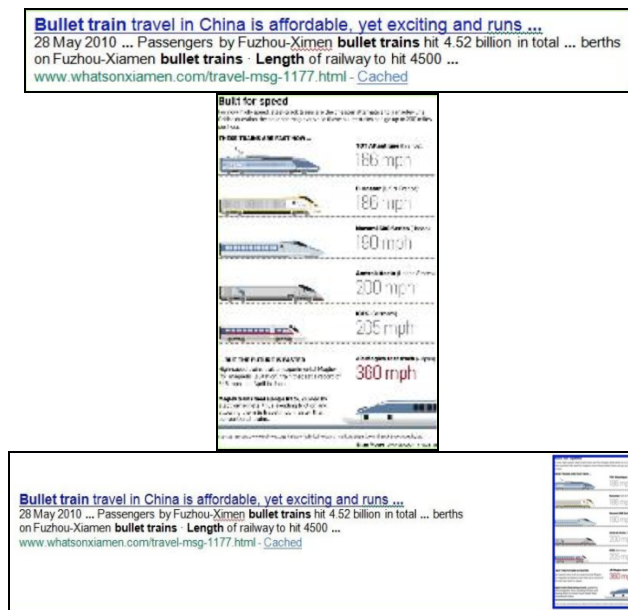


Figure 2. Cue treatments: Text-only cue (top), Image-only cue (middle), Text+Image cue (bottom) for *Bullet train* question.

Table 1 – Query keywords used to retrieve cues for questions.

Question	Query keywords
What is the problem with the antenna of the iPhone 4 and how can it be fixed?	iPhone 4 antenna problem
What is the length of a Japanese bullet train?	Bullet train length
Where does the blobfish live?	Blobfish
How many stairs does the Eiffel tower have?	Eiffel tower stairs
What is the maximum speed a cheetah can reach?	Cheetah
When (day, month, year) did Martin Luther King deliver the famous "I have a dream" speech?	Martin Luther King I have a dream
Who invented the pen Bic and how did they come up with the idea?	Bic pen
What is the address of the Google offices in London?	Google London

Text Cue	Image cue
Let's work together. - UK jobs - Google Our UK offices are based in London and other locations around the UK ... Please do not forward CVs to our jobs alias, Google employees or any other company ... www.google.co.uk/intl/en/jobs/index.html - Cached	
London - Student Jobs - Google Human resources, finance, legal, facilities, marketing, communications, business development and corporate development staff also work at Google London. ... www.google.com/jobs/students/emea/spotlightlondon/ - Cached - Similar	
Google Maps of London street gangs: Do the Mozart Bloods prowl ... 30 Mar 2010 ... Is your neighbourhood blighted by the Ghetto Diplomats or the Manor Park Goons? Do the Mozart Bloods walk your West London street? blogs.telegraph.co.uk/.../google-maps-of-london-street-gangs-do-the-mozart-bloods-prowl-your-neighbourhood/ - Cached	
Revealed: the environmental impact of Google searches - Times Online 11 Jan 2009 ... Alex Wissner-Gross describes his research in detail Google's response to th... 894646 Registered office: 1 Virginia Street, London, E98 1XY. technology.timesonline.co.uk/tol/news/tech.../article5489134.ece - Similar	

Table 2. Example of Text cues and Image cues for *Google London* question.

- *Text+Image* cues, where a cue showed both text and an image.

The three cue treatments are displayed in Figure 2. In order to counter-balance practice effects in this study, we randomised the order in which cue treatments appeared to participants.

For each question, we selected four text cues and four image cues as a basis for investigating scent. Image and text cues were found through Google Search and Google Images using keywords associated with the questions that a researcher generated (Table 1). For each cue, we tried to choose a range of items that would span the spectrum of scent, i.e. some potentially high-scent and low-scent cues. Table 2 shows the four text cues and four image

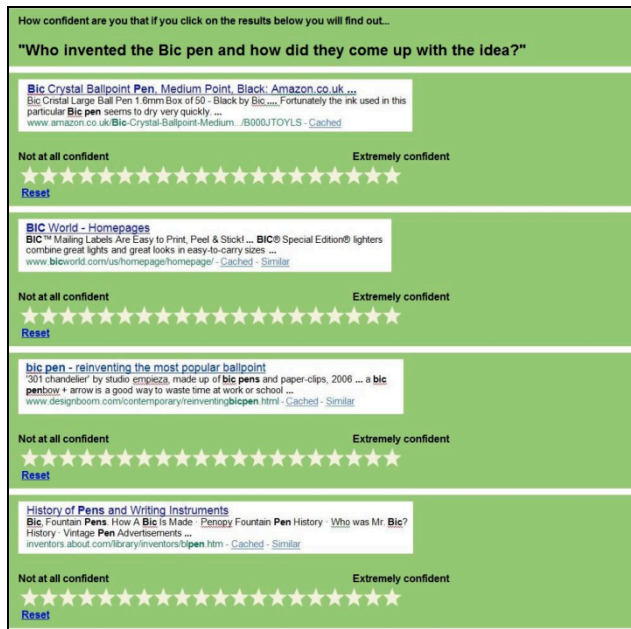


Figure 3. *Bic pen* question in Text-only cue treatment.

