2010 Senior Thesis Project Reports

Iliano Cervesato[*] **Majd Sakr**[*] **Mark Stehlik**[†]
Bernardine Dias[*‡] **Kemal Oflazer**[*]
Noah Smith[†]

May 2010
CMU-CS-QTR-103

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

[*]Qatar campus. [†]Department of Computer Science. [‡]Robotics Institute.

The editors of this report include the members of the Senior Thesis Committee on the Qatar campus and the students' advisors.

Abstract

This technical report collects the final reports of the undergraduate Computer Science majors from the Qatar Campus of Carnegie Mellon University who elected to complete a senior research thesis in the academic year 2009–10 as part of their degree. These projects have spanned the students' entire senior year, during which they have worked closely with their faculty advisors to plan and carry out their projects. This work counts as 18 units of academic credit each semester. In addition to doing the research, the students presented a brief midterm progress report each semester, presented a public poster session in December, presented an oral summary in the year-end campus-wide Meeting of the Minds and submitted a written thesis in May.

Keywords: Natural Language Processing, Entity Type Recognition, Web-Based Education, Technology for the Developing World, Mobile Education, Game-Based Education, English Literacy.

Contents

RICH ENTITY TYPE RECOGNITION IN TEXT

Senior Thesis

Rishav Bhowmick
rishavb@cmu.edu

Advisors

Kemal Oflazer
ko@cs.cmu.edu

Noah A. Smith
nasmith@cs.cmu.edu

Mentor
Michael Heilman
mheilman@cs.cmu.edu

May 1, 2010

ABSTRACT

Many natural language processing (NLP) applications make use of entity recognition as preprocessing step. Therefore there is a need to identify nouns (entities) and verbs in free text. The task boils down to using machine learning to techniques to train a system that can perform entity recognition with performance comparable to a human annotator. Challenges like lack of large annotated training data corpus, impossible nature of listing all entity types and ambiguity in language make this problem hard. There are existing entity recognizers which perform this task but with poor performance. An obvious solution is to improve the performance of an existing entity recognizer. This Senior Thesis will analyze the existing features, through a series of experiments, that are important for the recognizer. This project will also suggest usage of additional features like Word Cluster features and Bigram features to improve the performance of the system. At the same time, experiments will show that lack of large annotated training data may not be as big of a problem it might seem at first.

ACKNOWLEDGEMENT

Tons of thanks to my advisors Dr. Kemal Oflazer and Dr. Noah A. Smith, and mentor Michael Heilman for their constant support throughout the year. Without their help and guidance, I would not be able to experience the field of natural language processing (NLP). At the same time, I would like to thank Dipanjan Das who provided the word clusters for this project. I also thank Dr. Smith for granting me access to Malbec cluster where most of the work was done. Not to forget, I thank Dr. Behrang Mohit to access his VM in Qatar. I want to take this opportunity to thank Dr. Brett Browning, Dr. Bernardine Dias and Dr. Majd Sakr who let me use the Student Robotics Lab to set up my workspace.

I would also like to thank Dr. Dudley Reynolds for his constant help in making me think about the writing aspect of the thesis by providing me with comments and suggestions.

Last but not the least, I would like to thank my parents for their everlasting support and understanding my need to stay at university late nights and early mornings. Finally, a big hug to my friends and class-mates who were with me during these times.

TABLE OF CONTENTS

1 INTRODUCTION

Common applications of natural language processing (NLP) include summarization of text, classifying documents or automatic answering of questions posed in natural language. Each of these applications require entity type recognition in the text as a pre-processing step. Here, "entity" refers to concrete and abstract objects identified by proper and common nouns. Entity recognition focuses on detecting instances of types like *person, location, organization, time, communication, event, food, plant, animal,* and so on. For example, an entity recognizer would take the following sentence as input:

George Washington was the first President of the United States of America.

and output:

<noun.person> *George* Washington **</noun.person>** *was the first* **<noun.person>** *President* **</noun.person>** *of the* **<noun.location>** *United States of America* **</noun.Location>**.

Humans generally have no problems finding out what type a noun belongs to. In the example above, a human would look at "President" and know that it is of type *Person*. He/she would also know a location or organization can have a President. Additional knowledge about the country, makes him/her think it is a *location*. Finally, "George Washington" has to be a *person* as a president can only be a human.[1] The way a human figures out the entity types could be summarized in the following points:

- Recalling what entity type a word most likely belongs to
- Looking at the context the word appears in.
- Looking at features like word capitalization, punctuation marks. For example, the use of an upper-case letter after punctuation marks like periods or question marks does not indicate that the first word of the sentence is a proper noun. But in general, the use of capitalization does suggest a person, organization, or location.

Our task is to use machine learning techniques to train a system that can do entity type recognition with a performance comparable to human annotator. This problem is hard for a variety of reasons. In general, it is not possible to list all possible instances of a single entity type and feed it to the machine. The lack of a large annotated data corpus for training is another major impediment. Due to these reasons, existing entity recognizers are not very accurate (F1 ranging 70%-80%) (Shipra, Malvina, Jenny, Christopher, & Claire, 1900; Carreras, Màrquez, & Padró, 2002).[2] The obvious task then is to improve the performance of existing machine tagging systems. This would be achieved by looking for features (new as well as ones used with existing taggers) that affect the performance of the tagger the most. Additionally, finding out how much training data is needed, can help solve the problem of the lack of a large annotated training data corpus.

This Senior Thesis performs analysis on an existing entity recognizer by considering removal and addition of certain features that affect the performance of the recognizer. It will also address the issue of

[1] unless it is a line out of a fantasy novel, where an animal (other than a human) presides.

[2] Please refer to the section 3 for the definition of F1

whether large training data sets are necessary or not. The outcome of this project will be a step forward in making an enhanced entity recognizer which in turn will benefit other NLP applications.

2 Background and Related Work

The entity recognizer that we are analyzing is the Supersense Tagger (SST) (Ciaramita & Altun, 2006). The tagger performs sequence tagging with a perceptron-trained Hidden Markov Model (HMM). The following section will describe the tag set used, the type of sequence tagging and the model used for training.

2.1 Supersenses

The entity-type tag set we use in this research project contains types referred to as *supersenses* (Ciaramita & Altun, 2006; Ciaramita & Johnson, 2003; Curran, 2005). As opposed to the usual entity types *Person, Location* and *Organization,* and sometimes *Date* used in earlier Named Entity Recognition (NER),[3] the supersense tag set includes 26 broad semantic classes. These semantic classes are labels used by lexicographers who developed Wordnet (Fellbaum & others, 1998), a broad-coverage machine readable lexical database which has proper and common nouns, verbs, adjectives and adverbs interlinked via synonym, antonym, hypernym, hyponym and variety of other semantic relations. **Table 1** (from (Ciaramita & Altun, 2006)) shows the supersense labels for nouns. Wordnet is used to lemmatize a word and provide the most frequent supersense for the word.[4] Because this tag set suggests an extended notion of named entity, this particular process of recognition is called *supersense tagging*.

Furthermore, supersenses have been used to build useful latent semantic features in syntactic parse re-ranking (Koo & Collins, 2005). Supersense Tagging, along with other sources of information such as part of speech, domain-specific NER models, chunking and shallow parsing, can contribute a lot to question answering and information extraction and retrieval (Ciaramita & Altun, 2006).

NOUNS			
SUPERSENSE	NOUNS DENOTING	SUPERSENSE	NOUNS DENOTING
act	acts or actions	object	natural objects (not man-made)
animal	animals	quantity	quantities and units of measure
artifact	man-made objects	phenomenon	natural phenomena
attribute	attributes of people and objects	plant	plants
body	body parts	possession	possession and transfer of possession
cognition	cognitive processes and contents	process	natural processes
communication	communicative processes and contents	person	people
event	natural events	relation	relations between people or things or ideas
feeling	feelings and emotions	shape	two and three dimensional shapes
food	foods and drinks	state	stable states of affairs
group	groupings of people or objects	substance	substances
location	spatial position	time	time and temporal relations
motive	goals	Tops	abstract terms for unique beginners

TABLE 1 NOUNS SUPERSENSE LABELS, AND SHORT DESCRIPTION(CIARAMITA & ALTUN, 2006)

[3] *Named Entity* here refers to proper nouns only. Some of the earlier works include (Borthwick, 1999; Carreras et al., 2002; Finkel, Grenager, & Manning, 2005; Florian, Ittycheriah, Jing, & Zhang, 2003; Mikheev, Moens, & Grover, 1999; Zhou & Su, 2002)

[4] *Lemmatizing* refers to finding the root of a word. For example, the lemma of *ran* is *run*, while the lemma of *said,* is *say* and for *teachers,* it is *teacher*.

2.2 SEQUENCE TAGGING

In NLP, people often seek to assign labels to each element in a sequence. Here, sequence generally refers to a sentence where the words are the elements. Let $X = \{x^1, \dots, x^k\}$ denote the vocabulary of sequence elements, and $Y = \{y^1, \dots, y^m\}$ the vocabulary of tags. The task of sequence tagging is to assign lexical categories $y \in Y$ to words $x \in X$ in a given natural language sentence. NER and part-of-speech (POS) tagging are two such tasks which involve sequence labeling or tagging.

Label assignment may involve simply matching the element to a dictionary entry. For many NLP applications, however, the process can be improved upon by assigning the labels sequentially to elements from a string (usually a sentence). This allows the choice of a label to be optimized by considering previous labels.

The tagging scheme used in the Supersense Tagger is *begin-in-out (BIO) tagging scheme*. In this scheme, each token/word in a sentence is either marked as beginning of a chunk (B), continuing a chunk (*I*) or not part of any chunk (O) based on patterns identified on the basis of the training data. In the following example,

George Washington was the first President of the United States of America.

"George" would be labeled as *B-noun.person* and "Washington" as *I-noun.person*. This is because "George" is the beginning of the *noun.person* phrase and "Washington" continues that supersense. Similarly, "United" would be labeled as *B-noun.location*. Following this, "States", "of" and "America" would be labeled as *I-noun.location*. The remaining tokens are labeled *O*.

2.3 PERCEPTRON-TRAINED HMM

As mentioned earlier, the SST uses a perceptron-trained Hidden Markov Model (P-HMM)(Ciaramita & Altun, 2006, 2005). This model uses Viterbi and perceptron algorithms to replace a traditional HMM's conditional probabilities with discriminatively trained parameters. It has been successfully implemented in noun phrase chunking, POS tagging and Bio-medical NER (M. Collins, 2002; Jiampojamarn, Kondrak, & Cherry, 2009) and many other NLP problems.

The advantages of using this kind of model are that it does not require uncertain assumptions, optimizes the conditional likelihood directly and employs richer feature representation (Ciaramita & Altun, 2006). These kinds of models represent the tagging task through a *feature-vector representation*.

A feature represents a morphological, contextual, or syntactic property and typically looks like this for example

$$\Phi_{100}(y, x) = \begin{cases} 1 & if\ current\ word\ w_i\ is\ \textbf{the} \\ & and\ y = Determiner \\ 0 & otherwise \end{cases} \quad (1)$$

A vector of these features is represented as:

$$\Phi(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{d} \sum_{j=1}^{|y|} \Phi_i(y_j, \boldsymbol{x}) \quad (2)$$

Here *d* is the total number of features, x is the token that is being tagged and y is the label. The task of tagging can be represented as learning a discriminant function *F* which is the linear in a feature representation Φ defined over the space:

$$F(x, y; \boldsymbol{w}) = \langle \boldsymbol{w}, \Phi(x, y) \rangle \qquad F: X \times Y \rightarrow \mathbb{R} \tag{3}$$

$\boldsymbol{w}$, is the parameter vector of *d* dimensions. For an observation sequence **x**, the SST makes predictions by maximizing *F* over the response variables:

$$f_{\boldsymbol{w}}(\boldsymbol{x}) = \text{argmax}_{\text{y}\in\text{Y}}\, F(x, y; \boldsymbol{w}) \tag{4}$$

The process involves Viterbi decoding with respect to $\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{d}}$ **.** The complexity of the Viterbi algorithm scales linearly with the length of the sequence (Manning & Schutze, 1999).

In a nutshell, the perceptron is not a probabilistic model. It keeps scores for sequences and decides labels based on the scores.

The performance of perceptron-trained HMMs is competitive and comparable in performance to that of Conditional Random Field models (Ciaramita & Altun, 2006; Collins, 2002).

2.4 Incorporating Word Cluster Feature

The addition of new features such as word clusters in a more restricted task of NER has shown considerable improvement in performance in the system (Lin & Wu, 2009). A Word cluster is a grouping of words which fall in similar context. An example of a word cluster could be [“pet”, “cat”, “dog”,…].

The use of word clusters alleviates the problem of lack of annotated data. Word clusters are generated from unlabeled data, which is available in plenty. Once word clusters are created, the feature that holds “word belonging to a particular cluster”, can be used in a supervised training setting. Hence, even when a word is not found in the training data, it may still benefit from the cluster-based features as long as the word belongs to the same cluster with some word in the labeled data. For example, if the word “cat” is in the training data and the tagger encounters a new word “dog” in the test set, the tagger would not know what to do with the unseen word. But if a word cluster contains both of these words, the word cluster feature will be fired and the two words can share information. This improves the tagged output of the unseen words. In this project the word cluster used were created using Distributed K-means clustering (Lin & Wu, 2009).[5]

2.5 The Baseline Tagger

The baseline tagger used for comparison with modified taggers is a reimplementation[6] of the SST. It uses the same feature set as that of the SST to tag words which include both proper and common nouns, and verbs. The experiments conducted also involve tagging verbs along with nouns. The training data for the verbs is extracted the same way as for the nouns.

[5] Word clusters are generated using Distributed K-means clustering, by Dipanjan Das (LTI).

[6] This re-implementation has been done by Michael Heilman (LTI, CMU)

3 Evaluation Metrics

The following evaluations metrics are used to evaluate the performance of our tagger.

3.1 Precision (P)

Precision measures the percentage of the supersenses identified by the tagger that are correct. Large precision indicates almost everything the tagger tags are correctly tagged.

$$overall\ precision = \frac{number\ \ of\ correctly\ \ tagged\ \ phrases\ \ by\ the\ tagger}{number\ \ of\ tagged\ \ phrases\ \ by\ the\ tagger} \quad (5)$$

Note that a word is a phrase of size 1 token .

3.2 Recall (R)

Recall measures the percentage of the supersenses in the test set that are actually correctly identified. In other words, it tells, how few errors of omission are made by the tagger.

$$overall\ recall = \frac{number\ \ of\ correctly\ \ tagged\ \ phrases\ \ by\ the\ tagger}{number\ \ of\ correct\ \ hand\ labeled\ \ phrases\ \ from\ \ the\ test\ \ data} \quad (6)$$

Incorrectly tagging a phrase as *Y* that should have been labeled *X* will lower recall for *X* and lower the precision for *Y*.

Here is an example to illustrate precision and recall for a sentence (*O* refers to no supersense for the particular token):

	Hand Labeling	Machine Tagging
John	B-noun.person	B-noun.location
Smith	I-noun.person	I-noun.location
is	B-verb.stative	B-verb.stative
in	O	B-noun.place
Doha	B-noun.location	B-noun.person

Here, number of correctly tagged phrases is 1. Number of tagged phrases is 4. Hence, overall precision is 1/4. Number of labeled phrases in this case is 3. Hence, overall recall is 1/3.

Another example:

	Hand Labeling	Machine Tagging
John	B-noun.person	B-noun.person
Smith	I-noun.person	B-noun.location
is	B-verb.stative	O
in	O	O
Doha	B-noun.location	B-noun.location

Here, number of correctly tagged phrases is 1, as “John Smith” is wrongly tagged. Number of tagged phrases is 3. Hence overall precision is 1/3. Number of labeled phrases is 3. Hence, overall Recall is 1/3.

3.3 F1

F1 (F-score) is simply the geometric mean of precision and recall, and combines the two scores.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

EQUATION 1 F1

4 APPROACH

Our approach towards improving the performance of the SST involves two trials. For both trials, tagger performance is evaluated based on affects on F1.

4.1 VARYING TRAINING DATA SIZE

In the first trial, we experiment with different sizes of training data to determine if a threshold exists after which additional data does not improve the performance of the tagger. Therefore we train the system using different size training data and evaluate the trained model with test data.

4.2 FEATURE ADDITION, REMOVAL AND MODIFICATION

In the second trial, we investigate the effect of adding, removing and modifying features used in the tagging process and gauging which affect the performance of the tagger.

For this, we devise a series of experiments which involves removing one feature at a time and evaluating the tagger output. This task is termed as *feature ablation.* When a feature affects the F1 by +/-2 points, we mark it for future experimentation. As for the other features, we take conjunctions of these features and check if they collectively affect F1. Some of the baseline features from SST include most frequent sense (from Wordnet), POS tags, word shape (upper-case or lower-case, upper-case after period and so on) and label of preceding words.

Context is an essential feature while tagging words. As shown in the example in the section 1, while tagging *George Washington*, the knowledge about a President being of type *person* helps with tagging *George Washington* as *person*. The baseline tagger only looks at +/- 2 words around the current word being tagged. We perform additional experiments by reducing the context to not looking at any words (removing the existing context features) and then increasing the context to +/- 4 words (adding new context features) to see if the extra context helps.

Like any other entity recognizer, the SST can encounter words that it has never seen in the training corpus. A plausible solution to improve performance of the tagger to tag these unseen words correctly is to incorporate word cluster features (refer to section 2.4). As a result, we add new word cluster features for the current word and words (+/- 2 words) around it. We conjoin other *strong* features along with the word cluster features. These strong features are shortlisted from the feature ablation experiments and contextual analysis mentioned earlier.

5 Experiments and Results

5.1 Setup

We tested our tagger on the Semcor corpora (Miller, Leacock, Tengi, & Bunker, 1993) containing syntactically and semantically tagged text (articles) from the Brown Corpus. This includes the nouns and verbs labeled with their supersense. The Semcor data was split into 3 parts: Training, Development and Testing. The three parts were created by randomly selecting the articles. The size of the three parts were as follows:

	Training data	Development Data	Testing data
Number of sentences	11,973	4113	4052
Number of Tokens	248,581	92,924	93,269

5.2 Feature Extraction

The Tagger extracts features with their names and values for a particular token in a sentence. All of these are aggregated to get the feature vector or score for the whole sentence. From the programming point of view, if we want to add a new feature, we create a mapping of the feature name to the value and add it to the feature vector. If we do not want to include a feature, we simply do not add the mapping to the feature vector.

5.3 Experiment 1- Training Data Size

In order to find out how the size of training data affects the performance of the tagger, multiple percentages of training data were made. The tagger was then trained on each of these parts and evaluated. The F1's were calculated at intervals of 5% of the training data.

Figure 1 shows the results for the varying F1 with respect to the amount of training data used.

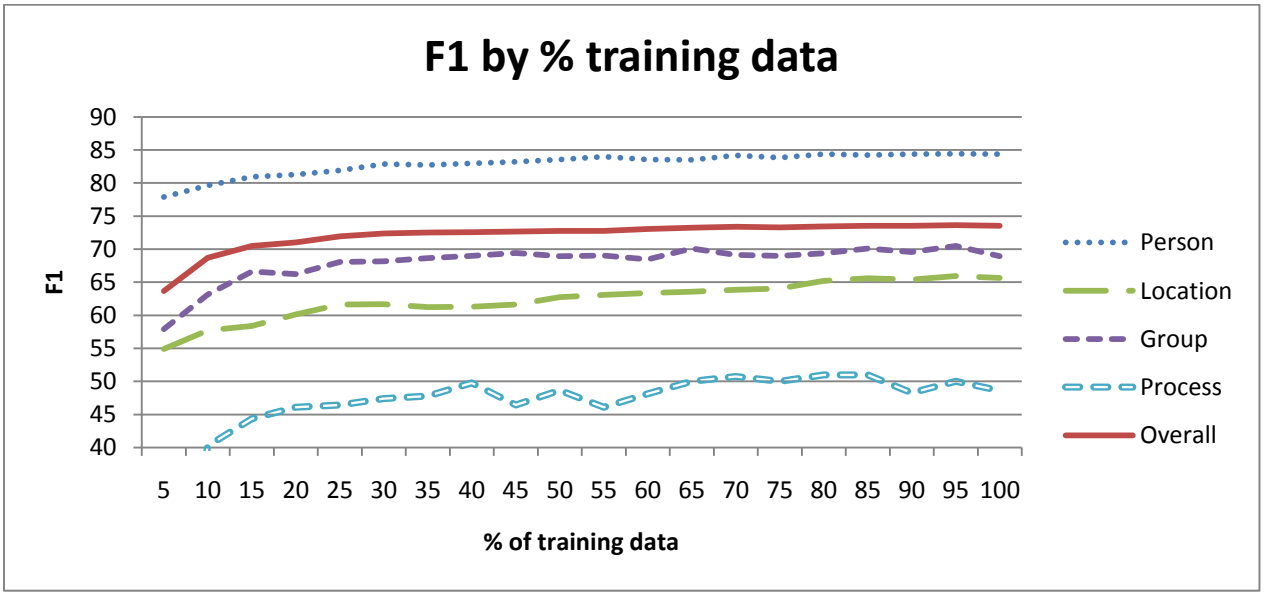


FIGURE 1 F1 BY % TRAINING DATA

The results of the experiment indicate that after about 33% of the training data is used, the overall F1 does not increase drastically. Results for the individual supersenses (a subset of supersenses shown in **figure 1**) are similar. However, there are fluctuations in the graph for *process,* while smoother curve for *Person*.

5.4 Experiment 2- Feature Ablation

The experiment to find out which feature impacts the performance of the tagger the most is conducted by removing one baseline feature at a time. These features include:

- The first-sense feature which refers to the most frequent sense of the word or token being labeled
- The POS feature which includes the part-of-speech label of the current word and the two words in front and back of the current word or token
- The word shape feature which includes capitalization of the current word/token, presence of punctuation as the previous token along with the next two and previous two tokens
- The previous label feature which is simply the label of the previous token

Feature Removed	F1	Precision	Recall
Baseline	**73.54%**	**73.12%**	**73.98%**
First Sense (Most frequent sense)	**57.11%**	57.12%	57.09%
Part of speech	73.11%	73.11%	72.50%
Word Shape	73.51%	73.13%	73.90%
Previous Label	**73.51%**	73.15%	73.89%

TABLE 1 FEATURE REMOVED AND RESULTING F1

Table 2 shows that the First Sense feature has the greatest impact on the performance of the SST, as the performance of the tagger suffers severely when it is removed. But at the same time, the removal of previous label feature affects the performance minutely. This is striking considering the fact that the tagger performs sequence tagging.

