

Seek-n-Tag: A Game for Labeling and Classifying Virtual World Objects

Bei Yuan*

Manjari Sapre†

Eelke Folmer‡

Department of Computer Science and Engineering, University of Nevada, Reno
Reno, Nevada, USA

ABSTRACT

We identified that virtual worlds that rely on user generated content often lack accurate metadata for their objects. The apparent lack of metadata is a problem for users who are visually impaired, as they rely upon textual descriptions of objects to be present for accessing virtual worlds using assistive technology, such as a screen reader or tactile display. This paper presents a scavenger-hunt game for the virtual world of Second Life -called SEEK-N-TAG- that allows sighted users to label objects as well as collaboratively develop a taxonomy for objects. SEEK-N-TAG aims to build a set of objects with accurate metadata that can be used as training data for an automatic object classifier. Our approach is novel due to its internal approach where the game is implemented in the virtual world itself as to improve its own accessibility. A user study with 10 participants revealed that labeling objects with a game is more effective and accurate than manually naming objects.

Index Terms: H.5.1 [Multimedia Information Systems]: Artificial, augmented and virtual realities—

1 INTRODUCTION

Virtual worlds such as Second Life [4] and World of Warcraft [7] have enjoyed increasing popularity over the past years with millions of participating users [8, 2]. The immersive graphics, large amount of user-generated content, and the social interaction opportunities offered by virtual worlds, could eventually make for a more interactive and informative World Wide Web [25].

Unfortunately, a significant number of users are excluded from accessing virtual worlds because of a disability [9, 16], especially users who are visually impaired [22, 35]. Whereas the Internet is still largely text based, for reasons such as search engine optimization [13] or to accommodate low bandwidth handheld mobile devices [14], virtual worlds are almost entirely visual. Though the Second Life viewer offers some features for supporting users with low vision, such as scalable fonts, it lacks features that allow users who are severely visually impaired to access using non-visual forms of feedback such as audio or tactile feedback. Individuals who are severely visually impaired often use screen readers to access word processors and Web browsers, but as Second Life is entirely visual and lacks any textual representation [9], screen readers and tactile displays cannot be used for accessing Second Life.

One of the most distinguishing features of Second Life is that everything in Second Life is created and owned by its users. Access to Second Life is free, but users must buy or rent land in order to be able to create content on that land, such as a house for an avatar. There are various ways to create objects but most commonly this can be done using the Second life viewer's built-in 3D modeling tool. Objects are created using a set of basic shapes called prims, which can be molded together and textured to create the desired

objects. Second Life also supports importing objects from external 3D modeling tools. This will allow for creating more elaborate content, such as clothes or skins for avatars. Object interaction is supported through a scripting language called the Linden Scripting Language (LSL). User created content can also be sold to or bought from other avatars using a virtual currency called Linden Dollars, which can be exchanged for real currency. As a result, a thriving economy has emerged, where a number of people are able to make a living by selling virtual content in Second Life [17]. The Second Life world is vast and consists of millions of user generated objects.

A problem with user-generated content is that it often lacks accurate metadata. Objects in Second Life can be given a name and a description, but as most content creators assume other users can see, they frequently leave these properties with their default value, e.g., “*object*.” The Second Life viewer does not enforce any object labeling conventions when objects are created. The lack of metadata is a serious problem to our approach [22] and related approaches [1, 6] towards making virtual worlds accessible to users who are visually impaired using synthetic speech. All these approaches rely on a textual representation of objects to be present which can be read with a screen reader or tactile display. Similar to how web images lacking “alt” tags remain invisible to a screen reader user, objects called “*object*” provides no useful information. This problem is actually worse in virtual worlds, since the meaning of a web image can often be derived from the text around the image or even from the filename [11]. Objects in virtual worlds typically lack such contextual information.

This paper identifies that millions of Second Life objects lack a name and adding names manually may be a costly and time consuming enterprise, not to mention tedious and error prone as different users may label objects differently. The specific way objects are defined in Second Life can be exploited to allow for large-scale automatic object recognition using appropriate shape descriptors [29, 28], which can be constructed and compared efficiently while discriminating between objects. For an automatic labeling approach to work, a library of examples for different object categories is required. As content in Second Life is inherently unreliable (as it is created by a diverse and unmanaged population), we must create a set of accurate metadata for objects manually.

Labeling virtual world objects shares many similarities with labeling web images that lack “alt” tags [12]. Humans significantly outperform computers in recognizing images and involving humans in labeling images is a form of human-based computation [19, 15]. The “games with a purpose” (GWAP) paradigm [32] seeks to make tedious, human-based computation tasks more attractive to humans by wrapping them into a game. This approach has been successfully used in both labeling web images [10, 31] and transcription of ancient text [34].

We present SEEK-N-TAG, a GWAP game implemented in Second Life itself that can label objects and which can make collaborative virtual environments such as Second Life more accessible to users who are visually impaired. A user study with 10 people was conducted that compares the efficiency and accuracy of SEEK-N-TAG with manual labeling. This paper concludes with a discussion of the broader impacts of this research towards automatic object recognition in virtual worlds.

*e-mail: bei@cse.unr.edu

†e-mail: msapre@cse.unr.edu

‡e-mail: efolmer@cse.unr.edu

Graphics Interface Conference 2010
31 May - 2 June, Ottawa, Ontario, Canada
Copyright held by authors. Permission granted to
CHCCS/SCDHM to publish in print form, and ACM
to publish electronically.



