Messaging Beyond Texts with Real-time Image Suggestions

Joongyum Kim
School of Computing, KAIST

Taesik Gong
School of Computing, KAIST

Kyungsik Han
Department of Computer Science, Ajou University

Juho Kim
School of Computing, KAIST

JeongGil Ko
School of Integrated Technology, Yonsei University

Sung-Ju Lee
School of Electrical Engineering, KAIST

ABSTRACT

While people primarily communicate with text in mobile chat applications, they are increasingly using visual elements such as images, emojis, and memes. Using such visual elements could help users communicate clearly and make chatting experience enjoyable. However, finding and inserting contextually appropriate images during the chat can be both tedious and distracting. We introduce MilliCat, a real-time image suggestion system that recommends images that match the chat content within a mobile chat application (i.e., autocomplete with images). MilliCat combines natural language processing (e.g., keyword extraction, dependency parsing) and mobile computing (e.g., resource and energy-efficiency) techniques to autonomously make image suggestions when users might want to use images. Through multiple user studies, we investigated the effectiveness of our design choices, the frequency and motivation of image usage by the participants, and the impact of MilliCat on mobile chat experiences. Our results indicate that MilliCat's real-time image suggestion enables users to quickly and conveniently select and display images on mobile chat by significantly reducing the latency in the image selection process (3.19× improvement) and consequently more frequent image usage (1.8×) than existing solutions. Our study participants reported that they used images more often with MilliCat as the images helped them convey information more effectively, emphasize their opinion, express emotions, and have fun chatting experience.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**; *Empirical studies in interaction design*; *Ubiquitous and mobile computing systems and tools.*

KEYWORDS

Mobile Chat; Image Recommendations; Visual Communications

ACM Reference Format:
Joongyum Kim, Taesik Gong, Kyungsik Han, Juho Kim, JeongGil Ko, and Sung-Ju Lee. 2020. Messaging Beyond Texts with Real-time Image Suggestions. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20), October 5–8, 2020, Oldenburg, Germany.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3379503.3403553

MobileHCI '20, October 5–8, 2020, Oldenburg, Germany

ACM ISBN 978-1-4503-7516-0/20/10…$15.00
https://doi.org/10.1145/3379503.3403553

1 INTRODUCTION

Among numerous smartphone applications available in the market, mobile chat applications are usually ranked top on the most-downloaded and most-popular lists. Despite the fact that a mobile phone's original purpose was to make voice calls, changes in people's lifestyles and the increase in device/network capabilities have made mobile chatting one of the most (if not the most) widely used methods of communication [49].

Despite their active usage, the functionality of many popular messaging applications is still limited in the sense that they are yet text-oriented. Such text-based information exchange can be effective in many cases, but when delivering specific information or for emotional empathy, text-based conversations can take long or be misleading [10, 54].

The phrase "a picture is worth a thousand words" suggests that images can convey the meaning and essence more effectively using a visual element than a text description. Images have been shown to be more effective than text in many areas. For example, in education, using images helps students understand information better than using text alone [9, 33, 34, 46]. Similarly in advertising, using images to emphasize information or persuasively deliver a message is considered to be more effective and faster than using text alone [19, 39]. Images are also effective in showing social intimacy [40, 42, 50]. We argue that such effectiveness of using images can also be applied to mobile chat environments.

Despite the potential benefits of enriching mobile chat experiences with images, for many popular mobile chat apps, using images is limited to emojis, memes, and personal photos. Emojis are good to express emotions and are often used in social communication [26, 27, 30–32, 35, 41, 56, 59]. Memes, or animated GIFs, are popular for trendy and funny images [5, 22, 23]. Personal images [11, 29] are useful in sharing deep personal and social context. We argue that public images on the Internet could be useful for sharing information and emphasizing opinion on mobile chats [28]. Moreover, current mobile chat apps require users to spend tedious steps to include public images in a chat; the user typically goes through other applications to find, store, and retrieve a desired image, then uses the original mobile chat app's features to include the image in the chat room. While many mobile apps try to offer an easy-to-access shortcut to streamline this process, putting explicit manual effort to find and attach an image can distract the users

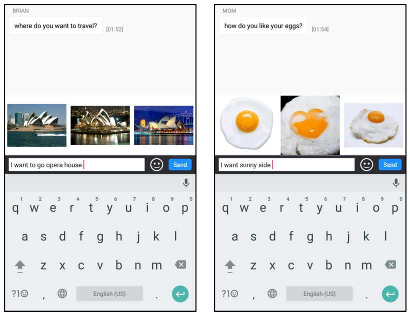


Figure 1: Image suggestion examples: from the images suggested in real-time based on the typed text, users can select an image to use during the chat. On the left, MilliCat suggests images of "opera house". On the right, MilliCat suggests images of "sunny side". Note that the suggestions are triggered with concrete nouns as a search query. This approach works even with grammatically incorrect (left) or incomplete (right) sentences, both of which are common in mobile chat.

from their chat conversation and discourage image usage during chat.

We believe that having autonomous real-time image suggestions can alleviate such user inconveniences when using images within a mobile chat. Moreover, it was reported in a survey [28] that 81.57% of survey participants prefer to be autonomously suggested images from the Internet, based on the chat application input text. To validate such hypothesis, we present the design and implementation of MilliCat, a system that autonomously and selectively suggests images with respect to the user's input text in a mobile chat application. MilliCat suggests images in real-time and allows the users to select a most suitable image that matches the context presented in the input text. The core of the MilliCat design is in analyzing the chat content to suggest relevant images in real-time when the user inputs their message. MilliCat uses part-of-speech (POS) tagging and dependency parsing to extract relevant keywords (or key phrases) from an input sentence and query external image archives to enrich information delivery within chats. Figure 1 shows examples of image suggestions that can help deliver information quickly and visually. Our image suggestion could be considered as an *auto-complete with images*.

We perform two user studies with a total of 45 participants to evaluate and understand the effect of MilliCat on mobile chat behavior. First, a small-scale controlled lab study is conducted that allows in-depth analysis of how frequently and why users actively use autonomous image suggestions. Second, a longer term in-the-wild user study is run to overcome limited external validity and possible novelty effect of the in-lab user study. Here, participants used MilliCat's real-time image suggestion and manual image search for 8 to 10 days on their own smartphones for all chats with the designated partner(s). Results from our studies suggest that with MilliCat participants used 1.8x more images and the system reduced the image usage delay by 3x than manual search-based image sharing.

2 RELATED WORK

2.1 Diversifying Expressions on Mobile Chats

Mobile chats have diversified their communication support by integrating various forms of visual context to the user discussions. A representative example is the use of emojis in mobile chat and social networking platforms. Emojis assist in the delivery of emotional expressions, and many studies have investigated the effect of emojis in online communications [13, 14, 32, 41, 52, 58]. While new emojis are being continuously designed to represent various objects and events, they primarily focus on representing emotions. Animated GIFs are another popular way of sharing emotions [5, 23]. However, animated GIFs mostly focus on funny memes and might not be supportive enough for expressing diverse chat scenarios. Beyond emojis and animated GIFs, we believe there are a more wide variety of visual context on the Internet that the mobile chat users can utilize in their conversation to effectively deliver information. Furthermore, by contextually analyzing the chat data, we believe that automated suggestions of such diverse resources can be made autonomously without explicit user requests.

2.2 Context Analysis for Chat Suggestions

Recognizing context in computer-mediated communication and providing users with suggested actions has been a subject of steady research. Remembrance Agent [45] keeps track of the user's behavior in e-mail conversations and helps users remember whether they have sent the e-mail or replied to a specific message. Short reply suggestions (e.g, "Yes, it's done", "Sounds fun") for receiving emails [18, 24, 57] have been shown to reduce time in replying.

