After Action Review Tools For Team Training with Chat Communications

Dr. Sowmya Ramachandran, Randy Jensen, Oscar Bascara, Dr. Tamitha Carpenter
Stottler Henke Associates, Inc.
San Mateo, CA
Sowmya@stottlerhenke.com, Jensen@stottlerhenke.com, Bascara@stottlerhenke.com, Tamitha@stottlerhenke.com

Todd Denning
AFRL/HEA
Nellis AFB, NV
Todd.denning.ctr@nellis.af.mil

Lt Shaun Sucillon
AFRL
Mesa, AZ
Shaun.sucillon@mesa.afmc.af.mil

ABSTRACT

As networked digital communications proliferate in military operational command and control, chat messaging is emerging as a preferred communications method for team coordination. In areas where chat messaging provides fundamental support to command and control processes, training methods must incorporate techniques to associate and analyze chat room content to determine effectiveness of the communications. Chat room logs provide a rich source of data for analysis in after action reviews affording considerable insight into the decision-making processes among the training audience. They are also relatively unstructured and replete with competing lexicons, abbreviations, and shortcuts. The employment of multiple chat rooms and multiple interleaved dialogs within them introduces high likelihood of missed or misinterpreted communication. This presents a number of challenges for near-real-time analysis. Recent joint-service guidance in chat room protocols and structures opens new opportunities to revisit investigation of techniques for near-real-time feedback in training programs. In this paper, we describe a research effort to develop chat analysis and filtering methods for after action review tools. The investigation focuses on operational planners tackling time-sensitive problems by employing chat communication for both intelligence assessment and mission planning coordination among a diverse set of expert domains. Our research employs combinations of sorting capabilities using organizational and temporal context given by the new joint service guidance, keyword filtering techniques, and informed analysis considering statistically paired dialog participants. These techniques, when combined with visual timeline based presentation of scenario ground-truth and key milestones in the planning processes, promise to provide a more cogent and effective use of time in after action reviews.

ABOUT THE AUTHORS

Dr. Sowmya Ramachandran is a research scientist at Stottler Henke Associates, a small business dedicated to providing innovative Artificial Intelligence solutions to real-world problems. Dr. Ramachandran's interests focus on intelligent training and education technology including intelligent tutoring and intelligent synthetic agents for simulations. She is also interested in issues of motivation and metacognition. Experience with military and private industry gives Dr. Ramachandran a unique perspective on the needs and requirements of the ultimate end-users and their constraints. She contributes expertise in AI, instructional systems, probabilistic reasoning, and knowledge management. She has developed ITSs for a range of topics including reading comprehension, high-school Algebra, helicopter piloting, and healthcare domains. She has participated in workshops organized by the Learning Federation, a division of the Federation of American Scientists, to lay out a roadmap for critical future research and funding in the area of ITSs and virtual patient simulations. She has developed a general-purpose authoring framework for rapid development of ITSs, which is currently being used to develop an intelligent tutor training Navy Tactical Action Officers. She has also developed tools and technologies for training emergency first responders.

Todd Denning is a training research investigator for AOC training research for the Air Force Research Laboratory's Warfighter Readiness Research Division in Mesa, AZ and an instructor/subject matter expert for dynamic and deliberate planning training with the 505th Operations Squadron, Nellis AFB, NV. He has extensive experience in fighter and air operations planning and operations to include combat operations in Southwest Asia and the Pacific region.

Randy Jensen is a group manager at Stottler Henke Associates, Inc., working in training systems since 1993. He has developed numerous Intelligent Tutoring Systems for Stottler Henke, as well as authoring tools, simulation controls, after action review tools, and natural language analysis methods. He is currently leading projects to develop an embedded training Intelligent Tutor for the Army, an after action review toolset for the Air Force, and an authoring tool for virtual training demonstrations for the Army. He holds a B.S. with honors in symbolic systems from Stanford University.

Oscar Bascara is a software engineer at Stottler Henke Associates, Inc. His interests include training systems, authoring tools, and user interface design. He holds an M.Eng. in Electrical Engineering from Cornell University and an M.A. in Mathematics from the University of California at Berkeley.

Dr. Tamitha Carpenter is a research scientist with Stottler Henke Associates, Inc., specializing in natural language understanding and collaboration support. Her recent work has focused on the analysis of dynamic text, such as online chat and e-mail, in support of a larger activity context. Other work includes mining online text for threat indicators, supporting research activities through context-aware search assistance, and automating human workflow management. She is currently researching techniques to recognize expertise contained within organizational documents to support the formation of just-in-time collaborations. Dr. Carpenter holds a Ph.D. in Computer Science from Brandeis University.

