Instant Messaging Semantics (IMSem)

*a messaging client that incorporates indexed images to enrich semantic meaning***.**

Hypothesis: Images improve text-based messaging by restoring context lost without images. I will conduct an analysis of current methods and then create a program, IMSem, that automatically suggests images to include while composing messages.

Background: Communication over technological media is compressed and distorted compared to physical interaction. Without nonverbal expressions, text based content often leads to ambiguity and lose a human touch. Nevertheless, there are many reasons for people to prefer indirect forms of communication in our high tech society. People have created means to restore this nonverbal communication, such as sending encoded expressions (e.g. :-)) or images that convey their intent[1]. These fixes emerge ad hoc out of the social consciousness, but limitations in the framework they are transmitted over limit their potential. I will examine the comparative usefulness of these social innovations and build a framework (IMSem) to enhance social messaging using catalogued images to replace their encoded representations.

Other approaches have been done to restore socially significant communication in primarily text-based environments. For examine, kinetic typography instant messaging provides richer context by animating sentences based on the underlying emotional context mined from key words[2]. Although it provides a richer environment to display emotion, this approach can only provide redundant information. The key element missing is the social context of the message that words cannot clearly transmit alone. As the technological capabilities of phones and other devices improve, the usage of visual media to convey meaning is key, potentially even preferred over personal communication.

Ethnography: I will first examine existing forms of visually enriched communication through two user studies to analyze the differences between messaging mediums. People may manually send recent face pictures, pictures from their photo library, and Internet generated memes. Besides attaching images to standard messages, many popular chat clients focus on sending images, such as Snapchat and Instagram. These methods provide an array of interfaces for people to communicate in a more complete form than text alone.

It is important to establish a baseline for this communication to quantify the effectiveness of and preference for other image media services. The first of two stages will have a series of interviews of a cross-demographic group exploring the qualitative trade-offs of each service and what is lacking. Questions will focus on the retention rate and perceived richness of interaction. The analysis will place manual IM solutions and services on a continuum of simplicity-specificity. The purpose of these initial interviews will be to inform a second study that will quantify the usage of each method. Particular use cases may vary dramatically, so an online survey with participants in the range of thousands will provide meaningful comparative results. My experience as a clinical research coordinator prepares me well for conducting these studies.

---

[1] Suttles, Jared, and Nancy Ide, 2013. Distant supervision for emotion classification with discrete binary values. In Computational Linguistics and Intelligent Text Processing, 121-136. Springer Berlin Heidelberg.

[2] Zhiquan Yeo. 2008. Emotional instant messaging with KIM. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08). ACM, New York, NY, USA, 3729-3734.

IMSem: While interviews and surveys are underway, I will design and write the Instant Messaging Semantics system. In order to seamlessly integrate images to message composition, the program will suggest replacing text encodings with images from the user's library. As the user types, a box will appear suggesting an image to use instead of the text, just as if the text program was correcting a misspelled word. For instance, as shown in Figure 1, the user makes a smiley face ":)". Instead of displaying an impersonal emoticon, they can opt to send a small image of themselves smiling. The advantage of this substitution is that it makes the emotion in the message clearer and more personal.

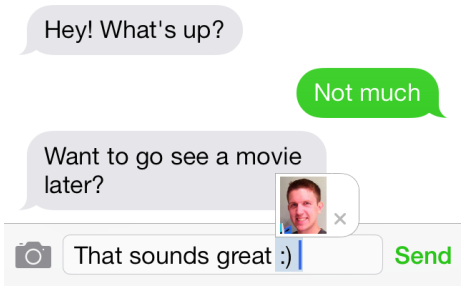

Figure 1: Autocorrect suggesting using a picture rather than text.

The program will have a dictionary of consisting of textual encodings and real images as the key/values. This dictionary can be implemented as a separate program from a text-based messenger.

Year One, Expanded Keyboard: While conducting the user studies, IMSem will be implemented as an image library application and an expanded keyboard in the message editor (with a version for the iPhone and Android). The expanded keyboard will contain manual input and operate in the same way the Emoji system is used. Emoji contains a library of emoticons that can be selected by the user while composing a message. Instead of impersonal images, the user could input their own images they have selected to represent different expressions.

Year Two, Autocorrect Emotions: With a stable application and analyzed studies, I will spend the next year writing a system that implements the expressive image dictionary in the message program itself. On a phone system this will have to be written as a separate program. On a computer instant messenger, this can be a plug-in for one of the open source programs such as Pidgin. This application will contain the autocorrect feature, suggesting expressive images in place of text whether the text is a smiley or any other phrase that exists in the image dictionary. For example, when the user writes a message and refers to their dog "Rover", IMSem will suggest attaching an image of the user's dog instead. IMSem will be a deployable application on at least one mobile and one computer system. Then I will conduct a third study, comparing the new application with existing systems.

Year Three, Automatic Suggestions: The last year of the project will use techniques from NLP and computer vision to make an intelligent expressive image suggestion system. As opposed to manual key-image pairs, the program could match keywords typed by the user and match them to photographs with metadata. The scope of this implementation will be rather complex so the best way to capture it will be to start with faces. An expression and emotion detection system[3] will scan the users library and suggest using these images when matched with emoticons or text emotions.

[3] Such as CMU's Intraface, http://www.humansensing.cs.cmu.edu/intraface/about.html