5.5 Experiment 3- Context Size

In this experiment, we look at how context size affects the performance of the tagger. The Baseline Tagger looks at the current word and two words before and after it. The tagger extracts all other features like most frequent sense of current word, POS and word shape of the current word along with two words before and after it, and label for the previous word before the current word. If the context is limited to “no words”, then none of the other features are considered. Obviously the current token to be tagged is visible. When the context is “current word”, it includes the other features along with the current token. For larger context (+/-2, +/-3, +/-4) the POS and word shape of the surrounding words are the additional features incorporated into the feature vector.

Experiments on context size are tabulated in **table 3**:

Context	F1	Precision	Recall
Baseline - Current word +/- 2 words	**73.54%**	**73.12%**	**73.98%**
No words	70.19%	69.09%	71.32%
Current word	70.94%	69.23%	72.75%
Current word +/- 1 word	73.34%	72.67%	74.02%
Current word +/- 3 words	73.50%	73.11%	73.89%
Current word +/- 4 words	73.27%	72.84%	73.71%

TABLE 2 CONTEXT SIZE AND RESULTING F1, P AND R

As shown, the highest F1 results from the baseline current word +/- 2 words context size. Reducing the context lowers the F1 and increasing does not affect the F1 much.

5.6 Experiment 4- Addition of Word Cluster Features

The first step to adding word cluster feature, is to fetch the word cluster ID of the word that is being tagged. The word cluster ID is simply a pointer to which cluster the word belongs to. This experiment considers a context of current word +/- 2. Hence, the cluster ID for these words are also needed. The next step involves adding the new features for each of these words as mentioned in section 5.2.

Word clusters of various sizes K = 64, 128, 256, 512, 1024, 2048 were used for this experiment. **Table 4** shows the results for the above experiment.

K (word cluster size)	F1	Precision	Recall
Baseline	**73.54%**	**73.12%**	**73.98%**
64	73.69%	73.26%	74.13%
128	73.77%	73.37%	74.18%
256	73.77%	73.39%	74.15%
512	73.73%	73.37%	74.09%
1024	**73.91%**	73.52%	74.30%
2048	73.80%	73.46%	74.15%

TABLE 3 WORD CLUSTER WITH VARYING K AND RESULTING F1, P AND R

Generally, the addition of word cluster feature has not led to poor results. As for K = 1024, the performance of the tagger shows promising results.

The obvious next step is to conjoin some strong features with the word cluster feature to expect good results. So following this experiment, we added the “First Sense” feature in conjunction with the existing word cluster feature. We chose the “First Sense” feature as it is the strongest of the features as shown in Experiment 2. We again evaluated the tagger with the same set of clusters and led to the following results in **table 5:**

K (word cluster size)	F1	Precision	Recall
Baseline	**73.54%**	**73.12%**	**73.98%**
64	73.69%	73.32%	74.02%
128	73.74%	73.32%	74.16%
256	73.80%	73.38%	74.22%
512	73.76%	73.39%	74.13%
1024	73.74%	73.42%	74.07%
2048	73.73%	73.34%	74.12%

TABLE 5 WORD CLUSTER AND FIRST SENSE FEATURE WITH VARYING K AND RESULTING F1, P AND R

These results also suggest the addition of word cluster features leads to slightly improved results but "First Sense" feature did not provide any extra help.

In both of the above evaluations, the SST had cluster features for all the words (current, previous and next, previous 2 and next 2). Next step, we took the case which had the best performance- the one with K=1024 and removed the cluster features for two words to the right and left of the current word. We evaluated the trained model and the result was as follows:

F1 = 73.98% P = 73.58% R = 74.37%

After this, we removed the cluster features for one word to the right and left of the current word. The result was as follows:

F1 = 73.83% P = 73.45% R = 74.21%

5.7 Experiment 5- Addition of Bigram Feature

We also trained the tagger with Bigram feature (the feature considers groups of two words; for e.g.: current stem = "are" AND next stem ="going") . The F1 for this turned out to be **74.10%**. The precision was 73.60% while the recall is 74.53%

The downside of training with the Bigram feature is that in the worst case, it would add $|V|^2$ features to the model, where *V* is the vocabulary. This eventually leads to more time for training. This experiment took around four hours as opposed to the previous ones which took only about two hours. Also note that the result achieved by including Bigram feature (with F1 = 74.10%) is almost equivalent to the result achieved by including word cluster feature for the current word and two words around it (with F1 = 73.98%).

6 Inference and Error Analysis

6.1 Looking At The Baseline

Figure 2 shows the F1 for some of the supersenses after using the baseline tagger.

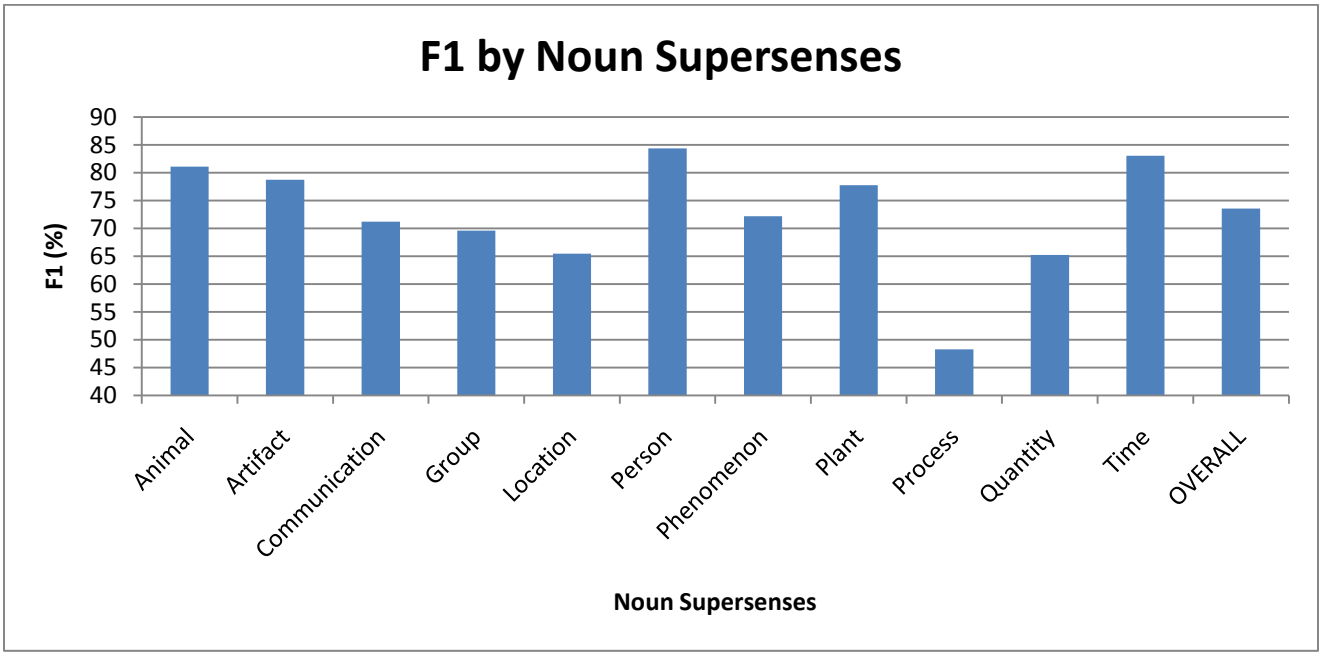


FIGURE 2 F1 BY SUPERSENSES

We can clearly see that supersense *noun.process, noun.location, noun.quantity* and some other supersense have low F1. This could be attributed to the fact that these supersenses have lesser number of training instances. Another possible reason could be that the most frequent sense for certain words from Wordnet is not the right sense for these words. The following sub-sections will address the two issues in hand.

6.1.1 Input Data Analysis

When we look at the number of instances of a supersense in the training set and compare with the F1 (which is also present in **figure 1**). **Table 6** contains the noun supersenses along with the number of instances in the training set, F1 and number of those instances in the test set. The table is sorted by the size of the instances in training set. The general trend would be – more the number of training instances, higher the F1. But this is not true in some cases like in that of *noun.shape, noun.motive* and *noun.process.* While *noun.process* has more instances in training set, its F1 is way lower than that of *noun.shape* or *noun.motive*. *T*he small size of the test set affects the F1 vastly, even if a minor change in number of correctly tagged changes. This explains the fluctuations for *noun.*process curve in **figure 1.**

Noun Supersense	Number of instances in training set	F1 (%)	Number of Instances in test set
noun.person	**8534**	**73.03**	3080
noun.artifact	5498	80.99	1859
noun.act	4516	79.35	1870
noun.communication	**4134**	**67.44**	1448
noun.group	3840	83.39	1096
noun.cognition	3574	73.54	1530
noun.location	2773	70.82	887
noun.attribute	2600	62.83	990
noun.time	2367	75.54	988
noun.state	1912	80.24	727
noun.body	1560	70.27	662
noun.quantity	1133	65.62	409
noun.possession	1127	69.23	209
noun.substance	1081	61.41	594
noun.event	1051	84.35	398
noun.object	905	72.19	300
noun.phenomenon	647	77.71	304
noun.animal	638	71.24	339
noun.relation	573	48.62	157
noun.feeling	481	64.82	163
noun.food	410	58.21	157
noun.plant	**350**	**50.00**	87
noun.process	317	68.65	119
noun.shape	219	81.98	53
noun.motive	**107**	**83.39**	19

TABLE 6 NUMBER OF INSTANCES OF NOUN SUPERSENSE IN TRAINING AND TEST SET ALONG WITH F1

6.1.2 First Sense Not The Winner Always

Digging deeper into the test data, words like “reaction” were tagged as *noun.act* or *noun.phenomenon* (which is the most frequent sense) while the right supersense was *noun.process*. Similarly, for “air”, the tagger marked it as *noun.substance* which is the most frequent sense while the original label for it is *noun.location*.

6.2 Context Size Analysis

Experiment 3 led to the conclusion that further away the word is (for larger context) the less likely there will be any semantic relation. Therefore context of current word +/-2 seems to be optimal.

6.3 Word Cluster Feature Benefits

Although the improvements with addition of word cluster features did not result in a very high F1, there are many instances where the word cluster feature has helped while the baseline tagger failed. Some of the examples are:

Sports Writer *Ensign Ritche of Ogden Standard Examiner went to his compartment to talk with him.*

The Baseline Tagger and Tagger with word cluster feature using cluster of size 1024 labeled "Sports Writer" as:

	Baseline	With Word Cluster Feature
Sports	B-noun.act	B-noun.person
Writer	B-noun.person	I-noun.person

In cluster 81, "Sports Writer" is with other occupations like "chemist", "cartographer", "scholar", "meteorologist" and many more. Another example:

Jim Landis' 380 foot ***home run*** *over left in first inning…*

The tagger with word cluster features recognizes "home run" as *B-noun.act* and *I-noun.act* respectively. On the other hand the baseline missed out on "run" after tagging "home" as *B-noun.act.*

7 Conclusion And Future Work

In this work, we highlighted how syntactic, contextual and word cluster features affect the performance of a system for tagging words with high level sense information. This project will help further research by suggesting areas to explore or not:

- We have demonstrated that lack of large annotated data is not a major issue. Nevertheless, this does not mean more annotated training data is not needed. But this suggests that a big project to annotate more data would likely be fruitful.
- The fact that previous label did not greatly affect the performance of the tagger seems to suggest that a sequence labeling approach is not necessary for good performance (as long as the constraint of proper BIO output is satisfied).
- Feature ablation methods like the ones described in the experiments help find out which features are important and hereby suggest areas to work (e.g.: new features to extend or add). Addition of word cluster and bigram features is an option to be considered.
- More research can be encouraged in creating word clusters using different techniques and of different granularities.

More importantly, it boils down to finding out which features are significant and when they should be used so as to achieve high performance standards in an entity recognizer or supersense tagger.

REFERENCES

Borthwick, A. E. (1999). *A Maximum Entropy Approach to Named Entity Eecognition*. New York University.

Carreras, X., Màrquez, L., & Padró, L. (2002). Named Entity Extraction using AdaBoost, proceeding of the 6th Conference on Natural language learning. *August*, *31*.

Ciaramita, M., & Altun, Y. (2005). Named Entity Recognition in Novel Domains with External Lexical Knowledge. In *Proceedings of the NIPS Workshop on Advances in Structured Learning for Text and Speech Processing*.

Ciaramita, M., & Altun, Y. (2006). Broad-coverage Sense Disambiguation and Information Extraction with a Supersense Sequence Tagger. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing(EMNLP)*.

Ciaramita, M., & Johnson, M. (2003). Supersense Tagging of Unknown Nouns in Wordnet. In *Proceedings of EMNLP* (Vol. 3).

Collins, M. (2002). Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms. In *Proceedings of conference on EMNLP-Volume 10*.

Curran, J. R. (2005). Supersense Tagging of Unknown Nouns Using Semantic Similarity. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* .

Fellbaum, C., & others. (1998). *WordNet: An Electronic Lexical Database*. MIT press Cambridge, MA.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the 43rd Annual Meeting of the ACL*.

Florian, R., Ittycheriah, A., Jing, H., & Zhang, T. (2003). Named Entity Recognition through Classifier Combination. In *Proceedings of CoNLL-2003*.

Francis, W. N., & Kucera, H. (1967). *Computational Analysis of Present-day American English*. Brown University Press Providence.

Jiampojamarn, S., Kondrak, G., & Cherry, C. (2009). Biomedical Named Entity Recognition Using Discriminative Training. *Recent Advances in Natural Language Processing V: Selected Papers from Recent Adavances in Natural Language Processing 2007*.

Koo, T., & Collins, M. (2005). Hidden-variable Models for Discriminative Reranking. In *Proceedings of the conference on Human Language Technology and EMNLP*. Vancouver, British Columbia, Canada.

Lin, D., & Wu, X. (2009). Phrase Clustering for Discriminative Learning. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*.

Manning, C., & Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Mikheev, A., Moens, M., & Grover, C. (1999). Named Entity Eecognition without Gazetteers. In *Proceedings of EACL*.

Miller, G. A., Leacock, C., Tengi, R., & Bunker, R. T. (1993). A Semantic Concordance. In *Proceedings of the 3rd DARPA workshop on Human Language Technology*.

Shipra, D., Malvina, N., Jenny, F., Christopher, M., & Claire, G. (1900). A System for Identifying Named Entities in Biomedical Text: How Results from Two Evaluations Reflect on Both the System and the Evaluations. *Comparative and Functional Genomics*.

Zhou, G. D., & Su, J. (2002). Named Entity Recognition Using an HMM-based Chunk Tagger. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*.

Final Report for Senior Thesis Research Program 2009-2010
Carnegie Mellon University
School of Computer Science

Education E-Village:
Empowering Technology Educators in Developing Regions

Mohammed Kaleemur Rahman
School of Computer Science, Class of 2010
Carnegie Mellon University

Advisor:
M. Bernardine Dias, Ph.D.
Assistant Research Professor, Robotics Institute
Carnegie Mellon University

Table of Contents

1. Abstract

There exists a significant need for relevant, accessible and useful resources to enhance technology education in developing regions [1] [2]. Currently, access to courseware for technical subjects such as Computer Science is available several online resources [3] [4]. However, these resources are designed for developed communities, where technology is ubiquitous, technology infrastructure is robust, and educators have easy access to a variety of academic publications and other helpful guides. Therefore, the available online resources do not provide sufficient avenues for educators in developing regions to understand the courseware or discuss alternative ways of teaching it based on their specific constraints. To address this deficit, the TechBridgeWorld group at Carnegie Mellon University initiated the "Education e-Village" (E-Village) project. E-Village is an online community where educators from around the world will be able to share ideas, experiences, expertise, educational resources, and strategies to promote and enhance technology education in developing regions.

This senior thesis project enhances the search functionality and user experience of E-Village. We analyzed existing search solutions and chose the Open Source search engine Lucene for integration as it met our needs best. To enhance the user experience, we followed both heuristic evaluation and user-testing approaches. In order to perform user testing, we created electronic mockups of features based on structured essential use cases [5]. These included modified screenshots and mockups of the current E-Village design and commonly used websites. Finally, we conducted these usability tests with a representative sample of 18 users. We compiled a list of problem areas and user preferences, and addressed them in a set of recommendations. The focus of the improvements was to make the user interface (UI) as intuitive as possible, while staying consistent with user expectations.

2. Introduction

There exists a significant need for relevant, accessible and useful resources to enhance technology education in developing regions [1] [2]. Currently, access to courseware for technical subjects such as Computer Science is available several online resources [3] [4]. However, these resources are designed for developed communities, where technology is ubiquitous, technology infrastructure is robust, and educators have easy access to a variety of academic publications and other helpful guides. For example, they assume the presence of resources such as a good internet connection to download course materials such as videos and lecture slides, and materials needed to build robots in a robot programming course. Although these resources have a mechanism for general feedback, there are no avenues to collaborate and adapt a course to a region, or to figure out if substitute materials can be used. Upon facing any issues with courses, there are no avenues to get in touch with the authors. Therefore, the available online resources do not provide sufficient avenues for educators in developing regions to understand the courseware or discuss alternative ways of teaching it based on their specific constraints.

To address this deficit, the TechBridgeWorld group at Carnegie Mellon University initiated the "Education e-Village" (E-Village) project. E-Village is an online community where educators from around the world will be able to share ideas, experiences, expertise, educational resources, and strategies to promote and enhance technology education in developing regions. Educators will benefit from course materials and curricula made available by members of the E-Village community, contribute their own resources or ideas towards

extending and evaluating existing resources, share best practices relevant to teaching computing technology in under-resourced settings, seek or offer advice on particular topics or issues, and learn about publication opportunities, conferences, funding sources, professional organizations, and other opportunities for advancing their educational offerings and professional growth.

The goal of E-Village is to empower younger generations in developing communities to become create technology solutions that would be useful in their immediate communities. We are focusing on post-high-school level because in most developing regions, students are not introduced to technical courses until they reach University level. We are focusing on technology courses due to the ease of their application and impact in the immediate communities [1] [2] [6]. Core costs of communication and computing have dropped significantly over the recent past and are at a point where they can be deployed to have immediate and large-scale impact [1]. Currently, a team of researchers at both Pittsburgh and Qatar campuses is working on different aspects of E-Village. This project is being sponsored by Yahoo! under their Yahoo Technology for Good grant.

3. Thesis Goals

My thesis is involved with 2 critical aspects of E-Village - search functionality and User Experience (UX).

3.1 Designing Search Functionality

Users of E-Village should have the ability to search for specific information and obtain accurate results in a reasonable amount of time. E-Village will contain information in different formats such as courseware, a discussion board, and general information. A further complication for E-Village is that many people who access this content from developing regions will be using low-bandwidth and often flaky internet connections so the search capabilities need take these parameters into account. Finally, internet access in developing regions can often be very costly. Hence, an efficient and effective search capability is essential for the success of E-Village.

As E-Village grows, it will become especially important that search results are quick and accurate. For example, a user who is looking for information on mobile robots might expect to see information under courseware, relevant topics that have been discussed on the discussion board, and a method to contact other technologists who have experience with this topic. Users might also require an option to look for things within a certain realm of information such as limiting the search to a specific geographical location or range of dates. A simple iterative search will often be ineffective in such applications and instead, a search function with indexing and data mining components will become essential to maintain effectiveness.

Therefore, to select an effective search option, we analyzed the different parameters and constraints for E-Village and rated the different options for search based on this analysis. We first determined what areas of E-Village should be searchable, how relevance should be assigned after performing user studies (see section 7.2). Other important aspects were determining how search results should be presented to users, when "Advanced Search" functionality should be presented, and what kinds options should be provided for advanced search.

3.2 Designing User Experience (UX)

According to the Nielsen Norman Group,

> "*User experience encompasses all aspects of the end-user's interaction with a company, its services, and its products. The first requirement for an ideal user experience is to meet the exact needs of the user, without fuss or bother. The next requirement is that products are defined by simplicity and elegance making them a joy to own, a joy to use. True user experience goes far beyond giving users what they say they want, or providing checklist features. In order to achieve high-quality user experience in a company's offerings there must be a seamless merging of the services of multiple disciplines, including engineering, marketing, graphical and industrial design, and interface design"* [7].

The User Interface (UI) design greatly influences how a user experiences web applications. In the case of E-Village, it will be critical that information is organized effectively and made available to the user in an easy-to access format. A preliminary interface for E-Village with basic functionality has already been implemented by TechBridgeWorld students and staff. We critiqued this initial design to assess its effectiveness in characteristics such as conveying information in a useful manner, ease of navigation, and efficient use of space. Our critique is informed by two types of Human-Computer-Interaction methodologies and by initial feedback from usability tests with relevant TechBridgeWorld partners who will ultimately be the primary benefactors of E-Village. These tests were conducted to determine the effectiveness of specific areas of the current prototype, and to evaluate potential features. Once our critique was complete, we identified areas for enhancement, and made detailed recommendations for the new design.

4. Related Work

In this section, we review existing solutions and methodologies that are relevant to our work. First, we describe work done on education in online communities, and the challenges that come with it. Second, we describe specific projects that are currently being used to serve purposes similar to that of E-Village. Third, we describe work that guides the design of interactive technology, specifically in web usability. Finally, we outline work that has been done on usability testing. As a major proportion of E-Village users are expected to be in remote developing regions, we focus on literature relevant to remote usability testing methodologies.

Regarding learning in online communities, Renninger and Shumar assert that learning and change in virtual communities is increasingly interdependent with learning and change in the participants' local institutions [8]. Stakeholders from these local communities are needed to channel them into a virtual community. The authors cite that using external mentors incentivizes students to learn. For example, a student may be given an assignment to write a business plan, and the critique by his teacher will be expected and likely be ignored as part of the grading system. But the critique received from an external businessperson will be viewed as authentic. Due to the experience of the external person, the critique is likely to be taken more seriously by the student. Also outlined are reasons for failure of online education communities. Some of these are that such communities are slow growing and need time to mature, traditional internet tools do not facilitate collaboration, the presence of technological gaps and limitations, and the lack of experience in teachers

planning or leading online activities. Finally, the authors conclude that access to collaborative tools such as discussion boards provide social support but do not create a sense of community. In most cases, a sense of community emerges only when educators, researchers, and scientists start working together on compiling educational materials.

In another study related to virtual classrooms and communities, Neal finds that communication in class has several aspects including engagement and involvement [9]. The use of a variety of collaboration technologies provides richer communication than any one of them alone, and helps to foster a sense of community as found in a physical classroom. The technologies evaluated include videoconferencing, audio conferencing, Internet Relay Chat, NetMeeting, virtual worlds, and modes of asynchronous communication. Neal mentions benefits of distance learning such as being able to experiment with technologies, minimal travel for the instructor, and the ability to bring in guest lecturers with no additional travel expenses. However, the amount of time needed for the instructor to prepare was higher due to the overhead caused by having to schedule meetings, contact students, and updating materials on the website.

