A Game with Many Purposes: Improving Information Retrieval Through Pleasurable Competition

Miles Efron
Graduate School of Library and Information Science
University of Illinois, Urbana-Champaign
501 E. Daniel St., Champaign, IL, 61820
mefron@illinois.edu

ABSTRACT

Besides an obvious instrumental motivation, people often engage in information seeking because it is pleasurable. Recent projects by major search engine companies bear this out. For many people, finding information is fun. This paper describes a system that is currently under development whose design is predicated on this idea. The system is both a collaborative information seeking environment and an interactive game. The game has two goals: it supports collaborative information retrieval while aiming to help its players become more skillful users of IR systems in general.

Categories and Subject Descriptors

H.5 [**Miscellaneous**]: Interaction Framework

General Terms

Human Factors

Keywords

Information retrieval, collaborative information seeking, human computation, games with a purpose

1. INTRODUCTION

Recent work by major search engine companies suggests that people use information retrieval (IR) systems not only to find information when they need it, but also as an avenue for deriving pleasure. The daily *New York Times* puzzle *A Google a Day*[1] presents solvers with an information problem that they must solve by crafting Google searches skillfully. In a similar vein, Microsoft's now defunct *Page Hunt* game allowed people to use the Bing search engine for play. Page Hunt players were presented with a random website. The object of the game was to submit a query that would rank that page in the results' top position. The idea that information seeking is often pleasurable in its own right has been explored in recent literature as well [2, 9].

In this paper I describe a game that is currently under development. Details of the game are described in Section 3. However, the game is intended to have two positive effects with respect to information seeking and information retrieval:

1. Provide a platform for collaborative information seeking.
2. Help people become more skillful users of search engines.

2. MOTIVATION

The motivation underpinning this work's system design stems from the literature of so-called games with a purposes (GWAP) [8]. Games with a purpose comprise a sub-domain of the more general paradigm of "human computation" [7]. Human computation brings the attention of people to bear on problems that computers struggle with. Crowdsourcing platforms such as Amazon's Mechanical Turk and Crowdflower exemplify this approach to work allocation[2].

Like these crowdsourcing systems, games with a purpose divide a large task into many small tasks, each of which is solvable with minimal effort. Additionally, games with a purpose entail an environment where "working" towards the system's goal is intentionally fun. Instead of compensating workers monetarily, games with a purpose typically offer pleasure as an incentive for people to work towards a common goal. Games with a purpose–and human computation more generally–have begun to play a role in IR and related domains [1, 3, 4, 5].

While this work looks to human computation as a general design principle, the problem I address derives from a different line of work. In particular, recent research has focused on helping people become better searchers (i.e. users of IR systems) using minimal intervention [6]. While most IR research has focused on making better search engines, we can alternatively improve search effectiveness by helping people become better searchers. That is the rationale behind goal 2 listed above. The proposed game is intended to help people learn to search effectively.

Though a chief motivation behind the proposed game is helping people become better searchers, the game is also intended to comprise a novel platform for collaborative IR.

[1] `http://googleaday.com`

[2] `http://mturk.com`, `http://crowdflower.com`

Collaborative Information Seeking–Bridging the Gap Between Theory and Practice. ASIS&T 2011 New Orleans, LA USA

The outcome of ongoing game play is a "persistent search" that has been created by mediated competition among individuals. Thus the game is instructive and instrumental insofar as it enlists competitive play to build a high-quality document ranking in response to a single information need.

3. THE FLOW OF GAME PLAY

The game defines two types of searches: *seed* searches and *seek* searches. Corresponding to these searches, people interact with the game in one of two roles (or possibly both): *seeders* and *seekers*. A game is an ongoing process. It is initiated when a person, in the role of a seeder, creates a seed search. This is a web query on a topic of his or her choice. The seeder then identifies a returned document that is highly relevant to the seed query. The document marked as relevant takes on a special role as the *target* of seeking.

This seed search drives the game from that point forward, acting as a type of persistent query. After seeding a game, all future interactions are of the seeking variety.

Once a seed and a target have been established, people play the game by issuing queries that attempt to rank the target document as highly as possible. As they participate, players create an evolving set of search results for the initial query.

Figures 1 and 2 outline the unfolding of a game. During the flow shown in Figure 1 a person with an information need creates a *seed* consisting of a query and a relevant document. Putatively, this seeder would like other players of the game to help build a high-quality set of search results for the seed query.