2 RELATED WORK

Adding metadata to virtual world objects is related to approaches that add alternative (“alt”) text to web images.

2.1 Image Labeling

Two approaches to labeling images can be identified:

- **Automatic labeling:** web images can be labeled based on heuristics or modeling methods such as semantic concepts and statistical relationships between images and words [11]. Automatic labeling approaches can be applied on a large scale and are efficient, but the accuracy of their inferences can be problematic. Mechanisms to add supplementary, manually-entered metadata are sometimes needed for practical use [30].
- **Manual labeling:** allow sighted users to add metadata to web images. Manual labeling efforts are often used to create training data for automatic labeling approaches, which is also what we seek to achieve.

Several (both manual and automatic) projects have been developed over the years:

- **The ALIPR** (Automatic Linguistic Indexing of Pictures - Real Time) [27] is constructed for fully automatic, high-speed annotation of web images. The system uses a categorized image database for training the algorithms and annotates any on-line image specified by its URL. This work proves that algorithms can be developed to annotate general photographs with certain accuracy by learning from a large manual collection of example images.
- **WebInSight for Images** [12] combines both automatic and manual labeling methods to provide alternative texts for web images to improve web accessibility. WebInSight consists of three modules: (1) contextual analysis of linked webpages; (2) enhanced Optical Character Recognition (OCR); and (3) forms of manual labeling that allow users to provide alternative texts in a local database. New labels for web images are added after the user downloads the webpage for the first time.

Most people typically do not find manual labeling particularly engaging. Games, however, can be used to make such tedious tasks more enjoyable. Gameplay, built around labeling, normally utilizes competition and reward systems to make the process fun. Within the GWAP paradigm, spearheaded by Louis von Ahn [32], the following image labeling games have been developed to improve web accessibility:

- **The ESP Game** [31] does not directly ask the players to label the images. Rather, two players are randomly paired for 2½ minutes and both players must try to enter the same word as their partner for each image on the screen. When both players agree on a word, a new image is displayed. Points are given based on the matching labels. The goal is to agree with the partner on words for as many images as possible.
- **The Google Image Labeler** [23] allows its users to label images and help improve the quality of Google’s image search results. It is very similar to the ESP game, except that the results of the partner’s guess are displayed at the end of the game to help the player learn the game better. This game also provides a timeout for each image besides the ‘pass’ condition as provided in the ESP game.
- **Phetch** [33] is a game developed to annotate images with descriptive paragraphs. This game can be played by 3 to 5 players, with one of the (randomly selected) players as the “Describer”, while the others are the “Seekers.” The data generated with this game has proven to be more accurate for web images than the ESP game and the Google image labeler.

As an extra incentive for players to perform well, it displays the high scores of the different players on their websites.

2.2 Virtual World Object Labeling

In an effort to make virtual worlds accessible to users who are visually impaired [24], IBM’s Human Ability and Accessibility Center developed a web-based interface for Second Life that can be accessed with a screen reader. IBM provides a plug-in for the Second Life viewer, which allows users to manually add descriptive information to Second Life. Sighted users equip an annotation object to their avatar. When clicking on the objects in Second Life, a Web browser is invoked, allowing sighted users to enter information, such as names and descriptions of items or locations. This information is then saved in an external database and IBM’s client can retrieve this data whenever a user encounters an object with missing meta-information. This approach can improve the metadata in virtual worlds but adding all this data manually for millions of objects could be a tedious and time-consuming task. In this paper a method is proposed that will allow for large scale labeling of virtual world objects.

3 METHODS

This section presents data from a detailed study on the lack of metadata of objects in Second Life, presents our approach for the automatic labeling of objects and discusses the need for a library of object examples to be present.

3.1 Metadata Study

In related research, an accessible client for Second Life was developed, called TextSL [22], which allows users who are visually impaired to access Second Life using a screen reader. Users interact using a command-based interface that is inspired by multi-user dungeon games [26] and which allows screen reader users to explore the Second Life world, communicate with other avatars, and efficiently interact with large numbers of objects and avatars. TextSL can be downloaded on: <http://www.textsl.org>. During the development of this client, we noticed that many objects in Second Life were called “object”. Though TextSL ignores these objects when screen reader users query their environment, we performed a detailed study to understand the magnitude of this problem.

To gauge the current accessibility of virtual world objects, a modified version of TextSL was developed that allows an avatar to remain logged into Second Life and collect metadata on objects to measure the prevalence of meaningful object labels in SL. The Second Life world is partitioned into smaller chunks called regions (13,543 regions currently exist [3]). An avatar logged in with a modified version of TextSL is teleported to a randomly selected region, retrieves the names of its objects, and stores them in a local database. Our database contains over 337,000 object names. 433 regions were sampled and we found that with 95% confidence, $31.3\% \pm 2.02\%$ (SD=21.4%) of the objects in Second life were called “object”. Unfortunately it was not possible to access all Second Life regions, because: (1) Second Life is growing fast (700 regions existed in 2006); and (2) many regions in Second Life are difficult to access from time to time [36].

It is important to understand what types of objects are called “object” to understand whether the lack of metadata is really a problem. A closer analysis found that users sometimes build large objects -such as a house- out of smaller objects but do not group these related objects properly into a single larger object. However, we also found many objects, such as buildings, furniture, vehicles, pets, billboards, trees, bushes, which are abundant in Second Life, to be called “object”.