The MIT OpenCourseWare (OCW) project is a web-based dissemination of MIT course content. Courseware is freely available to anyone in the world. OCW was started in 2002 by the Massachusetts Institute of Technology (MIT) to spread knowledge and educate students through the Internet [3]. By 2009, the site had 1950 published courses in more than 35 academic disciplines. Although there is a mechanism for general feedback, OCW does not provide access to any MIT faculty. According to the OCW site, each published course requires an investment of $10,000 to $15,000. This is used to compile materials from professors, ensure proper licensing, and converting the materials in a format that can be globally-distributed. Courses with video content are estimated to be twice as expensive as regular ones. OCW is being used successfully by educators, students, and self-learners for a wide variety of purposes.

The Open.Michigan project is an initiative by the University of Michigan (U-M) aimed at creating and sharing knowledge resources and research with the global community [4]. It was started in 2007 by the U-M Medical School as a move towards a more open education process. 10 academic departments within U-M are currently participating in the initiative. The site provides general directions to contact various departments within U-M regarding materials, and pointers to avenues to share educational resources. There is a facility to provide general feedback on a course but no platform to discuss issues with courses. Open.Michigan uses open content licensing, and encourages the use, redistribution and remixing of educational resources.

Project aAqua is a Q&A forum where people can post agriculture-related questions and get them answered by certified experts [10]. It was started in 2003 by Developmental Informatics Laboratory (DiL) of IIT Bombay, to provide an avenue for farmers in India to get their questions answered by experts. The certified experts include professors, deans and other credible authorities for information. As of September 17, 2009, there were 9393 members, 17 forums and 8769 topics being discussed. There are forums on crops, animals, agriculture recommendations, market information, prices and farmer schemes. In a poll conducted on the site, 85% votes indicated that users wanted to use aAqua on their cell phones. Registered users are sent "free" crop tips to registered users via cell phones.

In *The Design of Everyday Things*, Norman discusses usability principles by discussing poorly designed objects that are encountered in daily life [11]. He argues that good design "must explain itself," and describes four principles of good design. These include (i) Visibility - user can tell what the state of a device is, and what actions are available; (ii) Good conceptual model - the presentation of operations and results is consistent, and feels part of a natural process; (iii) Good mappings - relationships between actions and results should be easily determinable; (iv) Feedback - user should receive full and continuous feedback about the results of actions. Although these principles are described with commonly-used physical devices, they can be applied to anything that requires human interaction.

A good amount of work has been done on web usability. Most notable are works done by Krug, Nielsen, Loranger, and Tahir. In *Don't Make Me Think*, Krug lays the foundation for his first law of usability stating that websites should not have elements that make users think and distract from the task at hand [12]. Actions available on a site should be self-evident and intuitive. He also provides frameworks for conducting quick and cheap usability tests. In *Homepage Usability*, Nielsen and Tahir argue that the homepage is the most important page on any site, and make the case that special attention should be given to it serves as the entry point to a site [13]. They provide detailed descriptions of homepage usability, and an evaluation of the homepages of 50 commonly used websites. In *Prioritizing Web Usability*, Nielsen and Loranger report results of their extensive user testing, and critique real-world sites for legibility, navigability, searchability, appropriate design and other usability factors [14]. Finally, Nielsen outlines ten general principles for user interface design that can be followed as guidelines [15].The combination of these resources gives a good understanding of web usability principles.

In terms of remote usability testing, the literature is very sparse. In their work, Dray and Siegel outline the advantages and disadvantages of both synchronous and asynchronous modes of remote testing [16]. In synchronous methods, the test facilitator manages the test and receives data from the participant, who is remote, in real-time. In asynchronous methods, there is no interaction with the facilitator and there is no real-time date being received. One major disadvantage with asynchronous methods is that they do not collect real-time observations, and are hence limited to self-reporting and the biases that come with it. In related work, Thompson, Rozanski and Haake make the case that synchronous remote testing using software such as NetMeeting, WebEx, Lotus Sametime, and WebQuilt, can be as effective as traditional in-person testing to identify usability problems [17]. Finally, Nielsen describes a mathematical model for finding usability problems, which can be used to plan the quantity of user testing to achieve varied levels of problem-finding [18].

5. Initial Implementation of E-Village

The initial E-Village implementation was performed on the OpenCms Content Management System [19]. We switched to Drupal [20] due to better developer-support and ease of maintenance. The preliminary implementation of E-Village had an initial UI design and contained materials relevant to two courses for testing purposes. We envision the development of E-Village through three distinct stages:

(i) Pre-Pilot:
This stage determines the high-level requirements for functionality of E-Village, and broadly

specifies its design. Usability tests and interviews with a representative sample of our target users inform the needs and preferences that drive this design. By the end of this stage, all critical functions such as search and course submissions are fully specified, and an informed prototype for the UI is designed. This thesis completes the pre-pilot stage of E-Village.

(ii) Pilot:
During this stage the design and functionality specified in the pre-pilot stage are implemented and this pilot version of E-Village is launched for longer-duration tests with a selected group of first users. The usability tests at this stage will determine the final design and steady state operations of E-Village (see 'Future Work').

(iii) Post-Pilot:
This is the steady-state stage of E-Village where the online community is available to everyone online with tested features and functionality.. Occasional usability tests and feedback from users will drive any further enhancements as the needs arise.

6. Search Functionality

On any website, having useful search functionality is essential. When users are looking for something on a site, they mainly use either the navigation menus or search. Most users type one, two or three words into the search box to look for something and expect useful results [14]. As search functionality is commonplace these days, users have also formed mental models of what search should return. With the wide use of highly-optimized web search engines such as Google, users' expectations of search have increased tremendously [21]. Hence, it is very important that users' search experience on a site is favorable.

There are a number of benefits of having a search engine to look for information on a site. They can be used to understand what is important to users and tune the site accordingly, to satisfy users' mental model of having a search box on each site, and to set up automatic indexing mechanisms for dynamic content [22]. Internal site search has several advantages compared to a world-wide web search. Some of these advantages described by Nielsen and Loranger are [14]:

- Site search deals with a smaller set of pages compared to search engines for the entire web.
- User studies can be performed to understand users and their intentions.
- The importance of documents is well known. So the relevance rankings can be prioritized as opposed to being computed by web search engines.
- More metadata can be accessed, allowing site search to learn more about document relationships.

Due to the advantages of having a site search, we determined it would be best to find a suitable search engine, and customize it for E-Village.

6.1 Solution Requirements

In order to determine which solution would be optimal for E-Village, we enumerated a list of requirements that addressed its constraints:

- *Efficiency*: Solution would consume a reasonable amount of resources, and be able to return results quickly.

- *Supported content types*: Solution should be able to work with common content types including file formats such as HTML, PDF, TXT, and Microsoft Office.
- *Inherent limitations*: Solution should not have any limitations that would hinder the growth of E-Village in the long run.
- *Cost*: Solution should preferably be free or have minimal cost of setting up and maintenance.
- *Platform dependency*: Solution should be able to run on popular operating systems including Windows, Macintosh and Linux.
- *Offline functionality*: Solution should provide mechanism to allow users to search and save search results offline. This could be useful in areas where internet connections are less reliable.
- *Availability of documentation*: Solution should have sufficient documentation available freely. It should have a mechanism for support in case we encounter problems. These could take the form of books, online tutorials, and developer communities.
- *Ease of integration and management*: The solution should be easy to integrate within popular CMS software.

6.2 Solution Analysis

In our quest for the best search solution for E-Village, we looked at both Open-Source solutions and various commercial site search service options. The Open-Source solutions consisted of free search engines that could be integrated into E-Village. In this category, we looked at Lucene, Sphinx, and the Xapian Project [23] [24] [25]. The site search services are free or paid search services that would index the content on a site and return results based on processing that is done at the service's expense. In this category, we looked at Google Custom Search Engine (CSE), FreeFind and PicoSearch [26] [27] [28]. Among all these solutions, we looked more closely at Lucene and CSE as they were the most popular and offered the most in terms of functionality in their respective categories. The following table illustrates the comparison between Lucene and CSE:

Feature	**Lucene**	**CSE**
Supported file types	Text, Rich Text Format, XML, HTML, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Adobe Portable Document Format	Adobe Portable Document Format, Adobe PostScript, MacWrite, Microsoft Excel, Microsoft PowerPoint, Microsoft Word, Microsoft Works, Microsoft Write, Open Document Format, Rich Text Format, Shockwave Flash, Text.
Search indexing limit	Unlimited	5000 annotations, which represent inclusion of URL or URL pattern e.g. including the URL www.foo.com/ and all its subpages would constitute one annotation; including the URL pattern www.foo.com/*bar (which includes all pages in this domain with 'bar' in the URL) also constitutes one annotation.
Cost	Free	Free

Platform dependency	Platform independent	Platform independent
Offline functionality	No	No
Availability of documentation	Documentation is highly fragmented and disorganized. Main sources are Drupal site, Lucene site, online forums, and books on Lucene.	Clear and detailed documentation available on CSE site for both beginner and advanced users. Also included are tools like Google Marker to include other sites in search results.
Ease of integration and management	Implementation will be more involved, and search will require optimization and maintenance once implemented.	Does not require much upkeep once implemented. Availability of easy-to-follow documentation facilitates this.

Both solutions supported the essential file formats, were available at no cost, were platform independent, and did not possess any kind of offline functionality. Therefore, the primary difference was that Lucene had unlimited indexing whereas CSE had a limit of 5000 annotations after which you have to either remove documents from your index, or monetarily increase the limit. Even though the CSE limit of 5000 annotations would easily satisfy E-Village needs in the near future, it would limit us in the longer term, especially if the activity and content on the site increased drastically. In terms of documentation, CSE had better centralized resources compared to Lucene whose documentation was fragmented. Also, integration and management in Lucene would be more involved. However, Lucene had a supportive developer community. We also determined it would be best to have control over the search on our server, as search services could change their policies unexpectedly. Given these factors, we selected Lucene as the search engine that would be the best fit for E-Village.

6.3 Selected Solution: Lucene

Lucene is a simple, high-performance and powerful full-featured text search engine. It has the ability to perform field searching and add new indexed content without regenerating the entire index. It is a software library that can be integrated into various applications. Being a Java library, it is very flexible compared to other applications [29]. It provides the ability to search in many different languages, to perform date-range searching, and extended field searching i.e. focusing on a certain field e.g. Title, Author, or Content. MIT's OpenCourseWare uses a Lucene search engine at its backbone [30]. Zend Framework has a PHP port of Lucene that can be plugged into Drupal search. As we are using Drupal to run E-Village, this is very useful.

In his Pisa lecture, Cutting does a good job of outlining the architecture of Lucene [31]. The Lucene Architecture consists of four major abstractions: Document, Analyzer, Indexer, and Searcher. A Document is a sequence of Fields, where each Field is a <*name*, *value*> pair. Here, *name* is the name of the field e.g. title, author, content, etc. and *value* is the text (data) that it maps to. An Analyzer is a TokenStream factory, where each TokenStream is an iterator over Tokens. A Token is a tuple <*text*, *type*, *start*, *length*, *positionIncrement*>. Here, *text* is the data text in the document, *type* categorizes the text, *start* and *length* are offsets in characters, and *positionIncrement* is typically set to 1. The Indexer maps Terms to <*df*, <*docNum*,

<*position*>*>*> tuples. Here, Term is a <*fieldname*, *text*> tuple, *df* is the Document frequency, *docNum* is the Document ID, and *position* indicates location within Document.

Lucene uses a B-tree based inverted indexing strategy [32]. This has the advantage of updating in place and fast searching once indexing has been completed. Inserts and lookups on B-Trees are O(log n) operations. Lucene takes a slightly different approach in the way it indexes. Instead of maintaining a single index, it builds multiple indexes and merges them periodically. An index is built for each document, and these indexes are merged periodically to keep the number of indexes small so that searches are quick [33]. In Lucene's algorithm, a stack of indexes is maintained. For each new document, an index is generated and pushed onto the stack. The following pseudo code [31] illustrates this incremental merging algorithm [31]. Here, b is the merge factor and M is set to infinity:

```
for ( size=1; size < M; size *= b ){
      if ( there exist b indexes with size docs on top of stack ){
            pop them off the stack;
            merge them into a single index;
            push the merged index back onto the stack;
      }
      else {
            break;
      }
}
```

Fig. 1 below illustrates how this works with an example. Here, b=3, 11 documents have been indexed, the stack has 4 indexes, and 5 merges have taken place. The grayed indexes have been deleted.

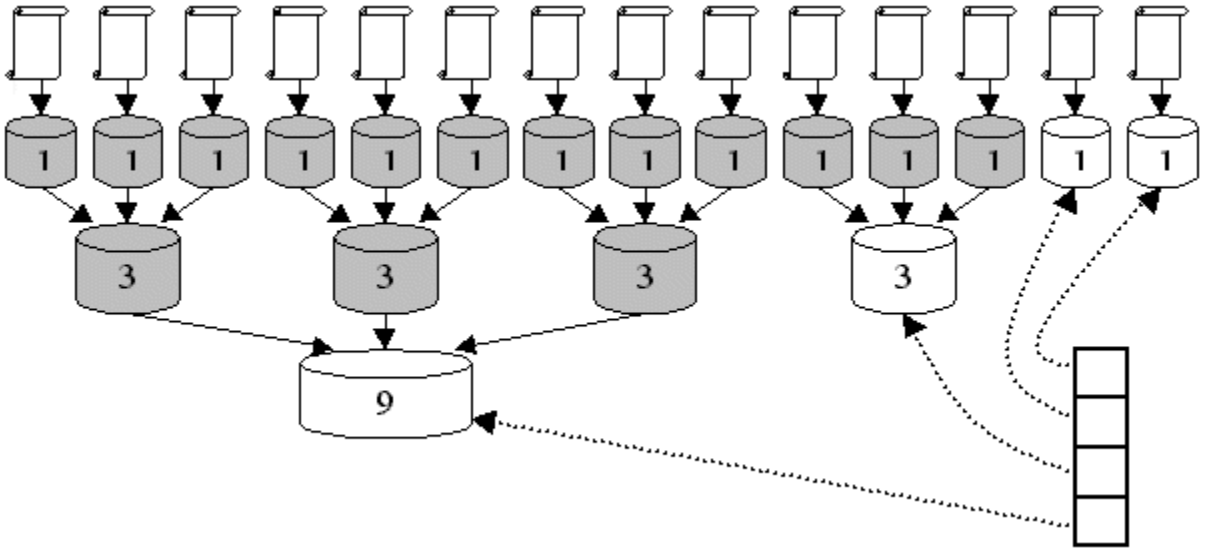


Fig. 1 Lucene Indexing Algorithm [31]

Lucene's search algorithm maintains a queue of posting streams, where each posting is a <*Term*, *Document ID*, *Weight of Term in Document*> tuple [34]. The following pseudo code was inferred from Cutting [34].

```
While ( there are posting streams remaining in the queue ){
      Calculate score for each posting in stream;
      Merge postings of each Term in query;
      Keep top k ranking documents only;
}
```

7. User Experience

The targeted E-Village users could belong to any of the following combinations of background and exposure to technology:

Developed region, high-exposure to technology	Developed region, low-exposure to technology
Developing region, high-exposure to technology	Developing region, low-exposure to technology

We expect most of the users to belong to the top-left and bottom-right quadrants. According to Jakob's Law of the Internet User Experience [13], users spend most of their time on other websites. The accumulated amount of time spent on other sites will greatly outnumber the amount of time spent on our site [14]. Users who have a good exposure to technology prepare their expectations from a site based on previous experiences on other sites. If they are accustomed to prevailing design standards, they will expect to encounter similar conventions on E-Village. Hence, it is not worth making them work hard with a deviant user interface. Additionally, for users who may not have had a good amount of exposure to technology, it will become important that the UI is intuitive and easy to follow. It should not have any elements that require the user to think for a while before figuring out how to accomplish a task, or elements that distract users from performing the task at hand [12].

Our principal goal in designing the E-Village user experience, therefore, is to ensure that the UI is consistent with expectations of users with good exposure to technology and intuitive enough for those with a lower exposure. Users should be able to navigate E-Village without a difficult learning curve. We followed a two-pronged approach on enhancing the UI of the preliminary implementation of E-Village. First, we take a heuristic evaluation approach which builds on work done by leading web usability experts. Second, we perform feature-specific usability tests to understand preferences and user habits of our target user group, and evaluate potential features. In both these testing methods, the goal is to find and document as many usability problems in a UI so that they can be addressed in future versions [18]. Finally, we use the results obtained from both approaches in conjunction with knowledge gained through our literature review to enhance the UI and the overall UX in E-Village.

7.1 Heuristic Evaluation

Heuristic evaluation is a discount usability engineering technique used for quick, cheap and easy evaluation of a user interface design [35]. It involves having a set of recognized usability principles ("heuristics") that can be used to evaluate the effectiveness of a UI. The UI is examined to see if it adheres to each principle as part of an iterative design process. It is the most popular usability inspection method. The idea here is to utilize work that has been done by leading web usability experts to enhance the UI of E-Village.

We used the work done by Krug to synthesize a set of usability guidelines that we could use as benchmarks [12]. For example, a sample guideline was designing pages for scanning, not reading. This included the attributes, using clear visual hierarchy, using conventions unless the new conventions do not confuse in any way, breaking up pages into clearly defined areas, making it obvious what is clickable, and ensuring low visual noise. Each attribute was evaluated on a 1-point scale. For each attribute, the scoring of E-Village was as follows:

- 1.0 - if the site met the attribute requirements completely
- 0.5 - if the site met the attribute requirements partially
- 0.0 - if the site did not meet the attribute requirements at all
- N/A - if attribute was not applicable to the site

We did not assign the number of points available to each attribute on any scale of relative importance. The reasoning was that the number of attributes that we could find under each guideline would indicate how important the guideline was. The goal in designing this scoring system was to figure out where E-Village is lacking in usability, and devise improvements for each attribute that needed them. The following table summarizes how the preliminary implementation of E-Village was evaluated against these guidelines:

Guideline	**Score (%)**
Designing pages for scanning, not reading	80%
Ensuring choices available to users are not ambiguous	50%
Using concise language	50%
Designing persistent navigation considering the 5 key elements: Site ID, Sections, Utilities, Home, and Search	80%
Designing page names	100%
Showing users where they are on the website through the use of appropriate indicators	0%
Using navigational indicators	0%
Using tabs for navigation	N/A
Designing content for homepage	50%
Getting the message across on the Home Page	0%
Designing pull-down menus	N/A

Although the prototype was at a rudimentary stage, it received a score of 62% indicating that there was room for significant improvement. These heuristics would also be influential when designing the UI, to ensure that it satisfies them.

7.2 User Testing

Although heuristic evaluations are good at judging an interface and uncovering problems with it based on design principles and past studies, they do not tell us anything about our target users. In order to make the UI effective for our target users, we need to understand their preferences, tastes and concerns. User testing is an interface debugging technique where real users are given a set of tasks and observed as they interact with

the system in order to perform the tasks [18]. In the case of E-Village where users have varying backgrounds and exposure to technology, it becomes important to conduct some form of user testing.

7.2.1 Testing Constraints

Our target user group consisted mainly of professors who have busy schedules, so in order to perform user testing, we had to consider 3 main constraints: (i) the tests should be conducted within a reasonable amount of time (ii) the tests should not require participants to perform any preparation on their side e.g. installing additional software, etc., and (iii) the tests should be consistent for both in-person and remote tests. The traditional method of user testing involves the use of a usability lab where test participants are audiotaped and videotaped as they perform tasks on a system [17]. This typically involves an administrator to run the test with the user, and a group of usability professionals who observe from behind a one-way mirror. This approach yields the best results but involves significant time investments, high costs and setting up infrastructure. Effective remote usability testing techniques mainly involve the use of software that allows shared screen capabilities [17]. This means that the participant's screen and cursor can be viewed by the test administrator. However, this involves the installation and setting up of additional software. Finally, as E-Village is currently rudimentary, we needed to determine ways of evaluating both the existing UI, and potential features that would be implemented during the Pilot stage.

7.2.2 Test Design

We had to use a methodology for testing that was feasible within the constraints outlined, but allowed us to test features that were not implemented yet. Hence, we decided to employ a synchronous technique [16] by using screenshots of our existing UI, and creating mock-ups of potential features that could be expected on E-Village. Our in-person testing method was inspired from the low-cost version of traditional user testing described by Krug [12]. To address testing users remotely, we decided to synthesize the mockups in a portable format that could be easily transmitted to the participant. We could then call the participant through low-cost calling services, such as Skype, and walk him/her through the mockups. In each testing method, we would use "thinking-aloud" as a way of testing our mockups [36]. Here, study participants would be asked to speak continuously about their perceptions of the mockups. Also, follow up questions could be asked on the mockups to get an understanding of user values.

At this stage, it is important to get both high-level and detail-oriented feedback from users. We wanted to get both open-ended feedback and answers to specific questions from the users. In both cases, it was important that users were honest and open in their feedback. So, we decided to purposely give the mockups an "unfinished" look and make users understand that the UI and features of E-Village are not finalized. These mockups would also include (edited) features found in commonly used websites. In this way, they would feel like they are a part of the design process and be comfortable in giving honest opinions on things. We created the mockups using Adobe Illustrator, and stored them in a PDF file. Each sheet would have a single mockup. In this way, users could be asked questions relevant to that particular mockup preventing the chance of jumping ahead and looking at other mockups.

7.2.3 Designing Test Cases

We set the test limit to 30-45 minutes, and hence it was important for us to scope our test areas down. To this extent, we selected the following areas of the UX for testing:

- *Overall Navigation*: This decides how easy it is for a user to find something on the website. It includes the four aspects, (i) navigation and menus, (ii) category names, (iii) links, and (iv) information architecture i.e. how information is organized.
- *Collecting User Information*: The kind of information collected from users will become critical for the success of E-Village. Ask for sensitive information, and users will become skeptical about the site. But with the right amount of information, a sense of community can be fostered.
- *Registration*: This determines how easy it is for users to sign up.
- *Login*: This determines how the user would log into the site, and the mechanisms that would need to be in place if users forgot their passwords.
- *Search*: This is a prominent part of the UX on any site.

We then created electronic mockups of features based on structured essential use cases [5]. These types of use cases are the most robust in the face of changing technologies because they "model tasks in a form closest to the essential nature of the problem" and do not mix design solutions with the problem description [5]. A key highlight of these use cases is the clear division between user intentions and system responsibilities. We used this form of use cases to avoid any inherent biases among ourselves in designing the solution, and to look out for unexpected solutions or paths taken by users to achieve their means. Fig. 2 below shows the header mockup of the existing UI. Fig. 3 below shows the mockup of the search results page taken from Monster.com.