cues used for the *Google London* question. For the Text+Image treatment we randomly paired text cues and image cues.

For their main task, we asked participants to rate how confident they were that this cue would lead them to the answer to the question i.e. how strong was the cue's information scent? The participants were able to rate the scent of cues on a 20-point Likert scale from "Not at all confident" (1) to "Extremely confident" (20). We chose a large scale instead of the more traditional five- or seven-point scale, because we believed this scale would give the users the freedom to intuitively select one point without spending much time thinking and trying to assign ratings into few categories. Figure 3 shows the presentation of question, cues and rating scale for the *Blobfish* question in the Text-only treatment.

Of the 81 people who took part in this study, 63 completed all rating tasks. Overall, we managed to obtain 6776 scent ratings across 96 individual cues (96 cues are derived from the 8 questions * (4 Text-only cues + 4 Image-only cues + 4 Text+Image cues)).

3.2 Study 1 Results

3.2.1 Do you think I smell?

We wondered first of all if scent is a reliable measure that could be utilised in SERPs. To investigate this, we calculated simple descriptive statistics for each cue.

Overall, we observed some divergence on scent ratings between users. Across all cues, there was a 25% spread of ratings, with an average standard deviation of 5.38. This was true for the majority of cues: 22.9% of cues received ratings from the whole 20-point spectrum of the scent scale, while 87.5% received ratings from a not much narrower 17-point spectrum. For example, Figure 4 shows the distribution of ratings for the "map" image for the *Blobfish* question. It can be seen that the majority of the participants rated the scent of this image very high; 58% rated it at 15 or above. However, there are 13 participants who rated the image as not very high, with a scent rating of less than 10. Thus, for the same cue, there was disagreement about whether it had a high scent or a low scent – participants' opinions diverged

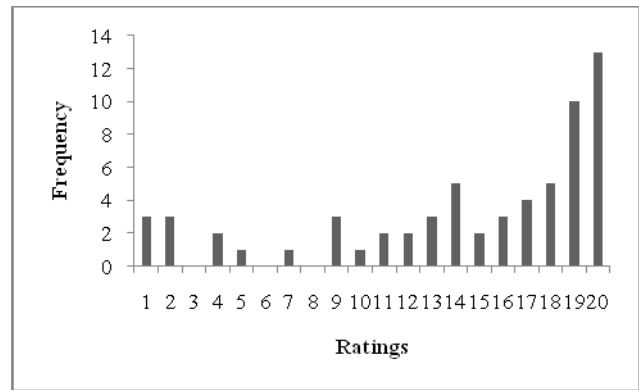


Figure 4. Frequency of scent ratings for the "map" image in the *Blobfish* question (mean = 14.4 , median = 17).

sometimes widely. This range pattern remained consistent across different cues and opinions diverged just as much for Text-only cues as for Image-only or Text+Image cues (Table 3).

Intuitively, it may be assumed that the more relevant the image is, the less disagreement there should be. However, standard deviation was *not* related to the median scent ratings of the images. This means that participants disagreed with each other consistently, even on images that were generally judged to be more relevant. From these results it appears that there is an individualistic component to ratings, in which one user's nose for information scent may be working differently to another user's.

However, fairly reliable scent ratings for cues can be obtained. In the best case (*Bic pen* question: "History of pens and writing instruments" text + "BIC think" image, SD=3.59), the spread of ratings was less than 20%. Although some images may be affected by outlier scent ratings, median values of scent rating can still reflect the most prevalent scent for an image.

However, reliable scent ratings may not be enough to justify adding images to SERPs; if image cues do not have a strong scent by themselves there may be no value in adding them to SERPs in the first place. In order to find out what scent dominates, we looked at the scent ratings of Text-only cues and Image-only cues across the questions in more detail (Table 4). In five of the eight questions, participants gave higher scent ratings to Text-only cues than Image-only cues. In fact, a t-test showed that Text-only scent was rated significantly higher than Image-only scent overall ($t(4483) = 2.85, p = 0.004$). Text-only cues therefore have a significant advantage in terms of scent compared to images. However, there are exceptions: For some image sets and questions (e.g. *Blobfish* and *Eiffel tower*), Image-only scent ratings were higher than Text-only ratings.