Instead of utilizing users' past behaviors for prediction, recent research has evolved so that the users' current communication content is analyzed to assist users' social activities. SearchBot [2] is an example of such system that listens to vocal conversations, extracts entities, performs a search, and proactively provides relevant information to the speakers.

In the context of mobile chats, the work by Buschek et al. provides design implications for augmented text messaging based on user context using heterogeneous sensing modalities, which include smartphone usage patterns, heart rate measurements and the smartphone's accelerometer readings [8].

2.3 Visual Support for Mobile Chats

Previous work have taken an additional step from analyzing discussion context for text-based support and have integrated visual aspects to their automated suggestions. meChat [29] analyzes and classifies mobile chat conversations and in-device photos. When users intend to share photos that are relevant to the chat content, the application presents appropriate photos to the chat. MessageOnTab [11] analyzes the content of mobile messaging applications and provides a shortcut interface to third-party applications that users could use (e.g., photo gallery, calendar, and contacts). While MessageOnTab is similar to MilliCat, we focus on recommending public images from the Internet while MessageOnTab suggests external applications for a possible next action. Moreover, meChat and MessageOnTab exploit personal, in-device photos that are triggered on-demand by the users and could be useful for personal and social context. On the other hand, MilliCat uses public images that are suggested autonomously based on the input text and can help improve information that may not be clear in text only and further fact sharing.

While the idea and challenges of image recommendation based on the currently typed word in mobile chat has been recently reported [28], contrary to our work, it does not suggest technical details in image search phrase extraction, word type selection, or dependency parsing that are required in realizing real-time image suggestions in mobile chat. Moreover, it lacks user studies that investigate how image recommendation affects mobile user chat behavior.

A number of mobile chat application products have added features to support using images in the chat. Facebook Messenger [36] and Dango [15], for example, allow users to send trending animated *memes* using an external application (e.g., GIPHY [17]) via keyword-based search. Other applications such as Google GBoard [16] and KakaoTalk [51] support web image search (within the application) using a "trigger" (e.g., a search key or keyword that enables image search). However, these services require an explicit user input request for image search and also outputs other non-image related contents such as restaurant information and location information, which leads to increased mobile device overhead. The goal of MilliCat is to eliminate such explicit user requests and autonomously suggest images in real-time with minimal system overhead.

3 MILLICAT DESIGN

MilliCat recommends images in real-time to alleviate users from inconveniences when using images within a mobile chat. Users also reported preferences in autonomously being suggested images from the Internet, based on the chat input text [28]. MilliCat takes in user's input text to the mobile chat application, analyzes the text to identify suitable words for image suggestion, fetches proper images from the Internet, and presents options that the users can choose from to embed image in their chats. The main design goals of MilliCat are as follows:

- **Real-time suggestion:** Images associated with the current input text should be suggested in real-time without requiring any additional user interaction. This includes the activity of changing applications (e.g., web browser or local image library), and the act of having to click additional buttons to explicitly request images for a word within the chat application. Images should also be presented within a short time frame so that they are suggested within the duration of the conversation topic.
- **Context-awareness:** Images should be suggested only in situations when an image can assist the conversation. Examples of image usage can be information delivery, emotion delivery, or nuance delivery. To satisfy such purposes, the image suggestion results should match the context provided in the chat conversation. The system should quickly analyze the chat message to find the proper keyword that fits the context for real-time image suggestion.
- **Resource efficiency:** Since mobile chat applications should mind the resource limitations of their platform, the communication and energy overhead should be minimized when analyzing text and exchanging data through its wireless connections for the image search and fetch operations.

We designed MilliCat to divide the functionalities between the mobile and the server so that the load on mobile devices is reduced. MilliCat leverages server capabilities when running NLP algorithms [28]. Below we detail our technical solutions for realizing real-time image suggestions in mobile chat.

3.1 Extracting Words for Image Suggestion

A major challenge for autonomous real-time image suggestion is to decide on an appropriate set of keywords for image searching. While topic extraction approaches such as Latent Dirichlet Allocation (LDA) [7] and Named Entity Recognition (NER) [37] are popular, they are not suitable for mobile environments due to their large processing overhead and latency. To overcome this challenge, MilliCat utilizes more light-weight NLP techniques, such as the part-of-speech (POS) tagger [38] and dependency parser [48], which we present in the following.

3.2 Word Type Selection

Performing an image search for every word would result in too many queries, causing not only inefficient use of bandwidth and energy resources, but also unnecessary distractions to the users. To address this challenge, using the NUS chat data [12] and 80 sample smartphone messaging conversations,[1] we examined image usage patterns in mobile chat applications to identify two important findings.

First, we found that the purpose of using images in chat is often information delivery, specifically for describing targeted *nouns*. For example, a user may send the text "I prefer fondue," with an image of fondue to deliver the meaning and visualize the target word. Second, among the nouns, abstract nouns are difficult to express as a single image due to their intangible nature. For example, for an abstract noun word "Catholic," one might vision a cross, a church, a nun, or the bible. On the other hand, concrete nouns are usually tied with a specific image. When we picture an image for "computer," we think of a PC-style or a laptop computer.

Based on these findings, we confine our image suggestion candidates to *concrete nouns*. Results from our pilot study suggest that the probability of a user selecting an image recommendation for a concrete noun (38.6%) was 2.4 times higher than that of an abstract noun (15.9%). Specifically, we use a part-of-speech (POS) tagger [38] to determine whether an input word is a noun. We then combine this with a Wmatrix [1, 43, 44], which semantically analyzes each word with an English semantic tagger and classifies a word into 21 major categories and 238 minor categories. From this categorization, we manually labeled each category as either abstract or concrete. This combination of a POS tagger and Wmatrix allows MilliCat to quickly identify a proper search keyword from an input text prior to the end of the sentence.

[1] These are WhatsApp, Facebook messenger, Android, and iOS chat conversations available on the Internet. Each conversation has an average length of 5.95 lines with each line on average being 175.7 characters (or 42.175 words) long.

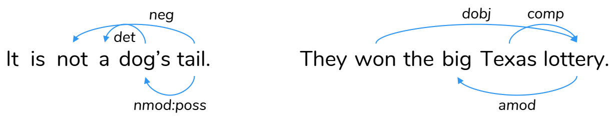


Figure 2: Dependency parsing result for the sentences "It is not a dog's tail." and "They won the big Texas lottery."

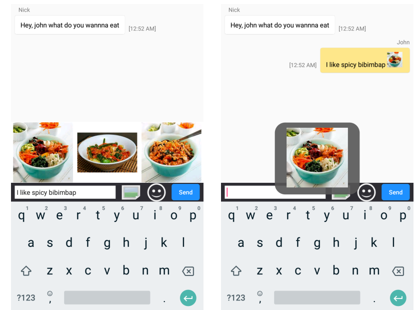
3.3 Dependency Parsing

Using only concrete nouns as the search phrase might not be expressive enough to extract a keyword phrase. For example, if a user inputs *raw* fish or *fresh* meat, instead of searching for fish and meat, adding the adjectives to the search phrase could improve the image suggestion accuracy. For this, MilliCat includes a dependency parser called Enhanced English Universal Dependencies [48] to capture dependency relations between words in a sentence. This dependency parser has shown high accuracy (93.9%) in determining dependencies with various data such as emails, newsgroups, business reviews, questions-and-answers, and web log data.

Figure 2 illustrates where the dependency relations are for two example sentences: "It is not a dog's tail." and "They won the big Texas lottery." The dependency relation is parsed and represented as a directed graph, where a dependency name is assigned to each edge. The parsed results are given as a form of *d*{x, y}, where *d* is the name of dependency from the word x to word y. The dependency relations are: *det*{dog, a}, *neg*{tail, not}, *nmod:poss*{tail, dog}, *dobj*{won, lottery}, *comp*{Texas, lottery}, and *amod*{lottery,big}. Here, *det* is the determiner, *neg* is the negation, *nmod:poss* is the possessive nominal modifier, *dobj* is the direct object, *comp* is the compound, and *amod* is the adjectival modifier.