Fig. 2 Mockup of header

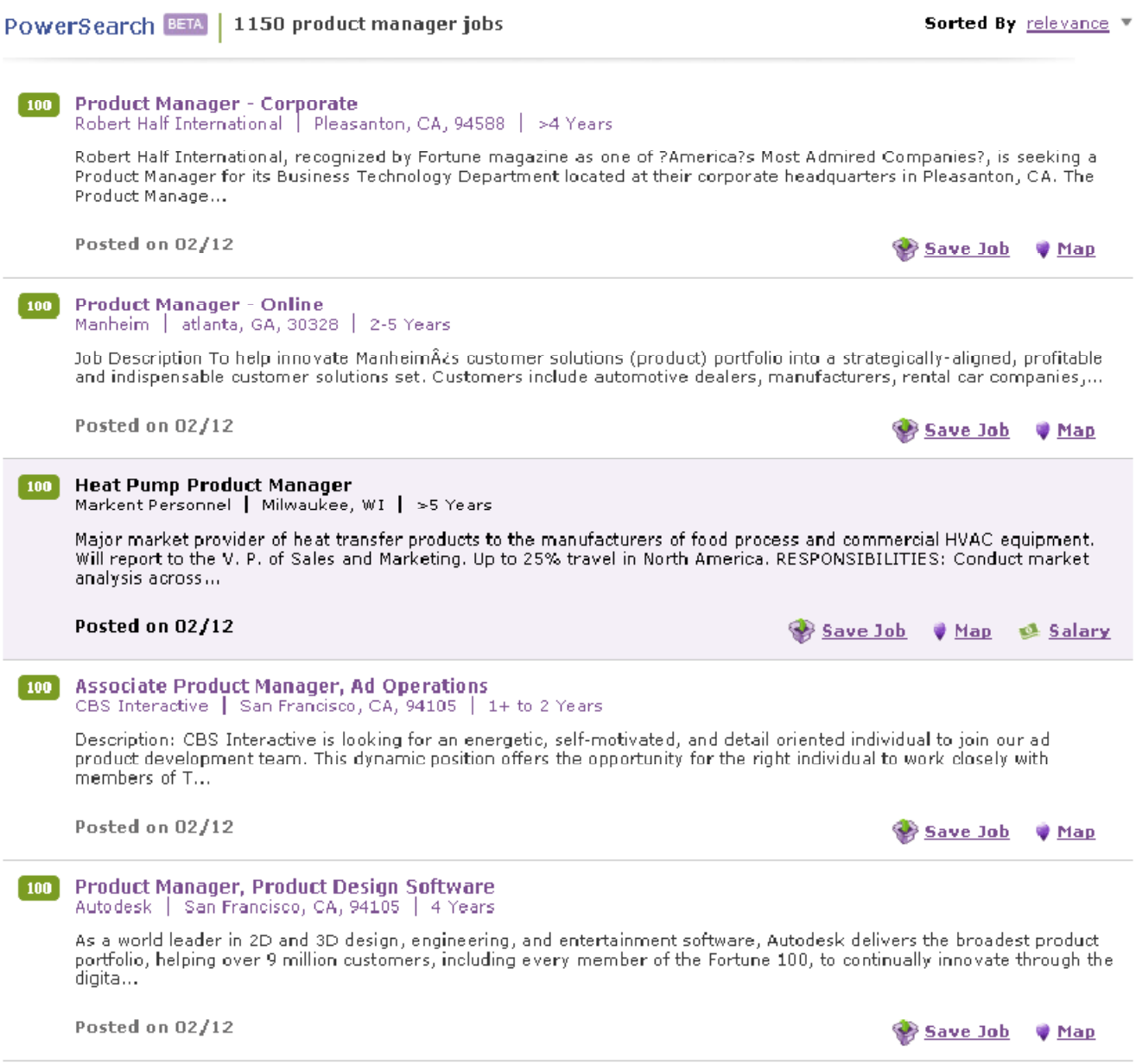


Fig. 3 Mockup of Search Results

7.2.4 Selecting Test Participants

In order to get useful feedback, we need to test with multiple users. By testing only a single user, there is the risk of being misled by the behavior of the user who may perform some actions by accident or in an unrepresentative manner. If the website has several distinct groups of users, testing should be performed with multiple users [37]. As E-Village is currently in its pre-pilot stage, we wanted to get as much feedback as possible. Hence, we contacted and setup user tests with 18 of TechBridgeWorld's contacts. These comprised a mixture of in-person and remote user tests, and were carried out in conjunction with Saurabh Sanghvi (ECE '10).

7.2.5 Test Setup

The in-person tests were conducted in a private setting in the form of a small room. The room was well-lit and a dual screen monitor setup was used. A video camera was placed to capture the monitor and to record the conversations. However, the video would only be used if the users consented to it. At the start the monitor, would be blank. The facilitator would then run over the administrative requirements, and obtain the participant's consent. The participant would then be asked a few background questions before delving into the mockups. The participant would then be shown mockups one-by-one and asked questions relevant to them. At the end, they would have the opportunity to ask any questions or offer feedback on the process.

8. User Testing Results and E-Village Enhancements

Among the three approaches to design, we are using a conservative approach i.e. we treat design as a scientific or engineering process involving methodology and control [38]. In order to ensure that we did not develop any personal biases while administrating the user tests, we did not look at any of the test data before we had collected it from all of our participants. After we collected all the data, we analyzed the data from each individual user in the framework that we had developed before testing. The following are the results obtained for the tests conducted on each mockup:

8.1 Header Mockup

For the header mockup shown in Fig __, 50% of users had negative comments, while 50% were indifferent towards the header. The fact that nobody had positive comments, and the large number of negative reactions indicates that the header will need major restructuring. Overall, users complained that the header took up too much space, the E-village title formatting was distracting, the top links were hard to read, and the placement of the Carnegie Mellon and TechBridgeWorld logos was "confusing." Hence, the following enhancements are recommended:

- The current header is taking up too much space. On a 15" laptop, it would occupy about 20% of the screen, which is a lot of valuable space. Reduce the current size of the header in terms of height.
- The navigation links are currently above the title. This causes a lot of users to miss them. Change the header so that the navigation links are below the title, closer to the body and content of the page. This would save the user that extra distance needed to be covered to click on the navigation links, as they would not need to cross the title each time.
- Users have mixed opinions about the ribbon. It does not seem to add any useful functionality or aesthetics, but adds some amount of complication in terms of the design implementation. Discontinue use of the ribbon and replace with a linear bar.
- People were irritated by the title due to the inconsistencies in coloring and formatting. The title includes both red and black colors, and E-Village is spelt with a lowercase "e." Change the title to read "Education E-Village" and use one color for the font. Instead of the current font, a standard font that is easy to read on the screen should be used e.g. Helvetica, Gill Sans, or Verdana.

- The links at the top are hard to read as they are too small and too close, and white text on red background is hard to read. Reading it is especially hard under poor resolutions. Increase the font size, lighten the background and use a screen-friendly font such as Helvetica, Gill Sans, or Verdana. Increase the width of the bar.
- The combination of logos is confusing. It is hard to infer if this is a Carnegie Mellon site or a TechBridgeWorld site. Reduce the size of the TBW logo and group the logos together so that there is no confusion.

8.1.1 Intuitiveness of Top Links

At least 83% of users found 'Home', 'About', 'Courses', 'Feedback', and 'FAQ' to be intuitive. However, 50% of users found 'Workshops' to be non-intuitive and 94% of users found 'Submit' to be non-intuitive. Hence, the following enhancements are recommended:

- Users understand that a 'Home' link brings you to the homepage. Retain the link. Also, ensure that clicking on the E-Village logo brings you to the site homepage, and that hovering over it says 'Home'.
- It is useful to have the 'About' link to help users get detailed information about the site. Retain this link. It could be relabeled to 'About E-Village' depending on how the text fits in with the design. It should include information about E-Village such as the project goals, who is working on it, and how it is funded. Each of these should be under separate headings as opposed to a paragraph form, for better readability.
- Most of the activity on E-Village is expected to be within courses. So retain the 'Courses' link.
- 50% of users found the 'Workshops' link to be non-intuitive, which raises concerns that it is unclear what the workshops are for. Either mention what the workshops mean on the homepage or place it at a secondary level of navigation. Also, if 'Workshops' is not a core feature, remove it from the top navigation and place it at a different level on the site.
- 94% of users found 'Submit' to be non-intuitive. Place the submit functionality under the courses tab and remove it from the top navigation.
- Although there were no major concerns with 'Feedback', it will likely not be used as much on the top level of navigation and does not have the same value as the other links. Remove 'Feedback' from the top navigation, place general feedback link at bottom of site, and course-specific feedback under each course homepage link.
- 100% of users found 'FAQ' to be intuitive. However, the general use of acronyms on websites is discouraged. Relabel 'FAQ' to 'Help' as it is consistent with conventions that are used on users' favorite websites. Under this link, it is essential to have most asked questions about E-Village. It should be placed at the top right of the page.
- For each link, when user is on the respective page, the link should be highlighted and unclickable e.g. on homepage, Home link should be highlighted and unclickable
- The number of top links is given by 5 ± 2. Although the number of links can be below this limit, it should never exceed this.

8.2 Sidebar Mockup

72% of users understood the sidebar model. 50% of users had negative comments on the sidebar, whereas 44% of users were indifferent to it. The low percentage of users who had a positive reaction (6%) indicates that the sidebar will need major improvements. Some of the user complaints were that the parts on the sidebar were not differentiated properly, the fonts were distracting, the search box was too small and in the "wrong" place, the bottom links were confusing, and it was unclear what could be done without logging in. The following enhancements are recommended:

- Overall
    - The fonts need to be consistent across the page. Use a screen-friendly font such as Lucida, Tahoma or Verdana.
    - Links are not justified. Users tend to read from left to right in most languages so format the links so they are aligned towards a left margin.
    - The curves at the bottom of the sidebar do not add any additional aesthetic or functional value but create complications in implementing the design. Discontinue use of the curves at the bottom and make the sidebar continuous till the bottom.
    - Consider using a different color as the font is hard to read on a poor resolution.
    - The sidebar should be used only when necessary. For example, it is useful to have the sidebar when user is viewing a course as you could have links based on the course on the sidebar. However, the sidebar is not very useful on the homepage as it takes up valuable space.
- Search Box
    - The current search box is too small. Upon typing one or two words, the user is unable to see what the first word typed was. Increase the size of the search box so that it can accommodate 25-30 characters. Inside the box, display 'Search'. This will be replaced by the query that the user types. Use a magnifying glass icon instead of a button called "Search".
    - Search box is currently in a non-intuitive place and users find it hard to locate it. Place it at top right of page preferably in the navigation bar, where it is easily accessible.
- Bottom Links
    - Bottom links are not indented correctly. They should be flush left i.e. indented with a left margin.
    - A number of users had issues understanding what the bottom links on the sidebar map to. Either discontinue use of the sidebar model completely and use top navigation or adapt the links on the sidebar to the content being viewed. The latter option is more useful as it gives the flexibility to provide context-dependent options to users i.e. when users would find them useful.
    - The presence of "Other materials" link creates a bit of confusion. Other materials should be paired with regular materials with an indication on why they are different materials e.g. a video icon next to material name if the material is in a video form.
- Login

- 'New User' meaning is unclear. The option should instead be called 'Sign Up', 'Create Account' or 'Register'.
- The Login button can be retained but the text boxes should be changed. We should have a text box with 'Email' written inside of it and another one with 'Password' written inside of it. This will make more efficient use of space. The fonts in the text boxes should have a relatively lighter color.
- Login box on sidebar seems unwelcoming as you are prompted to login to do anything. Remove the login box from the sidebar and place it on the top right side of the page. There are 2 options (i) place the login boxes on the top right corner with an option to register an account or (ii) place a link to login on the top right corner and either give a popup login box (as used by Twitter.com) or take the user to a different page. In this case, it will become important to ensure that the user does not lose any work if he/she was in the middle of a task.

8.3 Course Homepage (Layout)

17% of users had positive comments on the page, 50% were indifferent towards the page, 33% had negative comments. 89% of users were able to understand the layout quickly. For the most part, users appreciated the simple, clean layout. Key user complaints were that the text was too dense, it was unclear what this page was and where on the site the user was, the header and sidebar color combination was distracting, and the content of the page did not have a structured layout. The following enhancements are recommended:

- Users like the "fairly uncluttered" interface. In future designs ensure that the course homepage is clean and the number of distracting objects is minimized.
- Some users were unsure if they wanted to login or register account. Provide the benefits of logging when users are likely to register e.g. on a course page, users could encounter a 'favorite' feature. But being able to favorite courses requires the user to register.
- Either keep navigation vertically on sidebar or horizontally on top but not both. Having both sidebar and top navigation bar seems to cause confusion as to which one you should be following. The top navigation bar is preferable as it leaves space on the page for content. This allows for a sidebar to be shown where it could potentially be helpful e.g. on a course page.
- There is no way to find out where user is on the site. Highlight the specific tab on the navigation bar to indicate which area of the site user is on. Also, add breadcrumbs just above the body (content) heading to indicate where user is on site.
- The placement of logos is causing confusion as to who owns the site and what the role of each party is. Remove the TechBridgeWorld logo from the header, and place it in one of the information boxes on the site homepage. This could be a box introducing the user to E-Village and could feature the TechBridgeWorld logo in it. Additionally, have copies of the Carnegie Mellon and TechBridgeWorld logos at the bottom of the page in the footer. This gives us enough space to credit the agencies and not take up real estate on the content or navigation area.

- A majority of users did not like the color combinations. The site will be predominantly red to be consistent with the Carnegie Mellon and TechBridgeWorld colors. A color palette similar to those used on Cmu.edu or Cornell.edu could be used.
- Justify the links and text in the content to be flush-left i.e. aligned with a left margin. Especially in the case of the content text, this facilitates the flow of language and enhances readability due to the random edge on the right.

8.4 Registration Page Mockup

17% of users had positive comments on the page, 6% of users had negative comments, and 77% of users had indifferent comments. The fact that users did not have any major complaints about the page indicates that the page is fine for the most part. The most common irritation that users faced was the request for birthday information. Some users confused the Google reference thinking that they would need to login with their Gmail account. However, this confusion would be resolve upon explaining that some of the mockups were edited versions of features found on popular sites. The following enhancements are recommended:

- A number of users expressed irritation and concern upon seeing that the birthday was being asked. This indicates that users value their personal information, and anything that requests that kind of info without mentioning the reason behind it will lead to a drop in credibility of the site. The registration screen should only ask for information that is absolutely necessary for registration.
- Password strength checker is overkill. Instead just mention that password should be a certain number of characters.
- Some users had confusion whether each field is required. One option is to add a line saying that each field is essential but then valuable screen space is taken. Instead, do not mention that each field is required, display an error message if the user tries to proceed without filling in the required information. The error message should contain highlighted field(s) that the user did not fill in.

8.4.1 Entering Given Number of Fields

83% of users were fine with filling in the given number of fields. In the mockup, the user had to fill in 6 boxes including the Captcha. Keep the number of fields required at a maximum of 6-7.

8.4.2 Entering Captcha

72% of users were fine with entering the Captcha. These are necessary to protect against spam, and should be retained on the registration page.

8.4.3 Collecting User Information

Most users were comfortable most types of information. However, majority of users were not comfortable supplying information that could be considered personal, such as birthday or contact phone. The following enhancements are recommended:

- 89% of users were fine entering their real name. Keep this field but relabeled to “Full Name”.

- 89% were not comfortable entering their birthday. Remove this field as it does not seem to serve any kind of purpose. It adversely affects the UX with the users being skeptical of the site.
- 61% were comfortable entering this information on the site. Some users cited concerns of gender discrimination. This field should be removed as the information is not really useful for the purpose of effectively participating on the site.
- 94% of users were fine entering their profession, due to the 'professional' nature of the site. This field should be retained.
- 72% of users were comfortable supplying their employer information. This information could be optional, and users could be prompted for it at a later stage.
- Users were evenly split on supplying work address information. There seems to be no useful reasons to have this information so remove it.
- 100% of users were fine with supplying their contact information. This field should be retained.
- 72% of users were not comfortable sharing this information due to its personal nature and privacy concerns. Remove this field.
- Users were evenly split on supplying their profile picture. As having a profile picture could serve as a substitute for face-to-face communication, this field should be kept. However, it should be optional and the user should be requested for it at a later stage.
- Of the above fields, the ones that should be prompted during registration are Current Email Address (indicating that this will be user ID), Full Name, Choose a Password (indicate minimum length), Re-enter a Password, and Captcha. All of the fields should reveal what the purpose of the site is for. It is very important to keep it professional and only include things other users would find useful.

8.4.4 Reading Terms of Service (TOS)

72% of users indicated that they were unlikely to read the TOS. As the TOS is necessary for legal reasons, and to inform users how their content is managed, it should be retained. However, it should be concise and made scannable by breaking the text up into headings with paragraphs or bullets. Also, below the TOS box, there should be a statement such as “By clicking on Accept, I am agreeing to …”. This saves the user from having to perform an additional click.

8.4.5 Building Profile

55% of users were highly likely to build their profile and 22% would build it later on once they are comfortable with the site. Profile is an important step in bringing users back to the site, and establishing trust in the community. Hence, users should be able to enter certain pieces of professional information about themselves that could be useful for other professionals in the field. This includes information that was deemed as optional in section 8.4.3. 83% of users wanted to fill this information later on after registration. Allow users to fill in this information later under their account settings. They could be sent an email initially asking them to fill in this information. Another option is to display a blurb after registration asking users to build their profile. The user should have the choice of filling in this information at the time.

8.5 Login Page Mockup

61% of users had positive comments on the login box and 39% of users had indifferent comments. The fact that no user had negative comments on the mockup indicated that users are familiar with it, and like clean and uncluttered design. Following are recommendations for enhancements:

- The ‘Keep me logged in’ feature does not seem very beneficial at this point, and should be removed.
- Once user is logged in, the link at the top right should change to 'Account' to indicate that user is currently logged in.
- 33% of users preferred logging in on homepage, 33% preferred logging in on separate page, and 33% have no preference. In order to avoid having the login box occupy space at the top right, have a link to login and either display a popup as used by Twitter.com or take to separate page for login.
- In order to reset their password, 83% of users preferred to have a password email sent to their account. Following a series of steps on a site requires users to remember answers to "secret" questions. This is one more additional thing for the user to worry about. If a user forgets password, a password reset email should be sent to his/her email account. This should have a link that can be clicked on to reset the password.

8.6 Search Results Page Mockup

39% of users had positive comments, 50% of users had indifferent comments, and 11% of users had negative comments on the mockup. The low percentage of negative reactions indicates that there were no major issues with the layout. The following are recommendations for enhancements:

- 55% of users knew what the numbers in the green boxes meant (relevance scores), and 50% of users would take advantage of these scores. The relevance scores in the mockup were all 100, and hence did not add any significant help in figuring out the most relevant links. This is a case that could happen in which case no useful information is added by the relevance scores. Also, we will have options that the data is sorted by (relevance, date) so the user can see that and infer how the data is being sorted. Additionally, users who do not catch the relevance scores right away will have to think for a split-second about what the number means. It would be easier to just parse the blurb of text in the search result or look at the tagged attributes. Discontinue use of relevance scores.
- 83% of users indicated interest in marking courses that they could look up later on. Allow users to mark courses from both search results and course homepages so that they can view them on their account/profile page later on. The user should be able to view what courses he/she is teaching and what courses he/she is "following".
- Sorting options should include relevance (by default) and date posted.
- Subtly highlight alternate results block.
- Each search result will be displayed in a row form, as users have a mental model for it. Within each row, there could be 2 columns. One for displaying title and short blurb, and the other one for other useful information such as date posted, level of course, etc. Matching terms from the query should be bolded in the blurb text.

8.7 Search Filter Box Mockup

28% of users had positive comments, 38% of users had indifferent comments, and 33% of users had negative comments. Some of the major concerns raised included the categories being ambiguous and the possibility of being perceived as different things. Following are recommendations for enhancements:

- 'Region' is ambiguous. Change it to something like localization or language
- During the test, users seemed to look at the categories and think about them as if they were looking at a course hierarchy under the courses link. Hence, their assessment of the filter might have been inaccurate. We would need to do some user testing once E-Village has been launched to understand user behavior once they get search results, and if they use advanced search filters.
- Additionally, course categories could be perceived differently based on the background of the user, so some titles could become ambiguous e.g. is physics a Science or Engineering course? Since we are starting off with targeting technical courses, we could use categories that reflect different areas of technology e.g. computer programming, databases, software design, machine learning, etc.
- A number of users commented that there should be a string based search, and wanted to filter by level and type of content. This could be attributed to the fact that the search results mockup before this did not show the textbox where the search query was typed. If they had seen it, perhaps they would not have mentioned the string search as you could refine your search by modifying what you had typed in. Ensure that the search results page shows the query user had typed in within the box, and allow the user to modify the query and search again.
- One concern raised was that by 'filtering' results, we might be hindering multidisciplinary discovery of courses. This is something the E-Village team will need to think about (if they want to allow filtering into such categories). At the start, we think this filter not be necessary as the number of courses will be limited and such a filter only becomes useful when you have too many results to contend with. The other case is that users want to search within certain kinds of courses (by level, etc.). In this case, it might be more useful to provide a search option under the courses page. Even otherwise, if 20 results are displayed on a page, we would not need this functionality as most results would come in a page or two (for the near future). Hence, we would worry about this only once we have a number of courses that typing in "Intro" would return 3 or more pages worth of results.
- 77% of users preferred layout 1 with the vertical filter on the left. This is mainly due to the familiarity and visual appeal. The other reason is that in the horizontal layout, the filter is at the top of the page. So it takes some space even when users may not even use it, and if a user is dissatisfied with the returned results, he has to go back all the way up to the top of the page.

8.8. Course Homepage (Information Architecture)

77% of users wanted to use the search box, and 61% of users wanted to use the courses link to navigate down to the potential courses. The following are recommendations for enhancements based on our findings:

- Most users have a mental model of expecting a list of courses. The following information should be present under the 'Courses' homepage.
    - Short intro to courses (2-3 lines)
    - Featured course
    - Default categorization of courses
    - Additional ways to view courses by tags e.g. by level, localization
- Although it is tempting to throw in the kitchen sink in terms of categories, we should realize that the number of choices for a user is l1 * l2 * ... * ln, where li is the number of choices at level i, and n is the total number of levels. Once we have a certain number of course on E-Village, it will be useful to run usability tests to find out how effective the course navigation is.
- 33% of users wanted to search by university, 72% wanted to search by topic/title/area, 17% wanted to search by region, 33% wanted to search by professor, 17% wanted to search by keyword. Lucene should be configured so that title, author and content get higher proportions of relevance rankings. Of these title should get the highest ranking.
- In order to look for a person, 27% of users wanted to use the authors link although it was pretty unclear what they would expect. 61% of users wanted to use search where they would type in
    - Name of professor
    - University name
    - Courses taught
- 11% of users did not anticipate why they would be looking for a person
- Currently, we feel that looking for a professor might not be an activity that users would be doing regularly. Once materials are being uploaded, it is likely users will be looking for courses and dealing with other course-related things. As mentioned in Paradox of Choice, providing users with lots of choices reduces the quality of the overall experience. More choice is not necessarily better. It may be taking time that could be devoted to other matters.

