

# Verbosity: A Game for Collecting Common-Sense Facts

Luis von Ahn, Mihir Kedia and Manuel Blum

Computer Science Department, Carnegie Mellon University

5000 Forbes Avenue, Pittsburgh PA 15213

biglou@cs.cmu.edu, majin@cmu.edu, mblum@cs.cmu.edu

## ABSTRACT

We address the problem of collecting a database of “common-sense facts” using a computer game. Informally, a common-sense fact is a true statement about the world that is known to most humans: “milk is white,” “touching hot metal hurts,” etc. Several efforts have been devoted to collecting common-sense knowledge for the purpose of making computer programs more intelligent. Such efforts, however, have not succeeded in amassing enough data because the manual process of entering these facts is tedious. We therefore introduce Verbosity, a novel interactive system in the form of an enjoyable game. People play Verbosity because it is fun, and as a side effect of them playing, we collect accurate common-sense knowledge. Verbosity is an example of a game that not only brings people together for leisure, but also collects useful data for computer science.

## Author Keywords

Distributed knowledge acquisition, Web-based games, common-sense reasoning.

## ACM Classification Keywords

I.2.6 [Learning]: Knowledge acquisition. H5.3 [HCI]: Web-based interaction.

## INTRODUCTION

Over the past two decades, there have been several efforts devoted to collecting a large database of “common-sense” knowledge [3,5,6]. This knowledge consists of basic facts that a majority of humans accept as truth, such as “water quenches thirst.” The motivation for collecting a large database of true statements is the belief that such knowledge is necessary to create truly intelligent systems. There are also more immediate applications. For example, a search engine was prototyped that converts the query “my cat is sick” to “veterinarians, Boston, MA” by following a simple chain of reasoning based on an underlying network of common-sense facts [4].

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.

CHI 2006, April 22–28, 2006, Montréal, Québec, Canada.

Copyright 2006 ACM 1-59593-178-3/06/0004...\$5.00.

Efforts for collecting common-sense facts have demonstrated the promise of this approach. However, they have been unable to collect a large enough fraction of common human knowledge. After 20 years, much less than five million facts have been collected — far from the estimated hundreds of millions that are required [6].

This paper addresses the problem of constructing a truly large database of common-sense statements. Motivated by the ESP Game [1] (which collects labels for images on the Web), we introduce Verbosity — a fun game with the property that common-sense facts are collected as a side effect of game play. The design of Verbosity ensures that data obtained through the game are correct. As with the ESP Game, if our game is played as much as other popular games, we can collect millions of facts in just a few weeks.

Whereas previous approaches have relied on paid “experts” or unpaid “volunteers” [3,6], we put much stronger emphasis on creating a system that is appealing to a large audience of people, regardless of whether or not they are interested in contributing to Artificial Intelligence. We have *transformed* the activity of entering facts into an enjoyable interactive process taking the form of a game. Although some of the previous approaches have called their systems games to entice people to play, they have not transformed the mode of interaction into that of a real game.

Instead of asking users to “enter a true or false statement,” or to rate such statements (both far from enjoyable activities), we start with the realization that a popular party game called Taboo™ [7] already requires the players to state common-sense facts as part of game play. In Taboo™, one of the players gives clues about a certain word to be guessed without saying the word or any of the related words in a list of “taboos.” For instance, they might have to describe the word “apple” without saying “apple” and without saying “red,” “pie,” “fruit,” “Macintosh,” or “Newton.” This player has to give a good enough description of the word to get his or her teammates to guess it (he or she might say “you can make juice out of this,” or “they are sometimes green”). The key observation leading to our system is that Taboo™ requires players to say a list of common-sense facts about each word in order to get their teammates to guess it. Verbosity is based on this realization (but differs from Taboo™ in multiple ways).

By playing Verbosity, people help us collect data not because they feel helpful, but because they have fun.

Verbosity can be considered a “human algorithm”: given a word as input, it outputs a set of common-sense facts related to the word. Instead of using a computer processor, though, this “algorithm” uses ordinary humans interacting with computers throughout the Web. Our system therefore significantly contributes to HCI in two ways: it collects common-sense data that can improve HCI applications, and it addresses a typical AI problem with novel HCI tools.

