

Towards Measures and Models of Findability

Leif Azzopardi, Colin Wilkie
University of Glasgow
Glasgow, United Kingdom
{Leif.Azzopardi,
Colin.Wilkie}@glasgow.ac.uk

Tony Russell-Rose
UX Labs
Guildford, United Kingdom
tgr@uxlabs.co.uk

ABSTRACT

This poster paper outlines our current project which aims to develop models of how users search and browse for information. These will be used to create measures of findability, providing an estimate of how easily a document can be found within a given collection

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - *Search Process*

Terms Theory, Experimentation, Human Factors **Keywords** Retrieval Strategies, Simulation

1. PROJECT OVERVIEW

Designing and organizing a website or any other information space is a complex process where the goal is to create structure around content that enables users to complete their tasks (such as finding the relevant information and performing associated transactions) in a seamless and efficient manner. Typically, Information Architecture techniques are applied in an attempt to improve a site's usability and the overall user experience by optimizing "the structural design of an information space to facilitate task completion and intuitive access to content" [6]. Although there are numerous principles and heuristics that have been developed [6], Information Architecture, as a discipline, lacks formal models for evaluating or predicting whether such techniques will improve the usability of a website. In this project, we aim to develop formal models to measure, analyze and evaluate how easily a site can be navigated and the key resources within it can be retrieved. Such a model would provide a way to objectively measure what is colloquially termed "Ambient Findability", i.e. the ease with which a given information object can be found [6]. If the structure of a particular website precludes users from intuitively and easily locating key resources, then in competitive online environments users are likely to abandon the site in favour of alternative sites that provide competing services or information. For example, online retailers need to ensure that users can quickly and

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, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright is held by the author/owner(s). SIGIR 2013 Workshop on Modeling User Behavior for Information Retrieval Evaluation (MUBE 2013), August 1, 2013, Dublin, Ireland.

easily locate the products and services offered along with related information such as product reviews, help files, and other supporting information. Consequently, the ability to determine or predict the findability of information objects is vital.

While various measures have been developed to objectively quantify how easily a user can navigate a website, or how easily a user can retrieve a particular page [8, 12, 1, 2], these techniques are often very coarse grained in nature. They ignore the needs or goals of users in their calculations which directly influence how easily a user could locate an object. They are essentially agnostic to design, and thus ignore key issues such as layout, location and the visibility of links within web pages and the user's broader task context. Moreover, they assume each user is equally likely to select a given link (i.e. acting as a random surfer), and ignore well-known patterns and strategies of user behavior (see Figure 1). Thus, they provide little value to practicing Information Architects who require more sophisticated, behavioral measures that accommodate these dimensions. To this end, this project will attempt to model more accurately the interaction of users within websites to provide an estimate of how easily pages can be found based on specific user persona and scenarios (and thus the users underlying information needs and intentions).

2. BACKGROUND

Intuitively, the more findable content is, the more likely it is to be viewed and consumed, and vice versa such that: *if no one views your content/resource, then no-one will buy your products, solicit your services, or cite your papers*. The systems, structures and content used to provide relevant information to web users play a major role in shaping what information is findable. For example, when a user visits a website trying to find a particular product, they must either try to navigate to the product given the navigational structure, try to search for the product using the site-search provided, or undertake a combination of searching and browsing. If the structure is counter-intuitive or the site-search is not effective then the user will have difficulties in finding what they want. This often leads to abandoning the search, and the site, in favor of using other sites. As reported in [12], it was found that 56% of users were often confused by the structure of a website and had major difficulties in finding what they wanted. Sites that can ensure that the users are able to complete their goals and tasks in an efficient and effective manner are more likely to be successful [7].

However, a number of developments have been made recently which do provide ways in which particular aspects

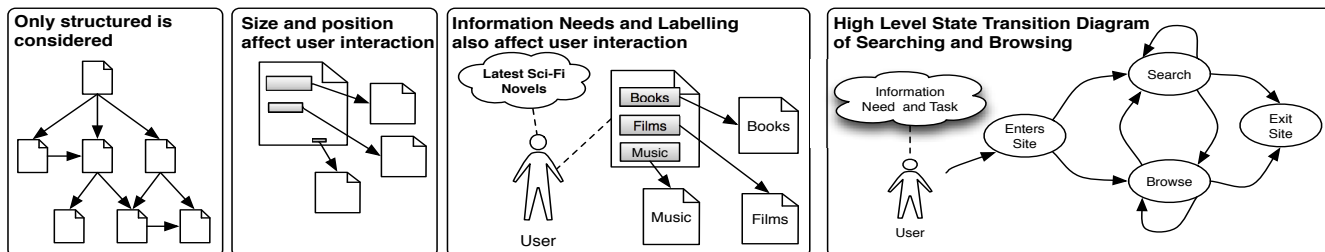


Figure 1: Left: Browsing assumed independent of content. Middle Left: Layout and Size affect browsing. Middle Right: Information Need affects browsing and searching. Right: High Level State Transition Diagram modeling a user with an information need, and their interactions within a website.

of “Ambient Findability” can be measured. One of the first measures employed was coverage which represents the amount of a site crawled by a search engine [5]. By ensuring that the information is visible to web crawlers, this increases the likelihood of the information being indexed by search systems. Dasgupta *et al* [4] referred to this as discoverability, while Upstill *et al* [9] referred to it as crawlability.