10. Conclusions and Future Work

In general, the approach we used to test users was effective. By splitting up the mockups into individual elements, we were able to gain detailed feedback on their usability, and ask open-ended questions to get a sense of high-level user preferences. Once the design has been implemented and content been published onto the site, usability tests will need to be run again (i) to confirm that the changes in fact fixed the problem, (ii) to ensure that new problems have not been introduced into the design, and (iii) to explore more deeply the usability of the structure of the site evaluating issues like information architecture and task flow [37]. According to Nielsen, the number of usability problems found through a usability test with *n* users is $N(1-(1-L)^n)$, where N is the total number of usability problems in the design, and L is the proportion of usability problems found while testing a single user. The typical value of L is 31%. Hence, testing with 15 users will uncover all the problems. Since testing with 5 users will uncover 85% of all problems, it is better to have 3 tests with 5 users each than to have one test with 15 users [37].

11. Acknowledgments

There is a group of people that I would like to thank, and without whom this work would not have been possible: Ameer Abdulsalam for being my mentor during this project and providing guidance towards resources; Alexander Cheek for providing invaluable feedback on the UI design sketches; Yonina Cooper for helping understand technical intricacies of search algorithms; Frederick Dias for helping setup the E-Village server and assisting with any issues regarding our Content Management System (CMS); Saurabh Sanghvi for helping conduct the usability tests and exchanging feedback on our respective work on E-Village; and Ermine Teves for helping gain IRB approval to conduct usability tests, and helping contact potential target users for them. I would like to thank Yahoo! for their support of this project, and all the users who participated in the usability studies.

12. References

[1] M. B. Dias and E. Brewer, "How Computer Science Serves the Developing World," *Communications of the ACM*, vol. 52, no. 6, June 2009.

[2] E. Brewer et al., "The Case for Technology in Developing Regions," *IEEE Computer*, 2005.

[3] MIT OpenCourseWare. [Online]. http://ocw.mit.edu/

[4] Open.Michigan. [Online]. https://open.umich.edu/

[5] L. L. Constantine and L. A. D. Lockwood, "Structure and Style in Use Cases for User Interface Design," in *Object Modeling and User Interface Design.*, 2001.

[6] M. B. Dias, B. Browning, G. A. Mills-Tettey, N. Amanquah, and N. El-Moughny, "Undergraduate Robotics Education in Technologically Underserved Communities," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007.

[7] Nielsen Norman Group: Strategies to enhance the user experience. [Online]. http://www.nngroup.com/about/userexperience.html

[8] K. A. Renninger and W. Shumar, *Building Virtual Communities: Learning and Change in Cyberspace.*: Cambridge University Press, 2002.

[9] L. Neal, "Virtual classrooms and communities," in *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*, 1997.

[10] almost All QUestions Answered - aAQUA. [Online]. http://aaqua.persistent.co.in/aaqua/forum/index

[11] A. N. Donald, *The Design Of Everyday Things.*: Basic Books, 2002.

[12] S. Krug, *Don't Make Me Think: A Common Sense Approach to Web Usability*, 2nd ed.: New Riders Press, 2005.

[13] J. Nielsen and M. Tahir, *Homepage Usability: 50 Websites Deconstructed.*: New Riders, 2001.

[14] J. Nielsen and H. Loranger, *Prioritizing Web Usability.*: New Riders, 2006.

[15] J. Nielsen. (2005) Ten Usability Heuristics. useit.com: Jakob Nielsen's Website. [Online]. http://www.useit.com/papers/heuristic/heuristic_list.html

[16] S. Dray and D. Siegel, "Remote Possibilities?: International Usability Testing at a Distance," *Interactions*, 2004.

[17] K. E. Thompson, E. P. Rozanski, and A. R. Haake, "Here, There, Anywhere: Remote Usability Testing That Works," in *Proceedings of the 5th conference on Information technology education*, 2004.

[18] J. Nielsen and T. K. Landauer, "A Mathematical Model of the Finding of Usability Problems," in *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, 1993.

[19] OpenCms, the Open Source Content Management System / CMS. [Online]. http://www.opencms.org/en/

[20] Drupal. [Online]. http://drupal.org/

[21] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," Computer Science Department, Stanford University, Stanford, USA, 1998.

[22] P. Morville and L. Rosenfeld, *Information Architecture for the World Wide Web*.: O'Reilly, 2007, ch. 8.

[23] Apache Lucene. [Online]. http://lucene.apache.org/java/docs/

[24] Sphinx. [Online]. http://sphinxsearch.com/

[25] The Xapian Project. [Online]. http://xapian.org/

[26] Google custom search. [Online]. http://www.google.com/cse/

[27] FreeFind. [Online]. http://www.freefind.com/

[28] PicoSearch. [Online]. http://www.picosearch.com/

[29] A. R. D. Prasad and D. Patel, "Lucene Search Engine - An Overview," in *DRTC-HP International Workshop on Building Libraries using DSpace*, 2005.

[30] D. P. Zhou. (2006, June) Delve inside the Lucene indexing mechanism. [Online]. http://www.ibm.com/developerworks/library/wa-lucene/

[31] D. Cutting, Lucene Lecture at Pisa, 2004.

[32] D. Cutting and J. Pedersen, "Optimization for dynamic inverted index maintenance," in *Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrieval*, 1989.

[33] B. Goetz. (2000) The Lucene search engine: Powerful, flexible, and free. [Online]. http://www.javaworld.com/jw-09-2000/jw-0915-lucene.html

[34] D. R. Cutting and J. O. Pedersen, "Space Optimizations for Total Ranking," in *Proceedings of RIAO*, Montreal, Quebec, 1997.

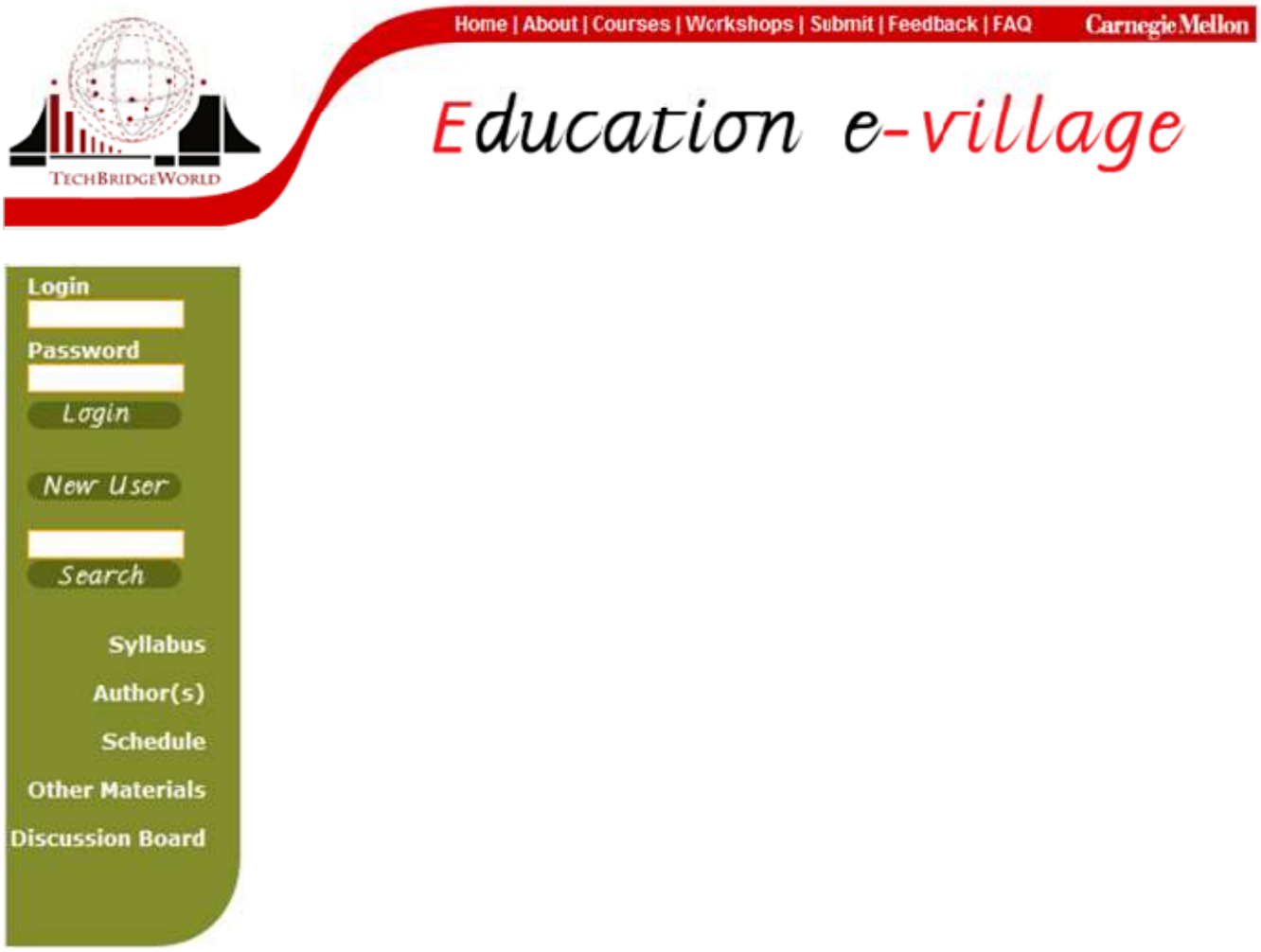
[35] J. Nielsen. Heuristic Evaluation. useit.com: Jakon Nielsen's Website. [Online]. http://www.useit.com/papers/heuristic/

[36] M. Hammontree, P. Weiler, and N. Nayak, "Remote Usability Testing," *Interactions*, 1994.

[37] J. Nielsen. (2000) Why You Only Need to Test with 5 users. useit.com: Jakon Nielsen's Website. [Online]. http://www.useit.com/alertbox/20000319.html

[38] D. Fallman, "Design-oriented Human--Computer Interaction," Department of Informatics and Umeå Institute of Design, Umeå University, Sweden, 2003.

Appendix A: Usability Test Mockups



TechBridgeWorld

Education e-village

Login

Password

Login

New User

Search

Introduction to Robotics

Syllabus

Robotics is a field that brings together computer science and engineering, resulting in systems that interact intelligently with their environment. With applications ranging from agriculture to factory automation, from healthcare to education, robotics is a fascinating and fun way to develop creativity as well as the design, implementation, and integration skills that are essential for computer scientists and technologists.

The main goal of the course is to challenge students to think creatively and to teach them to integrate diverse areas of knowledge such as Computer Science, Design, Electrical Engineering, Mechanical Engineering and Math to create innovative systems. In this course, students will work individually and in groups to build robots using Lego robot building kits and to program them using microcontrollers. Through these projects, they will learn how to write programs that control a physical device. They will learn to read and understand research papers, to give presentations to technical and non-technical audiences, and follow a project through from an initial idea through design to implementation.

Create an Account

If you already have a Google Account, you can sign in here.

Required information for Google account

Your current email address:

e.g. myname@example.com. This will be used to sign-in to your account.

Choose a password: Password strength:

Minimum of 8 characters in length.

Re-enter password:

Location: United States

Birthday:

MM/DD/YYYY (e.g. "2/2010/2010")

Word Verification:

Terms of Service:

Please check the Google Account information you've entered above (feel free to change anything you like), and review the Terms of Service below.

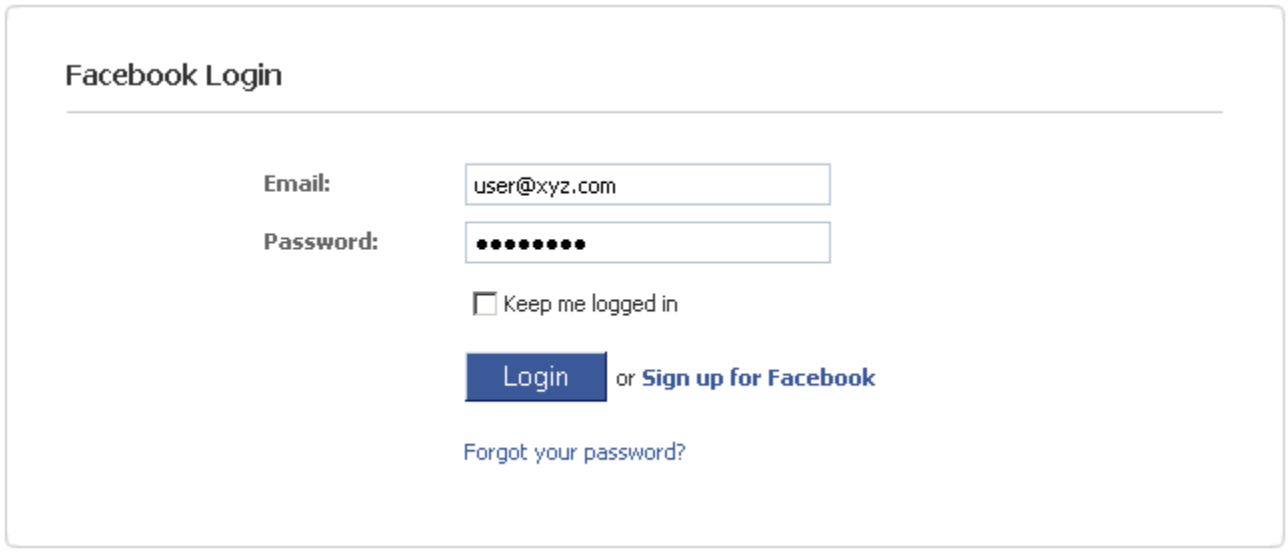
Printable Version

Google Terms of Service

1.1 Your use of Google's products, software, services and web sites (referred to collectively as the "Services" in this document and excluding any services provided to you by Google under a separate written

By clicking on 'I accept' below you are agreeing to the Terms of Service above and the Privacy Policy.

I accept. Create my account.


Facebook Login
Email:
user@xyz.com
Password:
Keep me logged in
Login
or Sign up for Facebook
Forgot your password?

PowerSearch BETA | 1150 product manager jobs

Sorted By relevance

100 **Product Manager - Corporate**
Robert Half International | Pleasanton, CA, 94588 | >4 Years

Robert Half International, recognized by Fortune magazine as one of ?America?s Most Admired Companies?, is seeking a Product Manager for its Business Technology Department located at their corporate headquarters in Pleasanton, CA. The Product Manage...

Posted on 02/12

Save Job Map

100 **Product Manager - Online**
Manheim | atlanta, GA, 30328 | 2-5 Years

Job Description To help innovate ManheimÂ¿s customer solutions (product) portfolio into a strategically-aligned, profitable and indispensable customer solutions set. Customers include automotive dealers, manufacturers, rental car companies,...

Posted on 02/12

Save Job Map

100 **Heat Pump Product Manager**
Markent Personnel | Milwaukee, WI | >5 Years

Major market provider of heat transfer products to the manufacturers of food process and commercial HVAC equipment. Will report to the V. P. of Sales and Marketing. Up to 25% travel in North America. RESPONSIBILITIES: Conduct market analysis across...

Posted on 02/12

Save Job Map Salary

100 **Associate Product Manager, Ad Operations**
CBS Interactive | San Francisco, CA, 94105 | 1+ to 2 Years

Description: CBS Interactive is looking for an energetic, self-motivated, and detail oriented individual to join our ad product development team. This dynamic position offers the opportunity for the right individual to work closely with members of T...

Posted on 02/12

Save Job Map

100 **Product Manager, Product Design Software**
Autodesk | San Francisco, CA, 94105 | 4 Years

As a world leader in 2D and 3D design, engineering, and entertainment software, Autodesk delivers the broadest product portfolio, helping over 9 million customers, including every member of the Fortune 100, to continually innovate through the digita...

Posted on 02/12

Save Job Map


Refine Your Search
Courses
All
Engineering
Science
Design
Literature
Humanities
Business Admin...
Other
go
Region
All
North America
South America
Africa
Asia
Europe
go

PowerSearch BETA | 1150 product manager jobs

Sorted By relevance

Refine Your Search

- Courses
  - All
  - Engineering
  - Science
  - Design
  - Literature
  - Humanities
  - Business Admin...
  - Other

go

- Region
  - All
  - North America
  - South America
  - Africa
  - Asia
  - Europe

go

100 **Product Manager - Corporate**
Robert Half International | Pleasanton, CA, 94588 | >4 Years

Robert Half International, recognized by Fortune magazine as one of ?America?s Most Admired Companies?, is seeking a Product Manager for its Business Technology Department located at their corporate headquarters in Pleasanton, CA. The Product Manage...

Posted on 02/12

Save Job Map

100 **Product Manager - Online**
Manheim | atlanta, GA, 30328 | 2-5 Years

Job Description To help innovate ManheimÂ¿s customer solutions (product) portfolio into a strategically-aligned, profitable and indispensable customer solutions set. Customers include automotive dealers, manufacturers, rental car companies,...

Posted on 02/12

Save Job Map

100 **Heat Pump Product Manager**
Markent Personnel | Milwaukee, WI | >5 Years

Major market provider of heat transfer products to the manufacturers of food process and commercial HVAC equipment. Will report to the V. P. of Sales and Marketing. Up to 25% travel in North America. RESPONSIBILITIES: Conduct market analysis across...

Posted on 02/12

Save Job Map Salary

100 **Associate Product Manager, Ad Operations**
CBS Interactive | San Francisco, CA, 94105 | 1+ to 2 Years

Description: CBS Interactive is looking for an energetic, self-motivated, and detail oriented individual to join our ad product development team. This dynamic position offers the opportunity for the right individual to work closely with members of T...

Posted on 02/12

Save Job Map

100 **Product Manager, Product Design Software**
Autodesk | San Francisco, CA, 94105 | 4 Years

As a world leader in 2D and 3D design, engineering, and entertainment software, Autodesk delivers the broadest product portfolio, helping over 9 million customers, including every member of the Fortune 100, to continually innovate through the digita...

Posted on 02/12

Save Job Map

PowerSearch BETA | 1150 product manager jobs

Sorted By relevance

Refine Your Search

Courses

☐ All ☐ Engineering ☐ Science ☐ Design ☐ Literature ☐ Humanities ☐ Business Admin... ☐ Other go

Region

☐ All ☐ North America ☐ South America ☐ Africa ☐ Asia ☐ Europe go

100 **Product Manager - Corporate**
Robert Half International | Pleasanton, CA, 94588 | >4 Years

Robert Half International, recognized by Fortune magazine as one of ?America?s Most Admired Companies?, is seeking a Product Manager for its Business Technology Department located at their corporate headquarters in Pleasanton, CA. The Product Manage...

Posted on 02/12

Save Job Map

100 **Product Manager - Online**
Manheim | atlanta, GA, 30328 | 2-5 Years

Job Description To help innovate ManheimÂ¿s customer solutions (product) portfolio into a strategically-aligned, profitable and indispensable customer solutions set. Customers include automotive dealers, manufacturers, rental car companies,...

Posted on 02/12

Save Job Map

100 **Heat Pump Product Manager**
Markent Personnel | Milwaukee, WI | >5 Years

Major market provider of heat transfer products to the manufacturers of food process and commercial HVAC equipment. Will report to the V. P. of Sales and Marketing. Up to 25% travel in North America. RESPONSIBILITIES: Conduct market analysis across...

Posted on 02/12

Save Job Map Salary

100 **Associate Product Manager, Ad Operations**
CBS Interactive | San Francisco, CA, 94105 | 1+ to 2 Years

Description: CBS Interactive is looking for an energetic, self-motivated, and detail oriented individual to join our ad product development team. This dynamic position offers the opportunity for the right individual to work closely with members of T...

Posted on 02/12

Save Job Map

Designing Mobile-Phone Based Educational Games to Improve the English Literacy Skills of Limited English Proficient (LEP) Adults

Senior Thesis Project
School of Computer Science
Carnegie Mellon Qatar Campus

Prepared by:
Aysha Siddique

Advised by:
M. Bernardine Dias, Ph.D.

Abstract

English is one of the most commonly used languages in international business, and therefore, some level of fluency in English becomes a pre-requisite for many employment opportunities. Due to their limited English proficiency and a lack of opportunity to improve their English skills, a variety of adult populations are disadvantaged in many ways including career advancement and societal acceptance. For example, low-skilled immigrant laborers in countries such as Qatar and the USA have limited English proficiency, which is often a barrier to their career advancement and creates communication problems with their supervisors. Similarly, limited English skills make it harder for refugee populations to find jobs and adjust to the local culture in their host countries. Also, the average deaf adult in the USA reaches only a 4th grade English reading level. Our work aims to address the problems of limited English proficiency among adults by providing these groups with a low-cost, easily accessible, fun tool for enhancing their English skills.

Mobile phones are the most prevalent and accessible computing technology for people of all ages and incomes. Related research efforts by several groups have demonstrated the success of mobile phone-based educational games in improving English literacy skills among primary school students. The goal of our work is to investigate the effectiveness of mobile phone-based educational games on adult English literacy. Our literacy tool consists of two parts: a single player game accessible on a mobile phone, and an online content authoring system which enables teachers to add useful educational content to the games. We incorporate proven techniques from expert teachers into these educational games, along with graphics and game concepts that motivate adults to play these games. The combined result is an effective and ubiquitous tool for enhancing English literacy skills among adults with limited English proficiency.

Acknowledgements

I can't imagine having done this project with anyone else, and I would like to acknowledge the support and motivation provided by my advisor, M. Bernardine Dias, at every point over the last one year. Her belief in me has gotten me through tough times and I have learned so many valuable things through this project. ☺. Special thanks also for lending me a working space in the Robotics lab and for all the patience through the missed deadlines and excuses.☺.

There are a multitude of people who were involved in this project that deserve more than a mention on this page. Dr. Yonina Cooper, for her valuable support; the TechBridgeWorld team (Sarah Belousov, Ermine Teves, and Frederick Dias), for all the help with IRB's and field testing (and a very fun time in Pittsburgh!); Dr. Silvia Pessoa and her students, for support, encouragement and for all the help with conducting field testing at the RAEL program in Doha; John Roberston and Robert Monroe, for support and funding for the phones; the Senior Thesis committee (Mark Stehlik, Majd Sakr, Iliano Cervesato, Kemal Oflazer) for valuable comments during the presentations; Ray Corcoran and Enrique Isidro, for helping with the field testing with the service attendants in Doha; Hatem Alismail and Frederick Dias, for the technical help and support with this project (and bearing with my emails at all inappropriate times ☺ , sorry about that!).

Table of Contents

1. INTRODUCTION

Globalization has propelled the need for a common language that can be used for communicating across international boundaries. English has become one of the most important languages used in international business, technology, and everyday life. Therefore, some level of fluency in English becomes a pre-requisite for many employment opportunities around the world. In today's globalized world, the ability to communicate in English can outweigh experience and other skills in priority when considered for career advancement and employment opportunities, and can allow a person to acquire a variety of skills where the medium of instruction is often English (for example, computer and software instruction), thereby further enhancing opportunities for career advancement. A reasonable grasp of the English language therefore allows an individual better career opportunities and standard of living in many countries, and was demonstrated through the positive correlation between earnings and English language ability found in the US Census of Data of 2000 [**1**].