### Why Not Use Google?

Computers now have the ability to search vast amounts of data in little time. This means that perhaps we could use a search engine to collect the information we need. Although such an approach yields some useful data [2], it is not good enough for our purposes for three reasons. First, most of the knowledge that we are trying to collect is so obvious that no one has bothered to record it. Second, there exists incorrect knowledge on the Web (for example, at the time of writing this paper, searching for “cars can swim” returns more results on Google than “cars can’t swim”). Third, the text on the Web is unstructured and turning it into a directly useful format is a non-trivial task. Our system does not suffer from these shortcomings.

### RELATED WORK

There have been a variety of other methods for collecting common-sense facts, which differ in nature from ours.

#### Cyc

Cyc (e.g., [3]) was the first effort at building a common-sense database. Cyc started by creating a seed database of common-sense knowledge using paid experts to enter facts in CycL — a proprietary, very precise language developed in order to avoid any ambiguity problems. Using this seed database as a starting point, more data can be collected (e.g., by combining facts to infer new ones).

One problem with this approach is that the number of common-sense facts required is orders of magnitude higher than that which could be entered by an isolated set of experts — over the course of a decade, Cyc has been able to populate its main database with around a million pieces of information [6]. As evidenced by the ESP Game [1], our approach has the potential to collect millions of pieces of knowledge within the space of several weeks.

#### Open Mind and Mindpixel

More recently, the Open Mind project has relied on “netizens” (Internet users) to enter common-sense facts. Open Mind has dozens of activities, each designed to gather specific types of knowledge: spatial, hierarchical, implication, etc. So far, they have gathered several hundred thousand pieces of knowledge. Another project, Mindpixel [5], is similar to Open Mind in that it relies on ordinary Internet users. They have a collaborative system in which many participants create and classify a statement as true or false, thus building up a large database of facts. Validation is a majority-based system, and it rewards those who consistently validate a fact inline with the other users.

The major improvement that we make over Open Mind and Mindpixel is that our system *transforms* the process of entering facts into a true gaming environment. Activities in Open Mind and Mindpixel require users to simply enter facts; we don’t need to explicitly ask users to do so.

#### Wikipedia

Another example of gathering knowledge from humans is Wikipedia.org, a tremendously successful online encyclopedia to which anybody can contribute.

In addition to the gaming aspect, our system is different from Wikipedia in that the knowledge we are interested in is significantly simpler. Encyclopedias are mostly useful for their little-known facts; we wish to collect facts that everybody knows (which are of little use in encyclopedias).

### BASIC GAME MECHANICS

Verbosity is meant to be played online by two players selected at random. One of the players is chosen as the “Narrator” while the other is the “Guesser.” The Narrator gets a secret word and must get the Guesser to type that word by sending hints to the Guesser. The hints take the form of sentence templates with blanks to be filled in. The Narrator can fill in the blanks with any word they wish except the secret word (or any string containing the secret word). See Figure 1. For example, if the word is LAPTOP, the Narrator might say: “it has a KEYBOARD.”

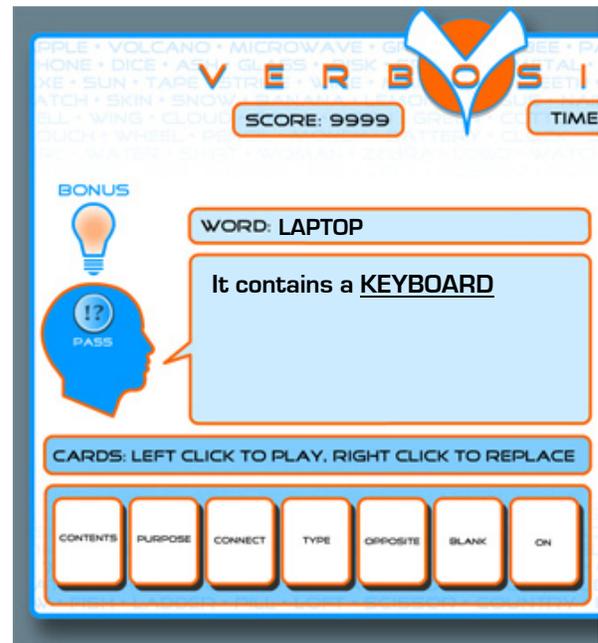


Figure 1. Part of the Narrator’s screen.

