

# Story Management Technologies for Organizational Learning

**Andrew S. Gordon**

(University of Southern California, Los Angeles, CA, USA  
gordon@ict.usc.edu)

**Abstract:** The stories told among members of an organization are an effective instrument for knowledge socialization, the sharing of experiences through social mechanisms. However, the utility of stories for organizational learning is limited due to the difficulties in acquiring stories that are relevant to the practices of an organization, identifying the learning goals that these stories serve, and delivering these stories to the right people and the right time in a manner that best facilitates learning. In this paper we outline a vision for story-based organizational learning in the future, and describe three areas where intelligent technologies can be applied to automate story management practices in support of organizational learning. First, we describe automated story capture technologies that identify narratives of people's experiences within the context of a larger discourse. Second, we describe automated retrieval technologies that identify stories that are relevant to specific educational needs. Third, we describe how stories can be transformed into effective story-based learning environments with minimal development costs.

**Keywords:** Storytelling, Knowledge acquisition, Natural language processing, Information storage and retrieval, Simulation-based learning environments

**Categories:** M.0, I.2.7, H.3.3, L.5.1

## 1 Story-Based Organizational Learning in the Future

Imagine that you, several years from now, are far from home conducting some business for your organization. Perhaps you are a businessman negotiating a new contract with a company in your supply chain, or perhaps you are an international aid worker trying to establish a clean water supply in an impoverished region. The work is gruelling, but is always interesting. The rapid pace of change brings you new experiences that you never expected when you first trained for this job, constantly forcing you to improve your skills in new areas. Keeping in touch with your colleagues back home is a priority for you. Each night, when both time zones are reasonable, you use your smart-phone to socialize with others in your organization - to tell them about the bizarre event that happened to you that day, or simply to hear the latest office gossip.

As usual after a long phone call, your smart-phone has some questions for you. In listening to the conversation you were having, it noticed that you told some new stories - different in some way than the sort of stories that it has heard from you in the past. It plays back some of the audio of these stories for your review. Listening to yourself tell them, you find that many of these stories are narratives of the personal encounters you have had on your journeys. However, some of them are specifically about the job you are there to do, ones that you would be happy to share more broadly

with the rest of the people in your organization. After sorting out which is which with your smart-phone, and re-recording the ones that could benefit from a better delivery, you authorize it to pass along the organizational stories into the story management pipeline.

Contributing in this way to your organization's story management pipeline is a bit more work for you to do, but you can see the value of it each morning as you open your laptop to prepare for the day's new challenges. For this job, preparation requires practice, so you spend an hour each morning honing your skills with a simulation-based training application. Each morning you are faced with new simulated situations that challenge your abilities, perhaps ones that present you with new problems for negotiating your supply chain contract or ones that introduce new considerations for laying out fresh water filtration systems. After completing each new simulation scenario, you always check to see what real-world story was used as the basis for the fictional situation. What was the lesson learned from real-world experience that made its way into the training application you are using this morning? Usually these scenarios are based on the experiences of people you know from your organization, often told into their smart-phones only days or weeks earlier - just like you had done the night before.

## **2 Story-Based Organizational Learning in the Present**

Many aspects of this vision of story-based organizational learning in the future, where real-world experiences are captured and transformed into fictional simulation-based learning environments, are already exhibited in the way that some organizations develop new training technologies. In their work to develop immersive training applications for the United States Army, the Institute for Creative Technologies at the University of Southern California has created a number of prototypes that transform the real-world experiences of U.S. Army soldiers into fictional video and virtual-reality training applications. In the Army Excellence in Leadership project [Hill, 04], real-world stories told by junior U.S. Army officers are fictionalized into Hollywood-style movies, then embedded in a leadership training application that allows trainees to interactively interview characters from the films. In the ICT Leaders project [Gordon, 04], real-world stories collected from U.S. Army captains are translated into decision points in a fictional branching storyline, allowing trainees to practice leadership skills by making decisions through conversations with virtual characters in a game-based simulation. ELECT BiLAT [Hill, 06] is a game-based simulation for soldiers to practice conducting bilateral engagements by negotiating with virtual characters, where scenarios are based on the experiences told to developers during interviews with subject-matter experts.