During the flow shown in Figure 2, many players in the role of *seekers* craft queries in response to a given seed. Seekers find seed queries that interest them through the game system's interface (by browsing, searching, etc.). The goal of the seekers is to create queries that score highly. Query scores are a function of their relationship to the target document and the results found so far by other seekers.

As shown at the bottom of Figure 2, the collective action of the seekers creates a ranking of documents related to the seed. If the game works appropriately, this ranked list should contain highly relevant documents. Additionally, the system displays those queries that have been most "successful" in the game so far. Thus the system presents results for the original seeder, as well as search strategies that yield good results.

3.1 The Rules of the Game

Of course it is easy to make a query that retrieves almost any page near the top of a search engine's results. To move the game forward, then, each *seek* query wins a numerical score according to the rules outlined below. Over time, people accrue points in the system as a whole through the accumulation of their individual query scores.

Two goals guide scoring in the game. High-scoring queries will:

1. retrieve the target document at a high rank
2. retrieve documents not already seen during the game.

Goals 1 and 2 present a fundamental tension. *Seek* queries must remain on-topic (c.f. goal 1). But if they are to score well, these queries must also retrieve "fresh" information (i.e. documents not already found by other seekers). To put it in terms of traditional IR, goal 1 rewards precision, while goal 2 rewards recall.

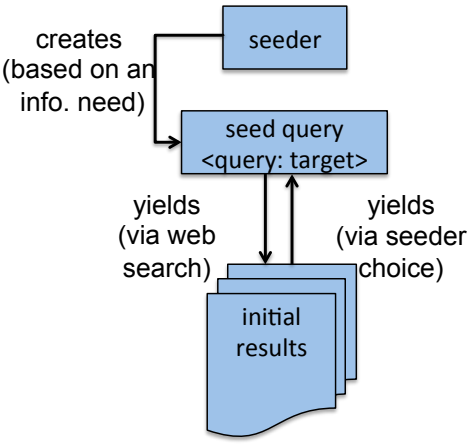

Figure 1: Flowchart of the first (seeding) part of the game. During seeding, a seed query and an associated target are defined by a seeder. The remainder of the game involves repeated queries in service to the seed.

The intuition behind goal 2 is that we reward seekers who try something different than other players. Of course as time passes, crafting queries that succeed on both goals 1 and 2 becomes more difficult. Thus points accrue more quickly as the game proceeds. As a game's search space becomes more saturated it becomes harder to articulate a novel query that still ranks the target highly. Thus searchers who participate in later stages of a game stand to win more points than those who play when the game is easiest (at the beginning).

Before defining the rules of the game, we note that when a seeker poses a query to the system, results are obtained from a standard Web search engine. For each query, we retrieve a search engine result page (SERP) containing s results. For purposes of illustration, we set $s = 100$.

With these mandates in place, for a game G with target T, a given *seek* query Q issued at time t is awarded a score according to:

$$score(Q, t, G) = p^{-1} * \frac{t^{\alpha} * n(R_Q \notin R_G^t)}{s} \qquad (1)$$

where p is the rank of T in the search results for Q, t is the number of *seek* searches performed so far for G, $\alpha \geq 0$ is a tunable parameter that guides the extent to which later *seek* queries are rewarded, and $n(R_Q \notin R_G^t)$ is the number of documents within the SERP retrieved by Q that have not yet been seen among the retrieved documents for G (i.e. R_G^t is the set of documents retrieved during the game so far).

Figure 3 outlines four hypothetical scoring scenarios for a given seek query Q issued at time t in response to target T. Three of the panels show scores derived under the case when $p = 1$; that is, Q_t retrieves the target at position 1. The upper left panel presents "worst case" performance for $p = 1$. This would occur if a player issued the original seed query over and over for times $t = 1 \ldots 100$. Here, no "fresh" documents are found from any seek query. All documents were found by the seeder during the process of initiating the game, so the score remains 0. In contrast, the upper right