These results are important for the design of SERPs in terms of choice of images. It shows that however relevant the image may

Table 3. Lowest and highest median scent ratings for individual cues observed and mean standard deviation by cue treatment

Cue Treatment	Lowest median cue rating	Highest median cue rating	Mean SD
Text	1	17	6.14
Image	2	17	6.14
Text+Image	2	18	6.21

Table 4. Mean scent ratings and standard deviation for Text-only and Image-only cues per question. Shaded cells indicate where Image-only scent is higher rated than Text-only scent.

Question	Mean scent rating		Standard Deviation	
	Text-only	Image-only	Text-only	Image-only
<i>iPhone</i>	12.00	11.25	2.16	2.63
<i>Bullet train</i>	10.50	10.50	3.79	5.20
<i>Blobfish</i>	8.25	11.75	4.19	4.11
<i>Eiffel tower</i>	8.33	10.25	6.66	4.65
<i>Cheetah</i>	10.50	6.75	2.12	4.11
<i>MLK</i>	12.25	8.50	1.89	3.11
<i>Bic pen</i>	7.25	5.67	6.08	2.52
<i>Google London</i>	11.75	9.00	4.79	6.48
Mean	10.10	9.21	3.96	4.10

seem to the designer (or an algorithm that extracts relevant images), it may not align with the scent that the user smells. However, cues from images appear to have a scent of their own and therefore adding images to text cues may be a viable option to explore.

3.2.2 Does text with an image smell even better?

We were interested in what happens to scent if images are added to text. How do text scent and image scent combine? By looking at the ratings of all three types of treatments, a pattern that emerged was that, in most cases, the Text+Image cues were rated higher by participants than both the Text-only cue or the Image-only cue. A t-test comparing the scent ratings of text with the scent ratings of Text+Image showed that the Text+Image were rated significantly higher than the text ($t(4028) = -6.68, p = 2.65e-11$). Moreover, a multiple regression test showed that the rating of the Text+Image result was predicted by both the text and the image ratings (multiple regression, $R^2=0.3426, p = 3.952E-184$). The text is the most influential element in the scent rating, with the image contributing a little scent by itself, shown by the regression function:

$$\text{Text+Image scent} = 0.53 * \text{Text scent} + 0.19 * \text{Image scent} + 4.05$$

It follows from this function that a very good image scent could overpower the scent of text. In other words, could a relevant image overcome irrelevant text? Table 5 shows the instances in our study when this happened. In all cases the Image scent ratings were significantly higher than the Text-only rating. Our findings show that participants rated some Text+Image ratings higher when high-scent images were paired with low-scent text.

Conversely, consistent with the scent regression function, when a high-scent text was combined with a low-scent image, participants usually ignored the low-scent image and rated the combined result highly. This shows that images played a mediating role when added to text and in some cases counteracted textual cues.

Our findings have implications for the design of novel SERP interfaces. Adding images to textual search results will affect what the user smells, however the scent of text will dominate – the role of the image in SERPs is generally of minor importance. In some cases however, adding an image can raise the relevance of results

Table 5 - The six results in which the high-scent image caused a significant increase in the Text+ Image results' score

Text+Image combination	Scent ratings		
	Text-only	Image-only	Text + Image
"Reinventing the most popular pen"+"BIC inventor"	6	13.5	10
"London street gangs"+"Google Map"	1	16	4
"Bullet train of Shinkansen"+"Blueprint"	8	15	11
"Blobfish" + "map"	11	17	15
"sometimes we all feel like the blobfish" + "WWF"	2	13	4
"Cheetah"+"Cheetah running"	10	14	14.5

and can lead it to smelling better overall. In other words: adding images, even if they are bad, probably will not matter that much. If images are good, they could make even a bad result appear more relevant.

4. STUDY 2: EFFECTS OF SCENT ON USER BEHAVIOUR IN SERPS

With the ratings gathered in Study 1, we designed a second experiment to investigate the effect of adding images to SERPs on user behaviour in terms of effectiveness, efficiency and satisfaction:

- We wanted to explore whether adding images to SERPs can lead users to choose results more accurately and whether users would be confused by placing images next to text cues where their scents conflict (Effectiveness).
- We also wanted to investigate whether adding images helps the user to choose more quickly or whether they are delayed by conflicting scents of images and text cues (Efficiency).
- We were interested in whether users liked having image cues added to SERPs (Satisfaction).