Lt Shaun Sucillon is a behavioral scientist assigned to the Air Force Research Laboratory's Warfighter Readiness Research Division in Mesa, AZ. He is the deputy program manager for the mission planning, brief, debrief, after action review training research program for the Continuous Learning Branch.

After Action Review Tools For Team Training with Chat Communications

Dr. Sowmya Ramachandran, Randy Jensen, Oscar Bascara, Dr. Tamitha Carpenter
Stottler Henke Associates, Inc.
San Mateo, CA
Sowmya@stottlerhenke.com, Jensen@stottlerhenke.com, Bascara@stottlerhenke.com, Tamitha@stottlerhenke.com

Todd Denning
AFRL/HEA
Nellis AFB, NV
Todd.denning.ctr@nellis.af.mil

Lt Shaun Sucillon
AFRL
Mesa, AZ
Shaun.sucillon@mesa.afmc.af.mil

INTRODUCTION

As networked digital communications proliferate in military operational command and control, chat messaging is emerging as a preferred communications method for team coordination. Like radio, chat facilitates instant communication among multiple people. In addition, it is less susceptible to loss of transmission quality, as is often the case with radio communications. Chat environments also maintain an electronic record of all team communications so participants can easily review past chat messages in real time.

Traditional team training has involved human observers for performance assessment, diagnosis, AAR, and other training intervention. However, with much of the communication and coordination happening electronically, key aspects of the interactions between team members are no longer accessible to these trainers. Analyzing these communications would mean poring over high volumes of raw electronic data to uncover patterns and events. Chat messages can be monitored, but when there are more than one or two dedicated chat rooms (as is often the case in large team-training exercises), monitoring all of them effectively is a challenge. Thus in areas where chat messaging provides fundamental support to command and control processes, training methods should incorporate techniques to associate and analyze chat room content to determine effectiveness of the communications.

Chat room logs provide a rich source of data for analysis in after-action reviews (AAR), affording considerable insight into the decision-making processes among the training audience. They are also relatively unstructured and replete with competing lexicons, abbreviations, and shortcuts. The employment of multiple chat rooms and multiple interleaved dialogs within them introduces a high likelihood of missed or misinterpreted communication. This presents a number of challenges for near-real-time analysis. Recent joint-service guidance in chat room protocols and structures opens new opportunities to revisit investigation of techniques for near-real-time feedback in training programs.

This paper describes a research effort to develop chat analysis and filtering methods for AAR tools. A common thread in team training is the goal of maximizing the level of instructor interaction with the training audience during exercises, but this has the consequence of reducing the time available for manual review of exercise data and preparation for after-action debriefings. Our investigation focuses on a training application where operational planners must tackle time-sensitive problems by employing chat communication for both intelligence assessment and mission planning coordination among a diverse set of expert domains. With this application, typically instructors have at most 30 minutes to digest a 3-5 hour exercise. Thus, tools are needed to aid instructors in quickly extracting useful information from the chat logs to inform the AAR.

BACKGROUND

Chat has been a consumer-market driven technology with much of the focus on its use for social interactions. Only recently has there been a trend towards using it for business communications, and even in this context it has served only as an informal communications tool. As a result, there has been little demand for tools to analyze chat communications. The exception is the research community, which is interested in developing techniques for mining social patterns from chat.

However, as chat becomes an increasingly recognized component of teamwork, there will be a need to visualize and analyze this data for various purposes. In this case, we are concerned with tools to support AAR of exercises where chat is the main mode of communication.

Previous related research involving multi-party dialog analysis has included much work to characterize spoken interactions in multi-party meetings, social structures, and collaborative learning environments. The most relevant work is being done by the CALO (Cognitive Agent that Learns and Organizes) project, a joint effort between SRI and Stanford University's Center for the Study of Language and Information. (Zimmermann 2006), and (Tur 2008) describe efforts within the CALO project to support multi-party meetings with transcription, action item extraction, and, in some cases, software control such as document retrieval and display updating. (Niekrasz 2004) describe an architecture in which the spoken conversation between meeting participants is processed using automatic speech recognition techniques, and grounded against the artifact being produced (e.g., a schedule, a budget) and the drawings made on an electronic whiteboard. All of these inputs are used to create an electronic version of the artifact. Although experiments with dialog models from spoken interactions are transferable to research with chat communications, there are also unique challenges with the chat medium.

Much chat-related research has focused on the inherent communication artifacts of the medium, such as the emergence of conventional abbreviations, emoticons, and other common stylistic practices. To a lesser degree, some research has yielded methods and tools to analyze or visualize chat communication patterns. Most require a coding step carried out by a human reader to tag messages or explicitly identify dependencies before analysis takes place in any automated form.

