

IBM Research Report

Design of Multimedia Instant Messaging System with Annotation

Zon-Yin Shae, Xiping Wang
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

Ferdinand Hendriks
Hitachi Global Storage Technologies
650 Harry Road
San Jose, CA 95120



Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Design of Multimedia Instant Messaging System with Annotation

¹Zon-Yin Shae, ¹Xiping Wang, ²Ferdinand Hendriks

¹ IBM T.J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532, USA

² Hitachi Global Storage Technologies
650 Harry Road
San Jose, CA 95120, USA

ABSTRACT

We have investigated enhancements to Internet based instant messaging (IM) system. Our system allows communication via handwritten information, mixed with other non-textual multimedia information. All information is presented to the user in a graphical user interface using a scrollable whiteboard metaphor. A compact timeline view, allows random access to recorded multimedia messages. Our system allows annotation of items entered anywhere in the IM record, as well as deletion and modification. All users are alerted to annotations by means of two hyperlinks; one at the end of the IM record the other at the annotation itself. The system is especially useful with the PDAs and Tablet PCs where handwriting input is currently supported. A prototyped IM system in iPAQ PDA with handwriting (ink) instant messaging and annotation will be presented.

1. INTRODUCTION

Instant messaging has its roots in well-known Unix commands such as “who” and “talk,” which allow users to see who is available as a communications partner, and to communicate nearly instantly. It is currently used more and more as a replacement for e-mail, and as a “control channel” among remote users running another collaborative application. An early protocol developed for instant collaboration is Internet Relay Chat [1]. Popular IM services are AIM (AOL Instant Messenger), ICQ (“I seek you” AOL), Microsoft Instant Messenger (MSN), Lotus Sametime, and many others. There is a growing number of enhancements to IM such as Moncreif et al. [2] which combines IM with live video reception. Despite the incompatibility of different proprietary IM protocols, IM is enjoying very robust growth. The growth would be accelerated into the enterprise environment with the development of SIP SIMPLE standard IM protocol. A typical IM session chat record, sometimes called history, would include a sequence of lines of text.

IM is currently available on many personal digital assistants (PDAs), such as Palm, Compaq iPAQ, and others. PDAs are not large enough to have physical keyboards, and their display screens are quite limited compared to those of desktop and laptop computers. The preferred mode of data input is handwriting with a stylus, or selecting individual characters with a stylus, using a virtual or software-defined keyboard, which is somewhat slow and cumbersome, although shortcuts and predefined phrases may help to speed up frequently used

1.10 Search button

tasks. A stylus produces stroke information. Sometimes referred to as electronic ink, or simply ink. Stroke input is, at the minimum, stored as an ordered set of quantized pen locations, but may include many more attributes such as the time at which each location was visited, the pen pressure and pen inclination, user-id of the pen, user-id of the writer and others. Despite the fact that ink is arguably the oldest method of data entry, standardization efforts have lagged. There is renewed interest in standardization by the w3C [6] consortium, owed in part to the increasing importance of handheld devices such as the Palm®[7], Compaq iPAQ®[8], Handspring Visor®[9] and others. Ink input is entered in a very natural way on PDAs and tablet PC, while entering text is more awkward. To be sure, other modalities such as speech and gesture also solve some of the problems associated with the lack of a keyboard, but due to a lack of processing power, the predominant data type for input is still text. Thus, the current state of IM lacks the expressivity that stroke input can provide. Furthermore, the record of the IM session is static, and does not allow a user to draw attention to particular messages, let alone add to them, or change them in any way, for all IM participants to see. It is also difficult to navigate among IM entries. This paper presents an ink instant messaging system to solve the above-mentioned problems.

The paper is organized as follows. Section 2 describes the system basic functions. Section 3 describes the scrolling whiteboard metaphor. Section 4 describes timeline user interface. Section 5 describes the hyperlinks mechanism for data and annotation and gesture based input is presented in section 6. An ink instant messaging and annotation IM system prototype is presented in section 7, and a summary is given in Section 8.