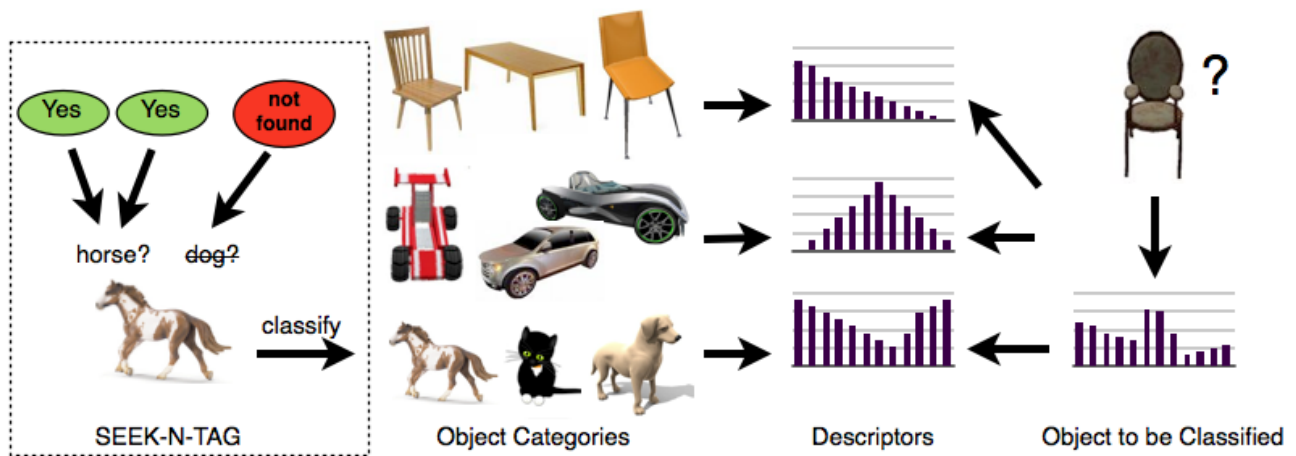


Figure 1: Automatic object recognition approach with SEEK-N-TAG providing accurate object labels and classifying objects into categories.

3.2 Automatic Object Labeling

Though manual labeling has some advantages, such as being able to provide a detailed description for an object, labeling millions of objects is tedious and error prone. We seek to develop a solution to the general problem of automatically recognizing virtual world objects using 3D object recognition. Metadata for objects in Second Life may also include descriptions, however, finding names for objects is more important, as descriptions for common objects could be retrieved from a dictionary. Automatic labeling has the potential to work better for virtual world objects than for web images, as virtual world objects are defined in isolation, which, unlike images may contain multiple subjects. Objects also have features such as topology, size, and orientation. Because Second Life is modeled after the real world, we try to take advantage of its apparent homogeneity. Rather than trying to recognize all possible objects, which may include objects that do not exist in the real world such as flying cars, we restrict ourselves to recognizing objects for the most common classes of objects in Second Life, such as furniture, vehicles, animals, etc. There are several ways to derive such object classes; either through analysis of large numbers of objects or by having sighted users define classes when they manually name an object. Object categorization is a natural way to deal with large numbers of objects sharing a common structure. Its key advantage is that it enables the recognition of previously unseen objects. This is in contrast to object recognition, which assumes that the objects to be recognized are drawn from a known database of object models.

A challenge for classifying 3D objects is to find a suitable shape descriptor that can be constructed and compared quickly while discriminating between shapes. 3D objects are typically represented by “polygon soups” - unorganized and degenerate sets of polygons lacking any topology or solid modeling information [28]. This is a problem for finding a suitable shape descriptor, as certain geometric features, such as volume, may be difficult to compute for degenerate meshes. Fortunately, Second Life’s objects are composed of seven different types of solid body entities called “prims” (primitives) and that are defined analytically (i.e., box, prism, sphere, cylinder, torus, tube, ring) with permutations such as twist and taper. The use of prims is motivated by the fact that they are compact, can be rendered more cheaply and they are more easily mapped to a physics model than meshes. We are developing a suitable shape descriptor that takes advantage of the fact that Second Life objects are composed of prims and which, for a small number of categories, has proven to be able to discriminate object categories efficiently. The discussion of this shape descriptor is outside of the scope of this paper.

3.3 Training Data

For our descriptor to work, we need to have a large number of categorized training examples available. Although it is tempting to create training samples using the data of objects that already have metadata in Second Life, this data may be unreliable and hence we must create a set of reliable data manually. SEEK-N-TAG seeks to tackle the first step towards establishing such set of training data by creating a set of objects with accurate names, which at a later stage will be classified into object classes. Figure 1 shows an overview of how the SEEK-N-TAG game contributes towards building a set of training data for automatic object labeling.

4 SEEK-N-TAG

Labeling thousands of objects manually is very time consuming and error prone. Therefore, we explore the use of a game to help create the set of training examples.

4.1 Requirements

As this game is designed based on the GWAP paradigm, we must first analyze what we want the game to achieve. Our game should achieve the following three important requirements:

- **Efficiency:** in order to be effective, the game must allow for more efficient labeling of objects than manually providing a name for each object. A strategy for achieving this could be to minimize or avoid players actually having to type names for objects.
- **Accuracy:** the game must also increase the accuracy of metadata for objects in Second Life. Many objects already have names but the accuracy of these names is unclear.
- **Gameplay:** the game must be fun and offer the player a challenge.

