Intentions: A Game for Classifying Search Query Intent

Edith Law

Carnegie Mellon University

5000 Forbes Ave

Pittsburgh, PA 15217 USA

edith@cmu.edu

Anton Mityagin

Microsoft Live Labs

One Microsoft Way

Redmond, WA 98052 USA

mityagin@microsoft.com

Max Chickering

Microsoft Live Labs

One Microsoft Way

Redmond, WA 98052 USA

dmax@microsoft.com

Abstract

Knowing the intent of a search query allows for more intelligent ways of retrieving relevant search results. Most of the recent work on automatic detection of query intent uses supervised learning methods that require a substantial amount of labeled data; manually collecting such data is often time-consuming and costly. Human computation is an active research area that includes studies of how to build online games that people enjoy playing, while in the process providing the system with useful data. In this work, we present the design principles behind a new game called *Intentions*, which aims to collect data about the intent behind search queries.

Keywords

Human Computation Game, Query Classification, Query Intent, Web Search

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

The classic article of [2] classified the intent behind search queries into three categories: navigational (to reach a particular site), informational (to gather information from one or more web pages) and transactional (to perform some web-mediated activities). In a study of a very large query log with

millions of queries [3], it was shown that more than 80% of web queries are informational in intent. Informational queries tend to involve more complex natural language phrases and a larger number of query resubmissions [3] than queries in the other two categories. This suggests that there is significant upside potential for better intention-detection algorithms in the search engine. In particular, if we could better detect the intention behind the often ambiguous queries, we might be able to provide a more efficient search experience that requires fewer query reformulations. For example, it would be useful to know that "how to make a pumpkin pie" corresponds to an intent to search for recipes and that "stores that have Wii consoles left" indicates an intent to buy a Wii console.

Query logs are a rich source of information for analyzing the intent of the most common search queries. Despite the availability of these logs, however, researchers still have to manually label hundreds [3,5,6] or thousands [1] of search queries in order to construct either the training data from which they create algorithms for predicting query intent or the ground-truth data against which they evaluate these algorithms.

The idea of human computation [7] is to gather useful labeled data quickly by involving people in an enjoyable task, such as playing an online game. For example, in the ESP Game [7], two players are asked to describe a common image. If their descriptions match, that description becomes a label for the image. The ESP Game (which is now also adapted as the Google Image Labeler) continues to be played on a daily basis, generating a huge number of tags to power image search. Since the inception of the ESP Game, many human computation games have been developed to tackle the shortage of labeled data.

In this work, we present *Intentions*, a game that collects data useful for understanding the intent of search queries. We describe some of its design

elements and discuss the particular needs and challenges associated with adapting the input-agreement mechanism [4] for this game, or generally, any games where the data is text-based.

Intentions to Queries: The Reverse Problem
In the absence of rich context, the intent of a search query is often ambiguous. For example, the underlying intent behind the search query "Greyhound Bus" could be to download the latest bus schedules, to check the latest stock price for the company, to locate a bus stop in a particular city, to find the customer service phone number, etc. As a result of this ambiguity, labeling the intent behind search queries is often cognitively difficult and prone to subjectivity [1].

Although learning the exact intent behind a search query is difficult, if not impossible, there may exist a set of intentions with which the particular query is most commonly associated. For example, it could be the case that the intent behind "Greyhound Bus" is almost always to download the latest bus schedule. In the original problem of mapping queries to intentions, it is difficult to infer these biases—in this particular example, the word "schedule" does not even appear in the search query.