2. SYSTEM DESCRIPTION

The client application window shown in Fig. 1 is divided into the following areas, starting at the bottom of Fig.1:

- 1.5 Private ink input field. Used for local previewing.
- 1.4 Awareness field showing registered users and their status.
- 1.3 Scrollable chooser for input modality: Text, ink, annotation, other multimedia objects.
- 1.2 Text input field
- 1.7 Send button. Causes input to be sent to all participants in the session.
- 1.8 Clear button. Clears the input area.
- 1.9 Annotation; written directly into the recording field.
- 1.1 Public recording field containing text, ink and annotation.
- 1.6 Scroll bar for the recording field.
- 1.11 Awareness icons that visually express the status of participants.

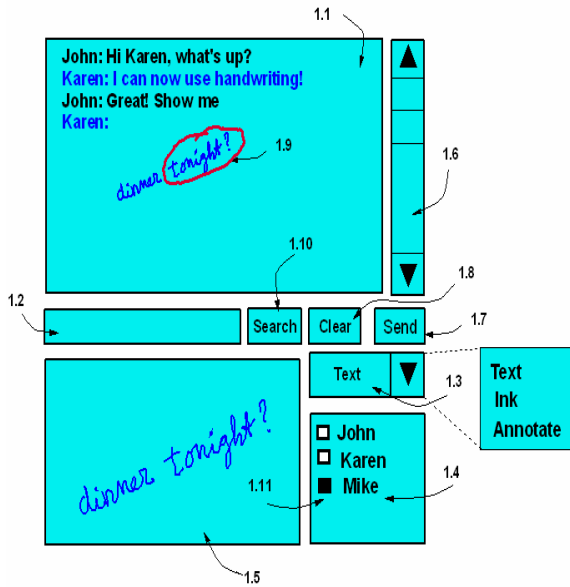


Figure 1. schematic diagram of the graphical user interface.

Field 1.3 is the modality selector, with which a user selects the type of instant message to be entered. Three examples are shown: text, ink and annotate. Other modalities can also be supported, in particular audio and video objects. In effect, the modality selector explicitly classifies the type of data entered in the IM record. The classification is a form of metadata associated with the object. This is useful for later browsing and retrieval of objects in the IM record. The awareness field 1.4 is a standard component of most IM systems. Fig. 1 shows a minimum version of such as field, containing the names 1.4 of those logged in to the same IM session, together with awareness icons 1.11 which have meanings such as “I am logged in,” “I am away,” “Do not disturb,” etc. In addition this field may contain a timestamp. Field 1.2 contains the input area for text based input. Clicking in this field is equivalent to selecting Text in the modality selector 1.3. Field 1.5 buffers stroke input and other IM objects for input and allows preview. Objects in this field are subject to geometric transformation when displayed in the recording field. This allows for a more compact IM objects that are not text. The “Clear” button 1.8 clears the last entered input object. The “Send” button causes the input to be sent to the server for distribution among session participants and storage in the IM record.

In addition, a user can annotate the recorded IM session with ink, which is in general of a different color compared with the normal ink input (ink entered in the ink input area). Annotations can also be general multimedia objects. The scrollable chooser 1.3 is optional. The intent of the user to choose ink, text or annotation can be expressed by clicking in the private ink input field 1.5, the text input field 1.2 or the annotatable recording field 1.1, respectively. The “Search” button causes the input to be used as key object for content-

based query. The content-based query for the ink media is currently a active research topic [10]. The emerging MPEG-7 [11] standard provides for a systematic way to describe the multimedia content of IM sessions so that browsing and querying become much more efficient. Stylus locations are reported in binary format, uncompressed or compressed. Compression of handwritten strokes is known in the art. However, progressive, scalable compression of strokes is an active area of investigation [3]. Progressive transmission of stroke information allows a user to gain an early visual impression of the strokes, while refinements are sent later.

