Proactively Suggesting Similar Past Stories Turns "Lessons Learned" Into "Lessons Used"

Eric Domeshek, Daniel Tuohy & James Ong
Stottler Henke Associates, Inc.
1650 S. Amphlett Blvd., Suite 300
San Mateo, CA
Domeshek|Tuohy|Ong@stottlerhenke.com

David Spangler
425 Flintlock Road
Chesapeake, VA 23322
spanky3745@gmail.com

Thomas Williams
3388 Jamaica Blvd S.
Lake Havasu City, AZ 86406
thomwilliams@hotmail.com

ABSTRACT

A common failing of existing lessons learned libraries is that important experiences are stored, but are then rarely retrieved when they might usefully inform decision making. Searching for lessons in existing systems is often seen as an inconvenient, time-consuming disruption when pressing decisions need to be made. We describe an approach to providing advice *proactively*, by extending task support tools to automatically formulate queries against a repository of past experiences. Retrieved text documents are offered *in context* to minimize disruptions and maximize connections with ongoing work. In order to improve recall precision relative to traditional text-based methods—which suffer from natural language's variability, vagueness, ambiguity, and gaps in what is explicitly stated—we experimented with matching structured representations of a current problem and solution under consideration.

In order to evaluate the utility of our approach, we built a prototype system to support operational planners working on Counterinsurgency (COIN) and Stability, Security, Transition and Reconstruction (SSTR) operations. To explore what an ideal information extraction system might usefully produce, team subject matter experts selected and annotated two-hundred experiential texts containing potential planning lessons. Then, a panel of forty-nine active duty and retired officers planned hypothetical missions using a simple web-based planning tool that was enhanced by integrating our automated lesson retrieval tool.

Based on user surveys, 69% of evaluators agreed that the system helped with planning; 92% agreed that the concept of experiential advice was valid and potentially useful for military planners; 88% agreed that the system should be further developed; and 82% agreed that work on the integrated planning tools should be continued. We asked users to rate the quality of the plans they developed, with and without automated story retrieval. A paired t-test showed that retrieval support led to a small but significant increase in their self-ratings of plan quality.

ABOUT THE AUTHORS

Dr. Eric Domeshek is an AI Project Manager at Stottler Henke Associates, Inc. where he leads and supports projects applying AI technology to problems in training and decision support. He has developed a wide range of knowledge management and collaboration tools, spanning applications to aerospace and other systems design, as well as military campaign design and planning. His training systems have emphasized application of automated Socratic tutoring techniques in domains such as tactical decision making, medical diagnosis, and technology troubleshooting. Dr. Domeshek received his Ph.D. in Computer Science from Yale University, focused on case-based reasoning. For his dissertation, he developed representations of decision rationale for social situations, intended to support case retrieval; this included extensive representations of characters' relationships, traits, and motivational structures. He served as research faculty at the Georgia Institute of Technology College of Computing where he contributed to the development of a line of case-based design aids. As an assistant professor at Northwestern University, he developed multimedia libraries and goal-based scenario training systems at the Institute for the Learning Sciences.

Mr. Daniel Tuohy is a Senior AI Engineer at Stottler Henke Associates, Inc. where he has served as technical lead on several projects that apply AI techniques and model-driven design principles to create robust collaboration tools. Those tools have focused on mission planning, campaign design, and lessons-learned indexing and retrieval. Mr. Tuohy has extensive research and development experience in AI, including knowledge representation, optimization, machine learning, and probabilistic models. His formal training in computer science and AI focused on machine learning and optimization using Neural Networks and Genetic Algorithms (GAs). His graduate thesis research applied GAs and machine learning to problems in graph theory and the automated arrangement of musical compositions for the guitar, and extended Hidden Markov Model algorithms for use in gene prediction from biological sequence data.

Mr. James Ong is an AI Group Manager at Stottler Henke Associates, Inc. He specializes in the application of artificial intelligence and information visualization technologies to create tools and solutions for autonomous systems, interactive and automated planning, decision and task support, and training. His experience at Stottler Henke, PPD Inc., Bolt Beranek and Newman, and AT&T Bell Laboratories spans applied research and advanced development, systems engineering, software products and solutions, management consulting, and product management. Mr. Ong received an MBA from Boston University, an MS in computer science (artificial intelligence) from Yale University, an MS in electrical engineering and computer science from the University of California at Berkeley, and an SB in electrical engineering from the Massachusetts Institute of Technology.