Note that there are approximately 50 possible dependency relations in the dependency parser. Among them, given that MilliCat extracts search keywords based on *concrete nouns*, this leaves us with 14 dependency relations. We also noticed that some noun-dependencies could be less informative. Take for example the dependency *det* (e.g., 'a' and 'dog' in the sentence "It is not a dog's tail."); searching for "a dog" would not improve the image search accuracy compared with searching for "dog". By eliminating such unnecessary dependencies, the following five types of dependencies remain: compound (comp), negation (neg), adjectival modifier (amod), direct object (dobj), and possessive (nmod:poss) [48].

To evaluate the effectiveness of dependency parsing we conducted an Amazon Mechanical Turk (MTurk) survey as part of a preliminary study. Each participant is given <sentence, image> pairs and asked to rate whether the suggested images match each sentence, ranging from 1 to 5, with a higher score being more appropriate. We used the NUS chat data [12] and 80 sample smartphone messaging conversations described in Section 3.2, and gathered data from 12 acquaintances who agreed to share their chat data. From this data, we selected a total of 25 sentences with five dependency relations equally represented. Two different methods are used for recommending images for concrete nouns; with or without applying dependency relations. For example, for sentence "He loves fried rice", images for 'fried rice' are suggested with the dependency parser and images for 'rice' without it.



(a) MilliCat's real-time image suggestion based on the user's typed text.

(b) When the user clicks on the image, a larger image appears.

Figure 3: Application developed for the study.

Our study results show that the average appropriateness score is 3.91/5 when dependency parsing is applied and 3.43/5 when not applied. Our analysis for different relations indicate that 'amod', 'comp', and 'neg' dependency relations are effective and therefore, MilliCat utilizes only the 'amod', 'compound', and 'neg' dependency relations. When considering only these three relations, the average rating was 4.25 compared to 3.18 when no dependency parsing is used. The reduction in the dependencies that our parser needs to process has a direct relationship with the image suggestion latency and communication overhead. For the three types of dependencies, MilliCat creates image search queries based on the number of valid dependencies it identifies.

4 EVALUATION

4.1 Experiment Overview

To see the effect of using real-time image suggestions in a mobile chat with real users, we conducted two user studies. We first conducted a controlled lab study to compare between MilliCat's real-time image suggestion and a baseline (i.e., without image suggestion). The goal of this study was to understand whether users actively use autonomous image suggestions on mobile chats, how frequently, and for what reasons. While the lab study gives us control in the chat environment and in-depth observational data, short chat sessions in the lab suffer from limited external validity and possible novelty effect. To overcome these limitations, we conducted an additional in-the-wild user study to capture the effect of image suggestion in a more realistic setting. Participants used MilliCat's real-time image suggestion and manual image search for 8 – 10 days on their personal smartphones for all chats with the designated partner(s). This study was designed to understand the benefits and limitations of autonomous image suggestions versus on-demand image search.

We implemented MilliCat as an Android-based mobile chat app for the study (Figure 3). When a user types text ("I like spicy bibimbap"), MilliCat suggests three images in real time. The user can select one of the suggested images to use in the message. Our implementation also includes common mobile chat features such as chat rooms, user nicknames, text-based chat, local gallery image

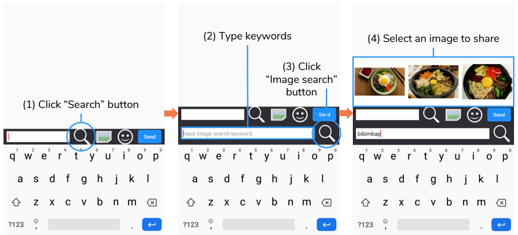


Figure 4: Image-sharing process in the manual search mode.

support, message notification, and emoji support. Our implementation operates on Android versions from 6.0 to 9.0. The MilliCat app also includes a logging feature for all usage behavior including conversations, image usage, and the image suggestion latency to analyze the users' image usage patterns.

In addition to MilliCat, our application includes two other operation modes for supporting comparative interface conditions: (1) *baseline* and (2) *manual search*. In the *baseline* mode, a user can manually select images only from the smartphone's local image gallery. For the baseline where there is no image suggestions, we selected a simple method that did not ask users to install any new app. In the *manual search* mode, users can manually search public images from the Internet through keywords, similar to image search functions offered in existing mobile chat applications [16, 51]. Figure 4 illustrates an example of the image search process using the manual search mode. We included this mode to analyze and compare the effectiveness of autonomous image suggestion of MilliCat against manually searching for images on-demand.

Table 1 compares the procedures of image sharing in different mobile chat apps, the baseline, the manual search, and MilliCat. Most apps that provide an image search feature such as Google GBoard [16] and KakaoTalk [51] have a button/text field to manually trigger the image search. As Table 1 shows, the manual search mode mimics the image search process of existing products and we thus believe it is a valid baseline to compare MilliCat with.

4.2 In-Lab User Study

The purpose of this controlled, in-lab user study is to investigate whether users actively use autonomous image suggestions on mobile chats, how frequently, and for what purposes.

4.2.1 Participants. We recruited 24 (21 males and 3 females; ages: 20-36 years, mean=24.87 years, stdev=4.26) participants by posting advertisements through campus and city-scale online communities. To create a natural mobile chatting environment with close acquaintances, participants were asked to sign up in *groups*. The groups (n=10) consisted of six groups of two members and four groups of three. The self-reported relationships among the groups were friends (n=8), a couple (n=1), and colleagues (n=1). Participants received $13 (in local currency) for their one hour of participation in the experiment.

4.2.2 Tasks and Conditions. To mimic a natural chat experience, we designed two tasks based on common topics in mobile chat scenarios: one is making a plan for a dinner get-together and the other is selecting a movie to watch together at a theatre. Using a within-subjects design, we instructed participants to discuss the two topics, each for 10 minutes. For each group, MilliCat is used for a topic and and the baseline (no image suggestions) is used for the other topic. The task order was fixed while the interface order was randomized. All participants were given a Nexus 5X smartphone with the MilliCat app installed to use during the study.

4.2.3 Procedures. We carried out the following steps in our experiments. After a brief tutorial of the app, participants spent 10 minutes to get familiar with the device and the MilliCat application. We asked them to casually chat with their group members in this phase. In the tutorial, we explained to participants how to use images in a chat. When using MilliCat, we explained how they can embed the suggested images in their chat. When using the baseline mode, we explained how they can send an image from the Internet to their chat by launching a Chrome browser with Google search, performing an image search, downloading the image to the local gallery, and then selecting the image from the local gallery function of the prototype application.

Next, participants were instructed to perform the two 10-minute chat tasks within their groups. We assured them that their chat history would be anonymized and only be used for research purposes as specified by the IRB. Note that participants in the same group were placed in different rooms during the chat sessions so that emotions or additional information could not be delivered using facial/verbal expressions.

After the experiment, participants answered a questionnaire on their real-time image suggestion experience and the effect of using images on mobile chats (c.f., Table 2). We also conducted both 1-on-1 interviews and group interviews. In 1-on-1 interviews, we asked participants to freely elaborate on their answers to the questionnaire.

4.2.4 Image Usage Results. We analyze the system logs, questionnaires, and interview data to understand how MilliCat affects image usage in mobile chats. Each group exchanged an average of 74.6 chat messages during a 10-minute chat task with MilliCat's real-time image suggestions and 88.6 chat messages without real-time image suggestions. The average number of images used per participant during the entire experiment for MilliCat and the baseline was 7.5 (std=6.05). The participant who used the most images included 25 images in the chat while one participant did not use any image.