4.1 Methodology

We decided to include only the questions from Study 1 where there were distinctive scent ratings for images and text, i.e. not all cues were rated to have roughly equal scent. Hence, we only gathered data about the following four questions (original numbers from Study 1 in brackets):

1. What is the problem with the antenna of the iPhone 4 and how can it be fixed? (1)
2. What is the length of a Japanese bullet train? (2)
3. What is the maximum speed a cheetah can reach? (5)
4. What is the address of the Google offices in London? (8)

To investigate user behaviour upon adding images, we developed four SERP versions to explore variations in combinations of scent in images and text:

V1 Baseline (Text-only): A SERP showing no images but giving textual cues and visual cues only in terms of styling. This is the traditional SERP presentation.

V2 (Text+Image scent-matched): A SERP which gives both textual cues and image cues and where the high-scent text is shown with the high-scent image.

V3 (low-scent Text+high-scent Image): A SERP which gives both textual cues and image cues but where a low-scent text is shown with the high-scent image. This will allow us to investigate whether an image that smells very good can outweigh "bad-smelling" text.

V4 (high-scent Text+low-scent Image): A SERP which gives both textual cues and image cues but where a high-scent text is shown with the low-scent image. Again, this allowed us to investigate if image scent would confuse users.

For each question, no matter the SERP version, the best result was shown in the same rank position. However, that position was not the same for all questions as we wanted to avoid a "training effect" that would help the user choose a result realising that the "best" result is always in the same position.

Each SERP showed seven results since we wanted to mimic a realistic SERP task, in which a user is confronted with a number of results that have an impact on effectiveness and efficiency. In addition to our four results from Study 1 for which we had scent ratings, we therefore had to insert additional "filler" search results

to make up the seven results. For these "filler" results, we chose three generic or unrelated text results and images, being careful not to introduce a new high-scent item.

We adopted a within-subject experiment design, in which each participant was faced with all four questions and all four SERP versions. We developed four alternative experiment treatments where we counterbalanced the SERP versions against questions using a Latin square design to remove ordering confounds (e.g. Participant 1 saw V1 with Question 1 whereas Participant 2 saw V1 with Question 2, and so on).

The SERPs were implemented using a static visual prototyping tool to resemble Google search result pages, using the same fonts and colours in order to feel familiar to the users (e.g. the title of the text result was coloured blue and underlined using an Arial 12pt font). The images were resized not to exceed 180pt in height and 200pt in width and an outer blue stroke was added to them to resemble Google linked images. All other features of the Google search results page (search box, number of pages, left bar faceting tools etc.) were removed so as not to distract the participants from their task. A search results page, as was displayed to the participants for V2, can be seen in Figure 5.

A total of 64 participants completed this experiment. Participants were recruited at a university in a main pass-through area, used by