Instead of asking users to label the intent behind a given search query, we consider the reverse problem: given a specific intention, we ask users to come up with search queries that will satisfy that intention. Intentions are phrased in the form of questions that users hypothetically have in mind when they perform the search. For example, instead of asking the users to label the intention of the search query "Greyhound Bus," we provide the question "What is the schedule for the Greyhound Bus?" and ask users to type a search query to retrieve search results that might help answer that question. Consequently, we may find that the query "Greyhound Bus" is most commonly associated with the intent to find schedules. Solving this reverse
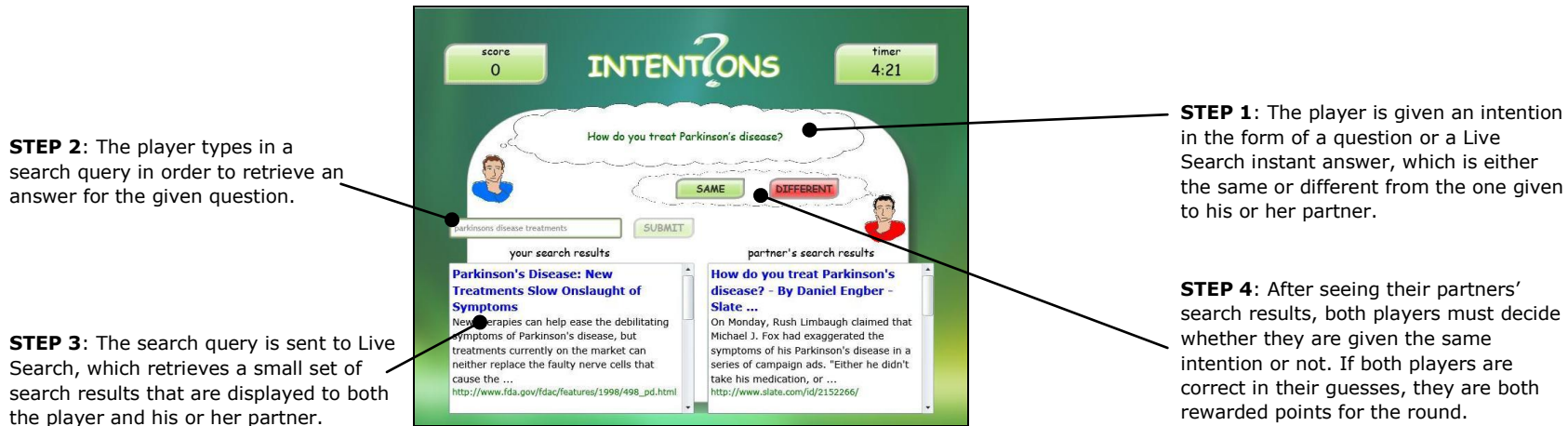
STEP 1: The player is given an intention in the form of a question or a Live Search instant answer, which is either the same or different from the one given to his or her partner.

STEP 2: The player types in a search query in order to retrieve an answer for the given question.

STEP 3: The search query is sent to Live Search, which retrieves a small set of search results that are displayed to both the player and his or her partner.

STEP 4: After seeing their partners' search results, both players must decide whether they are given the same intention or not. If both players are correct in their guesses, they are both rewarded points for the round.

Figure 1. Interface of *Intentions*

problem allows one to collect a set of intent-to-query mappings, which can then be used to tackle the original problem of inferring the intent behind a particular search query.

Intentions: Basic Game Mechanism

We introduce a multi-player online game called *Intentions* that will collect, for each of a number of given intentions, a set of search queries that will help satisfy those intentions. The game randomly matches a player with an anonymous partner and proceeds as follows. In each round, players are first given either the same intention or different intentions. Each player then types in one or more search queries that represent his intention and the corresponding search results from Microsoft Live Search (not the original search queries) are displayed to his partner. Upon seeing their partners' search results, both players must guess whether or not they share the same intention. If both players guess correctly, they are both rewarded with points; otherwise they *both* lose the round and get zero points. This scoring mechanism promotes cooperation and

ensures that both players have the incentive to provide queries that indeed match their given intention.

There are two methods by which we express the intentions to the players. The first method is to provide an explicit textual question, and the intention is to obtain an answer to that question using a search engine. For example, in Figure 1, the intention is to find information about treatments for Parkinson's disease, and the user's search query is "parkinsons disease treatments." In our current game implementation, intentions are expressed in the form of questions.

The second method of expressing an intention in the game, which we have not yet implemented, is to provide each player with a Live Search *instant answer*, which is a concise result for some informational need. Figure 2 shows examples of some instant answers currently provided by Live Search, which include weather information for a city, the area code for a city, stock quotes, movie show times, facts about planets, product information, dates for major holidays, news snippets, maps, and celebrity profiles.
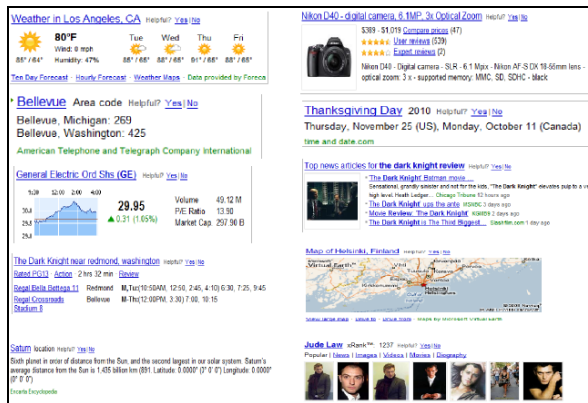
Figure 2. Examples of Instant Answers