Although each of these examples targets different training objectives and utilizes different delivery methods, each follows a similar philosophy toward the role of stories in organizational learning. In this context, stories are defined as a genre of discourse where the speaker provides a narrative description and interpretation of causally-related past events. The real-world experiences of members of the U.S. Army, told as stories, are used as the basis for fictional scenarios to be experienced by trainees in multimedia learning environments. As such, these projects might be characterized as an *extremely mediated* form of traditional storytelling, where the

lessons learned from the original experience are painstakingly transformed into products that can most effectively deliver these lessons to their target audience.

Unfortunately, all of this mediation comes at a price, both in terms of time and money. These projects, each of which required millions of U.S. dollars and years of effort by large development teams, cannot be endlessly replicated to process all of the stories told by members of an organization, even wealthy organizations that are only a fraction of the size of the U.S. Army. In accordance with the original intent of their development, these systems are extremely effective at delivering a modest number of lessons-learned for a modest number of training objectives. However, they do not represent a general knowledge management solution for story-based organizational learning, where the time between real-world experience and the training of others in the organization needs to be nearly instantaneous, with negligible costs.

In 2004 we began the Story Representation and Management project, a new research effort at the Institute for Creative Technologies at the University of Southern California to develop intelligent technologies for automating the labor required to create multimedia learning environments based on the real-world stories told by members of an organization. We focused specifically on three areas that required the greatest amount of time and effort in the development of the U.S. Army training applications mentioned above. First, *story capture* technologies were developed to automatically identify stories in discourse, obviating the need to collect these stories from practitioners through directed interviews. Second, *story retrieval* technologies were developed to automatically locate stories in a corpus that are directly relevant to particular learning objectives. Third, new designs for *interactive story-based learning environments* were explored, aimed at minimizing the costs of developing effective computer based training. Collectively, these technologies constitute a *story management pipeline*, and represent the first steps toward the realization of the vision of story-based organizational learning described in section 1. Our efforts in each of these areas are described in the next three sections.

### 3 Technologies for Story Capture

The primary methods that are currently used to gather stories for use in organizational knowledge management and training applications involve face-to-face interviews with subject-matter experts, e.g. as part of a cognitive task analysis [Clark, 07], in small group meetings [Snowden, 00], or with directed story-elicitation interviews [Gordon, 05a]. The facilitator/interviewer that participates in these methods is typically not a member of the organization, and their primary role is to ensure that a large number of stories are communicated and recorded that are relevant to specific knowledge management or training needs of the organization during the course of the interview. However, studies of casual storytelling in workplace environments have shown that stories relevant to the organization are common in everyday, non-facilitated conversations among co-workers [Coopman, 98]. If these stories could be captured directly from conversations among co-workers in an automated manner, then the costs associated with collecting stories relevant to organizational practices could be dramatically reduced. Although there are substantial privacy and user-interface challenges associated with extracting stories from conversations among co-workers, our research efforts in this area were specifically focused on technical feasibility.

In our first attempt in this area, we explored the use of machine-learning techniques for identifying stories in segments of conversational speech, using the words recognized with commercial speech-recognition software [Gordon, 05b]. We followed a traditional text classification approach, where a corpus of transcribed conversational speech was first hand-annotated (story / non-story) for use as training and testing data. By developing a clear definition of what counted as a story, our annotators were able to achieve reasonably high inter-rater agreement. Segments of training data were then encoded as high-dimensional feature vectors (word-level unigram and bigram frequency counts) and used to train a naïve Bayes binary classifier. To apply this classifier to test data, overlapping consecutive segments of test data were individually assigned to either the story or non-story class, with confidence values smoothed across segments using a simple mean-average smoothing function. Performance evaluations of our approach yielded low precision (39.6%) and low recall (25.3%), equal to random chance performance on this task. However, we observed substantially higher performance when using transcribed test data (as opposed to the output of a speech recognition system), with reasonable precision (53.0%) and recall (62.9%) scores. We concluded that significant advances in open-domain continuous speech recognition would be required in order to construct a usable story-capture system for casual conversations.