3. SCROLLING WHITEBOARD METAPHORE

The recording field presents a scroll metaphor to the user such that the entire scroll is accessible to users as if it were a whiteboard of continuously increasing length as a session progresses. Annotations entered in the recording field are public and distributed to all users. It will be understood that besides text and ink, other media types may be inserted into the recording area, such as images, video and audio files. The public recording field 1.1 provides a viewport of a generally very long logical recording strip. The recording area represents the contents of all that has been entered into the session. New content is added at the bottom. The graphical user interface gives the user the impression that he/she is looking at part of a very elongated strip of constant width through a viewport. It is clear that the restriction of constant width can easily be removed and a horizontal slider added to the interface so that a user can move left-right in the recording display. However, in most cases a fixed horizontal width will suffice.

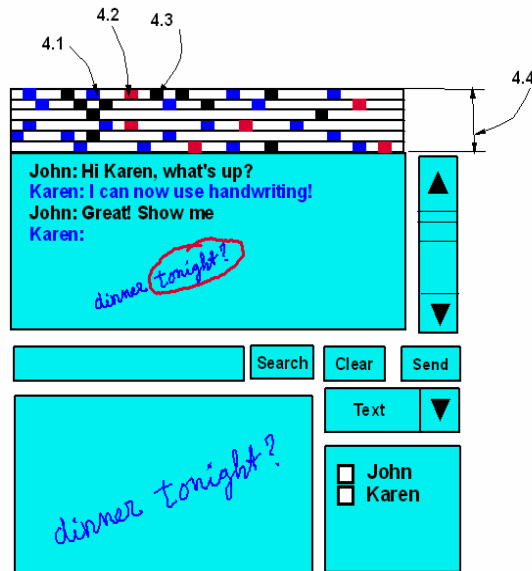


Figure 2. Graphical user interface consists a timeline field.

4. TIMELINE

In order to facilitate searching the recorded IM session, features such as skip-to-beginning/end and search by content may be part of the user interface. Fig. 2 shows a variation of the client graphical user interface. Some of the public recording area has been reserved to implement a “layered time line.” The frame shows the local user’s text as a mark 4.1, text from others as mark 4.3 and annotations (from anyone) as mark 4.2. Clicking on a mark causes the recording frame to show a section of the record corresponding to that mark (either ending at the mark, or beginning at the mark). The frame shows a timeline broken in sections, such that each section forms a layer. The intent is that frame 4.4 represents the whole instant messaging session. Distance along each layer represents an elapsed amount of time. Preferably the time is measured linearly and uniformly along each layer. However, especially when the session is particularly long, the distance may be measured non-linearly along each layer, for example, exponentially. This allows us to greatly increase the information content of the timeline frame (TF) 4.4, but at the expense of readability of the oldest items. The timeline frame allows the user random access to the recorded material, while the scrollbar allows sequential access. It will be understood that the marks of the timeline could themselves be icons that are specific for the modality of the input entered.

5. HYPERLINKS

All IM messages are objects (that is the object-oriented description of an abstract or concrete entity having attributes, methods and so on). In many cases, it will be necessary to store pointers such as a URL to the actual data describing the object. The present multimedia IM system allows a user to make stroke based annotations anywhere in the recording field. An annotation may or may not be associated with an earlier message. If an association is intended, the user is required to first select the message to be annotated, for example by clicking on it. For example, item 5.4 in Fig. 3 is to be annotated by Bob. Bob clicks on 5.4, indicating that the house will be the target for his annotation. He then enters the annotation 5.3. Message object 5.4 is now linked to (i.e. associated with) annotation 5.3. Message 5.4 is visible to the user as a bitmap of a house, but the message object may contain many attributes, such as cost, age, number of rooms and so on. In addition to the annotation message object 5.3 a “back link” 5.7 is added to annotation 5.3. The function of this “back link” will become clear later in this section with an application scenario. To facilitate navigation in the recording field, entering an annotation causes an entry at the end of the IM record. This entry is a hyperlink to the actual annotation, and is a message object in its own right. When clicked, it makes other users aware of the fact that another user made an annotation somewhere in the public record. Without this feature, users other than the one making the annotation would not be alerted to annotations. Because a user can scroll to any place in the recording field, what each user sees is not necessarily always the most recent part of the IM session. In addition to the hyperlink in the recording field, the hyperlink 5.1 is also entered as a symbol at the end of the TF 4.4. The actual