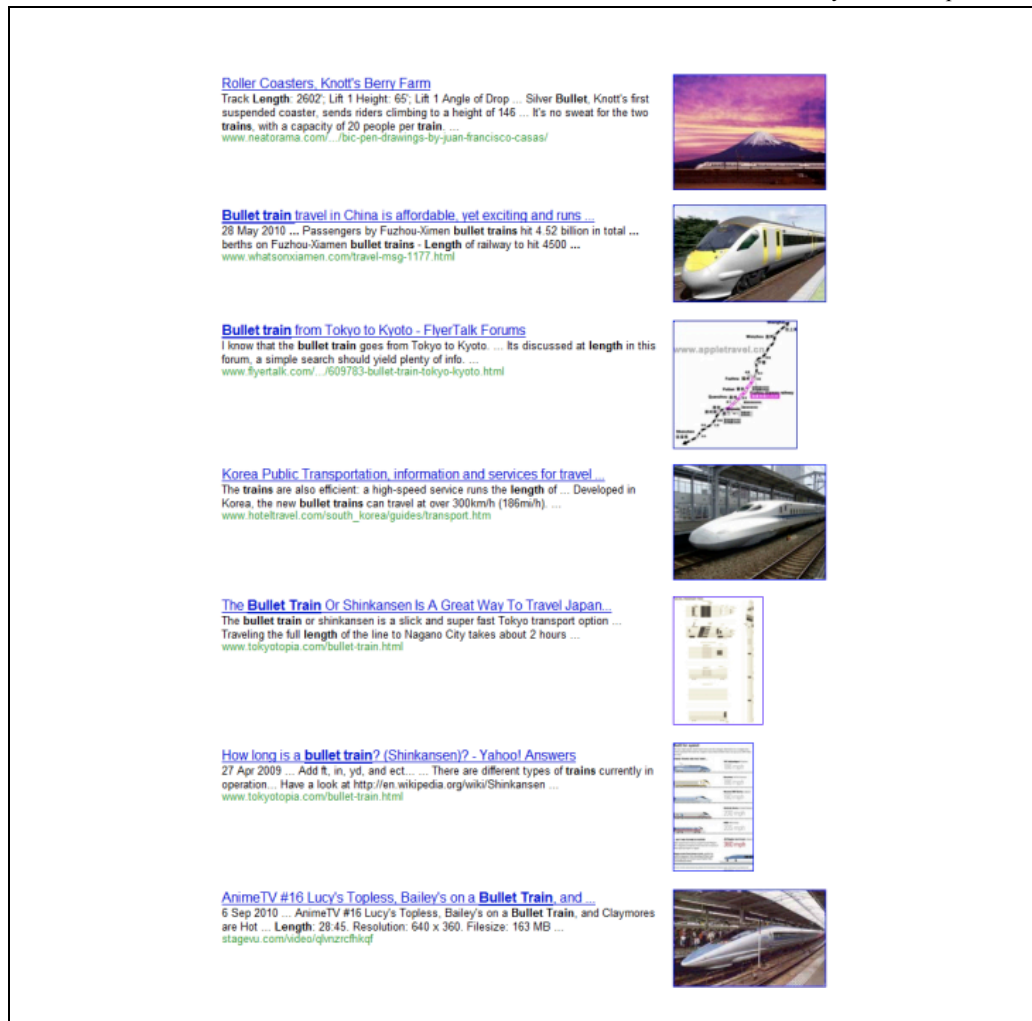


Figure 5. SERP version V2, Bullet train question

staff and students from different disciplines. The experiment lasted approximately 2 minutes. All participants were asked not to talk during the experiment and to use the mouse in order to keep the experiment conditions as consistent as possible. The experiments were conducted on a SONY VAIO laptop with an Intel® Core™ 2 Duo CPU P8400 at 2.26GHz, using a wireless Logitech mouse and the participants' screen was captured using CamStudio 2.0. All pages that were shown to participants were locally stored in the computer so there was virtually no delay between pages.

After a brief description of the experiment by the researchers, the participants were presented with a screen with instructions for the experiment. This was done in order to ensure that all participants would receive exactly the same information. This was followed by an example task. This example task's only purpose was to familiarise the participants with the procedure. The data from the example task was not included in the analysis.

Each question consisted of three screens. The first screen featured the question and a button which, upon being clicked, lead to the results page. The second screen was the SERP version. When a participant clicked on any link on the SERP (title of result or image) they would leave the SERP and continue on to a "thank you" page, asking them if they were ready to proceed to the next question. They would then click on a "next question" button which led to the next question, then the SERP, and so on. Once all questions and versions were completed by the participant, they were presented with a page asking them to rate their preference of SERP presentation.

For this study we captured three measures: effectiveness, efficiency and satisfaction. For effectiveness, we measured how many participants selected the "correct" highest-scent rated item out of the seven results shown. To measure efficiency, we took the time that participants spent between starting the task and selecting a result. The start of the task is defined by the moment a participant clicked on the "see the results" button on the question page. We analysed the selection of results and timing of the selections from the screen capture. Satisfaction ratings were obtained from participants at the end of the experiment using a 5-point scale, indicating their preference on a spectrum between 1 (Strongly prefer Text-only version) to 5 (Strongly prefer Text+Image version). Screenshots of the two versions were also shown on this page that acted as a reminder to participants.

4.2 Study 2 Results

4.2.1 How effective is image scent added to SERPs?

In order to test the effectiveness of adding images to SERPs, we set up two hypotheses:

(H1) Adding images to text cues will result in more correct selections – V2 (Text+Image scent-matched) will result in more correct selections than V1 Baseline (Text-only) version.

(H2) Placing image cues where their scent conflicts with the scent of text cues will cause more incorrect selections – there will be more incorrect selections with either V3 (low-scent Text + high-scent Image) or V4 (high-scent Text + low-scent Image) than with V2 (Text+Image scent-matched).