When using MilliCat's real-time image suggestion feature, a total of 172 images were used. Among them, 169 were selected from the suggested images by MilliCat. For the remaining three images, we asked the reason to the two participants who manually searched for and used images. In two of those cases, the participants wanted a very specific image for the search phrase "sad frog" and "Korean fried chicken" even though MilliCat suggested images similar to what the participants eventually selected. To avoid such manual search, MilliCat could suggest more than three images for each search, possibly with a horizontal scroll for images to not overly cover the chat screen. However, fetching more images would consume more power and storage. As for the third case, a participant used an image of a cat that he downloaded and stored during the introductory tutorial, as the final input to the 10-minute experiment session. This image was out of context to the chat and when we

Table 1: A comparison of image sharing process of existing mobile chat products, baseline, manual search, and MilliCat.

Name	Image sharing process
Google GBoard [16]	(1) Click the "Media" button, (2) Click the "Media Search" button, (3) Type keywords, (4) Click the "Search" button, (5) Select an image to share, and (6) Click the "ABC" button to return to chat.
KakaoTalk [51]	(1) Click the "#" button, (2) Type keywords, (3) Click the "Search" button, (4) Click the image tab in the new screen, (5) Select an image, and (6) Click the "Share" button.
Baseline	(1) Switch from the chat application to a browser, (2) Type keywords, (3) Click the "Search" button, (4) Save the image in local gallery, (5) Switch from the browser to the chat application, (6) Click the "Local Gallery" button, and (7) Select an image to share.
Manual search	(1) Click the "Search" button, (2) Type keywords, (3) Click the "Image search" button, and (4) Select an image to share.
MilliCat	(1) Select an image among autonomously suggested images.

Table 2: Questionnaires: we received responses on a 5-point Likert scale for all questions (1: strongly disagree, 5: strongly agree).

Evaluation of real-time image suggestions	AVG (SD)	Median
1. Image suggestions allowed me to use more images than without image suggestions.	4.41 (0.88)	5.00
2. Image suggestions were made at appropriate moments.	3.45 (1.14)	3.00
3. Image suggestions were made when I wanted to use images.	3.45 (1.10)	4.00
4. A right set of images were suggested.	3.87 (0.94)	4.00
5. The latency of image suggestions was acceptable.	3.95 (0.69)	4.00
6. I am willing to use the image suggestion feature in my smartphone chat applications.	4.08 (1.17)	4.00
Senders' experience using images during chat	AVG (SD)	Median
7. Images I used helped me effectively deliver information.	3.95 (0.90)	4.00
8. Images I used helped me effectively express my emotion.	3.50 (1.17)	4.00
9. Sending images in a chat was fun.	4.79 (0.41)	5.00
Receivers' experience using images during chat	AVG (SD)	Median
10. Images used by the chat partners helped me understand the conveyed information.	4.33 (0.81)	4.00
11. Images used by the chat partners helped me understand the conveyed emotion.	3.87 (1.11)	4.00
12. Receiving images in a chat was fun.	4.66 (0.48)	5.00

asked the participant for the reason of using this image, he said it was with no specific reason.

When using the baseline system on the other hand, participants searched, downloaded, and used a total of only eight images from the Internet. On a per-chat basis, with MilliCat, participants used an average of 7.16 images (std=5.15). MilliCat has increased the use of images by 21× compared to the baseline. These results show that real-time image suggestion can significantly lead to more frequent use of images in mobile chats. Most participants did not use images with the baseline method (i.e., the data was zero-inflated). We thus used a non-parametric comparison method, the Wilcoxon signed-rank test, and the results showed a significant difference (z=4.20, $p < 0.001$).

In the post-interview, participants mentioned that the reduced complexity of retrieving images (a simple click for selecting one from the suggested images) encouraged them to use more images. Some mobile chat app products have embedded image search features within the app (e.g., GIPHY, Kakao Talk #). Five participants explicitly compared MilliCat with these features and mentioned that MilliCat has advantages over these apps. P_{lab}^{18} (P_{lab}^{18} means the 18th participant from the in-lab user study) mentioned: *"With existing apps, to share an image to express my emotions, I must manually search for the image and select an image from the search results. But with the automated image suggestions, I could just click on a suggested image!"* P_{lab}^{12} also said: *"Image suggestion is definitely more convenient than the Kakao Talk feature as I don't actually have to search and look for images."*

We noticed that participants who do not enjoy typing long messages found MilliCat to be useful. P_{lab}^{12} mentioned: *"I think image suggestion is very good. Personally I am a lazy texter. So instead of texting long sentences, I just type in a few words and send images to keep the conversation going. I don't like using emojis or animated GIFs because I need to go and search for them."*

Participants reported that using MilliCat was an enjoyable experience. In some cases, several participants even used images that did not match the conversation context. P_{lab}^{18} mentioned: *"Some images I selected are out of context. I just found it fun to use in my chat."* P_{lab}^{1} noted: *"I liked using MilliCat. It's so simple; you just click on it and it happens, so I'm like, why not?"*

4.2.5 *Purposes of Using Images.*

During the post-interviews, we asked participants their intention of using every image during the study session. Based on participants' answers, the lead author identified five categories of purpose of image usage through open coding and the other authors verified. The categories are as follows: (1)

Emphasis, (2) Information delivery, (3) Fun, (4) Emotion expression, and (5) miscellaneous.

About 32% of used images were expression of emphasis. Nine participants reported using images to emphasize their opinions or persuade their partners. P^{22}_{lab} said: *"I thought the image would emphasize better than words. Instead of only the text of 'Game of Thrones' I also sent the image because I wanted to stress it. I really wanted to watch it."* Five participants tried to persuade their partners using images. P^{5}_{lab} said: *"I suggested Japanese ramen but my partner wanted sushi. I wanted to persuade him to go for ramen instead, so having that image helped me. I was trying to convince him: Look at this ramen. It looks so yummy!"*

About 26% of used images were expression of information delivery. Participants used images to convey information unknown to their chat partners (e.g., movie stars, movie information, food, etc.). Twelve participants said that using text to describe something their partners do not know requires a lot of words, effort, and time. But it was fast and convenient for them to explain with images. P^{1}_{lab} said: *"With words, my friend does not understand because he doesn't know it. But if I use an image with my words, he can understand what I'm trying to say."* In addition, nine participants thought that text could be misinterpreted when describing information they wanted to share and an image was a better choice in conveying the correct information. P^{21}_{lab} said: *"For images, it's quick and no need to type. But texts could have many meanings. I get so many messages from friends. Many times they have a lot of typing errors. It could be a different meaning. But images are more exact than texts."*

About 19% of used images were simply for fun. One group used many recommended images when entering the name of each other. P^{5}_{lab} said: *"I was just messing around with the image feature. The suggested images were sometimes funny and I think that is pretty fun with the image related to my chat partner."* In addition, although recommended images sometimes did not perfectly match the current chat context, participants reported it was still fun to use them in chat. P^{18}_{lab} said: *"We're like, 'Okay, let's get ramen.' I said then, 'Do you know any good place?' The suggested picture then was a screenshot of 'The Good Place' from the TV show. I thought that's funny and I just used that photo. We enjoyed it. It was kind of a joke that the word has the same meanings as another."*

About 14% of used images were expression of emotion. We also had an interesting use case where two participants used images of emoji, which are recommended by typing "thinking emoji" and "sad emoji". They said it was more convenient to use automatically recommend emojis than to find emojis from the emoji keyboard. Three participants used images of words such as "clap" for delivery of agreement.

