Using Real-Time Freeform Annotations as Qualitative e-Research Metadata

Mike Fraser[1], Muneeb Shaukat[1], Serguei Timakov[1], Phillip Smith[1], Jon Hindmarsh[2], Dylan Tutt[2], Christian Heath[2], Anne Manuel[3], Marie Gibbs[4] and Sally Barnes[4]

[1] Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, BS81UB, UK.

[2] Department of Management, King's College London, Franklin-Wilkins Building, London, SE1 9NH, UK.

[3] Institute for Learning and Research Technology, University of Bristol, 8-10 Berkeley Square, Bristol, BS8 1HH, UK.

[4] Graduate School of Education, 35 Berkeley Square, Bristol, BS8 1JA, UK.

fraser@cs.bris.ac.uk

Abstract. Qualitative analyses may emerge over long periods of time, sometimes through re-inspection of old data in a new light, and sometimes over the course of changing research groups and projects. However, predicting which materials and markings might be useful to retain for future use is often a difficult and time-consuming job which typically in e-Science has led to constraints on metadata categories through ontologies, or the use of automated tools for simple event logging. As a result, analysts may have to rely on memory to find and re-study data together, and may lose some potentially useful threads of analysis altogether. Within the MiMeG e-Social Science node, we have been working on ways of recording and storing a variety of annotations as metadata related to distributed analysis, such that they can be retrieved in later situations and repurposed to feed new ideas. Keeping track of research activities in this way requires i) that we maintain the use of freeform annotations and provide methods of retrieval partly based on skill and memory rather than automation ii) that the system should take into account ownership and authorship of annotation data and iii) that we understand what forms of freeform annotation, such as transcripts, scribbles, audio and even off-screen movements, may be used to support qualitative e-research.

Introduction

Recent work on qualitative e-Research has considered how to sustain relationships between researchers in real time and over large geographical distances (Fraser et al., 2005). There are an increasing number of researchers working on understanding the peculiarities of distributed collaborative research as it is conducted across domains and research groups (Zheng et al., 2002), and building systems which take these findings into account. One important goal has been to ensure that the process of creating metadata in such a setting does not increase the administrative overhead on that normally required for accountability purposes. In recent work as part of the MiMeG node, we have been interested in capturing analytic metadata as *part of the business of doing research*. We hope to challenge the 'minimal impact' approach to metadata production, and instead provide means for developers to treat certain forms of metadata production as an intrinsic part of existing research practice rather than a 'user chore', and to provide benefits that allow metadata not just to make existing work accountable, but to improve on areas of current difficulty without the need to undermine tried and tested procedures of research practice. We also contrast our freeform approach with the prevalent strategy across the e-Science community of predetermining domain ontologies for use in data sharing. In the following sections of our paper, we describe how our software uses analytic metadata to provide ways of logging selected activities and ways of retrieving annotations based on serendipity and similarity.

Annotations

Our approach to metadata has been to start by understanding the real-time communication which occurs through our distributed MiMeG application. MiMeG is an ongoing development of software we produced as part of the pilot VidGrid project (Fraser et al., 2005). Within VidGrid, our aim was to inform the design of tools to support the real-time analysis of video data by distributed groups of social scientists, and we produced an illustrative demonstration. Building on the stability and utility of the VidGrid pilot software, we released a first version of the MiMeG software earlier in 2006 at a workshop concerned with software support for video analysis in qualitative research. Broadly, among other functions, our software introduces support for two forms of annotation for video data: text transcripts and freeform scribbles.