The data gathered by the game can be used to learn a grammar (i.e., a set of linguistic patterns) of how people express intentions in search queries. For example, while "what is the weather like in Seattle" and "weather forecast in Seattle" are queries for finding out about the weather in Seattle, these same linguistic patterns can be used to detect the intention of seeking information about the weather of any other cities. Perfecting this grammar means that the intent behind queries can be more easily and accurately classified in real time, allowing users to get a quick and precise answers for the questions they have in mind.

Obfuscated Input-Agreement Mechanism

Whereas the specific design choices made for individual human computation games vary significantly, all of these games tend to fall into a remarkably small number of general mechanisms of input-output behavior. The recent music-annotation game TagATune [4], for example, introduces a mechanism called *input-agreement* [4]. In TagATune, two players are given either the same song or different songs. After describing their given songs to each other, the players must decide whether they have been given the same song or not. More generally, this mechanism can be

applied to games where the input data is not audio, but images, videos or text.

When we apply the input-agreement mechanism to *Intentions,* two issues emerge. First, motivated to win, players can cheat by typing in their given question exactly, thereby generating data that is essentially useless. Second, by revealing the outputs of the players (i.e. search queries) to their partners, the task of recovering the original question from the search queries becomes trivial. For example, it would be quite easy to guess that the question given to both players is "When was Michael Jackson born?" when the search queries "Michael Jackson birth date" and "Date of birth Michael Jackson" are revealed. This makes the game too easy to be fun.

To tackle these issues, our design of *Intentions* includes two important modifications to the input-agreement mechanism. First, we transform the questions given to the players so that even if they attempt to collude by typing in the questions exactly, the task of guessing correctly whether the pair of questions is the same is non-trivial. For example, instead of giving the same question "what are the symptoms for lung cancer?" to the two players, that question is transformed into a pair of questions that are equivalent paraphrases of each other (e.g. "How do I know if I have lung cancer?", "What are some signs of lung cancer"), which are then displayed to the players. The paraphrasing of questions ensures that even if both players copy the questions exactly in their submitted search queries, some discrimination is required of the players in order for them to arrive at the correct answer.

To address the second issue, the outputs of the players (i.e., search queries) are transformed into another format (i.e., search results) before they are revealed to each other. This has the effect of requiring players to perform a discrimination task that is more involved: they must guess the input of their partner by analyzing the subtleties of the more complicated output (search

results) as opposed to the more simplistic output (search queries) that was submitted by the partner. It is important to note that this transformation must be "correct", i.e. the search engine must be capable of returning search results that are at least somewhat relevant to a given search query; otherwise, players would have little chance of recovering the original question by analyzing the search results, causing them to perceive the game as being inconsistent or unfair.
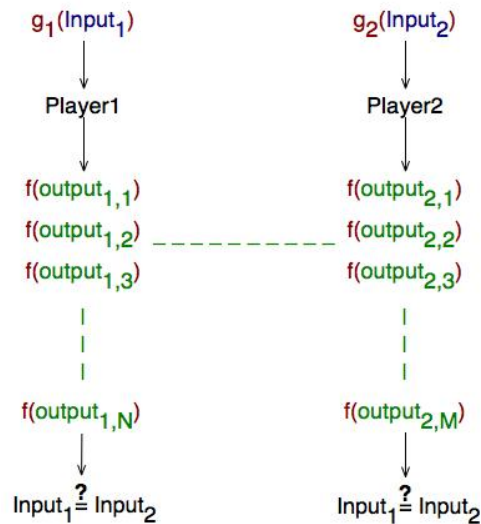

$g_1(\text{Input}_1)$ $g_2(\text{Input}_2)$

Player1   Player2

$f(\text{output}_{1,1})$ $f(\text{output}_{2,1})$
$f(\text{output}_{1,2})$ — — — — — — $f(\text{output}_{2,2})$
$f(\text{output}_{1,3})$ $f(\text{output}_{2,3})$

$f(\text{output}_{1,N})$ $f(\text{output}_{2,M})$

$\text{Input}_1 \overset{?}{=} \text{Input}_2$ $\text{Input}_1 \overset{?}{=} \text{Input}_2$

Figure 3. Obfuscated Input-Agreement Mechanism

We refer to these transformations of the input and output data by the general term *obfuscation*. The mechanism underlying *Intentions*, which we call *obfuscated input-agreement*, is depicted in Figure 3.