Once crawled and indexed, the findability then depends on the search engine itself, and how retrievable the search engine makes pages. Here retrievability measures were created to estimate how easily a document can be retrieved using a search engine [1, 2]. On the other hand, in [3, 8, 12], they considered findability in terms of browsing (i.e. navigability), and measured the ease with which users can navigate through websites¹.

3. PROJECT AIMS

In this project, our aims are two-fold:

1. Develop probabilistic models of how users interact within a website through searching and browsing
2. Develop measures that estimate how easily users can find content.

To this end, we have performed an initial analysis looking at the relationship between the usage of a site and various navigability and retrievability measures [11]. This was to evaluate the current baselines. Here we found that navigability measures held the strongest correlation with usage - providing a strong baseline. Our subsequent work will focus on making more sophisticated probabilistic models of interaction that use information needs, information cues, and page features to more accurately model browsing behavior (as shown in Figure 1). This is because these additional factors affect how users browse and search through a site. For example, if a link is at the top of a webpage it is more likely to be clicked than one at the bottom of a webpage, if the link is more prominent due to position, size and colour, it is more likely to be clicked. However, whether they perform such an action also depends upon the user’s information need and their underlying task. Consequently, our models will include each of these factors.

While in [10], we investigated the relationship between retrievability and retrieval performance. Here we found that the retrievability bias of system is correlated to retrieval performance - and that systems could be tuned using retrievability. We shall extend this work in two ways: (1) create a findability measure that considers both retrieval and navigation within the same model (i.e. combine navigability and

¹Note that PageRank and Hits can also be thought of as navigability measures that model a random surfer.

retrievability measures), and (2) perform a user study to determine whether this measure reflects the ease with which users can find certain pages by searching and browsing.

If we can validate our models and measures then it will be possible to develop tools to help Information Architects analyse websites: providing them with insights into how and what content is findable (and thus used), along with what features (terms/links/etc) makes pages easy or hard to find.

Acknowledgements: This work is supported by the EPSRC Project, *Models and Measures of Findability* (EP/K000330/1).

4. REFERENCES

- [1] L. Azzopardi and V. Vinay. Accessibility in information retrieval. In *Proc. of ECIR’08*, pages 482–489, Glasgow, 2008.
- [2] L. Azzopardi and V. Vinay. Retrievability: an evaluation measure for higher order information access tasks. In *Proc. of the 17th ACM CIKM’08*, pages 561–570, 2008.
- [3] E. H. Chi, P. Pirolli, and J. Pitkow. The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a web site. In *Proc. of the CHI’00*, pages 161–168, 2000.
- [4] A. Dasgupta, A. Ghosh, R. Kumar, C. Olston, S. Pandey, and A. Tomkins. The discoverability of the web. In *Proc. of WWW ’07*, pages 421–430, 2007.
- [5] S. Lawrence and C. L. Giles. Accessibility of information on the web. *Nature*, 400(6740):107–107, 1999.
- [6] P. Morville. *Ambient Findability*. O’Reilly Media Inc., Sebastopol, CA, USA, 2005.
- [7] J. Palmer. Web site usability, design and performance metrics. *Information Systems Research*, 13(2):151–167, 2002.
- [8] S. Pandit and C. Olston. Navigation-aided retrieval. In *Proc. of WWW ’07*, pages 391–400, 2007.
- [9] T. Upstill, N. Craswell, and D. Hawking. Buying bestsellers online: A case study in search & searchability. In *7th Australasian Document Computing Symposium*, Sydney, Australia, 2002.
- [10] C. Wilkie and L. Azzopardi. Relating retrievability, performance and length. In *To appear in SIGIR ’13*.
- [11] C. Wilkie and L. Azzopardi. An initial investigation on the relationship between usage and findability. In *Advances in Information Retrieval*, volume 7814, pages 808–811. 2013.
- [12] Y. Zhou, H. Leung, and P. Winoto. Mnnav: A markov model-based web site navigability measure. *IEEE Transactions on Software Engineering*, 33:869–890, 2007.