Figure 6 shows the breakdown of correctly and incorrectly selected results for all versions. As can be seen, participants in our study only chose marginally more correctly in V2 where images were added than compared to V1, the Text-only baseline version; this improvement is not statistically significant. This means that

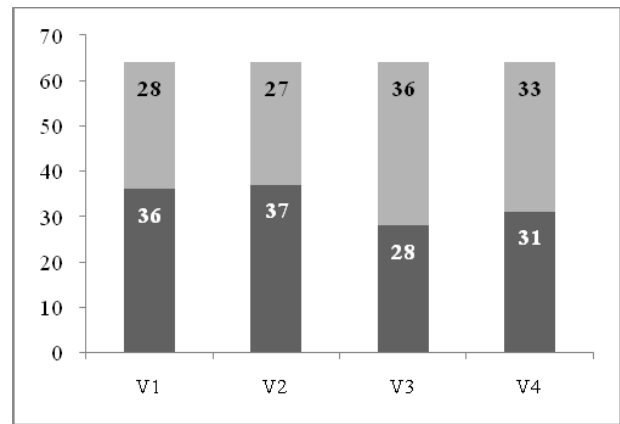


Figure 6. Number of correct (dark grey) and non-correct (light grey) selections by SERP versions indicating Effectiveness.

the high-scent image did not help users choose more accurately and therefore H1 is rejected.

Did confusing scent have any negative impact on participants' effectiveness? It would seem that there is a tendency towards more incorrect selections in both V3 and V4, where scents by text and image cues clashed, than in V2, where scent of text and images matched. However, the difference between V2 and V3 is not statistically significant ($\chi^2(3, N=64) = 1.63, p = 0.65$). Since V4 has a higher number of correct selections than V3, by implication there will be no significant difference between V4 and V2. Therefore, hypothesis H2 is also rejected.

For the design of SERPs integrating images, this means that adding an image to a SERP result will not significantly help users in identifying correct results but neither will it significantly hinder them if an image is placed with text cues where the scents may conflict.

4.2.2 Does image scent make SERP choices quicker?

We also investigated the effects of adding images to SERPs on users' efficiency. We surmised that

(H3) Adding images to text cues will result in selections being made in less time – Results in V2 (Text+Image scent-matched) will be selected faster than in V1 Baseline (Text-only).

(H4) Adding image cues to text where their scents conflict will confuse users and lead to longer selection times – Results in V3 (low-scent Text + high-scent Image) and V4 (high-scent Text + low-scent Image) will be selected slower than in V2 (Text+Image scent-matched).

The average timings of the four versions are displayed in Figure 7. According to H1, we expected that images would help with efficiency compared to the Text-only version. In our study, we found that, on average, participants tended to select results *slightly slower* in V2 where images were added than in V1 without images but this difference was not statistically significant ($t(123) = -0.45, p = .65$).

What about efficiency when scent confusion was added to the mix? We found that participant's selected a result *slightly quicker* in V3 (low-scent Text + high-scent Image) compared to V2 (without scent confusion) and *slightly slower* in V4 (high-scent Text + low-scent Image) than in V2. Again, this partially went against our initial expectations as we expected scent conflicts to have a detrimental effect on selection times. However, the time

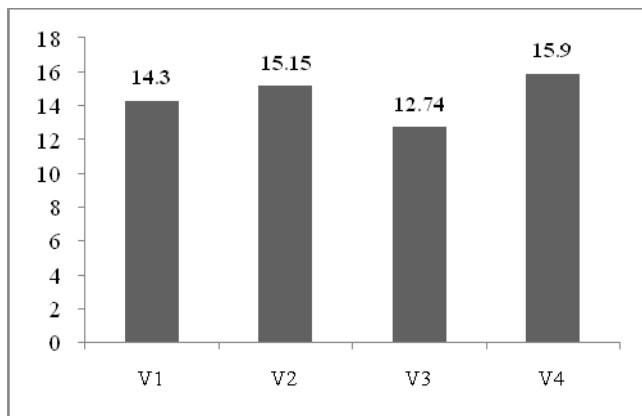


Figure 7. Average time (seconds) to select a result in the four SERP versions indicating Efficiency

that participants took is not statistically significantly different from V2 which is the best case where scents matched (Table 6).

Therefore, our hypotheses H3 and H4 are both rejected. This means that overall, adding images to SERPs will not lead to a substantial difference in user efficiency. Moreover, adding the "wrong" kinds of images to texts result also did not lead to significantly detrimental effects on participants in terms of efficiency.

4.2.3 Is added image scent appreciated in SERPs?

We also wondered if images had any effect on user satisfaction. In our post-study questionnaire, we asked participants to rate their preference for the Text-only version and versions with images added. The preference ratings were collated for the 64 participants who took part in the experiment and are presented in Figure 8.