Transcripts are a commonly used form of annotation in many types of video analysis, Layouts, format and schemas vary substantially, both methodologically and individually, but they all tend to illustrate features of the talk contained within the data. For some analytic orientations transcription of talk may be conceived of as an analytic step in itself, i.e. transcribing the talk gives the researcher insight into the data. For others, transcription may be considered a bulky and arduous step which might be devolved to more junior researchers. Such perspectives often depend on the detail, organisation and scale of the data that interests the researcher. Beyond the act of transcription itself, transcripts can provide an important collaborative resource to encourage others to find relevant moments in the action. By drawing attention to particular parts of a textual transcript, participants can occasion other researchers to notice action that occurs around the words or utterances that feature at those moments in the transcript. Such work can be crucial in reaching a shared perspective and transcripts often form the basis for coordination of perspectives in a co-present data session. Our software supports timestamped text events which can be coordinated into a transcript, as well as the ability to import existing text files as transcripts to be shown alongside the video. They can be shared between distributed researchers by normal electronic means (e.g. email, web), or through the application via a database (see below).

Freeform scribbles are the second form of annotation that have been introduced in our software. We have provided scribbles to allow researchers located in distributed sites to annotate video data to assist in their communication. A scribble is a set of individual points making up a line drawn freehand using a mouse, tablet computing pen or electronic whiteboard device. Scribble annotations are made directly over a video stream by any of the distributed users, who can all publish and subscribe to annotation events. Communication of these freeform annotations is via individual event notifications per pixel drawn, although we also anticipated the network load of per-pixel events to be significant, so created an option for packaged per-stroke transmission. Scribbles are provided so that participants in a real-time distributed datasession may highlight a phenomenon of interest for one another. They might indicate or circle the area of the data under inspection, or they might mimick a gesture or movement that features in the video data with the scribble. Our experience with co-present datasessions has shown that typical mimicking gestures to illustrate behaviour in videos are in many ways not concerned with providing ‘exact copies’ of on-screen conduct, but render both the relevant action visible and the analytic point that is being made about that action. They tend to exaggerate or transform the on-screen conduct in such a way that it becomes a proposed analytic point. These gestures may be later used again to make a further point about that action. These various embodied practices of revealing phenomena are critical as participants in datasessions highlight actions of interest and together develop characterisations of their organisation.

So, we have included two forms of data which act in different ways to mark-up primary video data with researcher-produced insights or information for the purposes of sharing perspectives with colleagues. As designers we have not invented the need which these kinds of annotation are addressing, we have derived them from studying our particular domain of interest. However, we suspect that there are few research domains which do not use transformed representations of primary data during analysis and require at least some communication around data with research colleagues or others – albeit perhaps in domain-dependent ways that would not need the particular transcripts and scribbles mark-up methods we have provided for supporting the work of distributed qualitative video analysis.

What is clearly relevant across domains is that our approach of deriving our annotation mechanisms from domain study rather than from technology, has produced mark-up tools unlike metadata systems as traditionally conceived within technology-oriented fields such as knowledge management and the semantic web (Berners-Lee et al., 2001). For the following section of the paper, we look at the particular characteristics of our metadata mechanisms, and compare and contrast them with those available in traditionally conceived e-Science applications.

Metadata Storage and Automated Concept Matching

In this work, we are assuming that researchers will actively decide to commit and retrieve freeform metadata during communication with one another. This process is typically termed *manual* metadata authoring, a term which reveals something about the way in which it is viewed as a design in e-Science. The term *manual* highlights the laborious efforts to which users will be forced in indexing their materials, and also reveals something of a hint to the potential for computers to relieve such labour through automation of indexing processes. It also reveals the natural tendency of computational scientists to dislike freeform human-generated records which may contain 'errors' or 'omissions' which preclude automated processing and matching. In practice, many e-Science projects which use metadata to index and retrieve data records do so with the aim of logging metadata automatically where possible and manually only where necessary (e.g. Buckingham Shum et al., 2006). This is unproblematic in support for low-skill collaboration scenarios, but in contrast for highly skilled researchers it is necessary for us to treat the creation of metadata as an important feature of the work of research. To automate that feature of the work would at best invest in research skills in inappropriate areas, and at worst remove the possibility of adequately conducting the research using the system at all.