4.3 In-the-Wild User Study

The user study in Section 4.2 was performed in a controlled setting in short duration and in a small scale to understand how image suggestions affect mobile chat usage behavior and deeply analyze the motivation of participants in using images in chat. We now perform an additional user study, this time in an "in-the-wild" setting for a 8 to 10-day period. The participants installed our chat application on their personal Android smartphone. The goal of this user study is to evaluate the impact of MilliCat on a more natural chat environment

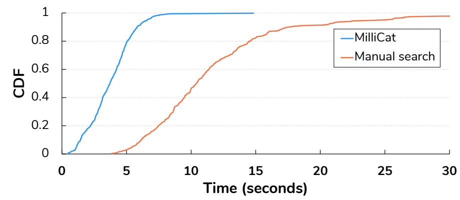


Figure 5: CDF of image usage process latency for MilliCat and manual image search.

on a longer term beyond the novelty effect. Moreover, we compare the usage of MilliCat and that of "manual search" (described in Section 4.1). This comparison could give insights into the benefits and drawbacks of autonomous image suggestion against on-demand search.

4.3.1 Participants. In this study in the wild, the participants were asked to use the MilliCat app on their personal smartphones for 8 to 10 days. We recruited 21 (15 males and 6 females; ages: 20-33 years, mean=23.76 years, stdev=3.51) participants by posting advertisements through similar channels to those mentioned in Section 4.2. A total of 10 groups participated, consisting of nine groups of two and one group of three. The groups consisted of nine friends and one couple. All participants are Android users and had an LTE connection with enough mobile data and Wi-Fi connections for the study. Participants received $63 (in local currency) for their participation.

4.3.2 Conditions & Procedures. We installed a mobile chat app on the Android smartphones owned by each participant and asked them to chat using the provided application when chatting with members of the group during the experiment period. We used a within-subjects experiment design, in which the half of the study period was for chats using MilliCat, and the other half using the manual search. The order of the image mode was counterbalanced across groups. Experiments for four groups lasted for ten days and six groups lasted for eight days.

Before the experiment, we walked the participants through a simple tutorial, in which we explained how to use MilliCat's image suggestion, the manual image search, and the prototype chat application. At the end of each group's experiment, an interview session was held. In the interview, participants answered a questionnaire on their experiences in using the two different image modes and the effect of using images in mobile chats (Table 3). We log-transformed the data and verified its normality. As Q1/3/8 satisfied normality, we applied a paired t-test for them and the Wilcoxon signed-rank test for the others (Q2/4/5/6/7). The result showed a significant difference in Q1 ($p < 0.001$), Q2 (p=0.001), Q5 (p=0.002), and Q6 (p=0.002). We also conducted 1-on-1 interviews so that participants could freely elaborate on their answers to the questionnaire and experiences.

4.3.3 Image Usage Latency. We analyzed how fast participants could select and use images with MilliCat compared to manually searching for the images. We define the delay as the following. For MilliCat, we measure the time between an autonomous image request (to the MilliCat server) and the time that the user selects an image to display on the chat. For the manual search, it is the

Table 3: Questionnaires for the in-the-wild user study: we received responses on a 5-point Likert scale for all questions (1: strongly disagree, 5: strongly agree).

Questions	MilliCat AVG (SD)	Manual search AVG (SD)
1. The image mode allowed me to use images quickly.	4.55 (0.60)	2.90 (0.78)
2. The process of using images for a mobile chat was simple.	4.50 (0.51)	3.35 (1.03)
3. I could use image at appropriate moments in mobile chat.	3.70 (1.17)	4.00 (0.85)
4. I could use appropriate image I wanted to use for mobile chat.	3.35 (0.87)	3.70 (0.65)
5. It was fun using images in a chat.	4.70 (0.47)	3.75 (0.96)
6. I used more images than my usual mobile chat.	4.30 (1.21)	3.50 (1.31)
7. The latency of loading images was acceptable.	3.60 (0.88)	3.65 (0.74)
8. I'm willing to use the image mode in my favorite mobile chat application.	3.95 (1.23)	3.45 (1.00)

time between when a participant types in the search keyword and when she selects a desired image. This roughly represents the steps introduced in Table 1. In case when multiple images were selected for the same search, we only measure the delay for the first image. As a result, we compute the image usage delay for 268 images for the manual search (88.16% of all images used in this mode).

Figure 5 plots the Cumulative Distribution Function (CDF) of the image usage delay for MilliCat and manual search. The median time is 3.72 sec for the MilliCat and 10.42 sec for manual search while the mean is 3.76 sec (stdev = 1.81 sec) with MilliCat and 11.97 sec (stdev=7.16 sec) for manual search. Due to the non-normality of the data, we used the Wilcoxon signed-rank test, and the result showed a significant difference in latency between the two modes (z=21.98, $p < 0.001$). Overall, the result suggests that MilliCat significantly reduces the image usage process latency from the manual search that mimics the current chat applications (3.19× improvement).

For manual search, an average of 8.65 sec (stdev=6.89) was spent on keyword input. With MilliCat's autonomous image suggestion, no such process is needed and thus users can quickly use images in a chat. Seven participants noted in the interviews that using images without typing keywords was convenient and fast. P^{14}_{wild} said, *"It was more comfortable with MilliCat. I can quickly give information on what I want to say. I'm satisfied with being able to use images quickly without typing search keywords."* We also notice from Table 3 that most participants responded that images can be used quickly using MilliCat, and the process of using the image was simpler than the manual search (Q1: 4.55 vs 2.90 and Q2: 4.50 vs 3.35).

4.3.4 Image Usage Frequency. A total of 848 images were used throughout the experiments, of which 544 were used with MilliCat and 304 with manual search.[2] That is, with MilliCat, participants used 1.79× more images than with manual search. In this experiment, 2,096 messages were exchanged while using MilliCat and 2,936 messages during the manual search mode. This suggests that on average, participants used an image for every 3.85 messages with MilliCat and for every 9.66 messages with manual search (2.51× improvement). We log-transformed the data, verified the normality, and applied a paired t-test. The t-test results show a significant difference in the image usage between the two modes (t(20)=3.55, p=0.002).

[2]In addition, 34 images from the local gallery were used; 16 while using MilliCat and 18 while using manual search.

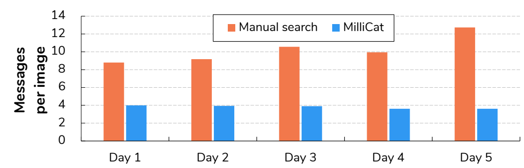


Figure 6: The number of messages sent per image for MilliCat and manual image search.

Analyzing questionnaire responses and interview data, we analyzed the reasons behind increased image usage with MilliCat. Responses for Q6 in Table 3 suggest that the participants generally agreed that they used more images with MilliCat. From the interviews, participants mentioned that the reduced complexity of retrieving images encouraged them to use more images. P^{4}_{wild} said, *"In the manual search mode, it has additional steps to find an image; we must click the search button and input the keyword. It was more distressing than with MilliCat."* P^{6}_{wild} mentioned, *"MilliCat gives more freedom. If I don't like the suggested image, I just continue with my conversation; I can seamlessly chat. But the manual image search process adds burden and increases complexity and time in using images."*

Participants also noted that even when they initially had no intention to use images, autonomous suggestion of MilliCat encouraged them to use images when appropriate images were recommended. P^{8}_{wild} said, *"Sometimes I used images even when I didn't have intention because images just pop up and some are eye-catching."* P^{5}_{wild} said, *"When chatting with MilliCat, it suggested some images related to emotion-expression, like a meme. For me, it was pretty fun to use these images. It was an interesting experience because I do not usually search for images related to my emotions. I usually search images for a specific noun. It was definitely fun being recommended and using emotion-explainable images in my chat."*

With the fun factor in image usage, one might wonder the impact of novelty effect in this study. Our analysis shows that participants used more images in the fist two days of experiments for each image mode (7.1 images per day per participant with MilliCat and 4.1 with manual search) than the rest of the experiment period (4.89 images per day per participant with MilliCat and 2.48 with manual search), but they also sent more messages in the first two days (28.14 messages per day per participant with MilliCat and 36.74 with manual search) than the rest (18.14 messages with MilliCat and 27.04 with manual search). Figure 6 shows the number of message sent per image for each image mode for each experiment day. We observe relative consistency in this metric. We conclude that although there is immediate novelty effect in image (and app) usage in the first two days of each image mode, there was no strong novelty effect that favors MilliCat.