(Cakir 2005) studied methods for assessing team problem solving with a chat environment and shared workspace. Essentially this employed a structure for organizing messages and identifying instances of interactions between two, three, or more participants as well as indices for factors like initiative. This is useful for learning research observations about how level and type of participation contribute to team dynamics and collaboration effectiveness.

(Shi 2006) introduce a conceptual framework for "thread theory," which suggests an approach for sorting out different chat threads based on topic or theme, and for characterizing defining features such as life, intensity, magnitude, and level of participation. (Herring 2006) describes VisualDTA, a tool designed to generate a visualization of a chat conversation that has been manually coded. In this visualization, messages are plotted in a descending tree, with temporal spacing represented on one axis, and semantic divergence represented on the other. The tool also accommodates the possibility of completely new topic threads appearing within the chat stream, resulting in new trees. This is useful for social interaction research, where plots of communication patterns reveal behavioral features.

While this work provides an excellent foundation for supporting automated analysis of natural dialog chat, the analysis goals in a training context can be quite different from those in the context of studying social behavior or other features of human interaction. With the specific objective of evaluating chat logs for team training, there isn't time for manual coding of chat logs, and visualization is most useful if it can help an instructor analyst drill down to observations about decision-making performance. In the pursuit of automated instructor tools, several areas of research must still be addressed:

- **Leveraging evolving context** – In most applications, both training and operational, the context of a dialog changes as the situation evolves. As a training scenario moves through stages of planning and execution, there is an implicit context that the training audience is aware of, which shifts over time. In other words, the concept of a thread takes varying meaning from a combination of several elements. A mission may be the central thread topic, but other factors such as the phase of operations and intelligence relating to multiple mission threads play key roles. In order to properly interpret chat communications with respect to any notional template of operational phasing context, analytical tools must either deduce contextual information from the chat stream, or rely on other sources of data. Deductions about the operational context must then be apparent to an instructor or reviewer who will be using such tools to sort through chat data.

- **Dialog modeling** – With informal chat communications, individual chat messages generally require the context of the containing dialog in order to be understood. Consider this simple exchange:

Player1: Do you see a motion sensor?
Player2: No

Only taken together can the meaning of the dialog (i.e., Player2 does not see a motion sensor) be determined. Modeling the structure of dialogs is particularly challenging, since even the simplest dialog situations, those where there are just two participants working within a limited domain, can give rise to complex dialogs. These complexities include:

- The presence of multiple sub-dialogs, each contributing to the meaning of the overall dialog;
- Suspension of a dialog while another topic is discussed, only to be picked up later, often without obvious cuing;
- Ambiguity about dependencies, meaning the links between a message and one or more preceding messages.

This complexity is magnified when there are more than two participants, since the range of possible dialog structures increases dramatically. Even the simple example given above regarding the motion sensor becomes more complex since there will potentially be a response from each participant.

- **Managing multiple conversations** – Unlike multi-party meetings, team-training situations can involve multiple conversations occurring simultaneously over a single communication channel. Again, consider a simple example:

Player1: Do you see a motion sensor?
Player3: How did you disable the door alarm?
Player2: No
Player4: I used a strong magnet.

Clearly there are two separate dialogs here. Each must be teased apart in order for the instructional system to recognize the outcome, relevance, or performance indicators in the dialogs.

Our research employs combinations of sorting capabilities using organizational and temporal context given by the new joint service guidance, keyword filtering techniques, and informed analysis considering statistically paired dialog participants. These techniques, when combined with visual timeline-based presentation of scenario ground-truth and key milestones in the planning processes, promise to provide a cogent and effective use of time in after-action reviews.

TRAINING DOMAIN

As a targeted training domain, the Air Force Research Laboratory's Training Research Exercise (TREX) provides a controlled research environment to investigate team performance dynamics in an air and space operations center. The environment allows mission-ready warfighters to practice their assigned duties using real-world systems in a scenario designed to test the full spectrum of decisions and coordination required in operational planning. The suite of systems includes collaborative planning tools, including chat rooms. As the warfighters conduct mission duties, researchers collect information on a variety of performance areas, leveraging chat as the complementary real-time communication mode in association with the suite of collaborative tools and shared situation awareness inputs available in an air operations center (AOC).