In the main, dislike of freeform manual metadata production stems from two original e-Science goals. The first is to provide and share data across geographical and institutional boundaries by engineering datasets which can be used by multiple stakeholders. This requires that the metadata must be of reproducible and generaliseable form too, and has led developers to demand the application of semantic ontologies of the domain to each new dataset to be deposited in a shared filestore. This approach requires significant investment in development work (of the ontology as well as the software) and then typically precludes research that does not fit the predetermined semantics of the domain without adapting the ontology independently of doing that research – a rather cumbersome approach when new research findings will by their very nature reorganise domain semantics. The second goal is to allow independent modules of data to be processed together in any organization of workflow, allowing the application of old data to new problems. To allow such module reorganisation, one must provide predictable inputs and outputs to the module, which in turn requires uniformity in that module's dataset. Again, conforming to a domain ontology guarantees this data integrity.

Unfortunately the first of these goals – arbitrary distribution of datasets – is completely inappropriate for the qualitative social sciences, in which data is rarely and carefully distributed in raw primary forms, due to ethical, legal and consent constraints. The second goal – reuse of old data in new ways by researchers – is more typical, for example digging out old data which in retrospect may have relevance to a new interest. However, this process already happens to an extent in cases where memory does not fail the researcher, and we feel there are more appropriate ways of supporting such work without the introduction of domain constraints or additional effort in identifying and reorganizing categories of metadata. Specifically, if we can rely on individuals who own or who have encountered data being the primary users of retrieval mechanisms at potentially significantly later dates (which is inevitable where ethical constraints on data distribution apply), then we may be able rely to a certain extent on their memory of themselves or a co-present colleague producing real-time annotations – as part of communication or transcription in our case – to recall the related data itself. This would allow researchers to attempt to search for freeform annotations by reproducing from memory the various annotations involved and pattern matching by adapting algorithms for that medium, such as text similarity (Lee et al., 2005) for transcripts or stroke similarity (Leung and Chen, 2003) for scribbles, narrowed by broad categories of browsing.

We have incorporated in our software means of storing and retrieving annotations according to such broad categories. Strokes are transmitted between analysts' client applications in real time. Our original application supports the transmission of these 'scribbles' over the video data which multiple distributed analysts are studying. Allowing local users to save and load their own annotations onto local storage media also provided annotation persistence in VidGrid. More recently, however, we have incorporated means of storing and retrieving annotations over an Apache Axis webservice (Axis, 2006) to a MySQL database. This integrated means of storage allows us to keep track of researchers as they log into the application, and to allow them to commit and retrieve annotation data (transcripts and scribbles) to and from the database.

Using this system, Figure 1 shows a typical example which we have implemented in MiMeG. A researcher, indicated by their current annotation colour (red) has retrieved annotations from their MySQL database created by another user (assigned the colour yellow). The other user produced the annotations during an earlier distributed datasession, but are not currently present. A window pops up in the system to identify the offline user whose annotations have been displayed, and provides means to contact them out-of-band (e.g. by email) if required. Their annotations are pulled from the database and merged with current on-line annotations in this datasession. In turn, the online user might decide to commit their own current annotations to the database to add to the pool.