annotation 5.3 also appears in the TF. The annotation can be associated with an earlier entry 5.4. Fig. 3 illustrates a scenario in which a hyperlink is used to choose a picture of a house 5.4. User Karen has entered several real estate listings on her PDA and has shown them to Bob. She then prompts Bob for the house that he likes best. Bob indicates his preference by scrolling to the house he likes best stopping at view 5.5 and circling the house (stroke-based annotation). Karen, if viewing the current view of the recording field immediately sees the hyperlink 5.1 on her screen. If she happens to be at a different view, she is still alerted via the new entry in the TF, which may flash to draw attention to it. To facilitate jumping back to the location in the record from which the annotation was reached, usually the end of the record, a “back” button 5.7 is added to the annotation 5.3. The bottom part of Figure 3 shows the Bob’s annotation hyperlink consists of pointer to the annotated object 5.4 and the ink data 5.3. The horizontal line in the upper part of Figure 3 presents the time line of the IM session, the small circles present the objects have been entered in the recording filed in the sessions. Object 5.3 (ink annotation data) and 5.7 (back link) will be created after the annotation of object 5.4. After the “Send” button is clicked object 5.1 (the hyperlink) will be sent and show to the all participants. Every participant can click object 5.1 to look at the annotation object 5.3, and jumping back from there to object 5.1 by clicking object 5.7.

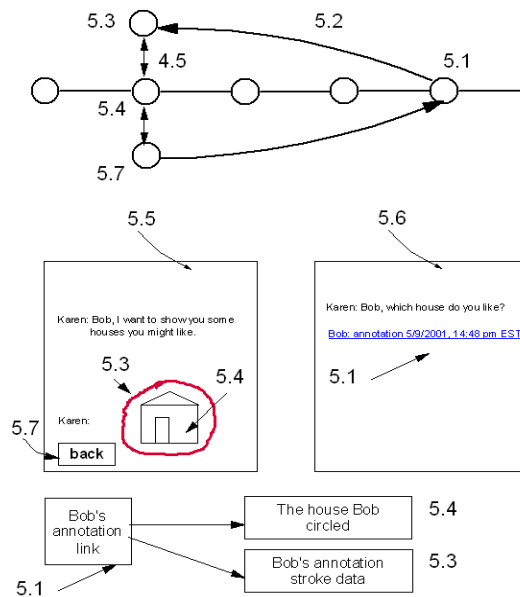


Figure 3. Annotation with hyperlinks, allowing quick jump to and back from an annotation

6. GESTURE-BASED INPUT

Stylus based information is sometimes used to indicate commands, referred to as ink gestures. Such input is immediately applicable to stroke based IM. But another

type of gesture, not stylus based, can be used to input information as well. Thus, besides input based on a stylus and keyboard, it will be understood that other input modalities may serve to create messages, or to navigate among the record, or to manipulate the record in other useful ways. Such other modalities can be based on human actions such as speaking and gesturing. To equip the device being used for IM to receive voice and gesture input, it must have the requisite transducers: a microphone for voice and a gesture capturing system. The latter can take many forms, depending on the physical principle on which they are based; such as systems that rely on acoustics, electromagnetics, optics and combinations thereof. While gesture input systems require considerable processing power and/or rely on hardware support, they are gradually becoming more practical.

7. INK INSTANT MESSAGING PROTOTYPE

We have implemented the proposed Ink instant messaging system on handheld devices, such as iPAQ. The IBM SameTime is used as the central controller for distribution and management of messages. The client is implemented in pure Java using the SameTime SDK for cross platform portability. The SameTime SDK provides programming interfaces with SameTime server. Fig.4 shows the software architecture of the prototype.

