Page Hunt: Improving Search Engines Using Human Computation Games

Hao Ma *
Dept. of CSE
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
hma@cse.cuhk.edu.hk

Raman Chandrasekar, Chris Quirk
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
{ramanc, chrisq}@microsoft.com

Abhishek Gupta *
Digital Media, LCC
Georgia Institute of Technology
Atlanta, GA 30332, USA
abhishek.gupta@gatech.edu

ABSTRACT
There has been a lot of work on evaluating and improving the relevance of web search engines. In this paper, we suggest using human computation games to elicit data from players that can be used to improve search. We describe Page Hunt, a single-player game. The data elicited using Page Hunt has several applications including providing metadata for pages, providing query alterations for use in query refinement, and identifying ranking issues. We describe an experiment with over 340 game players, and highlight some interesting aspects of the data obtained.

Categories and Subject Descriptors: H.3.m [Information Storage and Retrieval]: Miscellaneous
General Terms: Design, Experimentation, Human Factors.
Keywords: Web Search, Human Computation Games, Query Alterations, Relevance

1. INTRODUCTION
Search engines have become an integral part of our everyday lives. It is clearly important to evaluate the "goodness" or relevance of search engines, and this is typically done using hand-annotated evaluation corpora or measures such as clicks on results. These methods do not help if specific pages never get surfaced due to a variety of reasons; these pages will not get evaluated, nor will they figure in click data.

To obviate problems in finding pages relevant to a given query, we look in the other direction: given a web page, what queries will effectively find this web page? To study a large number of pages, we propose to use games to collect relevant data.

As reported in [3], more than 200 million hours are spent each day playing computer and video games in the U.S. Human computation games seek to harness the time and energy that people spend on games to solve computational problems that are otherwise difficult to tackle. Human computation games were made popular by the ESP game [2], where people tackle an image labeling problem. In these (typically) two-player games, players are matched randomly to prevent collusion. The players collaborate to find 'true' answers for each task. When there are an odd number of players, these games typically use previously recorded (two-person) games; this raises some issues. In addition, collaborative games tend to elicit very general tags.

*This work was done when the authors were on summer internships at Microsoft Research.

SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.
ACM 978-1-60558-483-6/09/07.

In this paper, we describe Page Hunt, where the player hunts down a given web page, trying to get that page in the top N results in a search engine using her queries. 'Truth' is determined by the search engine used in the game. We avoid the issues mentioned above, primarily by using a single-player model. The data elicited from players can be used to provide metadata for pages, provide query alterations for use in query refinement, and identify ranking issues. We describe an experiment we conducted with over 300 game players, and highlight some aspects of the data obtained.

One challenge here is to create a game that players find entertaining and captivating. The research challenge is to ensure that we get useful data comparable to data from other sources, and to develop efficient methods to use this data to improve search.

Figure 1. The Page Hunt game, after a successful 'hunt', showing the web page being hunted, and the operating panel

2. PAGE HUNT DESIGN
The basic idea of Page Hunt is to show the player a random web page, and get the player to come up with a query that would bring up this page in the top few results on a search engine. Fig. 1 shows a screenshot of this game. The web page being 'hunted' is shown in the background. The player types in queries and looks at results returned in the floating operating panel, initially in the lower right corner. This panel turns almost transparent when not in focus. The border of the panel provides feedback through animation and color changes. Game play goes thus:

1. The player is shown a random web page with URL U. The player can see the web page, but not the URL of the page.

2. The player types in one or more words as a query Q.

3. The system gets the top N search results for this query from a search engine (Live Search, in this case) and displays them.

4. This match is successful if the URL U is in the top N results for the query Q. Players get points based on the rank of U in the result set. The player's score gets updated, and the game advances to the next web page. If query Q does not lead to success, the player changes the query and tries again. The game continues, page after page, till the player hits a time limit or quits.

At each step, we record the player's screen name (or an anonymous id), the web page URL, the query that was tried, whether it was successful and if so at what rank position, the time, and the points the player got for this query. The player's query behaves as a tag or label for the page. When the player gets it right, this is valuable; but even when it is wrong, the label can be useful.

A transparent overlay on the web page prevents players from cutting and pasting long phrases from the page as queries. Players can type in long phrases, but they learn to be discriminating.