Exercise Objectives

The research objectives pursued in a TREX exercise are to: 1) Develop immersive scenarios to stimulate full team participation; 2) Develop tools to capture and validate team performance measures while conducting joint force planning for kinetic and non-kinetic effects; and, 3) Develop a synchronized suite of after-action review displays and tools to effectively communicate performance back to the team immediately after a training session. Immersive scenario development includes how to best present background material and real-time inputs to the team that mimic real-world operations tempo for planners. It also investigates the amount of material and efficiency of delivery to inform a planner sufficiently to execute their assigned tasks. Analysis of external environment and scenario control gives researchers insight on how to train focused teams in the AOC when they are not wrapped in the environment of the full operations center. The tools used to capture information about performance are both objective and subjective. Performance measurement applications mine data from collaborative tools regarding trainee actions and decisions and collect the full context of conversations and posted information. The *main effort* in research behind TREX is how to examine and effectively portray the large volume of data that is resident in joint operational planning.

Methodology

The research approach used in determining how to analyze and display information follows the operational planning methodology laid out in joint and USAF doctrine. The initiator for planning is normally a problem statement in the form of intelligence data or operational data reported to the team. The initiating report typically establishes a segregated planning approach to address the problem. The team then examines the problem in sequence with other planning tasks or a sub-team may be tasked to examine the issue in parallel with other team activities. In many cases, planning may be interrupted and take on an interleaved character. When a training session ends, trainees need to be able to see each problem in isolation, as well as in context with other workload. The isolation approach allows the team to review actual process versus doctrine, while the context of workload offers insight into time delays, distractions, errant information sources, and overall cognitive effort.

AAR Challenges

The most significant challenge to conducting an effective after action review of operational planning is to isolate processes efficiently for consumption by different members or subgroups within the training audience. Problems in operational warfare rarely involve an entire audience, since the team is composed of individuals with unique and non-overlapping areas of expertise. At the leadership level of the team, the decision makers must be able to track and review decisions in full view of the information available at the time to understand how well they acted on it. Planning specialists *involved in a process* will also want to segregate and review information pertaining only to the process in question. The specialists *not involved in a process* will want the review to move quickly enough to get to the next point in time where they are involved. After action review tools must help an instructor to sort and associate information with a unique process and be able to display information cogently to identify key areas that positively or negatively affected team and individual performance. This is true in the general sense, irrespective of the form that exercise data takes. Where chat logs are one of the primary sources of data indicating performance, tools for reviewing multiple chat logs in tandem become critical.

With the TREX exercise, the hardest information source to segregate and analyze is internet relay chat. In attempts to date, efforts to process chat data rely on query searches of chat databases or display of full context chat rooms with time synchronization. A single-level query string has limited value because it does not cleanly segregate an entire process for review. In practice, less than 10 percent of a process is typically captured using this method. Instructors conducting an AAR can sequentially query; however, each successive attempt dumps the previous data. Asking the instructor to become a query-building expert is beyond the scope of expertise and time available. Full-context chat room displays can employ limited room displays, but trainees typically will already have segregated full-context rooms they used during the mission available at their station during the AAR. At times, an instructor may find a full-context display useful to comment on distractions, workload, or cross-over information; however, for the majority of a review, only the chat associated with a process under review is required to keep the AAR effective and efficient.

CHAT ANALYSIS TOOL APPROACH

The goal for automated chat analysis tools oriented toward training is to ease the burden of the instructor by going beyond keyword-based search. Computers are at their best while crunching large quantities of data. However, in this case, most of the data is in the form of textual, natural-language dialogs, and understanding and analyzing natural language is a notoriously challenging problem for machines. While the communication standards for chat are being developed, participants still treat it as an informal communication medium, and use looser standards. Furthermore, analyzing the relationship between communications and operational mission outcomes requires a deep level of understanding of the content of the communications, their timing, and intent. With increasing numbers of participants, there is also an increased likelihood of distractions and tangential discussions. Automated, deep, content-based analysis of communications is still an open research problem, whereas tools for facilitating such analysis are needed immediately.

We hypothesize that a mixed initiative solution that leverages the strengths of the machine and the human is a feasible approach. The strength of the machine lies in data management, organization, filtering, presentation, and automated analysis for simple keyword-based and temporal-based patterns. The strength of the human lies in selecting analysis criteria and performing high-level, big-picture analysis. For example, with the TREX exercises, instructors have expressed a strong need for a tool that will classify the chat data according to the mission, further associate chat segments with different phases of a process, and provide complementary visualization that will clarify the communication flow within each process. Rather than supplanting instructional tasks, the goal is to facilitate them, so that instructors will be able to use their expertise efficiently to identify the training points and supporting data they wish to emphasize. Thus, the goal is to develop a tool that serves as a cognitive aid to instructors developing an AAR

Our approach to automating chat analysis for the purposes of an instructor's tool divides into capabilities to support two primary activities:

1. **Association and filtering**: In order to increase the speed and efficiency of putting together an AAR, automated natural language analysis and pattern recognition techniques produce a preliminary association between chat messages and specific processes and phases of the exercise. This association is the backbone of a filtering capability that instructors use to narrow the scope of the chat data they will be reviewing as they explore specific lines of inquiry into trainee decisions.