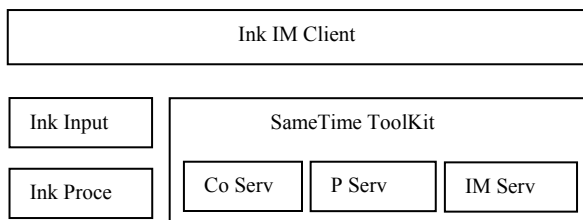


Fig. 4 Software architecture of Ink IM system.

All of the participants must first log into the same meeting place through SameTime's community service. After authorized, each participant can then communicate with all of other participants or have a private conversation with a desired participant. The communications among meeting participants is controlled and maintained by the SameTime's places service. In order to achieve a natural handwriting speed with the stylus on iPAQ, a special care must be taken on the stroke input. In our implementation, a simple decimation algorithm is used for simplifying the input strokes while maintaining a reasonable accuracy of strokes. The simplified strokes then needs to be

compressed before transmission for efficiency (details of streaming mechanism had been reported in [3]). We have asked different people to evaluate the system and found out that the proposed system is very suitable for PDAs and the speed of the ink messaging input on iPAQs is adequate for tracking natural human handwriting. Since handwriting Ink message is a natural choice for devices using stylus input, the Ink IM is much more desired than text-based IM. Especially the annotation feature and scrollable white board design are very attractive for discussion and history review.

8. SUMMARY

Instant messaging (IM) is often used "when e-mail isn't quick enough" or "spam e-mail is too much" for short text-based notifications and responses. IM is currently touted as the Internet's latest "killer app." This paper investigates the design of multimedia instant messaging system that extends the text instant messaging into multimedia instant messaging. Especially ink instant messaging with annotation and whiteboard metaphor that will enrich the effectiveness of the instant messaging in the PDA and Tablet PC devices. The scrolling whiteboard metaphor mechanism and timeline as well as hyperlinks user interface has been presented and an IM system using iPAQ PDA has been prototyped and evaluated. Users of this iPAQ IM system have expressed a quite favorable experience doing IM by ink writing. This system would also be very suitable in the Tablet PC. Inter-operability among various commercial IM systems can be achieved by SIP SIMPLE protocol and W3C ink data format standardization.

REFERENCES

- [1] J. Oikannen et al. "Internet Relay Chat Protocol," RFC 1459, May 1993.
- [2] U.S. Pat. 6,061,716 "Computer Network Chat Room Based on Channel Broadcast in Real Time," C.T. Moncreiff. This patent deals with a chat enhancement consisting of real time TV broadcast. It also deals with addition of commercials to the chat session.
- [3] Zon-yin Shae et al. "Multi-Resolution and Streaming of Whiteboard Stroke Media" IEEE International Symposium on Multimedia Software Engineering 2000
- [4] www.webdiving.com allows insertion of images into instant message. "Chat enhancer" analyzes chat messages and attempts to suggest related items, such as web pages to the chat participants.
- [5] www.openmash.org In particular MediaPad which implements MediaBoard (whiteboarding over the multimedia backbone: MBONE) on a PDA.
- [6] www.w3c.org/TR/InkML, inkXML group and document
- [7] www.palm.com Palm PDAs
- [8] <http://welcome.hp.com/country/us/en/prodserv/handheld.html> iPAQ PDAs
- [9] www.handspring.com Visor PDAs
- [10] Howard W. Leung, "Representations, feature extraction, matching and relevance feedback for sketch retrieval", PhD Dissertation, Carnegie Mellon University, department of electrical and computer engineering, June, 2003.

[11] ISO/IEC JTC 1/SC 29/WG 11, 15938-5, "Multimedia
Content Description Interface – Part 5 Multimedia

Description Schemes" 2002 .