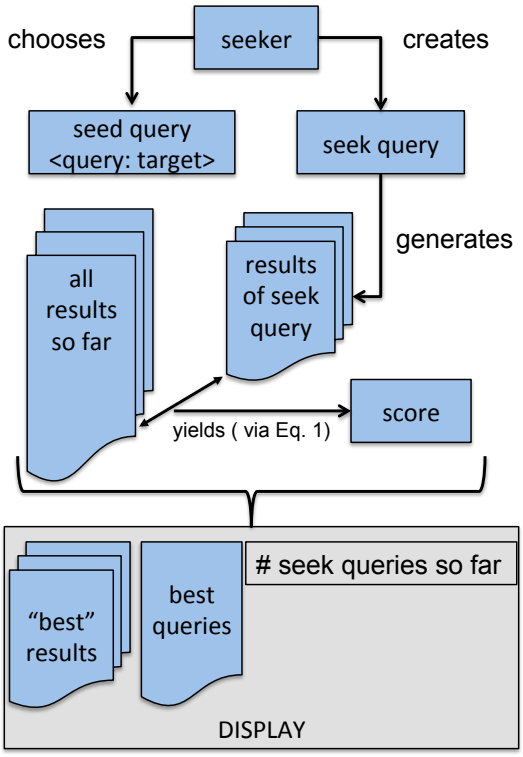

Figure 2: Flowchart of the second (seeking) part of the game. During seeking, multiple players issue "seek" queries in attempts to score points. Collectively, this activity generates the display shown at the base of the figure.

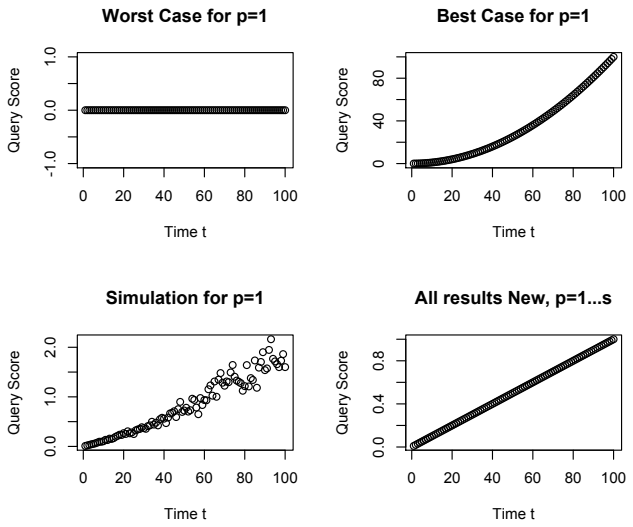

Figure 3: Four Scoring Scenarios Using Eq. 1. In each panel, the x-axis is the number of seek queries issued so far (i.e. time t) and the y-axis is the resulting query score.

panel shows best-case results. Again, for all Q_t, $p = 1$. Additionally, in all cases, all of Q_t's s retrieved documents are as yet unseen, contributing s new results to the collective retrieved set. The lower left panel results from a simulation where p is held at 1 but retrieving fresh documents becomes more difficult as t increases. This is the most plausible distribution of scores presented here. Finally, the lower right panel shows scores from queries where all retrieved documents are fresh, but p decreases linearly with t. That is, in this case, the queries are not so off-base that they fail to return T. They also succeed in pulling fresh content into the cumulative pool. But they suffer from increasing saturation of the search space.

An interesting dynamic emerges from Figure 3: the role that time plays in query scoring. Eq. 1 rewards participation late in the game according to the exponent α. This leads to the somewhat counterintuitive result shown in the figure's lower right panel. Even though queries in this panel are getting worse insofar as they rank the target T lower and lower, they achieve increasing scores by virtue of the temporal reward. Tuning α guides the extent to which we reward queries undertaken during a game's later stages. Higher values of α will lead to a steeper exponential increase in the case shown in the figure's upper right panel, also increasing the (admittedly low) magnitude of scores in the lower right.

I hypothesize that this temporal reward is crucial for the outcome of the game to be successful and for the game to present a plausibly entertaining challenge. Queries issued early in the game are likely to be obvious and easy to formulate. However, as the game unfolds, players need to show greater skill in crafting queries. While these queries are harder to devise, they hold the potential for an exponentially higher payoff than queries submitted at the game's early stages. Thus the game rewards ingenuity very highly.

4. OUTCOMES OF THE GAME

As Figure 2 shows, at any given time the game's interface displays two collectively created sets of data. First, users can see the "best" documents found by the group of seekers. I define document quality below. Second, users also see the best queries issued so far during the game. The ranked list of the game's best documents ostensibly serves the needs of the initial seeder. Additionally, this listing would be persistent, remaining web-accessible and indexible by search engines for discovery by other people with an interest in the seed topic.

The list of high-quality queries is intended to communicate effective search strategies. The motivation for showing these queries is to demonstrate to players (who are also likely to be search engine users in their own right) how sophisticated searchers solve a challenging information retrieval problem. The hope is that players will, over time, develop skills that will serve them in pursuit of their own information needs outside the realm of the game.