Page Hunt includes several features to increase fun in game playing, such as using timed responses, score keeping, having a top-scorers list, taboo queries, and bonus points.

The web pages we use in Page Hunt are those that have been identified by a search engine as requiring better labels, additional metadata, or have ranking issues.

3. PILOT EXPERIMENT

Before we released the game to the web, we conducted a pilot study within our company. During this experiment, 341 people played Page Hunt over a period of ten days, generating over 14,400 labels on 744 web pages in the system. On average every player contributed 42 labels, and every page has about 19 labels.

Using a session limit of 10 minutes of inactivity, we had a total of 681 sessions. Of these, 18% (123 sessions) were by anonymous players. About 47% of the sessions were from people who played 2 or more sessions. 16% of the people had 5 or more sessions, and only 240 sessions (35% of the total number of sessions) were single session players.

3.1 URL Findability Levels

From the data collected, we find that some of the pages are easily found (or 'hunted' down), while some others are difficult to get in the top search results. A 100% findability level indicates that a page can be easily located since every person shown this web page was able to bring up this page to the top 5 search results. In our experiment, about 27% of the pages in our database have 100% findability while roughly the same number of pages (26%) has 0% findability. We investigated this further and found that as the length of URL of the page (in characters) increases, the pages are harder to find. See Fig. 2, which shows the findability distribution of the pages as a function of URL length.

Findability can be used to compare goodness of search engines, and we intend to work further in this area.

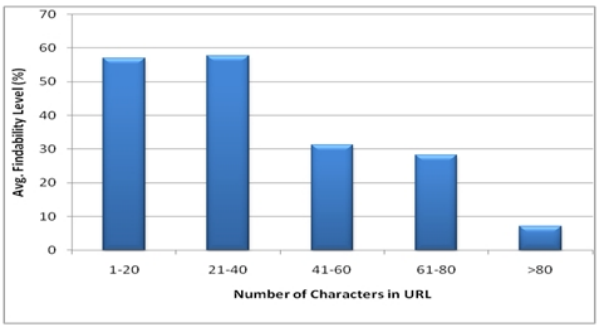

Figure 2. Findability as a function of URL Length

3.2 Query Alterations from Game Data

On most search engines, if you search for, say, [Wash. DC], the query is internally modified as if it were equivalent to a search on [Wash. DC] ORed with [Washington, DC]. This idea of modifying the user's query using other terms which may have been intended is called query alteration. In this section, we use bitext matching [1] to learn alterations from our game data.

For all web pages (URLs) in the Page Hunt pilot data, we extracted the queries that corresponded to winning trials, generated all pairs of queries as bitext data, and applied the bitext matching algorithm. This results in a large table of alterations (phrase pairs) with weights corresponding to the likelihood of each alteration. If we take the top scoring alterations, we find that they can be categorized into 4 bins, the last two of which are very promising:

1. Spelling or punctuation alterations: e.g. *JC Penny* vs. *JC Penney* or *J C Penny*.
2. Sitename to site alterations: *cbs4* vs. *cbs4.com*.
3. Acronym/Initialism-Expansion alterations: We can learn that *iht* is the International *Herald Tribune* etc.
4. Conceptual alterations: E.g. *capital city airport* is a valid alteration for *Kentucky airport*, and *jlo* for *Jennifer Lopez*.

Our plan is to garner additional data, improve the bitext ranking process, and run our bitext code on the new data to generate alterations. These alterations can then be used in a search engine to see if they cause an improvement in relevance measures.

4. CONCLUSIONS

In this paper, we present a page-centric approach to improving Web search using human computation games. We provided a description of Page Hunt and some applications of this game. To the best of our knowledge, Page Hunt is the first single-player non-collaborative human computation game. We plan to continue our work, creating games that can be used to improve search.

5. REFERENCES

[1] C. Quirk, C. Brockett, and W. B. Dolan. Monolingual Machine Translation for Paraphrase Generation. In Proc. EMNLP 2004, Barcelona, Spain, 2004.

[2] L. von Ahn and L. Dabbish. Labeling images with a computer game. In Proc. CHI '04, pages 319-326, New York, NY, USA, 2004. ACM.

[3] L. von Ahn and L. Dabbish. Designing games with a purpose. Communications of the ACM, 51(8):58-67, 2008.