Limited English language skills, which are due to reasons ranging from lack of opportunities and resources in education to lack of access to qualified teachers and also difficulties of the nuances of the language, can disadvantage individuals in many populations. This senior thesis project seeks to improve this situation for low-skilled immigrant laborers and deaf youth; two very different populations that face constraints in enhancing their English literacy skills.

Low-skilled immigrant laborers are people who migrate looking for job opportunities to countries like Qatar and the United States of America, which are countries with strong economies and job opportunities that exceed their human resources. The immigrant laborers' population is typically characterized by a limited educational background, low wages and low-skilled jobs in the construction, household, and service sectors. Their limited English skills cause them to face barriers in career advancement, perform poorly when communicating with supervisors and other authorities, and feel helpless in matters of negotiating their pay, vacation, and accommodation. These problems lower their standard of living and make it hard for them to assimilate and adjust into the foreign culture. The lack of societal acceptance or assimilation of the immigrant laborers into the host country's society is a serious concern as they make up a significant proportion of the country's population, 79.5% of Qatar's population [**2**] and 12.5% of the US population [**3**].

This senior thesis project explores the use of ubiquitous computing technology to provide the immigrant labor population with means to improve their English skills in a very affordable, engaging, and practical manner. Today, mobile phones contain the most accessible and ubiquitous computers. Almost everyone, in both developed and developing communities, has frequent access to a mobile phone. People use their mobile phones for many tasks beyond communication, including storing and accessing videos, music, and pictures, and playing games. Mobile phone games are popular among adults and children alike. This senior thesis will therefore explore the use of mobile phone-based games to improve the English literacy skills of immigrant laborers and other adults with limited

English proficiency. The approach taken will be to first understand how English is taught to non-native English speakers and/or to deaf populations, study what technology solutions already exist to help improve English literacy, and finally to incorporate both effective teaching techniques and lessons learned from previous projects into the implementation of an effective solution. The next section will describe the challenges faced by the above mentioned groups in learning English and some of the effective techniques used in English as a Second Language (ESL) instruction, the use of educational games to improve English literacy skills in children, and a brief literature review on the various technologies that try to help improve English literacy.

2. Literature Review

To make an effective tool for teaching English, we need to understand what are some of the challenges faced by adults in learning English as a second language and what are the most effective techniques used by professional ESL instructors.

2.1 English as a Second Language Instruction

There are four components of reading: Vocabulary, Alphabetic and Word Analysis, Fluency and Comprehension [4]. Some of the common techniques in teaching each of the components in reading will not work for adults who are not native English speakers. For example, some of common techniques used to teach vocabulary are teaching words in semantic sets and understanding vocabulary through context [4]. For non-native English speakers, this is difficult because when you present semantic set of words or similar set of words (for eg, days of the week or colors), the adult learner gets confused. Also, using context to understand vocabulary requires the adult learner to know about 98 % of the words in English which is not the case [4]. The adult learners could probably guess the meaning from context, but wouldn't gain any knowledge or understanding of the new vocabulary. Similarly, some of the common techniques used for teaching alphabetic and word analysis, which is the process of using letters to present meaningful spoken words, are to assess beginning reader's knowledge through pronunciation and to assess letter-sound knowledge [4]. The problem this presents for non-native English speakers is that they don't already have a vocabulary base in English and therefore, strategies relying on oral comprehension will not work [4]. For Fluency, one of the common techniques is to involve teachers to do repeated reading of texts. This presents problems for adult learners in that the native language might interfere in terms of stress, pause and intonation. Finally for comprehension, the common techniques are to give students cloze passage exercises and to require them to summarize short paragraphs. With non-native English speakers, there could be cultural differences in text that would make it hard for them to understand and summarize the text [4].

Keeping the difficulties faced by the non-native English speakers in mind, below are some suggestions and techniques, taken from the article "How should Adult ESL Reading

Instruction differ from ABE Reading Instruction?" by Miriam Burt [**4**], that could be used to make English learning easier for adults.

- Pre-teach vocabulary in a passage
- Avoid presenting synonyms or antonyms or words in the same semantic set
- Provide multiple exposure to specific words in multiple contexts
- Use of bilingual dictionaries, word cards, and regular tests
- Teach English letter-sound correspondences
- Identify parts of speech and their roles
- Make students listen to native speaker model of the reading to improve fluency in reading
- Build on learners culture and experiences whenever possible
- Pre-teach vocabulary and preview unfamiliar ideas and actions etc
- Use visual aids and physical objects in instruction
- Assess learner comprehension through short questions, summary writing after pre-teaching vocabulary, previewing cultural contexts and discussing the text

2.2 Educational Games for Literacy

In order to make ESL instruction more effective and engaging, teachers tend to use classroom activities and games. 'Games are highly motivating since they are amusing and at the same time challenging. Furthermore, they employ meaningful and useful language in real contexts' [**5**]. Games can be to used to introduce new concepts or to revise concepts learned in class, and should typically be used at all stages of the class. Games allow students to relax and have fun, therefore, helping them learn and retain words more easily [**6**]. Thus, games are really effective in motivating students to learn and decreasing their fear of a foreign language.

2.3 Use of Mobile Phones for Promoting Literacy

With the prevalence of technology in every sphere of life, it is only natural that technology will be used to try and improve literacy and education. In the recent past, a lot of projects have utilized mobile phones to promote literacy skills. Some examples are described below:

2.3.1 Mobile Immersive Learning for Literacy in Emerging Economies (MILLEE) [7]

FIGURE 1: STUDENT USING MILLEE GAMES

MILLE E is a project initiated at the University of California, Berkeley to promote English literacy among primary school students in rural India. Students in villages in rural India and slums cannot afford schooling and are not motivated to learn. Mobile phone games present a very engaging and motivating platform for primary school students to improve their English skills. The games focus on simple English language skills such as vocabulary, phonetics, sentence composition and spelling. Their field study in India showed that game play can produce significant learning benefits, and results for a set of 25 students show that the scores improved from 1.97/5 to 3.85/5 after playing the games for four months [**8**].

2.3.2 SMS-based literacy initiative

Pakistan has a high percentage of illiterate population and as a measure to increase the literacy levels, a pilot project that made use of mobile phone was started in 2009 where learners would receive informative text messages daily in Urdu. The learners will be evaluated every month to assess gain in knowledge and understanding. According to results, at the beginning of the program only 28% of the students managed an A and this number increased to a 60% at the end of the pilot program. This is an interesting use of SMS application and technology to promote literacy.

Literature review shows that mobile phones are an most ubiquitous technology that can be used in interesting ways to promote English literacy skills. In addition, educational games in class promote learning and keep the students motivated. The thesis looks to combine mobile phones and educational games and use the techniques provided by ESL instructors to come up with a viable solution to address the limited English literacy skills of immigrant laborers.

3. Thesis Goals

Most of the existing technology solutions that support English learning are targeted at primary or secondary school students in modern societies that have access to computers and resources that will support these technologies. For user groups like immigrant laborers, there are several constraints that make it hard to implement the same solutions. The immigrant labor group is characterized by long working hours, low wages and strict working conditions, which indicates that they do not have the time or resources to access many English classes. They also mostly do not have frequent access to high end technology like computers or smart phones or services like the Internet. Some workplaces have tried to address these issues by providing on the job training or ESL classes while others have tried to provide technology on the job (for example, Kelsa+ at Microsoft Research India [9]) where the laborers can access these technologies in their free time and seek to improve their skills. However, that is not the case in most places, especially in countries like Qatar and the USA. Therefore, there is a need to design a technology solution that is easily accessible and cost effective for these targeted users.

Moreover, many of the existing technology solutions for enhancing English literacy have user interfaces that are targeted towards primary and secondary school students. The same graphics and motivators will not be appropriate for the adult user groups, because of the difference in age, interests and cultural backgrounds. Hence, these user interface and graphics must be modified to better appeal to adults and be more relevant to their needs to encourage usage. Determining the various factors that motivate adults to use a tool to learn English in their own time will require some research as well.

The goal of this senior thesis is, therefore, **to design a low-cost and easily accessible educational and engaging tool that will enable guided practice of literacy skills for low-resourced adult users with limited English proficiency**.

Our approach to achieving the thesis goal is **to implement a mobile phone based educational game that is designed to improve the English literacy skills of the targeted adult user groups.**

The motivation for the "mobile phone" aspect of the solution comes from the fact that almost everyone, in developed and developing communities, owns a mobile phone. Various educational and income generation projects based on mobile phones, such as MILLEE [**7**] , Grameen Phone [**10**] and aAqua [**11**] , have also been successful in the past. Several of these projects have also been implemented using lower end phones and in societies with limited computing resources.

The "educational game" aspect is inspired by the fact that games are a fun way to practice English exercises, and educational games are employed by teachers in classrooms to motivate students. In addition, according to research conducted by the NPD Wireless Industry Market Research group in 2006, 29% percent of the mobile games were downloaded and played by adults aged between 25 and 34 [**12**]. Thus, there are indicators that show adults enjoy playing mobile phone games and our work seeks to leverage this fact to motivate adults to increase their practice time on guided exercises for improving English literacy.

All of the previous mobile phones for literacy projects have been aimed at primary school students, and therefore, modifications will be required to both the content and graphics of the games. Content should be presented at a level that is best suited for adult learners, and yet simple and effective at the same time so that it is accessible via a mobile phone. Also, the motivators for the game should appeal to an adult user.

The following sections will elaborate on the technical approach, user groups identified to participate with, needs assessment, implementation of the game and field testing results.

4. Needs Assessment

In order to make an effective tool that helps adult user groups learn English, it is important to understand their cultural and educational backgrounds, and customize the tool to meet their needs and interests. Needs assessment is a critical phase that will impact the technology development, and the researchers will need to identify and interview user groups to understand how this tool can be customized for their successful learning.

For this project, we identified several user groups that will benefit from this thesis work. As mentioned in the introduction, immigrant laborers who have limited English skills face barriers in career advancement and have problems in communicating with their supervisors.

Learning English will be beneficial for this group in order to advance their career ladders and to better seek and qualify for employment opportunities. In addition to immigrant laborers, we discovered that the project will also be beneficial for deaf youth, who have trouble grasping the English language due to the stark structural differences between English and American Sign Language. The average deaf adult in the USA reaches only a 4th grade English reading level, and only 10% (approximate) of 18 year olds can read at or above an 8th grade level [**13**]. Deaf individuals usually have "severely limited vocabulary and lack knowledge of the complex syntax of English that is critical for combining sentences together into cohesive text" [**14**] . Noun verbs, articles, noun-count nouns and verb tenses are some areas of English where deaf individuals have syntactical trouble. Their difficulties with communicating in English also add an extra layer of complexity for deaf employees at work places which are shared with both deaf and non-deaf individuals, in addition to limiting their opportunities for career advancement.

For our needs assessment phase we contacted several organizations that work with immigrant laborers and deaf youth. These groups are introduced next.

4.1 User Groups

The Literacy Tools Project titled "Mobile Phone Games to Improve the English Literacy Skills of Limited English Proficient (LEP) Adults" has received IRB approval with the IRB certificate number HS09-588. The groups ROTA Adult English Literacy Program, Service Attendants at CMU-Q, Western Pennsylvania School for the Deaf and Catholic Charities all belong under the same IRB Certificate.

Reach Out To Asia (ROTA) Adult English Literacy Program

FIGURE 2: ROTA LOGO

Reach Out To Asia (ROTA) [**15**], a non-governmental charity organization in Qatar, started an Adult English Literacy program where they teach English to immigrant laborers working with construction companies. In its second iteration, ROTA partnered with the Al Jaidah group to teach basic and intermediate English skills to some laborers. The laborers volunteer to join the classes and are awarded with certification at the end of the 8 week program. The interests of the user group are perfectly aligned with the goals of this thesis research, since they want to learn English and have taken initiative by enrolling in a structured classroom. Dr. Silvia Pessoa, an English Professor at Carnegie Mellon University in Qatar prepared the curriculum, pre and post tests for the basic and intermediate classes at the RAEL program. Dr. Pessoa has been extremely supportive of the research project and has agreed to allow the literacy tools project to be based on the curriculum she designed for the RAEL program. In addition, Dr. Pessoa is also teaching a class at Carnegie Mellon titled "Community Service Learning" where she teaches her students effective techniques to teach English to the migrant laborers. Her students, in turn, teach the laborers at the RAEL program.

Service Attendants at Carnegie Mellon Qatar Campus

The service attendants at the Carnegie Mellon Qatar campus are another user group that will benefit from this thesis work. They have to communicate with professors and students from different cultures where the common medium of communication is English, and therefore, learning English skills will be highly beneficial for this group. This user group was contacted and the project was explained to them with a request for voluntary participation in the project. The 100% positive response indicated high interest and willingness to learn English.

FIGURE 3: CMQ LOGO

This user group will not have a structured class environment where they are taught English concepts but they will be asked to play the games and will be tested for improvement in scores. This gives us the opportunity to test if the game itself has caused improvements in the English skills, as the learning happens only while playing the game and not in any class. We used the RAEL program content and tests for this group as well, with permission from Dr. Pessoa.

7th and 8th grade students at the Western Pennsylvania School for the Deaf (WPSD)

FIGURE 4: WPSD LOGO

The middle school students at the Western Pennsylvania School for the Deaf (WPSD) are another user group selected for this thesis project. Considering the difficulties faced by deaf individuals in learning English grammar concepts, this user group can potentially benefit from additional practice in English exercises through the mobile phone game. Since the individuals in this user group are teenagers, the game concept will be effective in motivating the students to participate in the project. The teacher, Ms. Joyce Maravich, provided the content for the game as well as administered the pre- and post-tests. This user group adds an interesting dimension to the group by adding another age range to the project.

Refugees at Catholic Charities, Pittbsurgh

FIGURE 5: CATHOLIC CHARITIES LOGO

The refugees at the Catholic Charities are another user group that will benefit from the Literacy Tools Project. The Catholic Charities in Pittsburgh hosts refugees from a lot of different countries who have limited English skills and face problems finding jobs and settling down in the United States. We are still communicating with the administrators at Catholic Charities to determine the content and curriculum for the

project; however, earlier discussions have determined that curriculum will deal with basic conversational and financial literacy.

4.2 Needs Assessment Outcomes

Before conducting needs assessment, we applied for and received IRB approval to work with each of the three user groups. The interviews conducted covered questions regarding their current English skills, their mobile phone usage and their hobbies and interests. This is so that the tool can be customized based on the needs and interests of the user groups. The interview questions used for the immigrant laborers are shown in **Appendix A** and those used for the deaf students are shown in **Appendix B**. Needs assessment with the Catholic Charities group has been delayed due to logistical complications on their part, however, needs assessment with that group will take place over the Summer 2010 through TechBridgeWorld as a continuation of this work.

ROTA Adult English Literacy Program (RAEL)

The needs assessment for this group was conducted in collaboration with Dr. Silvia Pessoa and her students who help teach the immigrant laborers. A majority of the immigrant laborers enrolled in the classes were from Egypt and Sri Lanka. All of the learners in the classes own mobile phones; some have multiple phones with different brands and different service providers. While many laborers in the basic class have very old models of Nokia phones, some in the Intermediate class had the latest Nokia phones and one or two owned Blackberry's and iPhone's.

Their hobbies and interests include talking about places in their home countries and sports. Soccer is a popular sport in all of the Middle East, and therefore, many Egyptians love soccer. The Sri Lankan and the Indian sub continent population enjoy cricket.

As part of the needs assessment process, sample questions for the basic and intermediate classes were also collected from Dr. Pessoa.

CMU-Q Service Attendants

We conducted the needs assessment for this group at the Carnegie Mellon building in Qatar. The service attendants, a total of 10, mainly came from three countries: Sri Lanka, Nepal and the Philippines. Since this is the only group that is not in a structured English class, they were asked what they would like to learn and the answers ranged from basic reading/writing, grammar, and questions to prepare for IELTS exam [**16**].

Most of them owned phones given to them by Carnegie Mellon, which is a Nokia 3000 series phone. Some of them owned a personal phone, usually later versions of Nokia that have graphics and more features. They use their phones to play games, mostly limited to the games already available on their phones, which include Sudoku, snake, and basketball.

Their hobbies and interests include playing sports. The service attendants from the Philippines often get together to play basketball, and those from Sri Lanka enjoy playing cricket.

7th and 8th grade students at Western Pennsylvania School for the Deaf

The needs assessment for this group was conducted at the Western Pennsylvania School for the Deaf in collaboration with TechBridgeWorld staff members Sarah Belousov and Ermine Teves. Interviews were conducted with two teachers and 15 students.

Teachers mentioned that some of the hardest English concepts faced by deaf students are articles, non-count verbs, verb tenses, conjugation, punctuation, wrong order of adjectives etc and they would like for the students to practice articles, non-count verbs and verb tenses using the literacy tools project. Sample questions were collected from Ms. Joyce Maravich as part of the needs assessment process.

Some of the students own phones but they are not allowed to use their phones in class. Students have a variety of hobbies which include reading, playing games, etc. and also enjoy playing word games. Students love challenges and wanted multiple levels and images in the game.

The purpose of the needs assessment was to make sure that the technology design is culturally relevant for the user groups. The next section will discuss the available technology and how the needs assessment conducted with user groups will lead to modifications necessary to make it an effective tool.

5. Technical Approach

The goal of the thesis is to develop mobile phone based games to help improve English literacy skills in adults. This senior thesis project builds on TechBridgeWorld's iSTEP 2009 Literacy Tools project [**17**], which is a tool that has been used to improve English literacy skills in children. This section will describe the Literacy Tools project in detail, and discuss the modifications that were necessary for deploying it with adult users.

5.1 iSTEP 2009 Literacy Tools Project

The Literacy Tools project was developed in Tanzania in the summer of 2009 during TechBridgeWorld's iSTEP [**18**] internship to give additional practice in English exercises to primary school students. Soccer is a popular game in Tanzania and the game, which is based on a penalty kick in soccer, is a motivator for the students in Tanzania to practice English exercises. A content authoring tool was also created to involve

teachers' input in the games and to motivate the teachers to be part of the project. The tool is meant to be used as a classroom activity. Therefore, the Literacy Tools project resulted in a two part tool which includes a single player game accessible on a mobile phone, and an online content authoring system which enables teachers to add useful educational content to the games.

5.1.1 Content Authoring Tool

The content authoring tool is available online for the teachers to add useful educational content to the games. The teacher can specify the question, answer, category of the question and the difficulty level. Once the teacher is done adding all the questions, an XML file is produced, which needs to be transferred to the mobile phone (via a USB or data cable) in order to be used in the game. A screenshot of the initial content authoring tool is shown below.

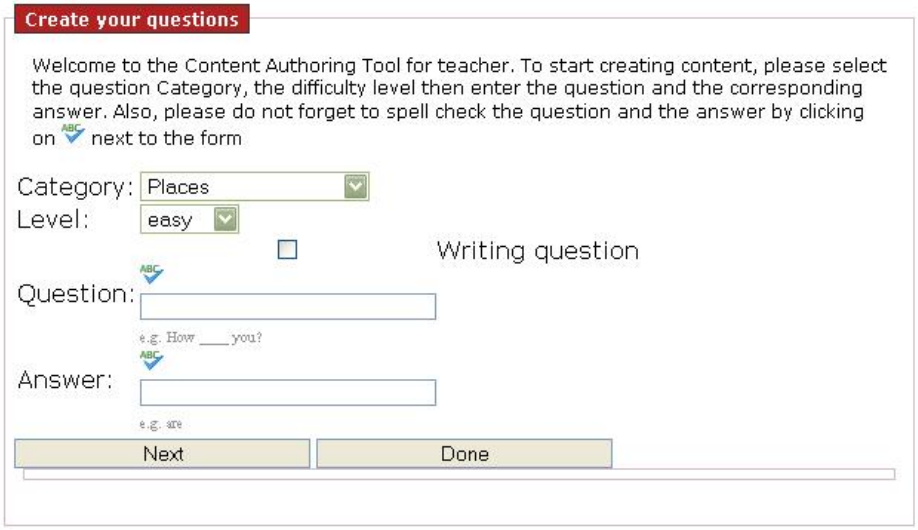


FIGURE 6: VERSION 1 OF CONTENT AUTHORING TOOL

5.1.2 Mobile Phone Game

The original mobile phone game was based on a soccer penalty kick concept, and has a quiz format where the screen shows a question and four options. If the user selects the right answer among the options, he/she scores a 'goal' and gets a point; else, he/she misses the goal and the phone scores a point. The user gets a maximum of three attempts at every question. If the user gets all three attempts wrong, the game displays the right answer on the screen. The number of wrong answers entered is measured by the phone's score. The goal and the missed goal are displayed as static animated gif images. Screenshots from the mobile phone game are shown below.

Figure 7: Question screen

Figure 8: Goal Animation

Figure 9: Miss animation

The game has an adaptive difficulty level feature, which automatically adjusts the difficulty level of the game based on the user's performance. 5 consecutive right answers will shift the game to a higher difficulty level, and 2 consecutive wrong answers will drop the game to a lower level. This scheme ensures that students get more practice before moving onto the difficult levels. The game uses the difficulty level specified by the teacher using the content authoring tool.

5.2 Technology Modification

This senior thesis focuses on modifying the Literacy Tools project for use with adult user groups. The modifications are based on the literature review and the needs assessment conducted with the user groups.

5.2.1 Content Authoring Tool

The content authoring tool has been modified to make it a more sophisticated tool for adding educational content onto the mobile phone games. Based on literature review and the needs assessment results of the different user groups, several modifications were made to the content authoring tool which are described below.

Categories

The iSTEP2009 design of the content authoring tool allows the teacher to specify a 'category' to each question. This will enable the mobile phone game to generate appropriate multiple choice answer options for the question. For example, if a question is categorized as "animals", the answer options displayed during the game for that question can contain options "cat, dog, cow, camel". This makes the question more challenging for the student with options belonging to the same category. This also removes the need for the teacher to re-enter the same answer options for each question.

In this initial version of the content authoring tool, the categories were hard coded and the teachers could not add new categories to the list. Our version of the content authoring tool enables the option to add, edit and delete categories. To add a category, the user specifies the name of the category and the answer options within that category. The user must enter a minimum of two answer options and a maximum of six answer options per category. Below is a screenshot from the content authoring tool that allows a user to view all categories and add, edit, or delete categories.

