

# GloBuddy, a Tool for Travellers

**Rami Musa**  
MIT - EECS  
Cambridge, MA  
rmusa@mit.edu

**Andrea Kulas**  
MIT - CMS  
Cambridge, MA  
kulas@mit.edu

**Yoan Anguilete**  
MIT - EECS  
Cambridge, MA  
anguilet@mit.edu

**Madleina Scheidegger**  
MIT - EECS  
Cambridge, MA  
mscheid@mit.edu

## ABSTRACT

GloBuddy is a common sense based application that assists English-speaking travellers in communicating in a foreign language. In addition to translating text, GloBuddy provides relevant words and phrases for the situation in which the traveller is. Common sense reasoning is done with OMCSnet, a semantic network representation of the OpenMind knowledge. GloBuddy could be easily adapted for use on handheld devices, making it a valuable tool for travellers in foreign countries.

## Keywords

translation, common sense, reasoning, languages.

## 1 INTRODUCTION

GloBuddy is a common sense based application that assists English-speaking travellers in communicating in foreign languages. It serves both as a dictionary and as a phrase book without being limited to a static set of words and phrases or a particular language. The program uses common sense from OMCSnet, a semantic network representation of the OpenMind database, to provide words and phrases that are relevant to the situation the traveller describes. In the future GloBuddy will be adapted to run on handheld devices.

There are two modes to GloBuddy. The first is just straight translation, which is always useful to have. The second provides words or phrases relevant to a situation described by the user. This paper describes how each of these modes work and presents some future extensions to GloBuddy.

## 2 TRANSLATION

All translations in the program are done by querying Altavista Babelfish, an online translation service. GloBuddy connects to Babelfish via an http interface.

The choice of using an online service was made in order to save on cost. However, it will be easy to change the interface to work offline in the future.

The user can set both the language of the text to be translated, and the language to which the text should be translated. The text is then passed to Babelfish, and the returned translation is displayed. Translation can currently be performed between English, French, German, Italian, Portuguese and Spanish.

## 3 ASSOCIATION OF WORDS

This functionality takes in a description of a situation in English. It then filters the description for keywords, using a predefined list of words to be excluded. GloBuddy then uses the OMCSnet through an http interface. Http was used for practical purposes, since an interface already existed and was easy to use.

GloBuddy then queries OMCSnet with each keyword to extract all nodes that are directly connected to the keyword. For each output word two scores are calculated. The first is a relevance score, which equals the number of keywords to which it is directly connected. The second is a frequency score, which equals the number of times the word appears in the total output across all keywords.

The output words are then sorted according to their relevance score. This ensures that outputs directly linked to a higher number of input keywords appear first. In other words, more relevant output words appear first. If multiple words have the same relevance score, they are then sorted according to their frequency scores. This deals with the situation where many of the output words are linked to the same number of keywords. In such a situation, the more often an output word appears in the total list of outputs, the more likely it is to be useful, and so it would be placed ahead of other words which appear less often.

The output is then translated to the language the user chose. The English output and the translated equivalent are then displayed, twenty at a time, to the user. The user can browse the output forward and back at will. In case GloBuddy fails to provide satisfactory results, the user can perform a more specific search. The user would choose a word, either a keyword or an output word, and an OMCSnet relation to perform a specific query.

## 4 ASSOCIATION OF PHRASES

This functionality currently uses two sources of phrases. The first is a collection of sentences and questions that are common in everyday conversation. Those are stored in GloBuddy under eight categories: "Food and Dining", "Accommodations", "Travel and Transportation", "Leisure", "Shopping", "Services", "Health and Safety" and "Emergency". The second source of phrases for this functionality is the OpenMind common sense database.

When using this functionality, GloBuddy starts by trying to guess the context of a situation that the user describes. More specifically, GloBuddy tries to guess whether the context of the user's situation fits in any of the eight categories mentioned above. If it does, GloBuddy queries the collection of phrases stored under the identified category for relevant phrases. If the user's situation does not fit any of the defined categories, GloBuddy turns to OpenMind to extract phrases.

### Guessing the Category

For the purpose of making a guess, GloBuddy maintains a list of relevant words and concepts for each of the eight categories mentioned above. Those lists are compared against a set of words relevant to the situation described by the user. A score for each category is calculated. The eight scores are then used to decide whether there is a fit between the user's situation and any particular category.

To generate a list of words relevant to the user's situation, GloBuddy first uses the filtering method described in section 3 to extract keywords from the user's description of the situation. GloBuddy then queries the OMCSnet with each keyword and extracts the words directly connected to it. The output for each keyword is then compared with the list of predefined words and concepts for each category. A percentage score is calculated for every keyword-category combination by dividing the number of words from the keyword's output which are present in the list of relevant words and concepts for the category, by the total number of output words. A final score for each category is then calculated by summing across the keywords. The category whose score is at least 15% larger than any other score is taken as the relevant category. If there is no score that satisfies this requirement, "Other" is chosen as the default.

### Selecting Phrases

After making the guess, GloBuddy proceeds to querying for phrases. If a category other than "Other" was identified, GloBuddy looks for phrases stored under

the identified category that contain any of the keywords, from the user's situation description, or any of the words that are directly connected to all of the keywords in OMCSnet. Phrases that contain more of the search words are provided to the user first together with their translations. If GloBuddy could not identify a relevant category, it defaults to "Other" and uses OpenMind. For each of the keywords obtained from the user's situation description, GloBuddy extracts fifty phrases from OpenMind and constructs a temporary phrase collection. GloBuddy then queries the new phrase collection with the keywords in a similar approach to that used in querying the predefined categories.

When GloBuddy is unable to identify a match between the user's situation and a predefined category or if GloBuddy mis-guesses the relevant category, the user can choose to manually pick the category that GloBuddy should query. The user can also choose to proceed with the "Other" category.

## 5 FUTURE WORK

The basic functionality intended for GloBuddy has been created, but many improvements and extensions can be introduced. For one, GloBuddy could be modified to work offline, using a local translation software, and a local copy of the OMCSnet. This would enable GloBuddy to be used on handheld devices.

Furthermore GloBuddy could be improved by learning from user input when it makes wrong guesses about context or when it mistakenly filters out important keywords from the input. Another learning opportunity would be for GloBuddy to track which words from an output users expand using the specific query functionality, and put those higher up on the list in future queries for the same situation.

It would be important to improve the quality of the word associations taken out of OMCSnet. There are still too many nonsensical associations. One way would be to include a user feedback to filter out all nonsensical words. Another way, if possible, would be to change the algorithm for getting relevant words.

An essential extension is to extend the sources of potential relevant phrases. For instance GloBuddy might have the ability to create new sentences from templates. Alternatively GloBuddy might be able to mutate sentences from OpenMind to take out irrelevant parts, or to rephrase them as questions. A third alternative is to construct a smaller OpenMind-like database, that contains common sense questions and sentences that are common to everyday conversation.