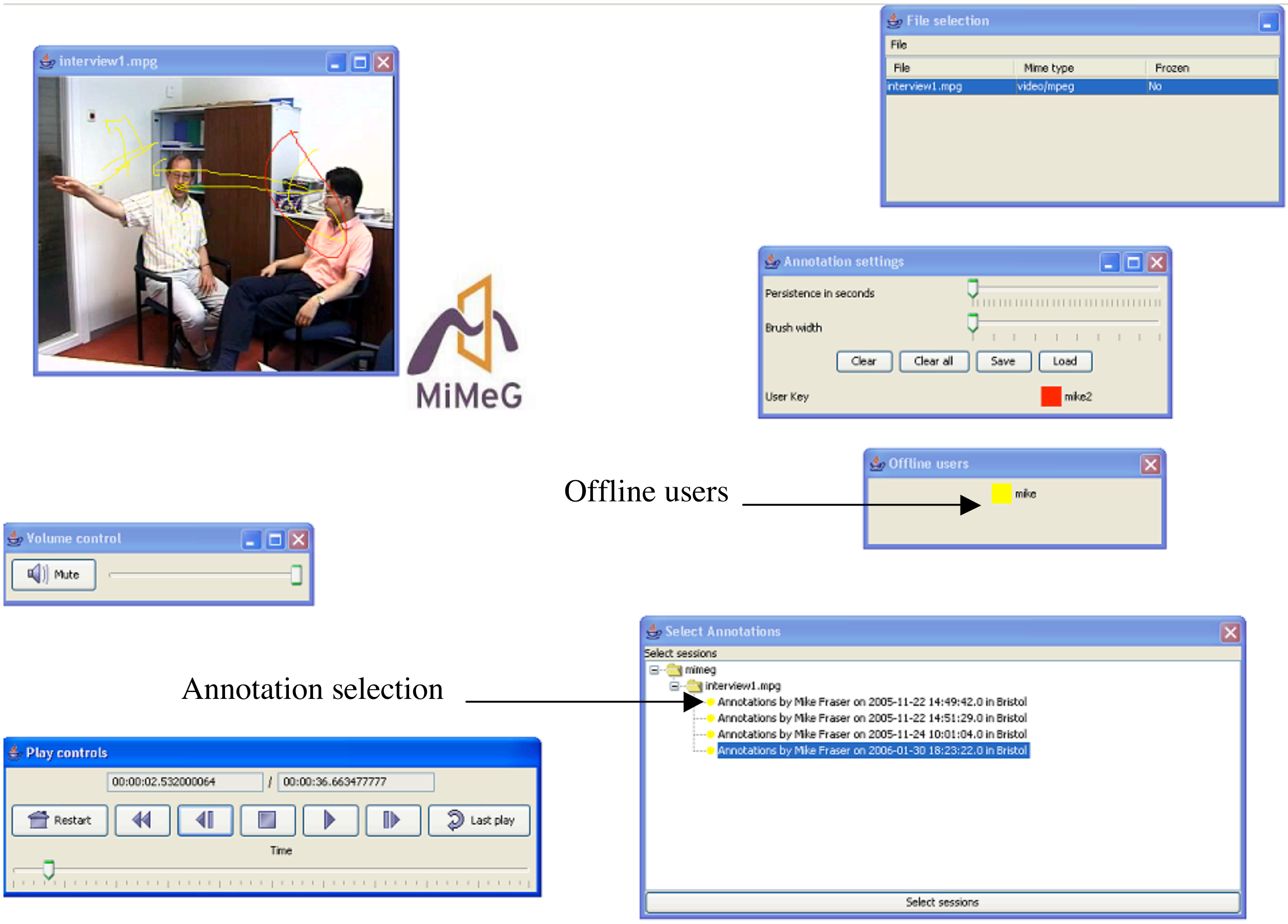

Figure 1 – Screenshot showing annotation selection and merged online/offline annotations

Thus far our interface stores scribbles according to the broad categories of user log-in, self-declared physical user location, project, data file annotated and time/date. This provides the means to browse for known previous annotations, for example by retrieving the annotations of the most recent previous data session working on this data as a means to remind the researchers of their interim conclusions or the process by which they reached them. We hope the simplicity with which such process data can be captured will improve the ability of the researchers to remember such positions, but may also make possible the narrowing of search strategies such as those described above to manageable matches.

There are further ways in which we might avoid constraining metadata away from freeform communication, for example by checkpointing time sections of data sessions to record annotations in. We might also narrow search possibilities by linking searches of annotations over video with transcript matching, or even allowing freeform annotations to be produced over transcripts as well as video and then pattern matching on those. In the next section, however, we go on to look at how further kinds of freeform annotation might provide more appropriate or detailed ways of indexing back into data for repurposing.

Complex Communication Metadata Production

Thus far we have described the storage of communication metadata with relatively simple devices, for example scribbles can be produced using two-dimensional devices such as a mouse or electronic whiteboard marker. We are currently investigating methods of storing and retrieving more complex data which features in communication between participants.