2. **Visualization and browsing**: Even with a filtered set of chat data, it is still a time consuming task to review synchronous conversation streams in multiple chat rooms and develop an understanding of the overall flow to identify performance indicators. This is the motivation for a tailored browsing capability that an instructor can use to review process-specific communications and visualize chronological relationships cross-referenced with exercise states. Typically, communications regarding a particular target will flow across multiple chat rooms, so synchronous browsing is a key feature. Additionally, the results of associations and filtering can be reflected in the browsing environment, as cues during the review process. For example, keywords related to a mission process that were detected in the filtering step will often be of interest to an instructor as highlighted terms while browsing.

The instructor uses these tools to focus on process-specific communications and draw their own conclusions about how the team's communication helped or hindered achieving the mission objectives. Notably, these two capability areas stop short of automated involvement in the instructional tasks of interpreting chat content for conclusions in support of training objectives. One might imagine an additional analytical capability for detecting and tracing certain kinds of failures reflected in the chat data. For example, weaponeering decisions that are apparent in the formulation of mission instructions could be the subject of automated review. And more complex performance measures such as communications effectiveness or even situational awareness can theoretically be gleaned via automated analysis. Although these forms of analysis are not contemplated in the research approach with this training domain to date, they would potentially be natural future additions, and may in fact carry added value for instructors in accelerating the AAR process.

PROTOTYPE IMPLEMENTATION

IDA (Intelligent Diagnosis Assistant) is a software tool that implements the approach described here. We are taking an incremental approach to the development, implementing increasingly complex rules of analysis. The system currently provides the following functionality.

Association and Filtering

An important analysis task is associating chat messages with specific processes and phases of the exercise. Once this association is made, instructors can filter the chat messages based on associated processes, thus narrowing the context of their analysis and discussions to relevant and meaningful units. Where automated methods can establish associations, one of the instructor's most time-consuming analytical tasks becomes much more palatable.

The objective of the associative mapping is to identify topics on the same thread, where a thread is defined for this domain as a mission. IDA first starts out with an untagged set of chat messages sorted in a chronological order. It incrementally tags the messages with associated missions, based on the patterns described below. Multiple passes are made over the message data to successfully refine the associations. It is possible for a message to be associated with multiple missions. IDA performs the following two types of analysis to recognize associations.

Keyword based: While keyword-based analysis can only uncover a fraction of the communication patterns of interest, they can uncover important cues to assist the human in the loop. IDA uses keyword-based analysis to classify chat utterances according to topics or missions. In particular, it uses mission-specific keywords to classify chat messages. The fact that each mission has a set of unique identifiers (e.g., mission numbers, code names for places or people, target types) is leveraged to tag chat messages. In the current version, mission-specific identifiers are specified in a configuration file prior to the analysis. A future step would be to automatically mine these identifiers from simulation data or even the chat stream itself.

In TREX, each mission is associated with a chat inject from a white force player (exercise controller). These injects can be analyzed automatically as they tend to pair mission identifiers with additional text which can contain unique mission characteristics such as target name, description, and location. The training use case also plays a role in how keywords are used for automated associations. Because the target use case is AAR, the analysis essentially has knowledge about the entire exercise, when considering any part of the exercise. So if a keyword associated with a mission appears in command and control software at a certain time during the exercise, this can potentially be used to match with chat messages composed at any time during the exercise, even preceding the “first known” correlated appearance.

While unique keywords help classify only a fraction of the chat messages by their mission, these tagged messages form the seed set for the next step of temporal pattern analysis.

Temporal pattern recognition: There are some types of temporal patterns that can be detected with reliable accuracy without the need to understand the content of utterances. An example is recognizing the pattern of a turn-by-turn interaction between two people in the same room (e.g. A says something to B and 3 minutes later B says something to A) and inferring that they belong to the same topic thread. Making an assumption of dialog coherence, one can say with a high degree of confidence that such conversation dyads refer to the same topic thread. The message classifications identified using the keyword-based approach are used as a basis to further identify and tag such pairs of massages.

Finally, the remaining messages are clustered according to the distribution of tags in the neighborhood of each message. A window around each message is analyzed, and the message is tagged with the most commonly tagged mission.