Based on these requirements the proposed game could implement some form of confirmation rather than typing in words, e.g., confirm the name that the object already has or confirm a user provided name for the object. One of the most important requirements of games is that they must be fun and offer the player a challenge. Several game mechanisms can contribute to these requirements. Mechanisms used in related games, such as the ESP game [31] and the Google Image Labeler [23], offer a reward system and competition between players. A reward system is usually implemented in the form of a score, which indicates a players performance in the game as a number or a time. Typically a higher number or a lower time





Figure 2: SEEK-N-TAG showing the spheres that are rendered above each object around the user and which the user has to click on to tag the object that needs to be found.

indicates a better score. Competition adds an extra dimension to challenge and may reveal a player's character and ability as well as making the challenge personal [18]. Challenge without competition may also be predictable. In labeling games players play against each other but competition is visible using a leader-board, e.g., a high-score on the website that is visible to other players, which may motivate other players to try to beat that score.

4.2 Design

Though it may be possible to extract objects from Second Life, export them as 2D images, and use these as input into one of the existing Web labeling games [23, 33], we are convinced that an automated object labeling approach will be more efficient.

Instead of developing our game for a 2D web interface, we opted for enticing users in the labeling of virtual world objects through an engaging third person interface that is already offered by Second Life. Our approach is further novel as our labeling game is directly integrated in the medium that we seek to make more accessible.

Rather than developing a new game with unproven elements of gameplay, we implemented a game whose mechanics most individuals will be familiar with: a scavenger hunt. The goal of a scavenger hunt game is to gather a number of specific items and the player who finds all the items first wins the game. SEEK-N-TAG is a scavenger-hunt-like GWAP game.

During the game, the player needs to “tag” objects that match with the name provided by the game. For example, one of the tasks could be: “Find a house” or “Find a cat.” The players of SEEK-N-TAG are initially shown a countdown clock in seconds, which starts with 900 seconds. Users have 30 seconds to find each object before the game provides the next task, and players score 50 points for each object tagged within this time. The faster the player can find an object, the more time the player has available to tag objects as the remaining time (30 seconds minus the time it took them to find the object) is added to their clock. When the player runs out of time, the game is over and we record the player's score.

4.3 Implementation

SEEK-N-TAG is made available and transferable to the player through

an interactive object programmed in Linden Scripting Language (LSL). The game renders a heads-up display (HUD) when the player attaches this object to their avatar. The SEEK-N-TAG object simply communicates through the instant messaging protocol with the agent. Upon activation the agent sends a message with names from the database to the SEEK-N-TAG object that the player needs to find. The agent uses the database of objects names for regions we previously collected in our Metadata study. The SEEK-N-TAG object renders this name in the HUD while keeping track of time. Players walk around and a colored sphere is rendered in front of each nearby object. Clicking on the sphere instead of the rendered object itself allows us to retrieve the identifier of the object without actually activating any scripts that this object may contain. When clicking on a sphere above an object that the player wants to tag, a small popup window will be displayed, which requires the player to confirm tagging the object. This is implemented to prevent players from accidentally clicking on the wrong object. In real Second Life environments, it is possible that the user may not be able to find the object or that the name is wrong. Thus, we set up a maximum time limit per each search task, so that the game can still continue when the player fails to tag the object. Players are not penalized for not finding an object. When tagging the object, we retrieve its identifier and send this information back to the agent as an instant message, which records this identifier in the database as well as the player who tagged it. To avoid players cheating and randomly clicking on objects, tagging attempts by one user are stored and later verified with other tagging attempts. Players' scores are updated based on the number of accurate labeling attempts. A working prototype has been implemented and a screen shot is shown in Figure 2.

5 USER STUDY

We evaluated SEEK-N-TAG through a user study that compares the effectiveness and accuracy of labeling with SEEK-N-TAG's (tagging) versus manually providing names for objects (naming). We also evaluated the usability of SEEK-N-TAG.

5.1 User Study Design

Our user study was conducted in a controlled environment in Second Life (Accessibility Test Island). A controlled environment was chosen in order to avoid interference from other avatars. To ensure the authenticity of a real Second Life experience, our island was modeled after a typical Second Life island and we used existing Second Life objects to populate our island. 50 of those objects were distributed in two sets (25 objects each), and randomly dispersed at two different locations on our Island, A and B. These locations were approximately 50 by 50 game meters and players could not see location A from location B. All objects were called “object”. A total of ten participants (9 males, 1 female), were randomly recruited. They were between 24 and 43 years of age ($M=30$, $SD=6.3$). Prior to the user test, participants received a brief tutorial on how to use the Second Life viewer, how to perform naming, and how to play SEEK-N-TAG. Users were then asked to perform the following tasks:

1. **Naming:** The user is asked to provide a name for each of the 25 objects in either environment, A or B, by clicking on the object and providing a name for it in the object's name field. We asked users to provide a literal name for the object and to refrain from describing any object properties. Users can name objects in any order they prefer.
2. **SEEK-N-TAG:** Players switched locations and played SEEK-N-TAG. Players were iteratively given the name of a object to find and to tag. All users had to find tags in the same order to allow for analyzing whether tagging efficiency increases over time. Players were told they could tag the same object multiple times if it matched the name that was looked for.