Of 21 participants, 19 used more images with MilliCat. P^{10}_{wild} and P^{12}_{wild} used more images with manual search. Both participants preferred finding the exact image they wanted to use through a manual search. P^{12}_{wild} noted that *"I use KakaoTalk's image search feature frequently. The manual search feature implemented in the app was more easy to use than KakaoTalk's image search."* P^{12}_{wild} added that a manual search is preferred over MilliCat because typing the search keyword is a more natural behavior.

4.3.5 Image Selection Ratio. Finally, we analyze how often participants used the images that were suggested (MilliCat) or searched for (manual search) in the chat. We believe this image selection ratio is related to the responses for Q3 and Q4 of Table 3. As participants actively search for images when they intend to use images in manual search, we expect it to have a noticeably higher selection ratio than MilliCat. With MilliCat, a set of three images was suggested for a total of 1,450 times. Of those, users selected images (among the suggested three) for 544 times, which yields 37.51% selection rate. We note that the image selection rate was similar in the in-lab user study (36.34%). For manual search, participants performed 596 image searches, of which 304 were used and thus 51.01% selection rate.

This selection rate of the manual search condition was surprisingly low given that even when participants actively performed the image search with the intention to use one, the actual selection of the images happened only half of the time. Participants noted that they did not use the image because (i) they were not satisfied with the search results or (ii) the conversation progressed too fast and using the image at that point was meaningless. The second reason indicates the potential usefulness of real-time autonomous image recommendation in mobile chat. P^{16}_{wild} mentioned: *"I tried searching for 'happy face' with manual search, but it took time to input a keyword and click the search button. After the images showed up, when I checked the chat thread, my chat partner had already moved on to another topic."* P^{18}_{wild} said: *"To perform an image search, I must click the search button to enable the search window, input the keyboard, and then click the search button to retrieve images. When an appropriate image is not given, I have to erase the search phrase and start over until I find the image I want. Sometimes I just decided to skip using images."*

The image search process in manual search requires more user effort than in MilliCat. While MilliCat's autonomous image suggestion has a lower selection ratio than manual search (MilliCat: 37.51% vs Image Search: 51.01%.), there is a smaller price to pay in terms of user effort when the images are not selected.

5 DISCUSSION

We discuss some of the key aspects of introducing autonomous image suggestion in mobile chat environments: what using more images means, how image suggestions could fail and improve, how autonomous suggestion compares against on-demand search, and what privacy issues need to be addressed.

5.1 Implication of Using More Images in Mobile Chats

Many studies and products are evolving towards the use of various visual media in mobile chats. This suggests that users are wanting to use more visual elements, similar to how emoticons were initially used in previously text-only communications. Previous research also suggests additional use of visual elements such as images and videos in mobile chats has assisted the chat experience by displaying intimacy and conveying previously difficult information [3, 4, 6, 47, 53, 55]. Our study takes a further step in amplifying such benefits by proposing a new way to exploit images in a mobile chat.

The results from our two user studies show that the use of real-time image suggestion led to an increase in image usage for mobile chats. Furthermore, the interviews conducted in our study show how the increased use of images affected the user's chat experiences.

Firstly, we noticed that images can be effective in emphasizing one's opinion with increased persuasiveness. In particular, images were effectively used when performing tasks such as selecting a menu for making dinner plans, in which an image can help express and evaluate one's opinion more effectively.

Secondly, we noticed that the use of more images adds an additional "fun" factor to the mobile chat experience. Many participants answered that the chat became more fun by using images, by diversifying visual elements in the chat. Some participants even reported that using images for expressing emotions was even more fun compared to using emojis in a chat, which were designed for that specific purpose. Many noted that they were excited to see a variety of new images that corresponded to the emotions they wanted to deliver. This is true given that while the number of emojis are fixed, image suggestion systems can offer more trending images that express a target emotion.

Finally, using images makes it easy to convey information. Twelve participants from the in-lab user study mentioned that using text to describe something their partners do not know requires a lot of words, time, and effort. However, it was fast and easy to explain the keyword/expression with images. P^{1}_{lab} said: *"With words, my friend did not understand. But by using an image with text, my chat partner could understand what I was trying to say."* P^{3}_{lab} used an image of a movie to convey that the new movie was upcoming and that they should plan for a movie night. These results suggest that the use of images in mobile chats possess a number of positive effects in assisting text-based chat communication using MilliCat.

5.2 Better Image Suggestion Timing

Image recommendations could diminish chat user experience when (i) images are unnecessarily suggested even when users do not want to include images and when (ii) images are not suggested even though the user wants to use an image. In case (i), MilliCat wastes network bandwidth, smartphone energy, and even pixel space. Two of the study participants mentioned that suggestions were occurring too frequently and in some cases distracted their chat conversations. A possible solution would be to use a finer level of word categorization within concrete nouns to improve the key-phrase extraction. Giving users more control in suggestion frequency is another possible solution.

For case (ii), such false negatives occur mainly due to two reasons. First, typos occur very frequently in mobile chats. Current MilliCat implementation does not recognize typos as a proper noun and does not suggest images. Such issues could be resolved by the use of a typo auto-correction engine [20, 21, 25]. Nevertheless, this was a limitation of our current implementation and the study participants experienced some difficulties from this limitation. Second, the use of proper nouns and abbreviations also cause false negatives. In our chat traces, there was a case when participants were discussing on going to a "Marvel" film. Since the word "marvel" is considered an abstract noun, MilliCat did not suggest images. Furthermore, abbreviations such as "LOL" and "OMG" were not successfully recognized by MilliCat. As these types of words are often used in mobile chats, we must maintain an up-to-date list of proper nouns and abbreviation words that are commonly used.

5.3 Autonomous Suggestion vs On-Demand Search

Commercially available mobile chat apps such as Facebook Messenger [36], Dango [15], Google GBoard [16], and KakaoTalk [51] use a manual trigger to perform an on-demand search for memes or images to be integrated into chats. This process requires explicit intent, and our user study results show that having autonomous image suggestions as in MilliCat encourages the use of relatively more image contents. The two approaches have their advantages and disadvantages on both system and usage perspectives. While autonomous suggestion does not require user intervention, an inaccurate understanding of the user's intention can lead to waste in the limited resources of the mobile platform and could also be a distraction to an ongoing chat. On the other hand, on-demand searching would lead to added manual overhead for the user.

Finding midpoints between the two can be a good way to use images. For example, an image recommendation button could be added, where clicking on the button displays recommended images based on the user's chat content. This system would not require the manual typing step of the on-demand image search approach and could also solve the timing problem of autonomous image suggestions. Another approach would be adding a "receiver-initiated" image suggestion feature where a message receiver clicks the image recommendation button when she can't understand the message.

5.4 Public vs Personal Images and User Privacy

MilliCat is designed to utilize public images autonomously within a mobile chat. This allows the system to access a variety of images available on the Internet. We discovered in our user studies that several participants did not prefer the chat app accessing locally stored images. Their main concern was related to privacy: that the chat app having access to local images would mean it will analyze and understand all photos taken by the user. We believe MilliCat's design choice of using only public images could alleviate the privacy concern. Apart from accessing local images, there are still concerns around the fact that the text input from the users is sent to an external server for an image search, but our user study participants indicated that this was less concerning than an app trying to analyze the context of images in the local photo library, regardless of their potential usage. Nevertheless, future work should carefully address privacy issues of the app using the user's chat text. One solution would be to give users more control over when external search occurs.

