Overall Project Goal

The goal of this project is to present a user with a simple interface with which they may use a visual language (a series of icons) to query for YouTube videos. While this interface could be useful in a variety of domains, for the sake of keeping the scope of this project reasonable, a small set of icons (ten nouns and verbs and five emotions) was used. This project should be considered a proof-of-concept, and any interface presented to real users would have to be much more stylish and sophisticated.

Implementation

The search interface was designed and implemented using HTML and CSS. There are two main panels: the search panel and the icon panel, both of which can be seen in Figure 1. The search panel is where the query is built, and this is made up of two sections. The first is for the query, which can be built using up to 3 Noun/Verb icons; the second is for the emotion/modifier icon, which is one icon that describes the general emotion of the query. For example, if you are looking for a funny video, the “funny” icon would belong here.

Figure 1. Basic Search Interface

The icon panel is where the possible icons can be viewed. Clicking on an icon will place it into the correct spot in the query, unless the query is full. The icon panel is split into two sections depending on the type of icon. Noun and verb icons, which can be used to make up the regular part of the query, are found in the left side. There are ten of these. Adjectives, which are emotions that can be used to modify the entire query, are found on the right. Figure 2 shows the icon panel, with the icons
numbered. Below this, there is a list of the icons, with the “intended meaning” of each below. Note that often, the intended meaning may be very different than how the user interprets it.

Figure 2. Icons and Intended Meanings

Finally, the search interface includes two buttons: Search and Clear. The Search button calls a JavaScript function which does a simple translation from icons into a query in the English language, which is used to query YouTube. The user is directed to the YouTube page with the results. If, instead of searching, the user decides to change or redo the query, the Clear button will remove all icons.

Example Queries

First, a simple query will be demonstrated. Figure 3 shows what the interface would look like when the user wants to search for funny videos of a dog and cat playing together. The “dog”, “cat”, and “playing” icons are selected, in addition to the “funny” emotion icon on the right.

Figure 3. Sample Query

When the Search button is clicked, these icons are translated into the string “funny dog cat playing”. The modifier is always moved to the front, to show that it describes the entire string (note that Google
probably does not use this order in the exact way for returning applicable videos). Figure 4 shows the YouTube results page. Please note in all YouTube results pages that returned ads (videos in the shaded yellow at the top of the page) are not always meant to be relevant to the query, so they should not be used to evaluate the usefulness or success of the system.

Figure 4. Results when Query from Figure 3 is submitted

It is possible to select more than one of a single icon, for example two of the “baby” icon. In this case, it does not make sense to search for “baby baby”, especially since the quality of results fall greatly when this is done. Thus, the plural form is used instead. Figure 5 demonstrates this query. In addition, it demonstrates that simple queries that do not include any emotion/modifier icon are also possible.

Figure 5. Sample Query without an Emotion

Figure 6 shows the results after this query is submitted. Note that the plural has been changed correctly, and the fact that no emotion is included is handled gracefully.
Evaluation

In order to evaluate my work, I had three people (None of them were familiar with visual languages) test my interface and answer a short survey. This gave me an idea of things that the interface did well, and also areas where there is room for improvement. First, I asked each person to label each icon with the meaning it represented. Most of their interpretations matched mine, with a few major exceptions. First, two people thought that the man and woman icons were male and female restroom signs; understandable, as these icons are often used on the doors. Second, the icon for “microphone” gave one person particular trouble and her best guess at the meaning was “turkey baster”. The icon for “playing” was apparently rather difficult, and two participants thought that it was two people pointing to either the moon or the sun. Finally, the emotion icons in general seemed to cause some problems for everyone. Perhaps in the future, these icons should be more detailed and in color to make them more distinguishable (I originally though simple would help people with the meaning, but clearly that was wrong). Ideally, adaptive symbols would help with this meaning difference.

One thing that became clear was that directions should probably be put on the interface to highlight some of its features. Participants did not seem to really understand the difference between the two types of icons. In addition, when asked what icons they would use to represent “children dancing”, none of them used the “child” icon more than once, suggesting that they would need to be taught how to use plurals.

In the final section of the survey, I gave the participants three sequences of icons. I asked them to describe the type of videos they expected the query to retrieve, and then rate the number of videos out of the top ten results that were applicable to the query. The numbers are below in Table 1. A * marks the cases in which I felt the participant was way, way off when it came to their query description. For example, the participant who thought the microphone was a turkey baster obviously got very different results (people singing) than she was expecting (people with turkeys).
Table 1. Relevance of Top 10 Videos to Interpreted Query

<table>
<thead>
<tr>
<th>Participant</th>
<th>Query 1 (“romantic man woman dancing”)</th>
<th>Query 2 (“cute baby dog playing”)</th>
<th>Query 3 (“funny women singing”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>2</td>
<td>0*</td>
<td>0*</td>
</tr>
<tr>
<td>Participant 2</td>
<td>1</td>
<td>10*</td>
<td>10</td>
</tr>
<tr>
<td>Participant 3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average</td>
<td>2</td>
<td>3.3</td>
<td>3.7</td>
</tr>
</tbody>
</table>

**Ideas for Future Work**

The status of the project outlined in the previous section is sufficient for showing the use of such an interface, but there are many possible directions for future work. First and foremost, it is important that more icons be available. Ideally, these icons should be adaptive, so that they may change based on the user’s interpretation of them and desired use. In addition to this, the interface would be more useful if the queries did not have a limit on length.

Another thing that could use improvement is the use of modifiers in the visual sentence. Right now, one modifier (the emotion icon) can be applied to the entire query, but the interface would be much more powerful if modifiers such as adjectives could be applied separately to different icons. For example, imagine that you’d like to search for a video of a big dog and a small cat. Right now, there is no way to effectively do this. It is easy to imagine, however, that there should be a way to apply “big” to the dog icon and “small” to the cat icon.

Finally, a major line of further exploration would be the linguistic side to this type of system; that is, the translation from icons and visual sentence into an English string, which YouTube takes as input. There are many sophisticated techniques that could be used in order to combine icons, add other semantic meanings, and predict most probable orders.

**Additional Resources**

A demo of the project can be found at [http://www.cs.pitt.edu/~friedberg/CS2310demo/](http://www.cs.pitt.edu/~friedberg/CS2310demo/). It has been tested in Chrome and Firefox. A YouTube Video demonstrating its use can be found at [http://youtu.be/9gxxLuKNNiU](http://youtu.be/9gxxLuKNNiU). Both of these links will remain up until January 2012; after that please contact me at friedberg@cs.pitt.edu for the files.