A certain level of confidence (n) is computed for each object name, which is indicated by the number of users (n) that agree on the name for that object. Confidence levels for tagging and naming, however, are computed differently. To achieve a confidence level of (n) for naming is straightforward: (n) users must agree on a name. However, as tagging is already seeded by a name, ($n - 1$) users must tag the object to achieve the same level of confidence for naming. Consequently, to be able to compare the effectiveness and accuracy of naming versus tagging in our user study, ten users provide names and eight users tagged objects for the two different locations. To further provide a more robust and fair comparison, instead of using accurate names for the objects provided by us in the SEEK-N-TAG game, we fed the names provided by the users in the naming process into the game for tagging. Thus, our user study was performed in two steps. First, ten users named the objects in locations A and B (five users in each location) then eight users returned to play the SEEK-N-TAG game at each location (four per location). If a user first named objects in location A, the user later plays SEEK-N-TAG in location B, and vice versa. Their combined efforts were compared. After completing the user study, users filled in a brief questionnaire evaluating the usability of SEEK-N-TAG.

5.2 Results

Each of the 50 objects is named once by five different users, which led to 66 different names (31 tags for location A and 35 tags for location B). We randomized the names for each location and fed all the 66 names to SEEK-N-TAG game. Compared with the user-provided naming process, SEEK-N-TAG utilized only 51 out of 66 names provided. A confidence level of 1 is achieved when all users name an object differently or if the object is not tagged at all (since initially the name is still provided by a user). A confidence level 2 indicates that two users provided the same name for the object or that only one user tagged that object. This pattern continues up to confidence level 5. All of the SEEK-N-TAG players finished tagging all of the objects within the given time.

Figure 3 shows the results of the combined naming and tagging efforts and shows the number of objects for each confidence level. SEEK-N-TAG outperforms manual naming for higher levels of confidence, and these results were found significantly different using Fishers Exact test using a 2×4 contingency table ($\chi^2_{2,96} = 20.26 p < 0.001$) for (n) > 1 .

In general, to achieve a higher level of confidence, more labeling or naming attempts must be undertaken. An explanation for the lower performance of naming, is that naming is a creative process and when more users are involved in the naming, the chance increases that a different names or synonyms will be provided for the object. Consequently, for a small number of naming efforts it may not be possible to feasibly discriminate between provided names for an object through analysis of their frequencies. Tagging just confirms a seeded name for an object and will always lead to consensus. To build a set of training data for our classifier, it may be more cost effective to be able to quickly confirm a candidate's provided name with few human tagging events, than to spend larger human naming efforts as to be able to disambiguate between different provided names. Synonyms for example could also be retrieved from a thesaurus. In our user study we asked users to provide a name for the object. An inherent risk with using an existing name for the object is that this name may describe an object properties such as color, e.g., "green", which may lead to false positives as it reflects the popularity and not the literal accuracy of the name. One possible solution to deal with this would be to filter names that are being looked for in SEEK-N-TAG using a list of common adjectives.

To more optimally compare naming versus tagging efforts, we fed different names that were provided for the same object into SEEK-N-TAG, to evaluate its ability to reach consensus.

Tagging and naming do not always lead to consensus as a confi-

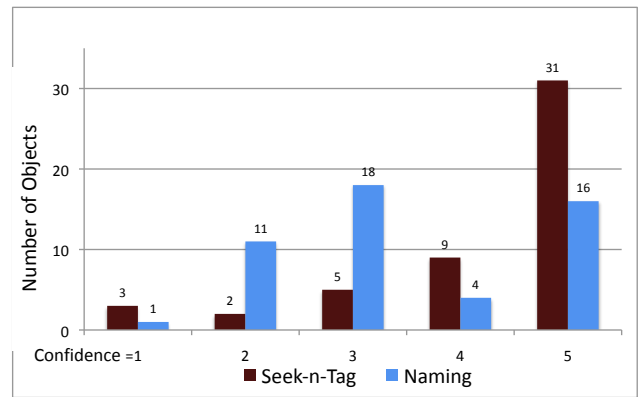


Figure 3: Accuracy of Naming vs. SEEK-N-TAG

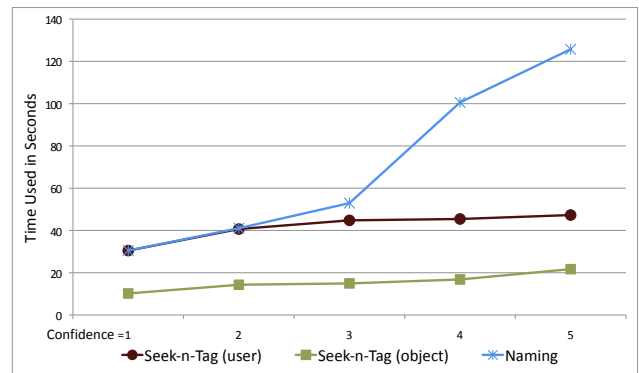


Figure 4: Efficiency of Naming vs. SEEK-N-TAG with user provided name and object provided name

dence level of 2 for an object could indicate that two groups of two users agreed on two different names. In this case, it still counts as a confidence level of 2 for the object. The same happens with a different user tagging the same object once for two or three different names that were provided by the object. With regard to achieving consensus we found consensus for 45 objects for both naming and tagging. Though these results were not found to be significantly different ($Z_{2,50} = 0 p = 1.0$), a better performance with regard to consensus for SEEK-N-TAG may be expected if we use the existing name of an object.