Commander David Spangler, U.S. Navy (Ret.) is Managing Member, Global Innovation and Design Associates. During his 26 years of active duty he had experience as a certified Navy Strike planner, operational-level planner, and joint planner, including teaching joint planning as an Assistant Professor at Joint Forces Staff College. He was the sole subject matter expert (SME) for the focused training event on strategic communication for all UNIFIED ENDEAVOR exercises to train unit rotating into Iraq and Afghanistan. All communication efforts to inform and influence a wide variety of audiences required detained coordination and planning across many interorganizational partners and integration into decision making at all levels. Airpower, C2, and targeting SME on real-world operations and exercises since 1995, led the NATO analysis teams for C2, civil-military affairs, and air integration. He has supported numerous highly successful SBIR projects at Stottler Henke.

Colonel Thomas Williams, USMC (Ret.) is the founder of Shamshir Technology Consultants, which supports early stage programs and tackles operational and technology risk challenges. He has a myriad of operational and supporting establishment experience in Naval, Joint and NATO environments. He also has an extensive background in supporting, managing, directing and transitioning S&T programs at the Missile Defense Agency as an Operational Concepts Developer, Office of Naval Research as a Military Deputy and a civilian State agency Utah Science, Technology and Research Agency (USTAR) as Managing Director.

Proactively Suggesting Similar Past Stories Turns "Lessons Learned" Into "Lessons Used"

Eric Domeshek, Daniel Tuohy & James Ong
Stottler Henke Associates, Inc.
1650 S. Amphlett Blvd., Suite 300
San Mateo, CA
Domeshek|Tuohy|Ong@stottlerhenke.com

David Spangler
425 Flintlock Road
Chesapeake, VA 23322
spanky3745@gmail.com

Thomas Williams
3388 Jamaica Blvd S.
Lake Havasu City, AZ 86406
thomwilliams@hotmail.com

INTRODUCTION

Complex decisions must generally be based on a synthesis of theory and experience. This is why many of the best military leaders are so interested in history. Social theories at the level that might be useful for military decision-making are generally incomplete and insufficient—especially when military leaders must cope with a broad range of missions, well beyond force-on-force combat. Questions about consequences of military action do not stop at whether we can destroy enemy forces. Planners must consider how our actions might affect Political, Military, Economic, Social, Information, and Infrastructure (PMESII) systems.

All the services have recognized the importance of learning from experience and have established formal "lessons learned" programs. However, most high-stakes decision-making—and especially, *military* decision-making—is carried out under time pressure. There is always too much to know, too much that is unknown, and too little time to fill in the gaps. Thus, most "lessons learned" systems are really just "lessons captured" systems. Under pressure to generate alternatives and settle on promising Course of Actions (COAs), senior leaders feel they lack the time to trawl through immense stores of long documents, hoping to find useful nuggets of information.

A repository of wisdom cannot contribute to wise decisions unless someone absorbs and applies the right lessons at the right time. Thus, we must shift the perceived balance of costs and benefits when it comes to exploiting lessons learned libraries—or really any source of decision-making advice. Some relevant strategies include: (1) minimizing effort and disruption by proactively offering planners advice-containing documents within a task-focused workflow; (2) minimizing the time required for relevance assessment by synopsizing retrieved items and their relationships to the current situation; and (3) improving retrieval so that more offered items are likely to be relevant to active decision topics. The work reported here explored all three strategies, emphasizing the first and third approaches.

CONTEXT

As our sample decision-making context, we explored early-stage military planning at the low operational level, with a focus on Counterinsurgency (COIN) and Stability, Security, Transition and Reconstruction (SSTR) operations. We chose COIN and SSTR for two reasons. First, such operations were the dominant activity for U.S. forces over the last decades, so there are many relevant documented experiences in the open literature. Second, COIN/SSTR operations require decision-making in complex, often unfamiliar, and generally hard-to-predict circumstances. Experience-based advice has tremendous potential utility when human hearts and minds—notably those of radically different cultures—are the primary targets of action and determinants of success.