Our findings show that there were a clear majority of participants who preferred Image versions over the Text-only version: 41 participants preferred images combined with text compared to 15 who preferred Text-only versions. A test confirmed this trend: there is a significant difference between ratings ($\chi^2(4, N=64) = 20.53, p = .0003$). This means that in terms of user satisfaction adding images to SERPs has a clear advantage – participants liked the added scent from image cues.

5. DISCUSSION AND FUTURE WORK

The results from our two studies on image information scent have important implications for the presentation of SERPs, and in addition, could also be of interest to the area of image retrieval in general.

Firstly, it appears that adding images to SERPs is, in general, a good idea. Combined scent from images and text cues resulted in higher scent ratings by participants in the first study, and participants in our second study preferred when images were added to SERPs.

However, our work also shows that some caution is advised in the choice of images. One immediate problem is that individual users may interpret the scent of an image differently to each other, to a

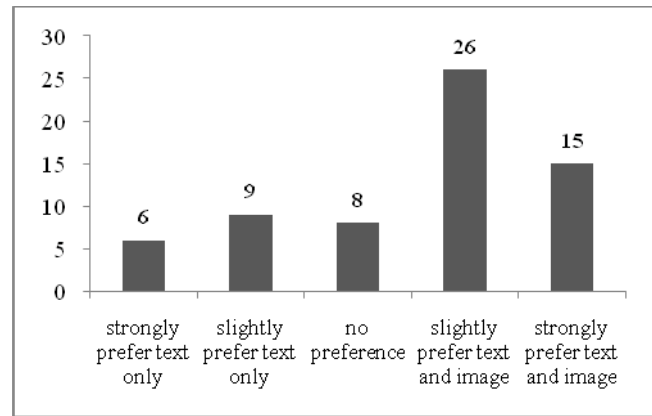


Figure 8. Frequency of participants' preference ratings for Text versus Text+Image

designer or to an algorithm that extracts suitable images from web pages automatically. This means that there will never be a "perfect" image to accompany a textual search result but neither is it understood what aspects contribute to the scent of an image. Some work on users' needs, when they are specifically searching for images, has started to shed light on this area [8, 23] but more research is required to determine what aspects affect the scent of an image in order to develop a suitable model of image information scent.

Secondly, our work shows that scent from images smells less than scent from text but we did not investigate the role that visual styling of text may contribute to scent. Previous research [4] has suggested that text styling does not override text cues but it is not clear how the scent from visual styling compares to the scent of images.

Thirdly, the regression function we developed and the scent rating results we obtained from our participants show that image scent in some instances may outweigh the scent of text. Although we did not find any evidence that these confounding images significantly affected user behaviour in SERPs in our study, previous research has found that text styling can increase users' efficiency but may also trick users into selecting the wrong result [21]. A careful investigation into the factors that could confuse users' sense of smell is warranted.

Finally, our results on user behaviour in SERPs in terms of effectiveness, efficiency and user satisfaction also yielded interesting insights. We found that our participants were not significantly better or quicker in completing search tasks when images were added than with traditional textual results pages, which contradicts previous research [14, 25, 26]. However, we also observed some very counter-intuitive trends when we added confusing scent cues: placing high-scent images with low-scent text made participants tend to select results slightly quicker. Future research employing larger field studies may want to shed more light on user effectiveness and efficiency in SERPs by investigating subtle interactions between scent strength and cue type, especially as users interactively refine their search needs.

6. CONCLUSION

This paper has investigated adding images to SERPs from an Information Foraging perspective. We conducted two empirical studies to explore the information scent of images and the effects of adding images to SERPs on user behaviour. In our first study, we found that:

Table 6. t-test result of timings

Comparison	t-test results
V2 - V3	$t(122) = 1.39, p = .16$
V2 - V4	$t(119) = -0.37, p = .72$

- Image information scent has user-dependent aspects and that users' scent ratings of the same image may differ.
- Overall, the combined scent of text and images is stronger than each on its own.
- Text has a stronger scent than images, however in some cases, the scent of images can outweigh the scent of text.

In addition, our second study showed that

- Adding images to SERPs may not have very much effect on changing user behaviour in terms of either effectiveness or efficiency, even when confounding images are presented.
- Overall, users appear to appreciate the addition of images to SERPs.

Our results have important implications for the design of SERP interfaces. We found that it will be extremely problematic to choose the "perfect" image to accompany textual results. In addition, although users may mainly rely on text cues to find search results, some images may throw users off the scent. Furthermore, it appears that there may not be a directly quantifiable benefit or detriment to users when adding images to SERPs – our participants did not select the correct results significantly better/worse, or slower/quicker. Further research into the basis of how users determine image information scent and the contribution of image scent to user behaviour in SERPs is therefore warranted. Our studies provide a valuable first step into image information scent and innovative SERP interfaces that take into account whether an image smells good.