Choosing Appropriate Intention Questions
Another important design element of input-agreement games is the selection of pairs of input data. In TagATune, for example, if the pair of songs given to the

players are different yet very similar, the game will be difficult. Likewise, in *Intentions*, the similarity of the questions in a pair directly influences the difficulty level of the game. For example, it is more difficult to tell that the questions are different if they are about the same subject (e.g. "How much do the tickets cost for *Superbowl 2009*?" and "Who won the *Superbowl 2009*?") than if they are about different subjects (e.g. "Where do I buy a *digital SLR*?", "Where do I find reviews for *Shrek 3*?").

Our current strategy for generating a large number of pairs of intention questions is as follows. We manually constructed a set of predefined topics, e.g., sports, medical procedures, companies, movies, celebrities, drug-condition interactions, products, etc. For each topic, we created a set of question templates, each of which is associated with a question type. Figure 4 shows some examples of question templates.

Questions with the same topic and question type (e.g., "What is the price of <drug>?", "How much does <drug> cost?") are paraphrases of each other and are considered to be representations of the same intention.

To construct pairs of questions that are different but easily distinguishable, one can sample (1) questions where the topics are different (e.g. "What are the effects of <drug>?" and "What is the latest stock price of <company>?"), or (2) questions with the same topic and question type, but with different entities substituted (e.g. "Can I take *Tylenol* when I *just had a baby*?", "Can I take *Viagra* when I *have a heart condition*?").

To generate questions that are different but more difficult to discriminate, one can select questions where the subjects are the same, but that the information sought about the subject is different. These are questions with the same topic and entity substituted, but with different question types, e.g. "Which Oscar award did the movie *Good Will Hunting* win?" and

Topic	Question type	Question Template
Drugs	Effects	What are the effects of <drug>?
	Cost	What is the price of <drug>? How much does <drug> cost?
Drug-Condition	Safety	Can I take <drug> when I <have-condition>?
Companies	Personnel	Who is the biggest individual shareholder of <company>?
	Stock	What is the latest stock price of <company>?
Movies	Award	Which Oscar award did the movie <movie> win?
	Cast	Who is in the cast of <movie>? Who are the actors and actresses in <movie>?

Figure 4. Question Templates

"Who is in the cast of *Good Will Hunting*?" In this scenario, players are each given an intention to find some information about the movie *Good Will Hunting*; however, the particular kind of information sought—i.e. award versus casting—differs. In order for the players to tell that the questions given to them are different, they must be able to judge from the search results that the questions are about different aspects of the same subject.

To make the game consistently challenging and fun, there should be an appropriate mix of easy and difficult pairs of intentions served in the game. This can be determined dynamically by observing the number of mistakes players have made so far, and selecting the pairs of questions accordingly.

Conclusion

In this paper, we highlighted the design principles behind *Intentions*, a human computation game that gathers labeled data about search query intent. Future work includes the deployment of the game to the general public, a thorough user study that determines the enjoyability of the game, an analysis of the quality of the collected data and the development of algorithms for automatically generating question templates.

References

[1] Baeza-Yates, R., Calderon-Benavides, L. and Gonzalez-Caro, C. The intention behind web queries. In *Proc. of SPIRE*. 98–109, 2006.

[2] Broder, A. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.

[3] Jansen, B.J., Booth, D.L. and Spink, A. Determining the informational, navigational and transactional intent of web queries. *SIGIR Forum*, 44(3):1251–1266, 2008.

[4] Law, E. and von Ahn, L. Input-agreement: A New Mechanism for Collecting Data Using Human Computation Games. To appear in *Proc. of CHI*. ACM Press, 2009.

[5] Lee, U., Liu, Z. and Cho, J. Automatic identification of user goals in web search. In *Proc. of WWW*. 391–401, 2005.

[6] Rose, D. E. and Levinson, D. Understanding user goals in web search. In *Proc. of WWW*. 13–19, 2004.

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