Clearly there is potential for the talk between sites (via voice over IP in MiMeG) to be logged and used as annotations at points of interest. Currently, we are working on extending the application's sensing of analytic interaction by providing additional real-world sensors around analytic displays. In previous work, we have noted the difficulty of adequately preparing remote sites that someone is about to produce an annotation. Despite the use of real-time strokes, there are aspects of annotating data for others which are lost by only transmitting screen-contact gesture and audio. Particularly, whilst co-located researchers are able to see the analyst prepare to produce a stroke in front of the screen, researchers at remote sites are only aware of the stroke *at the time it is being produced* (Fraser et al., 2005). It turns out that understanding the ways in which the display is approached, the trajectory of the gesture, is crucial to the organisation of perspective.

There are further temporal issues relating to ways of indicating features during video playback. The use of strokes over video data alters significantly when annotating a paused frame versus annotating at playback. The annotation of a single frame allows relatively straightforward reference to features of interest. However, during this process, participants tend to forget that the annotations they are producing have a variable persistence value which will result in those strokes continuing over subsequent frames. On playback, this persistence becomes noticeable, and as the frames change, the annotation loses its relevance whilst maintaining its presence. We might automatically reduce pause-frame annotations to very low persistence levels, but then those strokes would be barely visible during real-time playback of the sequence. Furthermore, annotation during playback introduces its own set of problems. Whilst persistence levels are more naturally configurable, the production of the strokes themselves is not, given each stroke has a particular start time and lifetime. For example, drawing an arrow to point at some feature results in two strokes being used – one for the line and one for the arrowhead. The line of the stroke will typically be produced first, and therefore disappear first before the arrowhead. We might address such issues by introducing particular shapes such as arrows as defined annotation options, but at the cost of both increasing interface complexity, and potentially reducing freeform flexibility.

To alleviate these problems, we have begun conveying some notion of where the annotating devices are with respect to the projected display interface, to both illustrate trajectory and to allow for temporal production of a stroke prior to screen contact. Our current system tracks an annotating electronic whiteboard marker pen's position in the space around a projected display using a less accurate low-cost three-dimensional ultrasonic positioning system (Randell and Muller, 2001) which tracks the pen to around 3cm accuracy, shifting to the pen's own tracking system which is far more accurate when the pen is in contact with the screen. We then present an appropriate visualization over others' displays at remote sites. If we can configure sensors to capture more detailed movement of participants' annotations, we have a better opportunity to record those movements to serendipitously retrieve data in which our analytic impression produced similar kinds of annotation.

Representations

We have implemented three different representations of the whiteboard marker's remote off-screen position to give remote users a sense of the trajectory of other participants' annotations. Whilst these are not finalized nor tested, we include them here to illustrate the kinds of annotation which might be possible with higher dimensional tracking at a remote site. The most simple representation is a crosshair spread over the entire display area with its opacity scaled according to the distance away from the board.

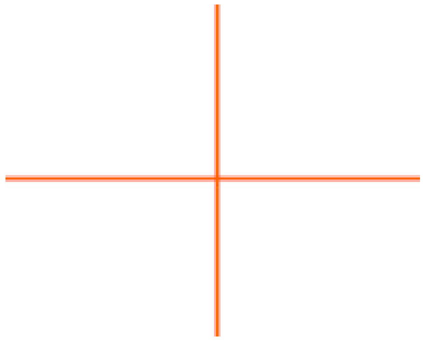

Figure 2 – Crosshair Representation

The next representation uses 'trails' (cf. Gutwin et al., 2004), leaving a line of varying width representing the distance away from the board and a varying opacity along its length fading from the least recent position. Two versions were implemented, the first uses a number of disjointed line segments, allowing each to have their own properties (figure 3), while the second is implemented as one line with the last N co-ordinates stored in a circular buffer.

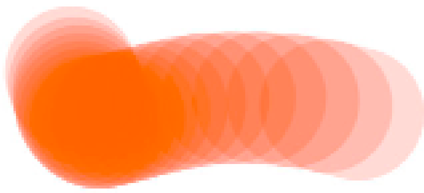

Figure 3 – Trail Representation (separate line segments)

Figure 4 – Trail Representation (continuous line)

Finally, using only the currently available position, we implemented a representation that consists of a 'circle of influence' showing a circle filled with a gradient colour. Both the opacity and diameter are scaled according to the distance away from the board, becoming smaller and less opaque the closer the user is to the display. This shows the user is likely to interact with an area with more certainty the closer they are.