5.5 Generalizability and Limitations

We conducted in-lab and in-the-wild user studies with a total of 45 participants to evaluate MilliCat. We acknowledge three major limitations of our participant demographics: it is (i) male-dominant (male: 36, female: 9), (ii) skewed towards a tech-savvy age group (avg age: 24.36, std: 3.93), and (iii) they have close personal relationships (mostly friends and couples). Fortunately, deeper analysis on a gender perspective suggests that the results do not differ noticeably for different gender groups. However, given institutional characteristics, we were not able to perform a study with a more diverse age population, and the per-age group analysis with the current population is unfortunately not meaningful. As part of future work, we plan to expand our study to a more diverse age group. Nevertheless we conjecture that, based on the feedback from our current user studies, the fact that MilliCat does not require any additional user intervention will induce minimal usability issues and the benefits raised by our current study population will hold for a less tech-savvy population as well. Lastly, our user studies were mainly performed in casual and friendly chat situations. Our participants mentioned in the interviews that for chats with close partners, they prefer autonomous image suggestions over manually searching. However, in other relationships such as customers, business partners, supervisors, results might differ. P^9_{wild} mentioned: *"I had fun chatting with my friends via image suggestions, and I used a lot of images. But for chats in which I talk with more formality, I would prefer image search over image suggestion. I don't want to send many images or memes during important discussions or send many images to my boss, for example."* Some participants also said that image suggestions occupy much screen space and could be distracting to the ongoing chat. How to present recommended images and maintaining/increasing user chat experience is an interesting direction for interface design. Some also complained about occasional poor image resolution. This issue can be handled by checking the resolution of candidate images while also considering the image size to balance efficient use of resources.

6 CONCLUSION

In this paper, we ask "would autonomously suggesting images for mobile chat in real-time improve users' chat experience?". With MilliCat, we propose an alternative way for mobile users to utilize images from the Internet in their chat by recommending the right images at the right time. To achieve this, we used a combination of concrete noun detection, POS tagging, and dependency parsing to analyze the chat content and extract relevant key phrases. Our user studies with 45 participants showed that MilliCat reduces image usage process latency by $3.19\times$ than existing chat applications, and in consequence, participants used images $1.8\times$ more. We discovered that users enjoyed using images in mobile chats for emphasizing their opinion, delivering information, and simply for fun. We believe our work shows potential for designing more visual, enjoyable, effective, and media-rich chat experience through real-time, content-adaptive recommendations.

ACKNOWLEDGMENTS

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2020R1A2C1004062), Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C4A7083534), and the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT (2018R1C1B6003869). The authors thank Alvin Chiang, Youngkyu Hong, Young Seok Kim, and Chia-Wei Wu for their early contribution.

REFERENCES

[1] Paul Rayson A, Dawn Archer B, Scott Piao B, and Tony Mcenery B. 2004. The UCREL Semantic Analysis System. In *In proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*. European Language Resources Association (ELRA), Paris, France, 7–12.

[2] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. 2018. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 1295–1307. https://doi.org/10.1145/3196709.3196734

[3] Ron Baecker. 2003. A Principled Design for Scalable Internet Visual Communications with Rich Media, Interactivity, and Structured Archives. In *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative Research (CASCON '03)*. IBM Press, Toronto, Ontario, Canada, 16–29.

[4] Craig Baehr. 2012. Incorporating user appropriation, media richness, and collaborative knowledge sharing into blended e-learning training tutorial. *IEEE Transactions on Professional Communication* 55, 2 (2012), 175–184.

[5] Saeideh Bakhshi, David A. Shamma, Lyndon Kennedy, Yale Song, Paloma de Juan, and Joseph 'Jofish' Kaye. 2016. Fast, Cheap, and Good: Why Animated GIFs Engage Us. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 575–586. https://doi.org/10.1145/2858036.2858532

[6] Frank Biocca. 2000. New media technology and youth: Trends in the evolution of new media. *Journal of Adolescent Health* 27, 2 (2000), 22–29.

[7] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[8] Daniel Buschek, Mariam Hassib, and Florian Alt. 2018. Personal Mobile Messaging in Context: Chat Augmentations for Expressiveness and Awareness. *ACM Trans. Comput.-Hum. Interact.* 25, 4, Article 23 (Aug. 2018), 33 pages. https://doi.org/10.1145/3201404

[9] Shana K Carpenter and Kellie M Olson. 2012. Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of experimental psychology: learning, memory, and cognition* 38, 1 (2012), 92.

[10] David Chen and Chengzheng Sun. 2002. Real-time text chat via collaborative editing systems. In *Proceedings. of the ACM CSCW 2002 Workshop on Collaborative Editing*. ACM, New York, NY, USA.

[11] Fanglin Chen, Kewei Xia, Karan Dhabalia, and Jason I. Hong. 2019. MessageOnTap: A Suggestive Interface to Facilitate Messaging-Related Tasks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300805

[12] Tao Chen and Min-Yen Kan. 2013. Creating a live, public short message service corpus: the NUS SMS corpus. *Language Resources and Evaluation* 47, 2 (2013), 299–335.

[13] Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. 2018. Through a Gender Lens: Learning Usage Patterns of Emojis from Large-Scale Android Users. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 763–772. https://doi.org/10.1145/3178876.3186157

[14] Henriette Cramer, Paloma de Juan, and Joel Tetreault. 2016. Sender-intended Functions of Emojis in US Messaging. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. ACM, New York, NY, USA, 504–509. https://doi.org/10.1145/2935334.2935370

[15] Dango. 2020. Dango. https://getdango.com/.

[16] Google Gboard. 2020. Google Gboard. https://play.google.com/store/apps/details?id=com.google.android.inputmethod.latin.

[17] GIPHY. 2020. GIPHY. http://giphy.com/.

[18] Matthew Henderson, Rami Al-Rfou, Brian Strope, Yun hsuan Sung, Laszlo Lukacs, Ruiqi Guo, Sanjiv Kumar, Balint Miklos, and Ray Kurzweil. 2017. Efficient Natural Language Response Suggestion for Smart Reply. arXiv:1705.00652 [cs.CL]

[19] Peter S Houts, Cecilia C Doak, Leonard G Doak, and Matthew J Loscalzo. 2006. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient education and counseling* 61, 2 (2006), 173–190.

[20] Y. Huang, Y. L. Murphey, and Y. Ge. 2013. Automotive diagnosis typo correction using domain knowledge and machine learning. In *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. IEEE, New York, NY, USA, 267–274.

[21] Yinghao Huang, Yi Lu Murphey, and Yao Ge. 2015. Intelligent typo correction for text mining through machine learning. *International Journal of Knowledge Engineering and Data Mining* 3, 2 (2015), 115–142.

[22] Jialun "Aaron" Jiang, Jed R. Brubaker, and Casey Fiesler. 2017. Understanding Diverse Interpretations of Animated GIFs. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. ACM, New York, NY, USA, 1726–1732. https://doi.org/10.1145/3027063.3053139

[23] Jialun "Aaron" Jiang, Casey Fiesler, and Jed R. Brubaker. 2018. 'The Perfect One': Understanding Communication Practices and Challenges with Animated GIFs. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 80 (Nov. 2018), 20 pages. https://doi.org/10.1145/3274349

[24] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, and Vivek Ramavajjala. 2016. Smart Reply: Automated Response Suggestion for Email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 955–964. https://doi.org/10.1145/2939672.2939801

[25] Mark D Kernighan, Kenneth W Church, and William A Gale. 1990. A spelling correction program based on a noisy channel model. In *Proceedings of the 13th conference on Computational linguistics-Volume 2*. Association for Computational Linguistics, New York, NY, USA, 205–210.