IDA uses quantitative confidence measures for the tagging. Since each mission has a unique set of identifiers, the keyword-based rules yield a 100% level of certainty in the identified tags. We are currently developing a set of heuristics to define the confidence measures for the temporal-based associations.

Chat Visualization and Browsing

To support causal analysis of exercise events, it is necessary to be able to review all the major events and communications in an exercise from various perspectives, including chronological, topical, role-based, and others. The information of interest will typically be spread over several chat rooms, and the visualization tool must make it easy for analyzers to follow the information flow across the various rooms of interest. Most of the existing chat visualization tools restrict their perspective to a single chat room and are limited to search-based analysis. There are none that support visualization of cross-room chronology and flow of information. One main objective of the IDA debriefing tool is to provide multi-perspective view of chat from multiple rooms for the purposes of analysis.

Figure 1 (Page 12) shows the primary visualization view implemented with the prototype. First and foremost, the IDA tool supports simultaneous, synchronous browsing of multiple chat rooms, while preserving chronology, thereby making it possible to follow the communications across time and across rooms. The user has the option of turning off the chronological synchrony when this gets in the way of analysis. With synchronous scrolling, the user can browse through the chat data exactly as it unfolded in the exercise. The tool is capable of showing one main chat window and four supporting chat windows. The rooms to be displayed in these windows are currently user-selectable. We are developing rules for automatically configuring the windows based on the phase of the exercise, the mission under consideration, and the density of chat traffic in the rooms.

Each window is called a **channel**, and it contains the chat lines for a particular room. A chat line consists of a time, speaker, and message. The chat lines across the channels are aligned in time, line by line. As a result, there may be many blank chat lines (other than a time) in a channel, in order to maintain a time alignment with the other channels. Channels can be scrolled synchronously or independently.

The IDA chat visualizer provides information at three levels of detail. A timeline, shown along the right side of Figure 1, provides a birds-eye view of the distribution of communication in the selected chat rooms, without providing the details of communications. The purpose for this is to establish the overall temporal context in a manner consistent with other tools utilized on adjacent screens during the after action review process. Individual channels are represented in the timeline, with independent markers to show the current temporal location of selected chat lines in each channel.

Figure 2 shows a closer view of an individual channel display. The chat channels provide the next level of detail, showing the time stamps, the sender, and the first line of the message content. The idea is to provide a summary reference of the channel that can be quickly and easily scanned. A movable magnifying lens within the channel display provides the third level of detail. It shows the entire contents of the selected chat line in larger text, using multiple lines if necessary.