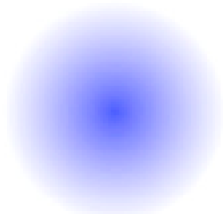

Figure 5 – Circle of influence Representation

Early experience shows the 'crosshair' is obstructive and hard to follow, especially when moving rapidly. 'Trails' are intuitive and the use of persistence means an area can be pointed out very easily, but occupies significant screen real estate when more than two people use the system at once. The 'circle of influence' provides a clear view of where and when a user is about to scribble and gives a very natural sense of distance by combining scaling with opacity.

When the marker pen becomes activated and emits its own ultrasound signal, our positioning system receiver also receives these signals. This causes the receiver to lose the correct position of the pen. The effects of interference are largely independent of the position of the pen in relation to the transmitters, and of the pen's current motion. The pen's position is currently reacquired correctly in under a second of interference occurring, provided the transmitters are not occluded. The erroneous readings by the receiver cause it to detect divergence, as the estimated point is a large distance from the previous point. This causes the software to enter a state where it tries to reacquire the position of the object to a high certainty. Once the pen has stopped emitting the signal, valid pulses are received allowing tracking to resume. The electronic whiteboard marker is, however, not affected by the emitters in the off-board positioning system, probably due to the intensity of the pulses transmitted. A method of updating the pen's 3D position from the point where it was last in contact with the board is being developed to allow the positioning software to find the position of the pen more rapidly after interference.

References

Apache Software Foundation, Axis Webservices (2006). http://ws.apache.org/axis/, verified 01/06/2006.

Berners-Lee, T., Hendler, J. and Lassila, O., The Semantic Web, in Scientific American, May 2001.

Buckingham Shum, S., Slack, R., Daw, M., Juby, B., Rowley, A., Bachler, M., Mancini , C., Michaelides, D., Procter, R., De Roure, D., Chown, T., and Hewitt, T. (2006). Memetic: An Infrastructure for Meeting Memory. Proc. 7th International Conference on the Design of Cooperative Systems, Carry-le-Rouet, France, 9-12 May.

Fraser, M., Biegel, G., Best, K., Hindmarsh, J., Heath, C., Greenhalgh, C. and Reeves, S. (2005). Distributing Data Sessions: Supporting remote collaboration with video data, in *Proc. ICeSS 2005*, Manchester, UK, June 2005.

Gutwin, C., Benford, S., Dyck, J., Fraser, M., Vaghi, I. and Greenhalgh, C. (2004). Revealing Delay in Collaborative Environments, in Proc. ACM CHI 2004, Vienna, Austria, pp. 503-510, ACM Press

Heath, C. and Luff, P. (1996). Documents and Professional Practice: "bad" Organisational Reasons for "good" Clinical Records, in CSCW Journal, pp. 354-363, 1996, Kluwer.

Lee, M. D., Pincombe, B. M. and Welsh, M. B. (2005). An empirical evaluation of models of text document similarity, in Proceedings of the 27th Annual Conference of the Cognitive Science Society, pp. 1254-1259, Mahwah, NJ, USA, Erlbaum.

Leung, W. H. and Chen, T. (2003). Hierarchical matching for retrieval of hand-drawn sketches, in Proceedings of ICME 2003, vol. 2, pp. 29-32, July 2003, Pittsburgh, PA, USA.

Randell C. and Muller, H. (2001). Low Cost Indoor Positioning System, in Proc. Ubicomp 2001, pp. 42-48, September 2001, Springer..

Zheng, J., Veinott, E. S., Bos, N., Olson, J. S. and Olson, G. M. (2002). Trust without touch: jumpstarting long-distance trust with initial social activities, in *Proc. CHI 2002*, pp. 141-146, ACM Press