To create these two lists we need criteria for assessing document and query quality with respect to a game G[3]. A simple count of the frequency with which a document or query has appeared during game play will lead to a tyranny of the majority problem, where results are dominated by documents that are easy to find and queries that are obvi-

[3]These lists will evolve as the game unfolds. They are not static.

ous in their construction. Instead, we might consider two approaches to defining quality. At time t during game play, a potential value for a document D in the pool of retrieved results is:

$$V_D(D,t,G) = \sum_{Q \to D} Score(Q,t,D) \quad (2)$$

where $Q \to D$ is the set of queries that return D and the sum is taken over the corresponding query scores from Eq. 1. Analogously, we can define the value of a query:

$$V_Q(Q,t,G) = \sum Score(Q,t,D). \quad (3)$$

Eqs. 2 and 3 could be normalized to reduce the effect of document (query) popularity, as well.

A more sophisticated measure of document and query value would define terms recursively such that valuation is computed by:

$$v_D(D,t,G) = \sum_{Q \to D} v_Q(Q,t,G) \quad (4)$$

and:

$$v_Q(Q,t,G) = \sum_{D \leftarrow Q} v(D,t,G) \quad (5)$$

where $D \leftarrow Q$ is the set of documents retrieved by Q. Here we take the idea that good queries return good documents and good documents are retrieved by good queries. Values for documents and queries could then be obtained by applying Eqs. 4 and 5 iteratively until convergence. This amounts to finding the steady state of a Markov process as is familiar from many bibliometric and hyperlink analysis algorithms.

5. RESEARCH QUESTIONS

This paper is highly speculative. But the motivations outlined in Section 2 raise several tractable research questions that I am currently pursuing. While these questions are addressable via empirical research, they are also, I hope, worth discussing less formally. In the context of the proposed game, we might ask:

- What productive roles can pleasure and competition play in collaborative IR?
- Does the query scoring function of Eq. 1 capture appropriate dynamics of game play?
- Does the game outlined here realistically hold the possibility of creating a valuable set of evolving search results for a seed query?
- Does interaction with the proposed game hold a realistic possibility of helping people become more skillful at searching in general?
- How could we measure the success of the proposed game at accomplishing its stated goals? What are suitable benchmarks for comparison?
- Does the proposed game offer an incentive that would draw a critical mass of players if deployed in the wild? For that matter, how many players are needed to make the game successful?

Of course, more general questions regarding the value of games with a purpose in collaborative information seeking also present themselves in the context of this brief discussion. I leave those for ongoing dialogue.

6. REFERENCES

[1] M. Ageev, Q. Guo, D. Lagun, and E. Agichtein. Find it if you can: a game for modeling different types of web search success using interaction data. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, SIGIR '11, pages 345–354, New York, NY, USA, 2011. ACM.

[2] D. Elsweiler, S. Mandl, and B. Kirkegaard Lunn. Understanding casual-leisure information needs: a diary study in the context of television viewing. In *Proceeding of the third symposium on Information interaction in context*, IIiX '10, pages 25–34, New York, NY, USA, 2010. ACM.

[3] M. Krause and H. Aras. Playful tagging: folksonomy generation using online games. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 1207–1208, New York, NY, USA, 2009. ACM.

[4] E. Law, P. N. Bennett, and E. Horvitz. The effects of choice in routing relevance judgments. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, SIGIR '11, pages 1127–1128, New York, NY, USA, 2011. ACM.

[5] H. Ma, R. Chandrasekar, C. Quirk, and A. Gupta. Improving search engines using human computation games. In *Proceeding of the 18th ACM conference on Information and knowledge management*, CIKM '09, pages 275–284, New York, NY, USA, 2009. ACM.

[6] N. Moraveji, D. Russell, J. Bien, and D. Mease. Measuring improvement in user search performance resulting from optimal search tips. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, SIGIR '11, pages 355–364, New York, NY, USA, 2011. ACM.

[7] L. von Ahn. Human computation. In *Proceedings of the 46th Annual Design Automation Conference*, DAC '09, pages 418–419, New York, NY, USA, 2009. ACM.

[8] L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51:58–67, August 2008.

[9] M. L. Wilson and D. Elsweiler. Casual-leisure searching: the exploratory search scenarios that break our current models. In *4th International Workshop on Human-Computer Interaction and Information Retrieval*, pages 99–102, 2010.