Though tagging and naming are different processes, it is meaningful to analyze the efficiency of each approach to determine an object with a particular confidence level. The average time in seconds to determine an object with a confidence level (n) is presented in Figure 4 for naming (Naming), tagging with a user provided name (Seek-n-Tag: user), and tagging with an existing object name (Seek-n-Tag: object). These values were computed by taking the average tagging and naming time and multiplying this by the observed success rate of each tagging or naming attempt to lead to an object of confidence level n . SEEK-N-TAG efforts relying on a user specifying a name had the cost of one naming effort added to it.

The efficiency of tagging versus naming depends on the spatial distribution of the objects around the user's start location. Both approaches were tested at two different locations where objects were randomly dispersed and areas were further confined to a 50 by 50 meter area. Because we average overall users for both locations for all objects, this spatial distribution can be considered to be invariant to our analysis.



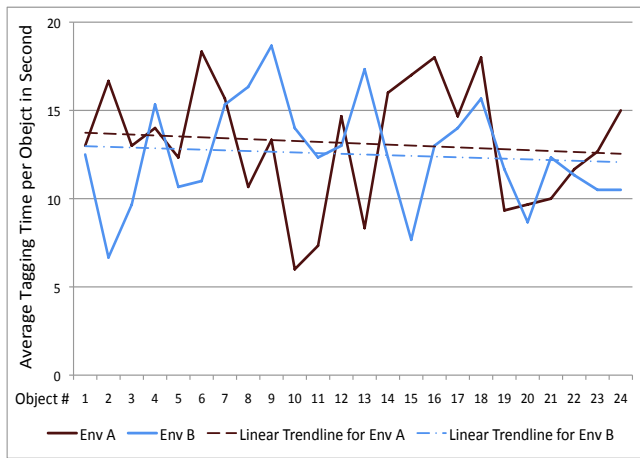


Figure 5: Time Used Tagging Each Object

To build a set of training data for our classifier, the choice which method to use depends on whether names are available for objects and the level of confidence that needs to be achieved. Figure 4 shows that up to a confidence level of 2, SEEK-N-TAG (user) and naming have the same efficiency. For higher levels of confidence, SEEK-N-TAG (user) is more efficient than naming. SEEK-N-TAG (object) is almost twice as efficient as naming per confidence level, which will be the mechanism we seek to deploy in future work.

To analyze the effectiveness of tagging while users play the game, we calculated the average time of each group to successfully place a tag in an location. No relative comparison could be made between naming and tagging as users name objects in any order but users tag objects in a fixed order. When a user fails to find a tag, their time is not included in the average. Figure 5 displays the average time used to tag each object in each of the locations. Trend-lines for both locations show a slight decrease in average tagging time, most likely because: (1) players familiarize themselves with their location and are able to locate objects more easily; or (2) they enjoy playing the game and are more motivated to achieve a higher score.

5.3 Usability

Heuristics for designing and evaluating games are mostly focused on three aspects of usability in games: interface (controls and display); mechanics (interacting with the game world); and gameplay (problems and challenges) [20]. Additionally, a game has to be fun to play. We used a 5-point Likert scale to measure these four aspects of the game. In the enjoyment part, we evaluate SEEK-N-TAG with manual naming and avoid bias through non leading questions. Table 1 shows the questions and the ratings we received, with the following scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree. With regard to the enjoyment factor of both labeling approaches, players enjoyed SEEK-N-TAG more than manual labeling ($T_{2,10} = 6.08 p < 0.01$).

6 DISCUSSION AND FUTURE WORK

6.1 Efficiency versus Accuracy

SEEK-N-TAG has shown to be more efficient method for object naming than manual labeling. Therefore, we can confidently start creating the required training set of data that is required for our automatic labeling technique. However, certain parameters in the game must still be determined. To spend human-based computation cycles cost effectively, inevitable trade-offs have to be made between efficiency and accuracy. What level of confidence constitutes adding an object to our set of training data that we seek to use for our automatic object labeling approach? For a fixed number of

Table 1: Usability Questionnaire

Aspects	Elements	Rate (1-5)	SD
1.Interface	The graphical design of the game is esthetically pleasing.	4.4	0.49
	The visual information displayed is not overwhelming.	4.2	1.2
2.Mechanics	The interaction with objects is simple and efficient.	3.2	1.3
	The controls are effective.	3.6	1.0
3.Gameplay	The game provides good amount of challenges.	3.2	0.75
	The rewarding system is fair and motivates the player.	3.6	0.8
4.Enjoyment	I will play the game in the future when its available.	3.6	0.80
	Labeling with Seek-n-Tag is fun	4.1	0.76
	Manual naming is fun.	1.8	0.73

computation cycles, a larger number of objects can be tagged at a lower confidence levels. Conversely, fewer objects can be tagged with higher confidence. It is difficult to determine the required confidence level of objects without an analysis of the effectiveness that can be achieved with the proposed automatic object labeling technique, therefore, once an estimate of this effectiveness has been determined using a reasonable set of training data, we can set these parameters and we will make SEEK-N-TAG available to users within Second Life. Furthermore, future studies will evaluate the usability of the names generated with SEEK-N-TAG for our TextSL users with visual impairment (See Section 3.1), though we consider this to be outside of the scope of this paper.

6.2 Object Taxonomy

In the future, we will also seek to research how SEEK-N-TAG can be used to establish a taxonomy for virtual world objects. This object taxonomy could have two purposes.