7. ACKNOWLEDGMENTS

We thank the participants of our study. We thank Hannah Locke for her help in preparing this paper.

8. REFERENCES

- [1] Blackmon, M.H., Polson, P.G., Kitajima, M., and Lewis, C. 2002. Cognitive walkthrough for the web. *Proceedings of CHI 2002*.
- [2] Borlund P. 2003. The concept of relevance in IR. *Journal of the American Society for information Science and Technology*. 54(10):913-925.
- [3] Broder, A. 2002. A taxonomy of Web search. *SIGIR Forum*, 36(2).
- [4] Buscher, G., Dumais, S.T., and Cutrell, E. 2010. The good, the bad, and the random: An Eye-Tracking study of Ad Quality in Web Search. *Proceedings of SIGIR 2010*.
- [5] Chi, E., Pirolli, P., Chen, K., and Pitkow, J. 2001. Using information scent to model user information needs and actions on the web. *Proceedings of CHI 2001*.
- [6] Chi, E., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J. and Cousins, S. 2003. The Bloodhound Project: Automating discovery of web usability issues using the InfoScent Simulator. *Proceedings of CHI 2003*.
- [7] Chi, E.H., Pirolli, P., and Pitkow, J. 2000. The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. *Proceedings of CHI 2000*.
- [8] Cunningham SJ, Masoodian M. 2006. Looking for a picture: an analysis of everyday image information searching. *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*. Chapel Hill, NC, USA: ACM 2006. p. 198–199.
- [9] Cutrell, E. and Guan, Z. 2007. What are you looking for? An eye-tracking study of information usage in web search. *Proceedings of CHI 2007*.
- [10] Fu, W-T. and Pirolli, P. 2007. SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Proceedings of HCI 2007*.
- [11] Jiao, B., Yang, L., Xu, J., and Wu, F. 2010. Visual Summarizations of web pages. *Proceedings of SIGIR 2010*.
- [12] Katz, M. A., and Byrne, M. D. 2003. Effects of scent and breadth on use of site specific search on e-commerce web sites. *ACM Transactions on Computer-Human Interaction*, 10(3).
- [13] Kitajima, M., Blackmon, M. H., and Polson, P. 2005. Cognitive architecture for website design and usability evaluation: Comprehension of information scent in performing by exploration. *Proceedings of HCI International 2005*.
- [14] Al Maqbali, H., Scholer, F., Thom, J.A., and Wu, M. 2010. Evaluating the effectiveness of visual summaries for web search. *Proceedings of the Australasian Document Computing Symposium*.
- [15] Pirolli, P. and Card, S. 1995. Information Foraging in information access environments. *Proceedings of CHI 1995*.
- [16] Pirolli, P. and Card, S. 1999. Information Foraging. *Psychological Review*, 106(4).
- [17] Pirolli, P., Card, S., and Van Der Wege, M. 2001. Visual information foraging in a focus + context visualization. *Proceedings of CHI 2001*.
- [18] Pirolli, P. and Fu, W-T. 2003. SNIF-ACT: A model of information foraging on the World Wide Web. *Proceedings of User Modeling 2003*.
- [19] Pirolli, P. 2007. *Information Foraging Theory: Adaptive Interaction with Information*, Oxford University Press, New York.
- [20] Rele, R.S. and Duchowski, A.T. 2005. Using eye tracking to evaluate alternate search results interfaces. *Proceedings of HFES*.
- [21] Tamborello II, F. P. and Byrne, M. D. 2005. Information Search: The intersection of visual and semantic space. *Proceedings of CHI 2005*.
- [22] Wen L, Ruthven I, Borlund P. 2006. The effects on topic familiarity on online search behaviour and use of relevance criteria. *Advances in Information Retrieval 2006* 456-459.
- [23] Westman, S. 2009. Image Users' Needs and Searching Behaviour. *Information Retrieval: Searching in the 21st Century*, Goker and Davies (eds.), Wiley.
- [24] Withrow, J. 2002. Do your links stink? Techniques for good web information scent. *American Society for Information Science and Technology Bulletin*, 28(5).
- [25] Woodruff, A., Rosenholtz, R., Morrison, J.B., Faulring, A., and Pirolli, P. 2001. A comparison of the use of text summaries, plain thumbnails, and enhanced thumbnails for Web search tasks. *Journal of the American Society for Information Science and Technology*, 53(20).
- [26] Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J.B., and Pirolli, P. 2001. Using thumbnails to search the web. *Proceedings of CHI 2001*.