The sentence templates that the narrator can use come in the form of cards. The Narrator starts with 7 cards in their hand, and each card contains one sentence template. To give hints about the secret word, the Narrator chooses a card, fills in the blanks in the card and sends it to the Guesser. Given the hints, the Guesser can type what they

think the word is. The Narrator can see all of these guesses, and can tell the Guesser whether each is “hot” or “cold.”

By observing the Narrator’s hints, we can collect common-sense facts about each word. For instance, when the narrator says “It contains a keyboard” about the word “laptop,” we learn that a laptop contains a keyboard.

Players take turns in narrating and guessing. Each session of the game lasts six minutes, and the players go through as many words as they can in that amount of time. The players can agree to pass on a word if they believe it is too difficult.

### Sentence Templates

As mentioned before, Verbosity is inspired by the popular game Taboo™. One major difference between Taboo™ and Verbosity, however, is the use of sentence templates. In Taboo™ players can describe the secret word using arbitrary language, while in Verbosity we restrict the Narrator to using only the types of sentences available to them at the time. There are multiple reasons for using sentence templates instead of natural language:

- **Disambiguation.** In English, the same sentence may have multiple meanings. By carefully choosing the templates, we avoid problems with multiple meanings.
- **Categorization.** We can use the templates to categorize the types of information we get (e.g. Purpose, Spatial Concepts, Causality). In addition, by providing specific lexical templates, we can control the type of information we receive. This allows us to obtain a variety of information about a specific word.
- **Parsing.** By using sentence templates, we don’t have to worry about parsing natural language sentences, some of which might have poor grammar.
- **Fun.** Requiring the Narrator to use sentence templates to describe the word adds an element of challenge and fun to the game. Instead of being constrained by “taboos,” players are constrained by their hint cards.

Our implementation currently uses the following templates:

- *is a kind of*    . Allows for hierarchical categorization.
- *is used for*    . Provides information about the purpose of a word.
- *is typically near/in/on*     (three templates). Provide spatial data.
- *is the opposite of*     /     *is related to*     (two templates). Provide data about basic relations between words.
- . In the game, this is a “wildcard” that collects related words; for example “dance dance” was a clue for the word “revolution.”

We remark that fresh templates can be easily added to our system (and adding them improves the game’s longevity).

### Scoring

The scoring system is cooperative — points are given to both the Guesser and the Narrator whenever the Guesser enters the correct word. In the current implementation, both players obtain 100 points for a correct guess. Points are not subtracted for passing, nor are incorrect guesses penalized.

The exact number of points given to the players for different actions is not important. However, we note that we neither charge nor give points for using templates. While giving points for each template used would encourage people to provide more facts per word, in user testing we discovered that players will sometimes fill a number of templates randomly to increase their score. Charging people for each template used would guarantee a higher quality of data, since people would try to describe the word using the fewest number of facts possible. However, it is counterintuitive to encourage people to use fewer facts, considering the purpose of the game. Thus, we do not assign any charge or bonus to using facts.

### THE SINGLE-PLAYER GAME

As described, Verbosity requires two players. However, it is entirely possible to use the data we have collected to create an automated player, or “bot.” This allows for a single-player game in which the player is paired with the bot. There are multiple reasons for doing this. First, and most importantly, we can use the single-player version to ensure that the facts we collect are *independent* and *useful*. By *independent*, we mean that the fact does not rely on previous facts mentioned during a particular game session; by *useful*, we simply mean that a Guesser can determine the word based on the facts given. Second, we can pair up players with automated bots when the total number of players is not even. Third, the bot can substitute a player who leaves the game so that their partner is not interrupted.

### Narrator

To emulate a narrator in an automated session, we simply display a subset of facts previously collected about a word. These facts usually come from different game sessions and are displayed in a random order. If the guesser is able to guess the word, we can verify the independence and usefulness of these facts.

More specifically, we assign a score of zero to each new fact collected. The automated player selects a subset of facts that have scores within  $\pm 1$  of each other — this ensures that we don’t artificially increase the score of bad facts by using facts that are known to be of good quality. We replay the facts, and if the word is guessed correctly, we increase the score of each fact by 1. If a fact is used in a number of different subsets where the guesser is unable to get the word, we discard it (or, perhaps less harshly, we can flag it for manual review). If a fact attains a high enough score, then we have confirmation of its validity.