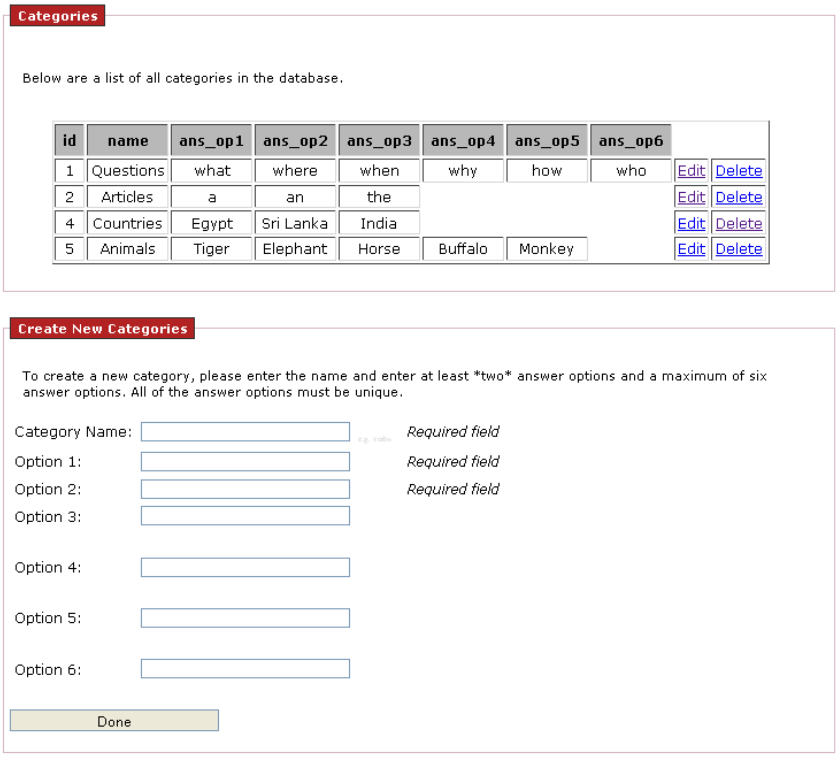


FIGURE 10: CATEGORIES IN CAT

However, there are examples of questions where having answers from a pre-defined category does not make sense. For example,

Q. Frogs _____ to croak. (likes to/like to)

In the above example, the answer does not really belong to a category as it does not make much sense to include answer options just to fill up the answer choices. Therefore, it was decided to allow the teacher with the option of not specifying a category and manually entering the multiple choice answer options. This is for cases where the answers are not easily re-usable. Note that in our example above, if the same answer option "likes to/ like to" is to be used for other questions as well, the teacher has the option to create a category that allows several questions to be entered without having to re-enter the answer choices.

Question Formats

As mentioned in the literature review, the most effective ESL teaching techniques include evaluating the students through regular tests, giving them classroom and homework activities that have sentence construction and cloze passage exercises, and the use of visual images. Also, the sample questions collected from the teachers during needs assessment included sentence construction and cloze passage exercises along with image questions. Some examples of image question and cloze passage exercises are illustrated below.

Q: Which of the images below describe the emotion 'sad'?

a)  b)  c) 

Q: Hi! _____ are you?

a) What b) How c) Where d) When

Thus, we further enhanced the content authoring tool to support the use of images in questions. The goal of the literacy tool is to teach English, and therefore, either the question or the answer needs to be in English. Following this model, two new question formats have been added to the content authoring tool: image question with text answer, and text question with image answer.

Based on the literature review and sample questions from teachers, the content authoring tool has been modified to include five different types of questions:

1) **Writing question**

   The teacher should specify the question, answer, and the difficulty level for the question.

   Example:

   Q: What __ your name?

   A: is

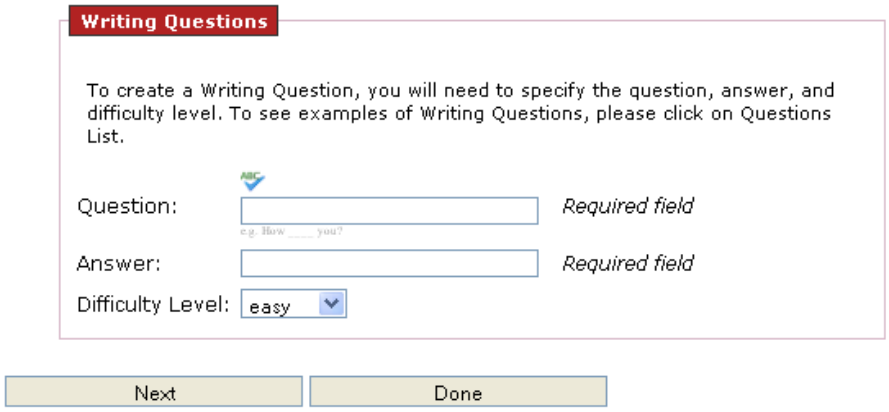


FIGURE 11: WRITING QS IN CAT

2) **Multiple choice question with categories**

The user should specify the question, answer, category and the difficulty level of the question.

Example:

Q: How old __ Beatrice?

A: is

Category: being verbs

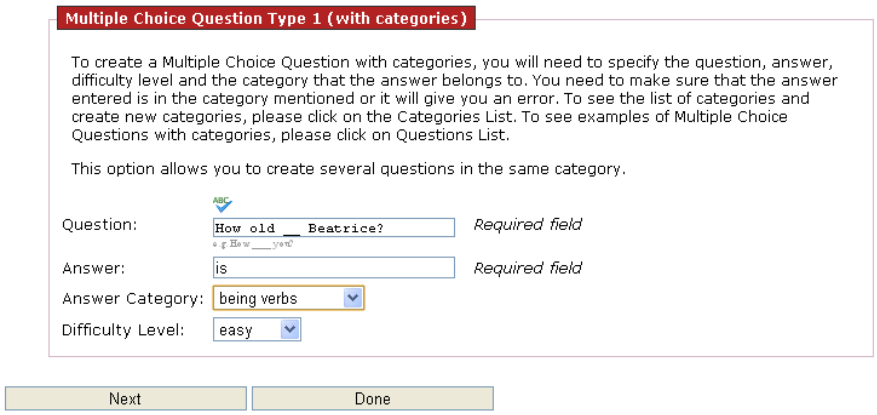


FIGURE 12: MULTIPLE CHOICE TYPE 1 IN CAT

3) **Multiple choice question with user defined options**

The user should specify the question, answer, the multiple choice options for the question and the difficulty level

Example:

Q: ____ you speak English?

A: can

Op 1: are

Op 2: can

Op 3: does

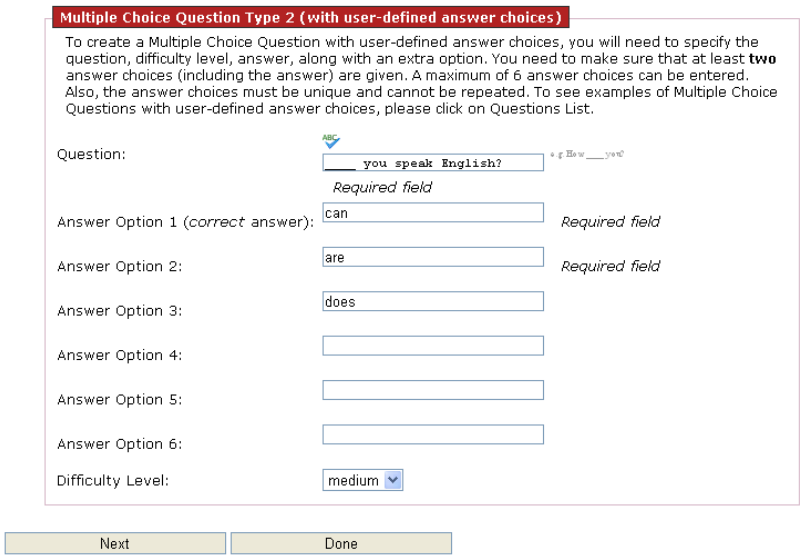


FIGURE 13: MULTIPLE CHOICE QS TYPE 2 IN CAT

.

4) Image question with text answer

The user should specify the question, the image related to the question and the difficulty level of the question. For the answer, the user can choose any answer type from the options of writing, multiple choices with categories, or multiple choices with user defined options.

Example:

Q: What emotion does the picture below describe?

Qs image: 

A: sad

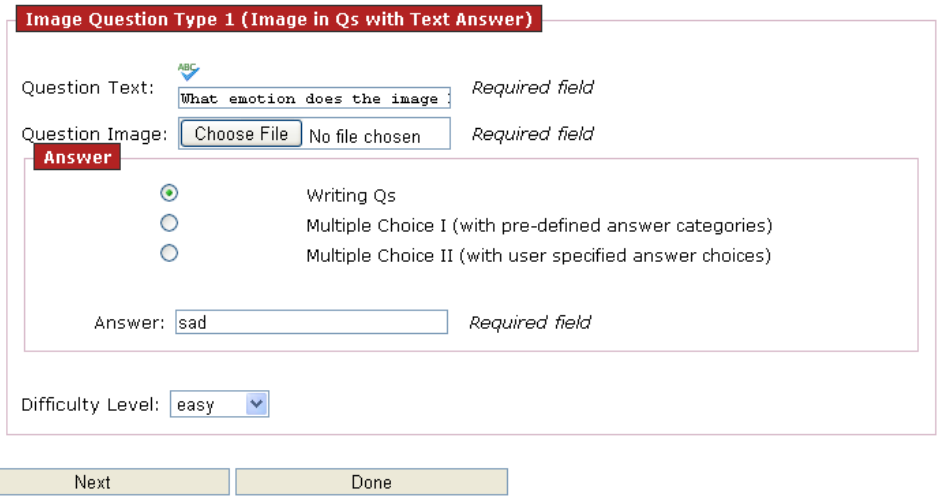


FIGURE 14: IMAGE QS TYPE 1

5) **Text question with image multiple choice answer**

The user should specify the question, difficulty level, the image answer, and the image answer options for the question.

Example:

Q: Which of these images describe the emotion sad?

A: 

Options:   

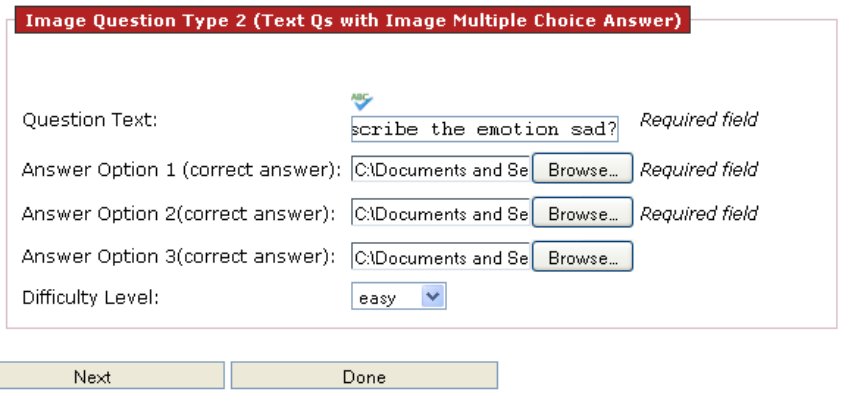


Figure 15: image qs type 2

Database Backend

In the previous version of the Content Authoring Tool (CAT), questions entered in the tool were not saved. At the end of a session, the questions are re-produced to the user via a XML file, which has to be manually saved. In order to add a question to the list of questions, the teacher would have to re-type the entire list of current questions in addition to the new ones. Moreover, the teachers were not provided with an option for editing questions that has been entered incorrectly. In order to address this inconvenience, a mySQL backend was added to the CAT. The questions entered into the database will get saved in appropriate tables, and they can be edited or deleted at any time. The XML file reproduces the list of all questions saved in the database.

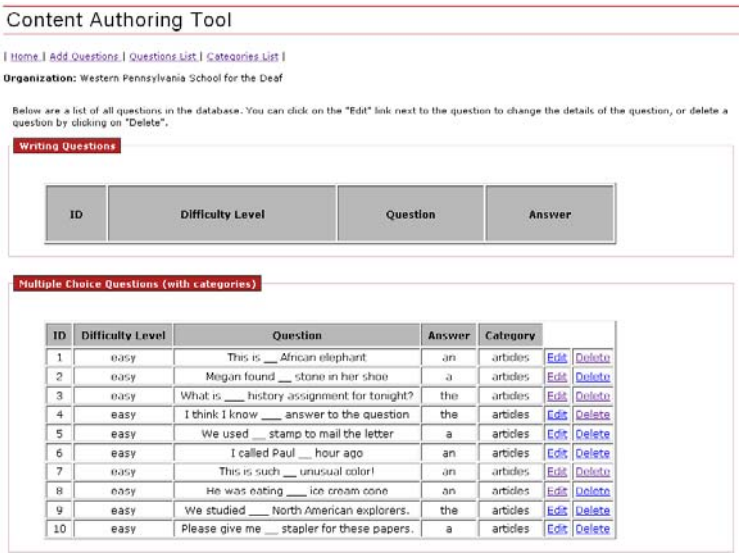


FIGURE 16: LIST OF QS IN CAT

Adding/Editing questions to the database

HTTP *POST* variables from the forms are processed by a PHP script and entered into the appropriate database based on the type of question. There are 6 tables in the database.

Table name: *answer_ops*

This table maintains the categories and the answer options in each category.

Field	Type	Value
Id	INT	Id of the category
name	VARCHAR	Name of the category
ans_op1	VARCHAR	Answer option 1
ans_op2	VARCHAR	Answer option 2
ans_op3	VARCHAR	Answer option 3
ans_op4	VARCHAR	Answer option 4
ans_op5	VARCHAR	Answer option 5
ans_op6	VARCHAR	Answer option 6

TABLE 1: TABLE FOR CATEGORIES

Table name: *writingqs*

This table maintains the list of writing questions in the database.

Field	Type	Value
Id	INT	ID of the question
w_level	VARCHAR	Difficulty level of the question
w_qs	VARCHAR	Question
w_ans	VARCHAR	Answer

TABLE 2: TABLE FOR WRITING QS

Table: *multiqs*

This table maintains the list of all multiple choice categories (with and without categories)

Field	Type	Value
Id	INT	ID of the question
multi_level	VARCHAR	Difficulty Level
multi_qs	VARCHAR	Question
multi_ans	VARCHAR	Answer (Right)
multi_op1	VARCHAR	Option 1
multi_op2	VARCHAR	Option 2
multi_op3	VARCHAR	Option 3
multi_op4	VARCHAR	Option 4
multi_op5	VARCHAR	Option 5
mutli_op6	VARCHAR	Option 6
multi_cat	VARCHAR	Category name, if applicable
isCat	INT	1 if category is specified, 0 if not

TABLE 3: TABLE FOR MULTIPLE CHOICE QS

There are two types of multiple choice questions, one with categories and one with user-defined answer choices. The *isCat* field specifies if the question has a category or not. The question, answer and difficulty level are specified by the user. If the user selects a category, then the *multi_cat* field name is set and the answer options are pulled from *answer_ops* table for the matching category name. If the user does not specify a category, then the category name is set to null and the table is filled with the answer options entered by the user.

Table name: *img_writqs*

This table maintains the list of image questions with writing answers.

Field	Type	Value
Id	INT	ID of the question
level	VARCHAR	Difficulty level of the question
qs_text	VARCHAR	Question
qs_img	VARCHAR	Filename of the Image related to the question
ans	VARCHAR	Answer

TABLE 4: TABLE FOR IMAGE WRITING QS

The uploaded image, which is initially stored in HTTP $_FILES variable, is given a randomly generated 13 character name and then moved to the data folder on the server. The filename of the image is then saved in the qs_img field.

Table name: *img_multiqs*

This table maintains the list of image multiple choice questions.

Field	Type	Value
Id	INT	ID of the question
level	VARCHAR	Difficulty Level
Qs_text	VARCHAR	Question
Qs_img	VARCHAR	Filenameof the image related to the question
Ans	VARCHAR	Answer (Right)
Ans_op1	VARCHAR	Option 1
Ans_op2	VARCHAR	Option 2
Ans_op3	VARCHAR	Option 3
Ans_op4	VARCHAR	Option 4
Ans_op5	VARCHAR	Option 5
Ans_op6	VARCHAR	Option 6
Cat	VARCHAR	Category name, if applicable
isCat	INT	1 if category is specified, 0 if not

TABLE 5: TABLE FOR IMAGE MULTIPLE CHOICE QS

Table name: *img2*

This table maintains the list of questions with image multiple choice answers.

Field	Type	Value
Id	INT	ID of the question
level	VARCHAR	Difficulty level of the question
qs	VARCHAR	Question
ans	VARCHAR	Filename of the image answer
ans_op1	VARCHAR	Filename of the answer option 1 image
ans_op2	VARCHAR	Filename of the answer option 2 image
ans_op3	VARCHAR	Filename of the answer option 3 image

TABLE 6: TABLE FOR IMAGE QS TYPE 2

Viewing questions in the database

All the questions in the database can be viewed under the "Questions List" link on the content authoring tool. It is done by executing a simple "*SELECT * FROM table*" query on the database.

Deleting questions from the database

The questions in the database can be deleted by clicking the "Delete" link next to a question . It passes the ID of the question to the PHP script which then executes a "*DELETE FROM table WHERE ID = 4*".

New XML file format

All of the new question formats are entered into the Text Questions XML format, which required the addition of many new tags.

Categories:

<catlist> - Indicates the beginning of the category list
<cat> - Indicates a new category
    <name> - Name of the category
    <opt> - Answer option in the category
</catlist>

Writing Questions:

<wqslist>- Indicates the beginning of the writing questions list
<wqs> - Indicates a new writing question
    <w-level> - Difficulty level
    <w-qs> - Question
    <w-ans> - Answer

Multiple Choice Questions with Categories

<m1qslist>- Indicates the beginning of the multiple choice with categories questions list
<m1qs> - Indicates a new multiple choice question with categories
    <m1-level> - Difficulty level
    <m1-qs> - Question
    <m1-ans> - Answer
    <m1-cat> Category

Multiple Choice Questions with User Defined Options

<m2qslist>- Indicates the beginning of the multiple choice (with user defined answer options) questions list
<m2qs> - Indicates a new multiple choice question with user defined answer options
    <m2-level> - Difficulty level
    <m2-qs> - Question
    <m2-ans> - Answer
    <m2-opt> - Answer option

This XML file can be downloaded by the users and used in the phone to add new questions.

There is also the ***Image questions XML*** file which cannot be used on the phone. The image questions are parsed in the following way:

Image question with Multiple Choice text answer (with category)

<i1_m1list> - Indicates the beginning of the image question with multiple choice (with categories) text answer

<i1_m1level> - Difficulty level
<i1_m1qs_text> - Question
<i1_m1qs_img> - Question image
<i1_m1ans> - Answer
<i1_m1cat> - Category

Image Question with Multiple Choice text answer (with user defined options)

<i1_m2list> - Indicates the beginning of the image question with multiple choice (with user defined options) text answer

<i1_m2level> - Difficulty level
<i1_m2qs_text> - Question
<i1_m2qs_img> - Question image
<i1_m2ans> - Answer
<i1_m2opt> - Answer option

Text Question with Image Answer

<i2_list> - Indicates the beginning of the text question with image answer

<i2_level> - Difficulty level
<i2_qs> - Question
<i2_ans> - Filename of answer image
<i2_opt> - Filename of answer option image

The content authoring tool provides a way for the teachers and/or administrators to provide content to be used on the games. The XML files serve the purpose of transferring the questions from the tool to the mobile phone game. The next section will look at the mobile phone game aspect of the Literacy Tools and the modifications necessary there to support the new question formats.

5.2.2 Mobile Phone Game

The original iSTEP2009 game was developed in Java Mobile Edition (JavaME) using the Light Weight User Interface Toolkit (LWUIT) according to the MIDP 2.0 and CLDC 1.1 specifications [**19**]. In this thesis work we modified this original mobile phone game to support the new format of questions and additional challenge modes. Both of these components are discussed below.

Support for new format of questions

The new XML file formats described above have to be parsed in the mobile phone game and new questions have to be created from it.

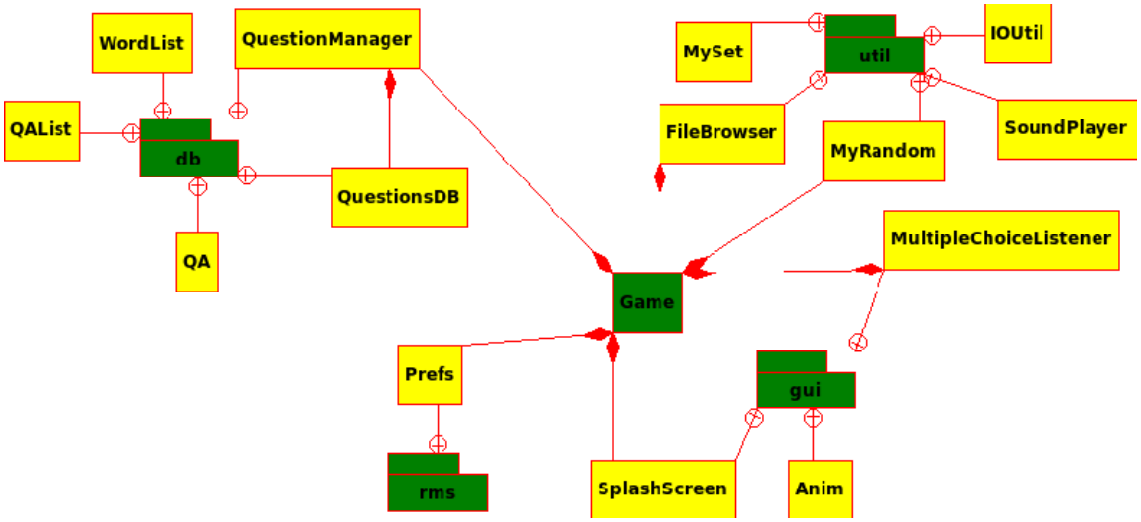


FIGURE 17:IMAGE SHOWING THE VARIOUS COMPONENTS OF THE MOBILE PHONE GAME

The above image, taken from the iSTEP 2009 Final Report [**19**], illustrates the various modules and components that form the literacy game. The part that deals with question formation and selection is the "db" module, and that has been modified to match the new XML file format to support the new format of questions.

The description for the modified components in DB, as adapted from the iSTEP 2009 Final Report [**19**], is shown below.

Name	Description
QuestionManager	Handles parsing and loading all the questions from the DB (This used to be done in QuestionsDB before). This class also constructs the actual question that is eventually posed to the user.
QuestionDB	Acts as the provider of questions based on the difficulty level.
QAList	This keeps track of the basic details as in the QA class as well as all the possible answers to the question, and also the index of the right answer. The QAList objects are constructed by the QuestionManager
QA	A class that logically represents a question. It holds the question and answer text, the category, as well as the difficulty level of the question, and specifies if it is a writing question and/or an image (type 1 or 2) question. QA objects are created by QuestionsManager.
WordList	A data class that merely holds list of words that belong in a particular category. WordLists are used by the QuestionManager while reading the XML files.