1. It would help create examples for the object categories that are required for our automated object labeling approach. Our object categories need to have multiple examples of real Second Life objects that can be used to exploit commonalities and to develop suitable shape descriptors that allow for automatically categorizing objects without a name.
2. As Second Life, and virtual worlds in general, are very densely populated with objects, a taxonomy could be used to provide a more “digestible” and usable form of feedback for synthetic speech based approaches towards making virtual worlds accessible [5, 6]. A problem with these approaches is that users may be easily overwhelmed with feedback when they query their environment as numerous object names are returned [22]. A taxonomy could reduce the amount of feedback with a minimum amount of information loss e.g., [bicycle, car, bus] → [3 vehicles].

The creation of a taxonomy could be tied in by making the game more challenging the longer you play. A small set of high-level categories for the most common objects in Second Life can be created including buildings, furniture, vehicles, animals, billboards, and trees. For each category, a few examples are added during each round of play. Instead of looking for a specific object, initially players in SEEK-N-TAG would tag objects in a specific object category, which may be easy to find. Based on objects that have been categorized before in an area, we can guarantee the user will be able to find at least one of such object. For example, the player may be



asked to find a “vehicle” if we know that a car can be found in that region. There is a chance that the user will indeed tag that particular vehicle, but there is also a chance that the user may tag another object such as a bicycle that fits that category. Depending on whether this object has an accurate name or not, when the result of a tagging attempt for an object category differs from what is expected it will allow us to expand our taxonomy with new candidate rules *Fact(bicycle→vehicle)*. Subsequent tagging events could validate the accuracy of such a rule if we let players look for specific car objects. The same rule could be also be inferred at another location with an object with the same name. SEEK-N-TAG starts out easy by looking for categories, and users may score fewer points for finding an object by category. The game gets harder progressively by having users look for specific objects. A form of adaptive difficulty adjustment [21] could be implemented where, as the player runs out of time, SEEK-N-TAG goes back to providing object categories until the player’s performance has increased. This has the benefit of tailoring our game to less experienced players as not to discourage them.

Ideally, SEEK-N-TAG could be extended so it can create a taxonomy by itself. What names need to be provided to the SEEK-N-TAG game is not a trivial problem as we seek to include as many differently named or shaped objects for each object category. If we start with the highest level category descriptions, SEEK-N-TAG could fill these with examples and split them into subcategories when possible. Categories should be split into smaller categories when our shape descriptor detects that a particular category starts to contain clusters of objects (such as cars) that are more similar to each other than to other examples for that category (bicycles). Some form of manual naming may be required, when no name match can be found between the examples that establish this subcategory.

When creating a taxonomy, creating an ontology could be useful *Fact(bicycle has 2 wheels)*. However, this may be difficult to achieve with SEEK-N-TAG as we cannot take advantage of the apparent inheritance relationship that certain objects have and which SEEK-N-TAG exploits to find new objects for the object class. For example, a bicycle is a vehicle but a wheel is not a bicycle. Furthermore, many parts of an object such as wheel of a car, do not exist as a clickable object but it is defined as a basic shape (prim), which makes the creation of an ontology for most objects in Second Life irrelevant because these are part of the geometry of other objects and not defined as objects in isolation.

6.3 Real Second Life Environments

The island, where we conducted our user study, was modeled after other islands that can be found in Second Life. The game area was relative small (50x50), compared to regions in Second Life that can be as big as 256x256 game meters. Thus, locating objects may be significantly more challenging as objects may be hidden in buildings or obscured from view due to larger obstacles. Objects could also float in the sky as users can build structures above their regions. A solution for this could be to change SEEK-N-TAG so that when the user’s time runs out for a particular object to start showing visual cues to the location of where this object is expected to be found. By following visual cues the player can quickly confirm or deny whether the visual appearance of the object corroborates with the name. Such a mechanism may be undesirable for tagging object categories, as it may not allow for the definition of new rules. Our participants provided the names we used for the tagging in our user study. We need to explore the use of real names from our database of Second Life objects, in order to be able to identify the true effectiveness of tagging. Our database currently contains over 337,000 object names for the 433 regions we have analyzed.

7 CONCLUSION

The lack of metadata for objects in virtual worlds imposes a significant barrier towards making virtual worlds accessible to users who rely upon such information. This paper presents a game called SEEK-N-TAG, which can be played within the virtual world of Second Life. SEEK-N-TAG is one of the first GWAP games with a third-person interface, which may offer a more compelling and immersive game experience than Web-based GWAP games. SEEK-N-TAG is further novel as it is part of the medium itself, whose accessibility we seek to improve. Modeled after a scavenger hunt game, SEEK-N-TAG offers sighted players a game where they will try to find and tag objects in Second Life. These tagging efforts help establish a set of objects with accurate metadata, which can be used for training an automatic object classifier, to further help make virtual worlds more accessible to users who are visually impaired. A user study with 10 participants compared SEEK-N-TAG with manual labeling and found that SEEK-N-TAG is more effective and accurate than manual labeling. The game mechanics of SEEK-N-TAG could also be implemented in other virtual worlds where metadata for objects is lacking. Future work will focus on expanding SEEK-N-TAG to allow for tagging efforts to create a taxonomy of objects that can be used for: (1) automatically creating examples for object categories for automated object recognition; and (2) synthesizing more usable and concise forms of feedback for exploring virtual worlds using a screen reader.

8 ACKNOWLEDGEMENTS

This work is supported by NSF Grant IIS-0917362.