## Guesser

Emulating a Guesser in a convincing manner is more difficult. If a real player enters useless descriptions of the word, we do not want the emulated Guesser to guess correctly. Although this is not a significant problem because (a) most Narrators enter accurate descriptions of words and (b) we later in the game verify that each fact is correct, we must nevertheless deal with this problem to protect the illusion that a real game is taking place.

Once we have collected enough facts, we can use them to aid us in guessing the word. For now, though, we have to rely on approximations. What we do is we compile a list of related words — for example, the word dog could have “bone,” “canine,” “cat,” “terrier,” etc. as related words. If enough related words are entered, we guess the correct word.

## DESIGNING FOR ACCURACY

As mentioned before, we use the text entered by the Narrator as common-sense knowledge about the word in question. We employ a set of design strategies to ensure the accuracy of facts entered.

- **Success of the Guesser.** We use the time taken by the Guesser to enter the proper word as an indicator of the quality of the Narrator’s statements. If the Guesser does not get the word, we discard the Narrator’s text.
- **Random pairing of the players.** Verbosity is meant to be played online by many people at once. By randomly assigning the players to different sessions of the game, we force players who want to poison the data to have a low probability of playing together.
- **Description testing.** Most importantly, we use the single-player game mentioned above to check the quality of statements entered. We replay a permutation of collected facts for the word, and if the single-player Guesser can still guess the correct word, we have a significant indicator that the facts are useful: multiple different people chosen at random were able to guess the word given these facts.

## EVALUATION

We collected evidence showing that Verbosity is fun to play and that people provide correct common-sense facts while playing. Since the game has not been formally released to the public, we present results from allowing random players of another game to get a sneak peak of Verbosity. A total of 267 people played the game in a period of 1 week, generating 7,871 facts. This means that, on average, each player contributed 29.47 facts. In terms of time, each person played for an average of 23.58 minutes in one sitting, and some played for over 3 hours. We believe these numbers show how enjoyable the game is.

### Quality of the Data

We conducted an analysis of the data collected in the period of time mentioned above. The analysis consisted of

selecting at random 200 facts collected using Verbosity and asking the following question about them to six different raters: Is this sentence true? Since Verbosity has not been formally released to the public, the sentences were chosen without having been verified using the automaton. Overall, 85% of the sentences collected were rated as true by all six raters. Something to note is that many of the sentences not rated as true by all were debatable — for example, “Buddha is a kind of god.” Thus, even without our mechanisms for validating facts, the collected data was extremely accurate.

## CONCLUSION

We have presented Verbosity, a game to collect common-sense facts. We have shown data indicating that Verbosity is enjoyable and that it produces correct data. Although the game has not been formally released to the public, we were able to collect a large number of facts from just a few test players over the course of a week. The major contribution we present is the transformation of the tedious process of entering facts in a database into an enjoyable game.

Verbosity is an example of an emerging class of games similar to the ESP Game that can be considered “human algorithms”: humans act as processing nodes for problems that computers cannot yet solve. By providing an incentive for players, we gain a large quantity of computing power that can be harnessed for multiple applications. Constructing a complete common-sense database would be a monumental achievement, and we believe Verbosity can be tremendously effective in doing so.

## ACKNOWLEDGEMENTS

We thank Laura Dabbish, Susan Hrishenko and the anonymous CHI 2006 reviewers for their insightful comments. This work was partially supported by the National Science Foundation (NSF) grants CCR-0122581 and CCR-0085982 (The Aladdin Center) and by a generous gift from Google, Inc. Luis von Ahn was also partially supported by a Microsoft Research Graduate Fellowship.

## REFERENCES

1. von Ahn, L., and Dabbish, L. Labeling Images with a Computer Game. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2004, pp 319-326.
2. Wetware: Gathering Common Sense. <http://wetware.hjalli.com/000016.html>.
3. Lenat, D. B. CYC: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM* 38(11), 1995, pages 32-28.
4. Lieberman, H., Liu, H., Singh, P., and Barry, B. Beating common sense into interactive applications. *AI Magazine*, Winter 2004, 25(4), pages 63-76.
5. Mindpixel Project: <http://www.mindpixel.org>
6. The Open Mind Common Sense Project <http://www.kurzweilai.net/articles/art0371.html>
7. Instructions, Taboo. [http://www.hasbro.com/common/instruct/Taboo\(2000\).PDF](http://www.hasbro.com/common/instruct/Taboo(2000).PDF)