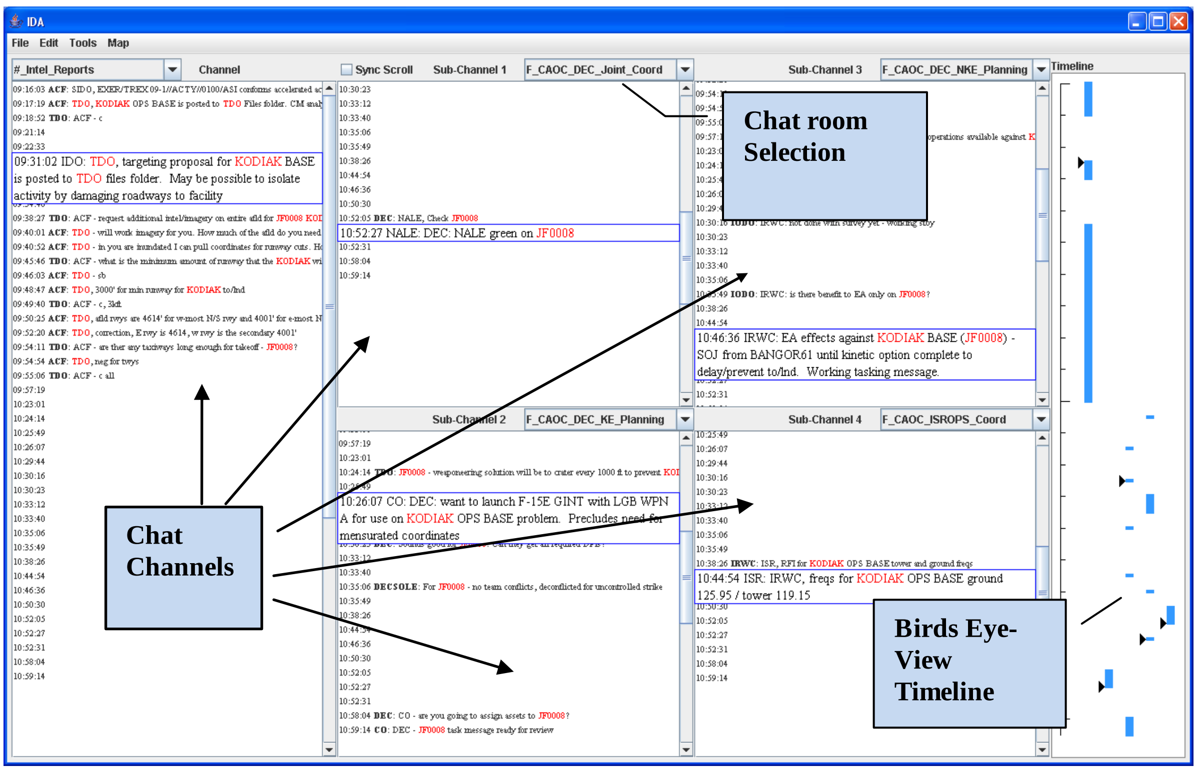

Figure 1. The IDA Visualization Tool

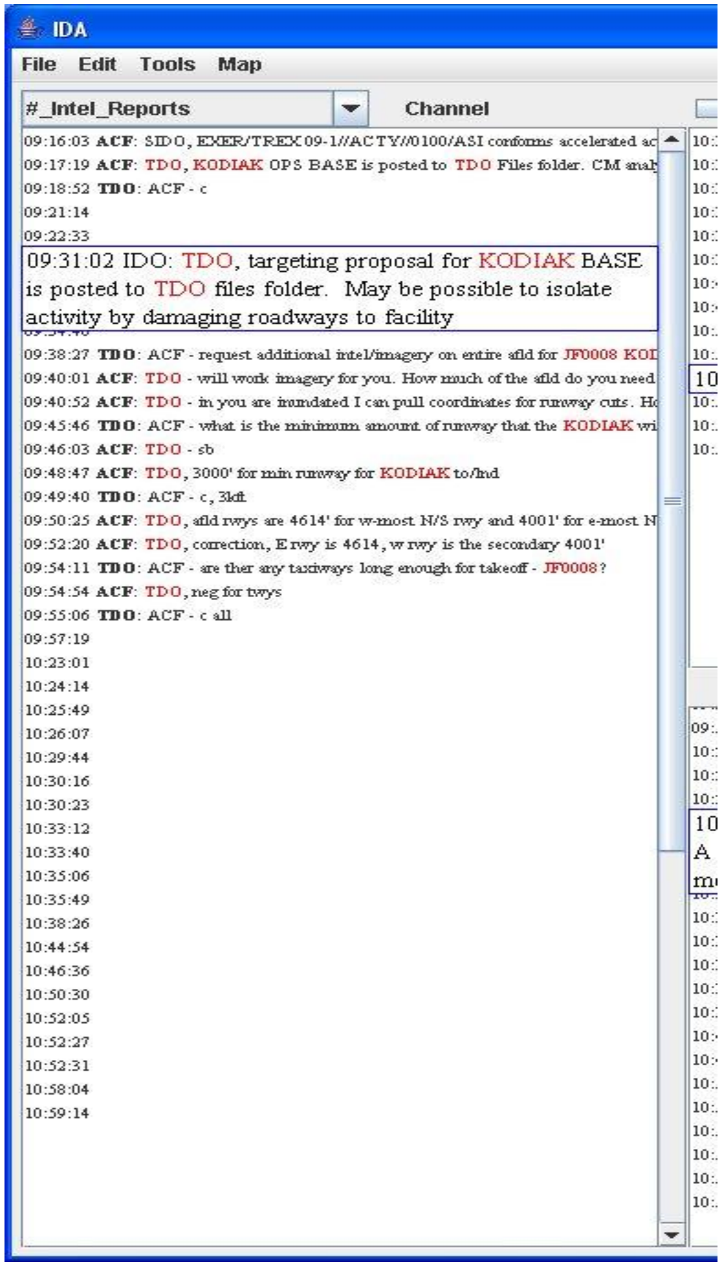

Figure 2. Zoomed View of a Chat Channel in IDA

This visualization leverages the analysis and filtering capabilities in several ways. Upon filtering a chat stream based on the tags automatically generated for each message, the user may elect to see only the filtered result, or may still prefer to see the entirety of the chat data. In the latter case, color coding differentiates the matching messages from non-matching. Where messages are included in the filtered result by virtue of a keyword match, the relevant keyword is highlighted. For similar rule-tracing reasons, dependency relationships between messages can be indicated to instructors through a menu option.

CONCLUSIONS AND THE WAY FORWARD

The current implementation of the prototype provides lessons learned not only on the technical challenges involved, but also for directions for future research.