In prior research for the U.S. Army, we had developed a suite of tools to support emerging doctrine and best practices for COIN/SSTR campaign design (FM3-24, 2006; TRADOC, 2008; SAMS, 2008). The result was a web-based collaboration environment, patterned on the idea of a wiki, but extended to manage relevant types of structured information, and to provide appropriate interactive and automated visualizations, tools, and analyses (Domeshek, Tuohy, & Spangler, 2015). A *wiki* is a web site that allows (authorized) visitors to contribute to shared content. Where a simple wiki manages generic web pages, a *structured wiki* is endowed with some range of defined page types, suited to the supported activity that may contain expected kinds of formal data as well as informal text. For instance, our wiki for campaign design defined page types structured to match doctrinally approved products such as "Commander's Design Guidance," "Environmental Frame," "Problem Frame," and "Operational Approach."

To support the environment modeling effort required by campaign design, the wiki also provided a range of page types (corresponding to Java objects and database tables) for factors such as *actors*, *resources*, *relationships*, *actions*. In addition to wiki pages, the system provided interactive graphical tools for constructing entity-relationship diagrams to visualize aspects of the environment—again tied to the wiki pages/objects representing those environmental factors. To support the research required to build up an environment model, the wiki included a content management system for storing source documents, and it applied automated analysis of those documents to extract entity mentions and recognize quoted references. To support the development of an operational approach, the wiki was extended with page/object types and a visualization editor allowing creation of Lines of Effort (LOE) diagrams per doctrine (JP 5-0). Finally, to provide sample content for the campaign design wiki, we used documents from the Army's Mission Command Battle Lab for a campaign design exercise called *Caspian Challenge*.

Thus, for the work described here, we established a *task context* (COIN/SSTR planning), an *operational context* (the notional *Caspian Challenge* campaign), and a *supporting tools context* (the wiki, with its LOE editor. In the next section, we describe the cognitive and technological underpinnings of our *approach* to providing high-quality in-context planning advice in this context. Later sections discuss system *design*, *implementation*, and *evaluation*.

APPROACH

The lessons of history are rarely clear-cut. Arguments can often be made about the influence of one set of special circumstances or another on historical processes and outcomes. There are two main responses to this reality: (1) attempts to abstract from specifics, e.g., using statistical analysis to find (probabilistic) rules and associations; and (2) attempts to compare and contrast a variety of past situations with current circumstances to generate insight about possible causal patterns (rather than specific predictions). We believe the second approach is the more realistic and valuable. That is, it is better to treat history as a prompt and guide to exploration, reflection, and analysis, rather than as an oracle expected to provide answers.

We position our effort in the context of prior work on case-based reasoning (CBR; Kolodner, 1993). CBR is a technology inspired by cognitive theories of human problem solving and memory. There is a significant history of case-based advisory systems—systems that emphasize retrieval of past cases to be used in support of human decision-making (e.g., Domeshek, 1992; Domeshek & Kolodner, 1993; McLaren & Ashley, 2001; Bach, et al., 2016). We note two key issues identified in prior work: (1) the utility of identifying pieces of extended historical experiences that may have value in supporting particular kinds of decisions; and (2) the utility of retrieving such episodes based on similarities in factors that influenced choices of actions or determined actual outcomes. We thus focus on the questions: *"What is a good advisory case?"* and *"What is a good basis for advice retrieval?"*

Our inspiration for scoping the contents of cases derives from the notion of *stories*. The 'story' in 'history' reflects the traditional importance of narrative in making sense of the world around us. While chance, fate, or circumstance may play their roles in any story, what really makes narratives hang together—what drives the selection of which events and conditions to include and which to exclude—is a set of implicit and/or explicit causal connections. Thus narratives can carry lessons and implicit explanations about the kinds of consequences that have been seen to follow from actions in the past. And thus narrative structure can also help determine which parts of a larger experience hang together as a meaningful, informative episode or advisory case.