Given the low performance of story capture from speech data we decided to shift our focus to written electronic discourse, specifically weblogs. By randomly sampling weblog entries on the Internet, we found that between 14% and 17% of the text in weblog entries consisted of stories, i.e. narrative descriptions and interpretations of the author's past experiences [Gordon, 07]. To apply our existing story extraction technology to Internet weblog entries, we created a new hand-annotated (story / non-story) corpus of weblog entries for use as training and test data, and achieved reasonable performance levels (precision = 30.2%, recall = 80.9%, F-score = 0.414). By incorporating techniques for automatically detecting sentence boundaries in the test data, utilizing a contemporary support-vector machine learning algorithm, and using a Gaussian function to smooth the confidence values, we were able to significantly improve the overall performance of this approach (precision = 46.4%, recall = 60.6%, F-score = 0.509). Although these performance levels leave some room for improvement, we believe that it is high enough to explore the integration of automated story capture technologies in productivity software applications used by organizations (including email, online forums, and general-purpose word processing).

## 4 Technologies for Story Retrieval

The second key challenge in automating the management of organizational stories is connecting people who have specific learning needs with the stories that are most relevant to these needs. This problem does not arise when a modest number of stories are collected through directed interviews as part of the development of specific training applications, where the learning objectives are known ahead of time. However, as the size of organizational story collections becomes very large, and where these stories may have relevance to unanticipated learning needs, it is critical to have some mechanism for automatically pairing available stories with new learning objectives. Previous work on this problem has focused on story *indexing* (e.g.

[Domeshek, 92][Johnson, 00]) where human analysts identify the specific points that are made by a story, and encode these points using formal representation to facilitate their retrieval. Point-based indexing of this sort is particularly suited for cross-domain retrieval, e.g. where a story about conducting contract negotiations with a company in a supply chain illustrates a point that is directly relevant to establishing a clean water supply in an impoverished region. In our own work, we have not pursued this style of point-based indexing. Aside from the difficulties in automating approach, cross-domain stories have not proven to be particularly useful in the development of the training applications mentioned earlier in section 2. Instead, we have seen that the stories most relevant to a learning need are within the same task domain or activity, e.g. the stories that are most useful for learning how to conduct contract negotiations are about the task of conducting contract negotiations. The technical challenge is to automatically select stories that describe experiences in the context of task or activity.

Our approach to this automation challenge was to use techniques for textual information retrieval, where textual descriptions of activities or tasks are used as queries for the retrieval of textual stories from a collection. To explore the feasibility of this approach, we constructed a large-scale retrieval system for finding stories automatically extracted from Internet weblogs [Gordon, 08]. In this system, called StoryUpgrade, users described an activity as a paragraph-sized description of events expected to occur in the task, written as a past-tense first-person narrative (a “boring story” of the activity). The system then encoded the query as a weighted vector of lexical features, and then ordered the relevance of stories in the collection using vector-based similarity measures. Two versions of this system were built that differed in the way that they collected stories from Internet weblogs. In the first version, we created a single collection of stories by applying our story capture technology (described earlier in section 3) to 3.4 million weblog entries, yielding a text corpus of over one billion words. In the second version, we constructed specialized story collections for each query by first sending fragments of the activity description to a commercial weblog search engine, and then applying our story capture technology to hundreds or thousands of the top search results.

We evaluated the effectiveness of the StoryUpgrade system (version 1) as a tool for finding stories relevant to the development of simulation-based training applications. The Intelligent Convoy Operations Support (ICOS) project at the University of Southern California’s Institute for Creative Technologies was an effort to develop a tutoring module to augment a training simulation for military convoy operations. Four members of the ICOS project team used the StoryUpgrade system to find relevant stories about military convoy operations told in Internet weblogs. In one forty-five minute session, these participants authored five activity descriptions of the convoy operations task and judged the relevance of 67 retrieved weblogs. There were 23 relevant stories, 6 relevant non-stories, 14 non-relevant stories, and 24 non-relevant non-stories (55% story precision, 43% relevance precision). We concluded that the performance of our approach to pairing task descriptions to stories in the collection was not high enough to fully automate this process, but that a developer could efficiently use this technology to find a sufficient number of task-relevant stories in very large collections in less than an hour.