REFERENCES

- [1] Ibm human ability and accessibility center, virtual worlds user interface for the blind, <http://services.alphaworks.ibm.com/virtualworlds/>.
- [2] Second life economic statistics, http://secondlife.com/whatis/economy_stats.php.
- [3] Second life region statistics, <http://stats.slbuzz.com/sims/browse/>.
- [4] Secondlife. <http://www.secondlife.com>. Last accessed October 2008.
- [5] Textsl; a screen reader accessible client for second life, <http://www.textsl.org>.
- [6] Virtual guide dog project, <http://virtualguidedog.org>.
- [7] World of warcraft, <http://www.worldofwarcraft.com>. Last accessed October 2008.
- [8] Blizzard entertainment, world of warcraft hits 10 million subscribers, <http://www.blizzard.com/us/press/080122.html>, 2008.
- [9] P. Abrahams. Second life class action, http://www.it-analysis.com/blogs/Abrahams_Accessibility/2006/11/second_life_class_action.html. 2006.
- [10] L. Ahn, S. Ginosar, M. Kedia, R. Liu, and M. Blum. Improving accessibility of the web with a computer game. In *CHI 2006: Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press, 2006.
- [11] J. P. Bigham. Increasing web accessibility by automatically judging alternative text quality. In *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces*, pages 349–352, New York, NY, USA, 2007. ACM.
- [12] J. P. Bigham, R. S. Kaminsky, R. E. Ladner, O. M. Danielsson, and G. L. Hempton. Webinsight:: making web images accessible. In *Assets '06: Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, pages 181–188, New York, NY, USA, 2006. ACM.
- [13] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, 1998.
- [14] O. Buyukkokten, H. Garcia-Molina, and A. Paepcke. Seeing the whole in parts: text summarization for web browsing on handheld devices. In *WWW '01: Proceedings of the 10th international confer-*



- ence on *World Wide Web*, pages 652–662, New York, NY, USA, 2001. ACM.
- [15] C. Caldwell and V. S. Johnston. Tracking a criminal suspect through “face-space” with a genetic algorithm. In *Proceedings of the Fourth International Conference on Genetic Algorithm*, pages 416–421. Morgan Kaufmann Publisher, July 1991.
- [16] W. S. Carter and G. D. Corona. Exploring methods of accessing virtual worlds. *American Foundation for the Blind AccessWorld*. 9(2), 9(2), 2008.
- [17] K. Craig. Making a living in second life, <http://www.wired.com/gaming/virtualworlds/news/2006/02/70153>, 2006.
- [18] C. Crawford. *Chris Crawford on Game Design*. New Riders Games, June 2003.
- [19] R. Dawkins. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*. W. W. Norton, September 1996.
- [20] M. Federoff. Heuristics and usability guidelines for the creation and evaluation of fun in video games. Master’s thesis, Department of Telecommunications of Indiana University, 2002.
- [21] E. Folmer. Usability patterns in games. In *Future Play ’06: Proceedings of the 2006 conference*, 2006.
- [22] E. Folmer, B. Yuan, D. Carr, and M. Sapre. Textsl: A command-based virtual world interface for the visually impaired. In *Eleventh International ACM SIGACCESS Conference on Computers and Accessibility*, 2009.
- [23] Google. Google image labeler, <http://images.google.com/imagelabeler/>, 2009. Last accessed March 2009.
- [24] IBM. Virtual worlds user interface for the blind, <http://services.alphaworks.ibm.com/virtualworlds/>, 2009. Last accessed March 2009.
- [25] D. Kirkpatrick. No, second life is not overhyped, http://money.cnn.com/2006/11/09/technology/fastforward_secondlife.fortune/index.htm, 2006.
- [26] D. Lebling and M. Blank. Infocom games, zork I, <http://infocom-if.org/games/zork1/zork1.html>, 1980.
- [27] J. Li and J. Z. Wang. Real-time computerized annotation of pictures. In *MULTIMEDIA ’06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 911–920, New York, NY, USA, 2006. ACM.
- [28] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Matching 3d models with shape distributions. In *SMI ’01: Proceedings of the International Conference on Shape Modeling & Applications*, page 154, Washington, DC, USA, 2001. IEEE Computer Society.
- [29] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Trans. Graph.*, 21(4):807–832, 2002.
- [30] H. Takagi, S. Kawanaka, M. Kobayashi, T. Itoh, and C. Asakawa. Social accessibility: achieving accessibility through collaborative metadata authoring. In *Assets ’08: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 193–200, New York, NY, USA, 2008. ACM.
- [31] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI ’04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, NY, USA, 2004. ACM.
- [32] L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51(8):58–67, 2008.
- [33] L. von Ahn, S. Ginosar, M. Kedia, R. Liu, and M. Blum. Improving accessibility of the web with a computer game. In *CHI ’06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 79–82, New York, NY, USA, 2006. ACM.
- [34] L. von Ahn, B. Maurer, C. Mcmillen, D. Abraham, and M. Blum. recaptcha: Human-based character recognition via web security measures. *Science*, pages 1160379+, August 2008.
- [35] G. R. White, G. Fitzpatrick, and G. McAllister. Toward accessible 3d virtual environments for the blind and visually impaired. In *DIMEA ’08: Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*, pages 134–141, New York, NY, USA, 2008. ACM.
- [36] N. Yee, J. N. Bailenson, M. Urbanek, F. Chang, and D. Merget. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology and Behavior*, 2006.