[26] Sujay Khandekar, Joseph Higg, Yuanzhe Bian, Chae Won Ryu, Jerry O. Talton Iii, and Ranjitha Kumar. 2019. Opico: A Study of Emoji-first Communication in a Mobile Social App. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) *(WWW '19)*. ACM, New York, NY, USA, 450–458. https://doi.org/10.1145/3308560.3316547

[27] Joongyum Kim, Taesik Gong, Bogoan Kim, Jaeyeon Park, Woojeong Kim, Evey Huang, Kyungsik Han, Juho Kim, Jeonggil Ko, and Sung-Ju Lee. 2020. No More One Liners: Bringing Context into Emoji Recommendations. *Trans. Soc. Comput.* 3, 2, Article 9 (April 2020), 25 pages. https://doi.org/10.1145/3373146

[28] Joon-Gyum Kim, Chia-Wei Wu, Alvin Chiang, JeongGil Ko, and Sung-Ju Lee. 2016. A Picture is Worth a Thousand Words: Improving Mobile Messaging with Real-Time Autonomous Image Suggestion. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications* (St. Augustine, Florida, USA) *(HotMobile '16)*. Association for Computing Machinery, New York, NY, USA, 51–56. https://doi.org/10.1145/2873587.2873602

[29] K. Kim, W. Ryu, J. Park, and S. Lee. 2019. meChat: In-Device Personal Assistant for Conversational Photo Sharing. *IEEE Internet Computing* 23, 2 (March 2019), 23–30. https://doi.org/10.1109/MIC.2018.2883059

[30] Philippe Kimura-Thollander and Neha Kumar. 2019. Examining the Global Language of Emojis: Designing for Cultural Representation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 495, 14 pages. https://doi.org/10.1145/3290605.3300725

[31] Miki Liu, Austin Wong, Ruhi Pudipeddi, Betty Hou, David Wang, and Gary Hsieh. 2018. ReactionBot: Exploring the Effects of Expression-Triggered Emoji in Text Messages. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 110 (Nov. 2018), 16 pages. https://doi.org/10.1145/3274379

[32] Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the Ubiquitous Language: An Empirical Analysis of Emoji Usage of Smartphone Users. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. ACM, New York, NY, USA, 770–780. https://doi.org/10.1145/2971648.2971724

[33] Richard E Mayer. 2003. The promise of multimedia learning: using the same instructional design methods across different media. *Learning and instruction* 13, 2 (2003), 125–139.

[34] Richard E Mayer and Roxana Moreno. 2002. Aids to computer-based multimedia learning. *Learning and instruction* 12, 1 (2002), 107–119.

[35] Qiaozhu Mei. 2019. Decoding the New World Language: Analyzing the Popularity, Roles, and Utility of Emojis. In *Companion Proceedings of The 2019 World Wide*

Web Conference (San Francisco, USA) *(WWW '19)*. ACM, New York, NY, USA, 417–418. https://doi.org/10.1145/3308560.3316541
[36] Facebook Messenger. 2020. Facebook Messenger. https://www.messenger.com/.
[37] David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes* 30, 1 (2007), 3–26.
[38] Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In *Proc. of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*. Association for Computational Linguistics, New York, NY, USA, 380–390.
[39] Larry Percy and John R Rossiter. 1997. *A theory-based approach to pretesting advertising*. Psychology Press, Mortimer Street, London. 267–281 pages.
[40] Matthew Pittman and Brandon Reich. 2016. Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior* 62 (2016), 155–167.
[41] Henning Pohl, Christian Domin, and Michael Rohs. 2017. Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication. *ACM Trans. Comput.-Hum. Interact.* 24, 1, Article 6 (March 2017), 42 pages. https://doi.org/10.1145/3039685
[42] Anabel Quan-Haase and Alyson L Young. 2010. Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of Science, Technology & Society* 30, 5 (2010), 350–361.
[43] Paul Rayson. 2008. From key words to key semantic domains. *International journal of corpus linguistics* 13, 4 (2008), 519–549.
[44] Paul Rayson, Roger Garside, and Pete Sawyer. 2000. Assisting Requirements Engineering with Semantic Document Analysis. In *Content-Based Multimedia Information Access - Volume 2* (Paris, France) *(RIAO '00)*. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, Paris, FRA, 1363–1371.
[45] Bradley Rhodes and Thad Starner. 1996. Remembrance Agent: A continuously running automated information retrieval system. In *The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology*. AAAI Technical Report SS-96-02, Palo Alto, California, 487–495.
[46] Ladislao Salmerón, Thierry Baccino, Jose J Cañas, Rafael I Madrid, and Inmaculada Fajardo. 2009. Do graphical overviews facilitate or hinder comprehension in hypertext? *Computers & Education* 53, 4 (2009), 1308–1319.
[47] Andreas Schmeil, Martin Eppler, and Mattia Gubler. 2009. An Experimental Comparison of 3D Virtual Environments and Text Chat as Collaboration Tools. *Electronic Journal of Knowledge Management* 7, 5 (2009), 637–646.
[48] Sebastian Schuster and Christopher D Manning. 2016. Enhanced english universal dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. European Language Resources Association (ELRA), Paris, France, 2371–2378.
[49] Aaron Smith. 2015. U.S. Smartphone Use in 2015. PEW Research Center Report. Available at: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/.
[50] S Shyam Sundar. 2008. *The MAIN model: A heuristic approach to understanding technology effects on credibility*. MacArthur Foundation Digital Media and Learning Initiative, Chicago, IL, USA. 73–100 pages.
[51] Kakao Talk. 2020. Kakao Talk. http://www.kakao.com/talk/en.
[52] Garreth W. Tigwell and David R. Flatla. 2016. Oh That's What You Meant!: Reducing Emoji Misunderstanding. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Florence, Italy) *(MobileHCI '16)*. ACM, New York, NY, USA, 859–866. https://doi.org/10.1145/2957265.2961844
[53] Tomoko Watanabe Traphagan, Yueh-hui Vanessa Chiang, Hyeseung Maria Chang, Benjaporn Wattanawaha, Haekyung Lee, Michael Charles Mayrath, Jeongwon Woo, Hyo-Jin Yoon, Min Jung Jee, and Paul E Resta. 2010. Cognitive, social and teaching presence in a virtual world and a text chat. *Computers & Education* 55, 3 (2010), 923–936.
[54] M. Uther. 2002. Mobile Internet usability: what can 'mobile learning' learn from the past?. In *Proceedings. IEEE International Workshop on Wireless and Mobile Technologies in Education*. IEEE, New York, NY, USA, 174–176.
[55] Zheng Wang, Prabu David, Jatin Srivastava, Stacie Powers, Christine Brady, Jonathan D'Angelo, and Jennifer Moreland. 2012. Behavioral performance and visual attention in communication multitasking: A comparison between instant messaging and online voice chat. *Computers in Human Behavior* 28, 3 (2012), 968–975.
[56] Sarah Wiseman and Sandy J. J. Gould. 2018. Repurposing Emoji for Personalised Communication: Why (PIZZA_EMOJI) Means "I Love You". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 152, 10 pages. https://doi.org/10.1145/3173574.3173726
[57] Ning Ye, Ariel Fuxman, Vivek Ramavajjala, Sergey Nazarov, J. McGregor, and Sujith Ravi. 2018. PhotoReply: Automatically Suggesting Conversational Responses to Photos. In *WWW '18: Proceedings of the 2018 World Wide Web Conference*. ACM, New York, NY, USA, 1893–1899. https://doi.org/10.1145/3178876.3186164
[58] Peijun Zhao, Jia Jia, Yongsheng An, Jie Liang, Lexing Xie, and Jiebo Luo. 2018. Analyzing and Predicting Emoji Usages in Social Media. In *Companion Proceedings of the The Web Conference 2018* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 327–334. https://doi.org/10.1145/3184558.3186344
[59] Rui Zhou, Jasmine Hentschel, and Neha Kumar. 2017. Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 748–759. https://doi.org/10.1145/3025453.3025800