FIGURE 18: DESCRIPTION OF THE MODIFIED COMPONENTS IN DB

The *QuestionManager* class now handles the parsing of the XML file and the creation of questions. For creating a question, it has a *parseQs* method that assigns the question, level of difficulty and answer for each question.

- For writing questions, it sets the category and word list to be null, *isWriting* = true, *isImage1* = false and *isImage2* = false, where *isWriting* indicates if the question is a writing question or not, *isImage1* indicates if it is a type 1 image question (i.e. image question with text answer) or not and *isImage2* indicates if it is a type2 image question (i.e. text question with image answer) or not.
- For multiple choice questions with categories, it sets the category name and loads the word list from the *wordCat* hashtable based on the key "category". It also sets *isWriting*, *isImage1* and *isImage2* to false.
- For multiple choice questions with user defined answer options, it creates a word list from the given answer options, sets *category* to null, and *isWriting*, *isImage1* and *isImage2* to false.
- For image questions with written answers, it follows the same procedure as for writing question and sets *imgQs* and *isImage1* to true. It does the same for multiple choice questions with categories and with user defined answer choices.
- For text question with image answers, it creates a word list of the image file names, and sets *category* to null, and *isImage2* to true.

For writing and multiple choice questions, the game already has forms that will display the question on screen, like in the image below.

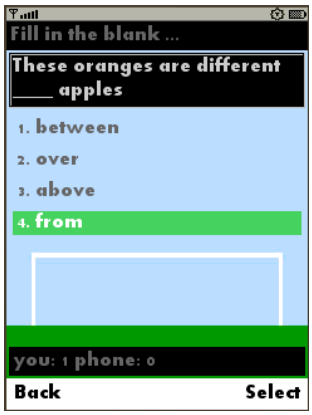


FIGURE 19: MULTIPLE CHOICE QS SCREEN

For image questions, new forms were created that would display the image question.

For image question type 1 (i.e. image question with text answer), the form displays the question along with the question image and the list of answer options. This is done with the help of a LWUIT Container component inside of the Form. The image is displayed in terms of a disabled button, which means that nothing happens if you hit the button.

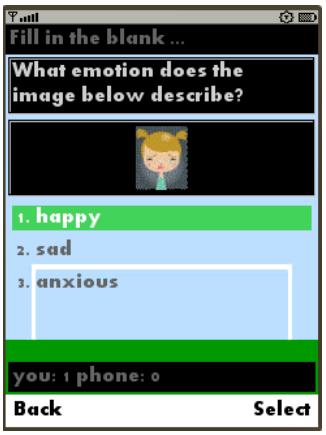


FIGURE 20: IMAGE QS 1 SCREEN

For image question type 2 (i.e. text question with image answer), the form displays a question and then a list of images that act as the answer options. This is also done with the help of LWUIT Button and Container components, with the Box Layout along the y-axis. As you can see in the picture below, the form is missing the 'Select' command. This is because the answer images are displayed as buttons, and to select the correct answer, the user would have to click on the button.

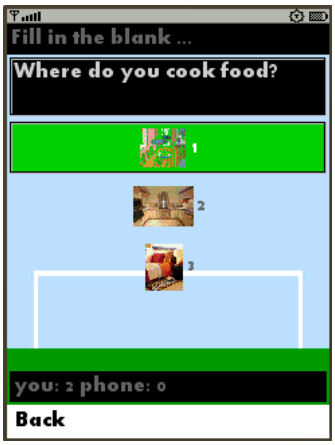


FIGURE 21: IMAGE QS TYPE 2 SCREEN

Challenge Component

Based on needs assessment, we learned that teenagers and adults want a challenge component in their games that motivate them to play the game for longer durations. For the literacy tools game, we decided to add a "challenge mode" to the game. In the challenge mode, the user is asked to attempt to reach a target score, which is a random number between 5 and the total number of questions loaded in the game. When the user has answered the target number of questions correctly, he/she wins the game.

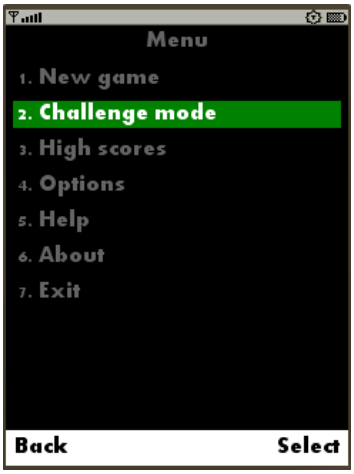


FIGURE 22: CHALLENGE MODE IN THE MENU

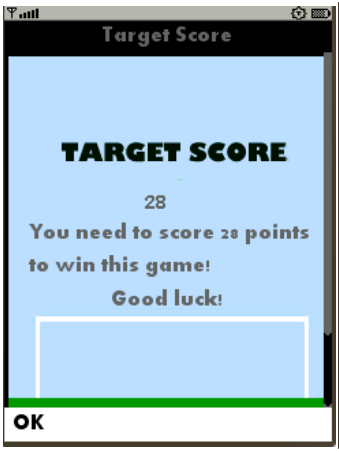


FIGURE 23: TARGET SCORE

FIGURE 24: CONGRATULATIONS SCREEN

Recommendations

Our needs assessment also revealed that adults like to see the progress they have made and feel a sense of accomplishment when they have completed specific levels of a challenge. Therefore, a suggestion would be to include graphical recognition of levels crossed. Currently, the game has an adaptive difficulty level that promotes or demotes the users to different levels based on their performance. However, it happens discretely without any appearance of level change. This modification can happen in two ways:

- Each time the game shifts between levels, the user is shown a screen that says "Good! Onto the difficult questions now!" or "Uh-oh, let's get some practice in the easy levels".
- Once the user has completed and gotten all the questions in the "easy" set right, the user is taken to the next level "Medium", and shown the screen "Easy Level Completed". In this method, the questions come only from that level of

difficulty. The user needs to complete all the questions in a certain level to proceed to the next level and the game ends when the user has completed the "Difficult" level.

In the future, the game can also have multiple levels and the game can go on until the highest level is completed.

Needs assessment conducted with the young adults from WPSD indicates that they like games with more interaction and that have images, videos and challenging levels. Also, most adults have various sports as their hobbies and interests and would be a good theme for future games. In order to meet the interests of the various user groups, it would be recommended to create additional games based on user demographics.

Game Ideas

1) **Soccer:**

   This comes from the needs assessment interviews of the RAEL students from Egypt. We can extend the existing soccer game to be more interactive.

   The player gets to choose his own team and the opposing team (For eg, Egypt vs. Algeria). The top three players from his team are on the soccer field. For each right answer, the first person scores the goal. For every attempt at a wrong answer, the ball gets passed behind to the next player. If all three attempts have been used, the opposing team scores a goal.

   This way, the scores are displayed as "Egypt vs. Algeria". The motivation in this game is for the person to do his best to make his team win! The game can be extended and made more interactive using J2ME tiled layers and sprites. The tiled layer would be the soccer field, and the various sprites would be the three players, soccer ball and the goalie.

2) **Cricket:**

   This comes from the needs assessment interviews of the RAEL students from the Indian sub-continent who enjoy cricket.

   Similar to the previous game, the player gets to choose his own team and the opposing team (For eg, SriLanka vs. India). The team is given a target score to achieve, for example 214/3. This means that they should score 214 right answers before losing all of their players. There are 11 players in each team. Each player is out on the field to bat. For every right answer, the player scores points based on the difficulty of the question. For example, for a hard question, the player scores 6 runs, 4 runs for a medium question and 2 runs for an easy question. If a player gets a question right after a wrong attempt, he scores only 1 run. If he gets all the attempts wrong, he is out of the game. The game goes on until all 11 players are out or until the target score is met.

Again, the motivation in this game is for the person to do his best and make his team win! The game can be made interactive be using a cricket pitch as a tiled layer and sprites with a batsman, bowler and the cricket ball.

3) **Navigation:**

The third idea for the game comes from the fact that the RAEL students love to talk about places in their home countries. This could also be an interesting international travel game for the WPSD students.

The idea of the game is that there are several local or international destinations stored in the game. Each destination has a score attached to it; the more famous or desirable places are worth more points. You need to navigate through the game and answer questions to score points. Depending on the points you score, you can choose certain locations or destinations, which will then display details about the new location. For example, if the location was Paris, it would show pictures of the Eiffel Tower etc. The navigation game will continue in Paris until the player chooses another location.

The motivation behind the game is to be able to visit all the destinations, more like "collect all the beans" and maximize the number of right answers. This game would involve more extensive images and graphics for the various locations.

The following section will outline the field testing details and results of the existing soccer game with the RAEL students.

6. Experiments & Results

The Phase 1 of the field testing consists of testing the modified soccer game with the user groups. Field testing started rather late and the researchers got only 2 weeks of testing data from the user groups. Out of the three user groups, it was possible to conduct field testing with two of them: RAEL program and the service attendants at Carnegie Mellon Qatar campus.

IRB Consent:

The research requires voluntary participation from all of the users. The purpose of the research, its potential benefits and expectations of participant involvement (interviews, tests, and field testing) were explained to the user groups. Their voluntary participation was requested and it was also explained to them that they could quit during any point in the research. Field testing was carried forward with those that gave their verbal consent to participate in the research.

Pre-tests:

Each of the groups was given a pre-test to fill out and the scores were collected by the instructors or administrators. During the pre-test, the users' phones were collected, studied and tested to see if it would support the literacy tools game. It turned out that most phones do not support the literacy tools game for a variety of reasons:

- Users had old Nokia models don't have the necessary requirements on the phone for the game
- Users had a lot of pictures, videos and music on their phones that took up a lot of memory and hence, the game wouldn't work.
- Users had phones from different brands for which data cables weren't accessible at that point.

Testing Period:

The original plan for testing was to install the game onto the users' phones and have them play the game in their own time or do homework through the games. However, considering the fact that the game doesn't work on most of the user's phones, it was decided that we would conduct testing with the TechBridgeWorld phones. The users will be given the phones for testing for the testing period and field testing will take place in class.

- **RAEL**

FIGURE 25: RAEL STUDENTS

RAEL classes happen every Monday and Wednesday. The last 5 or 10 minutes of the RAEL basic and intermediate classes were reserved for field testing the literacy tools mobile phone game. The users were asked to participate if they are interested and almost 90% of the users stayed back to check out the game. Almost all of them wanted to have the game installed on their phones.

Basic Class:

The basic class takes place every Wednesday. The basic class students enjoyed the game and expressed interest to play the game. They found the questions hard, and this was indicated also by their scores.

Average score for the first week: In a period of 15 minutes, they got an average of 17 questions right, and 9 questions wrong.

Average score for the second week: In a period of 10 minutes, they got an average of 14 questions right and 6 questions wrong.

Note: The questions used in both weeks were different; the questions were added onto the content authoring tool by Dr. Pessoa's students and these questions were used to test the students in the second week. The first week used standardized sample questions to test them. Also, not the same set of people played the game

both weeks. Therefore, the improvement or lack of, of the scores cannot be attributed to the class or the questions.

Intermediate Class:

The intermediate class takes place in two different locations on Monday and Wednesday respectively. The class enjoyed playing the game, however, found the boo's and yay's (audio feedback) annoying after a while. They also hated the writing question, as they would spend time typing in long questions and would be annoyed if they got the wrong answer because of missing the apostrophe or full stop etc.

Average scores for the first week: In a period of 15 minutes, they got an average of 11 questions right and 6 questions wrong.

Average scores for the second week: In a period of 10 minutes, they got an average of 26 questions right and 12 questions wrong.

Note: The questions used in both weeks were different; also, the game for the intermediate class also included questions from the basic class that the class was able to answer very easily. Also, whenever they would get an intermediate level question wrong, the game automatically adjusts the game to be of an easy level, at which point they get basic level questions that increases their scores.

- **Service Attendants**

Testing with the service attendants took place thrice a week for 30 minutes each. Interested participants showed up with 90% probability for the next class. Here again, the service attendants were provided with phones for testing and they would play for the entire duration of 30 minutes.

Average scores for the first week: In a period of 15 minutes, they got an average of 17 questions right and 11 questions wrong.

Average scores for the second week: In a period of 10 minutes, they got an average of 35 questions right and 14 questions wrong.

Note: The service attendants have varying levels of English skills, and so the average scores reported are not exactly accurate. Some service attendants got real good scores like 47 questions right and 7 questions wrong, while some others got scores like 12 questions right and 19 questions wrong. The average therefore does not represent the whole group.

- **Western Pennsylvania School for the Deaf**

Field Testing could not be conducted with WPSD as they had some policy changes and the school is going through standardized testing which forced us to put off testing until the summer.

Challenges faced:

Among the biggest challenges faced during testing is the lack of time. The RAEL classes have a packed schedule and trying to reserve the last 10 minutes for testing turned out to be harder than expected. Some time would also be gone in answering questions, or quitting the game by mistake, etc.

The other challenge obviously is that the game does not work on most of the users' phones. There is a text based version of the game; however, there is no standalone version which we can install on phones directly. However, having the 10 research phones for testing helped solve the matter and bring consistency into the testing environment.

7. Discussion & Analysis

The user groups were all really excited about the opportunities presented in playing this game. This was more so among the service attendants, who weren't enrolled in a structured English class and were excited to use a tool that would help them improve their English skills.

One of the important things observed during the field testing is that the adult learners gave more emphasis to the learning rather than the game component. For example, at the RAEL Intermediate class, if a student got the answer wrong repeatedly, they made it a point to stop and ask their instructors about the right answer and a short clarification about similar questions. Similarly, we observed the same case with the service attendants; if they got a wrong answer, they would make sure to stop by and understand why they got it wrong before proceeding to the next question. The challenge component of the game for the adult learners came from comparing their scores with that of their peers. In both the RAEL and the service attendants groups, the students were motivated to get the best score, and would frequently keep comparing between peers.

The challenge component in terms of target scores and levels were a result of the needs assessment conducted with the 7^{th} and 8^{th} grade students at the WPSD. They are teenagers who enjoy serious gaming and the challenge component is probably more important as a motivator for this age group. This age group will also require more graphics and interactive games to enjoy the games.

Initially, we intended this tool to be designed for self-learning, i.e. to be used in your free time to practice English exercises on your mobile phone. However, it seems that the literacy tools project is better designed for a structured classroom environment where English concepts are taught formally, and the mobile phone game can be treated as a non-traditional and engaging platform to practice the concepts. The teacher can regulate the questions that the students are practicing with new and more challenging questions every week, thus making the literacy tools project more sustainable in the long run.

The evaluation of the mobile phone game as a learning tool is to be conducted via pre and post tests. However, we haven't yet received post test results for the RAEL group. Also,

considering the fact that we got a total of 1.5 hours of field testing in two weeks, the pre and post test evaluation probably won't be an accurate measure of the effectiveness of the mobile phone games for learning English.

Statistically significant results cannot be derived from the field testing as the user groups practiced on the mobile phone game for a rough period of 1.5 hours over two weeks. Also, the pre and post tests were conducted with a long gap in between, and the students who are in the RAEL class have accumulated additional skills through the class, and therefore, improvement in scores cannot be attributed to the literacy tools mobile phone game.

The most significant observation from the field testing, however, is that the user groups enjoyed playing the game, did not get bored of it in less than 10 minutes and expressed interest to continually use it to learn English. They understood the benefits and opportunities of using the tool to learn and improve their English skills. This positive interest in the tool indicates that the user groups will continue to play the game leading to increased practice which should result in improved skills.

8. Conclusion & Future Work

The field testing could not yield statistically significant results as the user groups did not get enough time to play the game on the mobile phone, however, the users enjoyed playing the game and expressed interest to continually use it to learn English. This positive interest in the project indicates that there will be increased practice and therefore, improvement in their English skills.

If successful, the literacy tools project presents significant opportunities to the immigrant adult population to improve their English skills. They can use this tool in their free time to practice English exercises and improve their skills at their own pace. The ultimate goal of the tool is to motivate the user groups to want to learn English and minimize the barriers to it.

For future work, the recommendations for the mobile phone game regarding the more interactive games and challenging levels should be implemented and tested with user groups. More thorough and organized testing should be conducted with user groups with longer duration set aside for testing.

Over the summer, TechBridgeWorld will conduct field testing with WPSD and Catholic Charities in Pittsburgh, while in Qatar, a group at Vodafone that teaches English to Nepali workers are interested in deploying the tool in their class. Additionally, the literacy tools project will be continued via the iSTEP 2010 internship in Chittagong, Bangladesh.

Bibliography

[1] Miriam Burt. (2003, December) CAELA: ESL Resources. [Online]. http://www.cal.org/caela/esl_resources/digests/workplaceissues.html

[2] Nation Master. (2008, December) Qatar Immigration Statistics. [Online]. http://www.nationmaster.com/country/qa-qatar/imm-immigration

[3] Nation Master. (2008, December) American Immigration Statistics. [Online]. http://www.nationmaster.com/country/us-united-states/imm-immigration

[4] Miriam Burt, Joy Kreeft Peyton, and Carol Van Duzer, "How should Adult ESL Reading Instruction differ from Adult ABE Reading Instruction?," *CAELA Brief*, 2005.

[5] Aydan Ersoz, "Six Games for the EFL/ESL Classroom," *The Internet TESL Journal*, vol. Vol VI, no. No. 6, June 2006.

[6] Nguyen Thi Thanh Huyen and Khuat Thi Thu Nga, "Learning Vocabulary Through Games," *Asian EFL Journal*, December 2003.

[7] Matthew Kam. MILLEE: Mobile and Immersive Learning for Literacy in Emerging Economies. [Online]. http://www.cs.berkeley.edu/~mattkam/millee/

[8] Microsoft, "Mobile Language-Learning Tools Help Pave the Way to Literacy," *External Research Digital Inclusion Program*, pp. 1-2, 2008.

[9] Microsoft Research India. Kelsa+: IT Access for Low Income Workers. [Online]. http://research.microsoft.com/en-us/projects/kelsaplus/

[10] Grameen Phone. Grameen Phone Official Site. [Online]. http://www.grameenphone.com

[11] aAqua - Almost All Questions Answered. aAqua - About. [Online]. http://aaqua.persistent.co.in/aqualinks/aboutAqua.html

[12] Balan, Elena; Softpedia. (2007, April) Adults Play Games More than All Phone Owners. [Online]. http://news.softpedia.com/news/Adults-Play-Games-More-than-All-Phone-Owners-53198.shtml

[13] Kathleen F. McCoy and Lisa N. Masterman, "A Tutor for Teaching English as a Second Language for Deaf Users of Amercian Sign Language," CIS Department, University of Delaware,.

[14] Raymond D. Kent, "Language of the Deaf: Acquisition of English," in *The MIT Encyclopedia of Communication Disorders*., 2004, ch. Part III: Language, pp. 336-337.

[15] Reach Out To Asia. [Online]. http://www.reachouttoasia.org

[16] IELTS. International English Language Testing System. [Online]. http://www.ielts.org/default.aspx

[17] TechBridgeWorld. (2009, August) iSTEP 2009 - Tanzania. [Online]. http://www.techbridgeworld.org/istep/iSTEP_Tanzania_2009_Final_Report.pdf

[18] TechBridgeWorld. iSTEP 2009 - Tanzania. [Online]. http://istep2009.techbridgeworld.org/

[19] TechBridgeWorld, "iSTEP 2009 Final Report, Literacy Tools Project," Robotics Institute, Carnegie Mellon University, Pittsburgh, USA, August 15, 2009. [Online]. http://www.techbridgeworld.org/istep/iSTEP_Tanzania_2009_Final_Report.pdf

[20] Mathew Kam, Divya Ramachandram, Urvashi Sahni, and John Canny, "Designing Educational Technology for Developing Regions: Some Preliminary Hypothesis," in *Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, 2005.

[21] MaryAnn Cunningham Florez Miriam Burt, "Beginning to Work With Adult English Language Learners: Some Considerations," *Q & A*, 2001.

[22] Ermine Teves, "Shizzle3".

[23] Grameen Bank: Banking for the Poor. [Online]. http://www.grameen-info.org/

#
#
#
#
#
#
#
#
#
#
#
#
#

Appendix

Appendix A: Interview Questions for Immigrant Laborers

Senior Thesis: Mobile Phone Based Educational Games for Improving English Literacy Skills of Limited English Proficient (LEP) Adults

Needs Assessment Questions

General:

a. Name:
b. Age:
c. Nationality:

Mobile Phone Usage

1. What phone do you currently use? Please note down the brand and model number. (Nokia, Sony Ericsson, Samsung etc)
    a. What do you use your phone for?

        [ ] Local calls

        [ ] International calls

        [ ] Text-Messaging (SMS, MMS)

        [ ] Bluetooth

        [ ] Games

    a. Do you play any games on your phone?
        i. If yes, what kind of games?
        ii. Could you please show us your favorite game? (Take observation notes)
        iii. What do you enjoy about those games?

Hobbies/Personal Interests

1. What do you do during your free time?
2. Do you play sports? What kind of sports?
3. Do you watch TV? What kind of shows would you prefer watching?

Appendix B – Interview Questions for Deaf Individuals

Mobile Phone Based Educational Games for Improving English Literacy Skills of Limited English Proficient (LEP) Adults

Needs Assessment Questions – WPSD TEACHERS

General:

1. How many students does your class have?
2. What subjects do you teach?
3. What age-group of students does your class have?

English related:

4. What are the challenges faced by hard of hearing students in learning English?
5. What concepts do you think they would require additional practice in?
6. What, in your experience, motivates the students to learn English?

Technology related:

7. Do you use technology to support your teaching? If so, what do you use and how?
8. Are the students allowed to use computers in class?
9. How many students have mobile phones?
10. Are the students allowed to use their phones in class?

Teaching through games:

11. Have you experimented teaching exercises through games? If yes, please explain.
12. What are the challenges faced by hard of hearing students in learning?
13. What do you think about using educational games on mobile phones to improve their English skills?
14. Do you think the students would like to practice exercises via playing games on mobile phones?

Needs Assessment Questions – WPSD STUDENTS

General:

1. Name:
2. Age:
3. Grade:

English Proficiency Level:

1. Do you enjoy studying English?
    i. If no, why not?
    ii. If yes, why?
2. What do you find difficult about learning English?
3. What kind of English lessons do you like?
4. Do you read English story books?
    i. If yes, what kind of books do you like?
    ii. Please mention a few books that you have read.

Mobile Phone Usage:

1. Do you have a phone?
2. If yes, what phone do you have? (Nokia, Sony Ericsson, Samsung) What model number and brand?
    a. What features do you like about this phone? (camera etc)
    b. How do you mainly use your phone?
    c. Do you use text-messaging service?
    d. Do you use Bluetooth services on your phone?
    e. Do you play any games on your phone?
        i. If yes, what kind of games?
        ii. What do you enjoy about those games?

Hobbies/Personal Interests:

1. What do you do during your free time?
2. Do you play sports? What kind of sports?
3. Do you watch TV? What kind of shows would you prefer watching?
4. Is there anything else you would like to tell us?