## 5 Interactive Story-Based Learning Environments

Contemporary computer-based education and training systems are effective to the degree in which they can incorporate four complimentary design characteristics. Systems should support *situated learning* by presenting learners with situations that mirror the contexts in which their acquired skills will be executed. The acquisition of these skills is more effective when learners are given one-on-one *guided instruction* from an experienced tutor. *Multimedia* presentation can help people learn certain types of information, particularly when they encourage dual coding of the information, when the media support one another, and when learners have low prior knowledge or aptitude for the subject matter. Finally, educational experiences should be set in a rich *narrative environment*, supporting the learner's retention of educational experiences as remembered stories. The adoption of these four design characteristics has become nearly universal in the development of immersive learning environments, as in each of the systems described earlier in section 2. Unfortunately, the development costs (time and money) of these systems are also universally enormous, typically including the labor of computer programmers, scriptwriters, 3D graphic artists, animators, voice talent, and sound engineers. In our own research, we explored the opposite extreme of this design space: what is the minimal-cost design for computer-based education that supports situated learning and guided instruction in a multimedia narrative environment?

Our design solution is a web-based application called Fourth Frame Forums, which combines traditional four-frame comic strips with online web-based discussion forums [Gordon, 06]. In this application, users are presented with a four-frame comic strip that depicts a fictional problem scenario, where the protagonist is faced with a decision. The last dialogue balloon in the comic, where the protagonist decides what to say to resolve the problem, is left blank. By typing a statement directly into this dialogue balloon, the user creates a new discussion thread in an online forum, where the user's dialogue choice can be critiqued and discussed by other users of the forum, guided by an experienced mentor that moderates this discussion.

The central innovation of the Fourth Frame Forums application is the authoring methodology, which is designed to allow a training developer or instructor to quickly create new Fourth Frame Forum scenarios based on non-fiction stories of real-world experiences gathered from practitioners in the organization. The approach mirrors the techniques developed for authoring branching storylines for training applications [Gordon, 04], where stories are analyzed to identify the specific expectations about the task that they violate (the reason they are surprising and interesting). These expectations are then used to craft a fictional situation where a decision is to be made by the user, and where their choice will be dependent on whether or not they believe the expectations that were violated in the real-world story.

As part of our original work on the Fourth Frame Forums application, we created a series of fictional scenarios based on real-world stories involving leadership as told by U.S. Army Captains interviewed for the Army Excellence in Leadership project [Hill, 04]. To visualize these scenarios as four-panel comic strips, we gathered a number of screen captures from the game-based virtual-reality environment used in the ICT Leaders project [Gordon, 04], and used a standard graphical illustration application to arrange these screen captures into the four-frame comic format with the

appropriate dialogue balloons and narration boxes. As a more general solution for visualizing scenarios relevant to a broader set of tasks and organizations, we have found that contemporary “life simulation” computer games (e.g. *The Sims*, by Electronic Arts) provide an ample array tools for quickly creating scenes of customized characters in custom environments, and for exporting screen captures of these scenes for use in new Fourth Frame Forum episodes. Figure 1 shows a proof-of-concept comic authored in a couple of hours using this approach.