Additional Planned Development

There are several areas currently planned for further research, including the following.

Secondary Keywords

We intend to explore the use of probabilistic methods to identify additional keywords that may be uniquely associated with a mission, beyond those that are specifically retrievable from a reference list or from exercise injects. This essentially involves an iterative review of the space of messages that can be tagged as known matches. Using this method to find additional unique keywords can potentially refine the retrieval results with the inclusion of messages matching secondary keywords.

Probabilistic Communication Pattern Associations

The heuristics being implemented do not exploit other domain characteristics, such as the communications plan or the different mission phases, for the tagging of messages. IDA could use the comms plan for deeper analysis of the chat messages. For example, in the case of conversational triads, where more than two roles are involved in a dialog on the same process thread, a probabilistic approach that takes into account a baseline of knowledge about assigned roles and communication flow may be able to identify links between messages that don't have otherwise apparent ties. Also, when the responsibilities for different roles are well-defined for each phase of a mission, IDA could analyze the chat messages to detect patterns of people acting "out of role."

Instructional Interventions

Discussions with potential end users have revealed a desire for capabilities to support instructional interventions to adjust the automated associations for a specific chat data set. An instructor may want to select specific words within a chat message as keys to either include or exclude from the matching set for a given mission thread. For example, the word "scud" might appear in the set of target description keywords for more than one mission. In such a case, an instructor may want to manually adjust the association rules so that appearances of this word do not qualify alone as unique match keywords for a given mission. Similarly, conjoined words (e.g., "scud garrison") may have a more one-to-one relationship with missions, so an instructor may want to replace the consideration of either of these two words in isolation with matching on instances of their adjacent pairing. Further, this adjustment capability also suggests a need to determine the scope of resulting changes. In other words, an instructor may prefer for an adjustment to apply only to an individual message, or may prefer that it apply to all messages as well as any messages included by virtue of other temporal or dialog relationships with the adjusted message.

Analytics Targeting Training Objectives

As mentioned earlier in this paper, another future research goal is the expansion of automated analytical methods to make further progress with assessments of performance with respect to training objectives. Performance measures might range from the simpler observations of weaponeering choices and response times following injects, to more complex conclusions regarding communication methods and situational awareness.

Conclusions

This research aims to develop tools that can be used routinely for training exercises where chat messaging plays a major role, both as an operational medium and as a source for training evaluation. Initial experiments are planned for the coming year, to test out the utility of the chat analysis methods and visualization tools in the fundamental goal of helping instructors perform after action review quickly and incisively. The design is specifically intended to operate as one of potentially several applications used in after action review for different forms of playback or other indications of exercise events and trainee actions. The underlying methods for automatically detecting and differentiating threads of operational conversation could theoretically apply to many training domains, where common features of unique keywords and definable roles and communications patterns are present.

ACKNOWLEDGEMENTS

The research effort described in this paper is sponsored by the Air Force Research Laboratory.

REFERENCES

Cakir, M., Xhafa, F., Zhou, N., and Stahl, G. (2005), "Thread-based analysis of patterns of collaborative interaction in chat", paper presented at the *Conference of Artificial Intelligence for Education (AIEd'05),* Amsterdam, Netherlands.

Herring, S. C., & Kurtz, A. J. (2006). “Visualizing dynamic topic analysis”, *Proceedings of CHI'06*. New York: ACM Press.

Niekrasz, J., Gruenstein, A., & Cavedon, L. (2004). Multi-human dialog understanding for assisting artifact-producing meetings. In Proceedings of the 20th International Conference on Computational Linguistics (COLING).

Shi, S., Mishra, P., Bonk, C. J., Tan, S., & Zhao, Y. (2006). “Thread theory: A framework applied to content analysis of synchronous computer mediated communication data”, *International Journal of Instructional Technology and Distance Learning*, *3*(3), 19-38.

Tur, G., Stolcke, A., Voss, L., Dowding, J., Favre, B., Fernandez, R., Frampton, M., Frandsen, M., Frederickson, C., Graciarena, M., Hakkani-Tür, D., Kintzing, D., Leveque, K., Mason, S., Niekrasz, J., Peters, S., Purver, M., Riedhammer, K., Shriberg, E., Tien, J., Vergyri, D., & Yang, F. (2008). “The CALO meeting speech recognition and understanding system”, In *Proceedings of the 2008 IEEE Workshop on Spoken Language Technology*.

Zimmermann, M.; Liu, Y.; Shriberg, E. & Stolcke, A. (2006). “Joint Segmentation and Classification of Dialog Acts in Multiparty Meetings”. In *Proceedings of IEEE ICASSP*, Toulouse, France (2006).