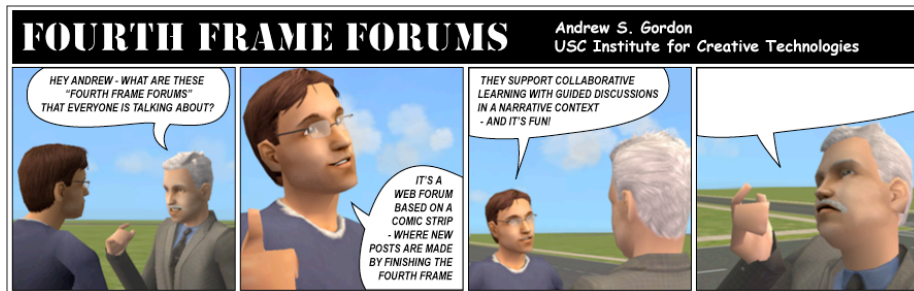


Figure 1: A Fourth Frame Forum visualized using scenes from a computer game

## 6 Conclusions

Achieving the vision of fully automated story management for organizational learning presented in section 1 will require overcoming many remaining technical challenges. However, there are key processes that can be automated to lessen the development costs of story-based learning environments, making them practical as an *extremely mediated* form of organizational storytelling. In this paper, we described research efforts to develop three complimentary story-management technologies for reducing the costs associated with collecting organizational stories, retrieving stories relevant to organizational learning needs, and transforming these stories into computer-based education that supports situated learning and guided instruction in a multimedia narrative environment. There remains a need for technology that can automatically capture stories from conversational speech, but we have demonstrated that stories can be extracted from extremely large text corpora with reasonable accuracy. Likewise, we have shown that finding relevant stories can be viewed as a text retrieval task using activity-based queries, but these techniques are not applicable to non-textual story collections and do not service retrieval needs based on the specific points of stories in the collection. Finally, Fourth Frame Forums offer an extremely low-cost solution for authoring story-based learning environments, but still requires the effort of a training developer or instructor. Collectively, these technologies allow a single training developer, over the course of a single day, to search through stories collected from the weblogs of members of an organization, identify ones that are relevant to specific organizational tasks, and author (and deliver) an effective story-based learning environment for immediate use throughout the organization.

## Acknowledgements

The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

## References

- [Clark, 07] Clark, R., Feldon, D., vanMerriënboer, J., Yates, K., Early, S.: Cognitive Task Analysis. In Spector, J., Merrill, M., van Merriënboer, J., & Driscoll, M. (Eds.) Handbook of research on educational communications and technology (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, 2007.
- [Coopman, 98] Coopman, S. & Meidlinger, K.: Interpersonal Stories Told by a Catholic Parish Staff. *American Communication Journal* 1(3), 1998.
- [Domeshek, 92] Domeshek, E.: Do the Right Thing: A Component Theory for Indexing Stories as Social Advice. PhD Thesis, Yale University, 1992.
- [Gordon, 04] Gordon, A.: Authoring Branching Storylines for Training Applications. 6th International Conference of the Learning Science, Santa Monica, CA, June 22-26, 2004.
- [Gordon, 05a] Gordon, A.: The Fictionalization of Lessons Learned. Guest Editorial for Media Impact Column, *IEEE Multimedia* 12(4):12-14. 2005.
- [Gordon, 05b] Gordon, A. & Ganesan, K.: Automated Story Extraction From Conversational Speech. 3rd Intl. Conference on Knowledge Capture, October 2-5, 2005, Banff, AB.
- [Gordon, 06] Gordon, A. (2006) Fourth Frame Forums: Interactive Comics for Collaborative Learning. 14th ACM Intl. Conference on Multimedia, October 23-27, 2006, Santa Barbara, CA.
- [Gordon, 07] Gordon, A., Cao, Q., & Swanson, R.: Automated Story Capture From Internet Weblogs. 4th Intl. Conference on Knowledge Capture, October 28-31, 2007, Whistler, BC.
- [Gordon, 08] Gordon, A. & Swanson, R.: StoryUpgrade: Finding Stories in Internet Weblogs. International Conference on Weblogs and Social Media, March 31-April 2, 2008, Seattle, WA.
- [Hill, 04] Hill, R., Gordon, A., and Kim, J.: Learning the lessons of leadership experience: Tools for interactive case method analysis. 24th Army Science Conference, Orlando, FL, 2004.
- [Hill, 06] Hill, R., Belanich, J., Lane, H., Core, M., Dixon, M., Forbell, E., Kim, J., Hart, J.: Pedagogically structured game-based training: Development of the ELECT BiLAT simulation. Proceedings of the 25th Army Science Conference, 2006.
- [Johnson, 00] Johnson, C., Birnbaum, L., Bareiss, R., Hinrichs, T.: War stories: harnessing organizational memories to support task performance. *Intelligence* 11(1): 16-31. 2000.
- [Snowden, 00] Snowden, D.: The art and science of Story or 'Are you sitting uncomfortably?': Part 1: Gathering and harvesting the raw material. *Business Information Review*, 17:147-56, 2000.