This view of coherent narrative structures underlying advisory cases also suggests ways to think about advice retrieval. Narrative structures emphasize story causality—why things happen the way they do. This includes both *physical causality* (e.g., the chicken got to the other side by walking across) and *intentional causality* (e.g., the chicken crossed the road because it wanted to get some exercise, or it wanted to meet another chicken). Thus a focus on narrative naturally leads to a focus on representing causally relevant situational aspects that underly episode coherence. It has long been realized that case retrieval for problem solving support benefits from attention to causally relevant features (Hammond, 1989; Kolodner, 1993). In CBR, the term *indexing* is used to describe the selection of minimal sets of important and discriminating features that best justify (or explain) the relevance or utility of a case.

Beyond establishing case content and indexes, retrieval implementation requires addressing questions of case/index representation format and matching algorithms. There are two major (though not entirely distinct) approaches to format and matching: *feature-based* and *structure-based*. Retrieval based on features is generally simpler and faster. A universe of allowed features is defined, resulting in an overall "feature vector" representation. Given fixed-length, fixed-order vectors, there is always a simple one-to-one alignment of features in a retrieval cue and features in all stored items. Retrieval based on structural similarity is more complex and is computationally slower. Matching structures is typically a recursive process that has to check many alternate possible correspondences between available pieces of nested situation descriptors.

Cognitive science studies of human memory in problem-solving contexts suggest that there is a continuum from "mere similarity" to "deep analogy" (Gentner & Maravilla, 2018). For problem-solving purposes, analogy is often more informative. Research suggests that analogies are most meaningful, convincing, and useful when they involve systematic mappings of relationships that lead to unique and consistent mappings of entities (Holyoak, Gentner, & Kokinov, 2001). The Structure Mapping Engine—an algorithmic implementation of these insights into analogy—has been tested across a range of applications over the course of thirty years (Forbus et al., 2017).

In our work, we adopted MAC/FAC as our starting point (Forbus, Gentner, & Law, 1995). MAC/FAC is a two-tier retrieval scheme. The first tier ("Many are Called" or MAC) is an approximate feature-based matcher whose job is to rapidly filter the full corpus down to a smaller set of plausible candidates. The second tier ("Few are Chosen" or FAC) is a slower but more careful structure-based matcher that chooses and ranks the final small set of retrieved stories. An algorithm like the Structure Mapping Engine is suitable for use as a FAC component. The MAC/FAC scheme for combining costly structure matching with more efficient feature matching has seem substantial application and testing (Forbus, Gentner, & Law, 1995) as a cognitive theory.

The final aspect of our approach is to embed advice retrieval in the context of ongoing work. Our sample work environment is the LOE planning tool embedded in the structured campaign wiki described earlier. User work to create and edit LOE diagrams—potentially referencing known actors and resources with known relationships and goals from the wiki's environment model—allows for behind-the-scenes automated composition of advice queries and consequent recommendation of retrieved stories. The goals of this integrated approach are to provide advice:

- **Precisely:** Recommended experiences must carry useful advice for the current situation. Analysis of experience texts and current work allows for semantic matching that improves retrieval relevance.
- **Efficiently:** The automated system provides advice without consuming experts' expensive time. It consumes less of its users' time as well by avoiding overheads typically associated with research, such as tool context switches and explicit query formulation in tool-specific formats.
- **Proactively:** By monitoring workers' activity, the system can offer advice, even when it has not been requested. Such monitoring allows for autonomous formulation and execution of queries as the user's focus and information needs shift.
- **Continuously:** As a technology solution embedded in the work environment, there is no need to wait for scheduled meeting times or an expert's availability.
- **Broadly:** An interface to an organizational or community knowledge store is not limited to the experiences of any individual or any small panel of experts.

SYSTEM DESIGN

This section describes the design of a system intended to address the issues identified above, using the approach just described. With the focus on *task-embedded advice-bearing story retrieval*, the required pieces of a solution included: (1) computer-interpretable story encodings; (2) some method and tool to construct such encodings; (3) a representative corpus of stories so encoded; (4) a task support environment including some kind of planning tool; and (5) algorithms for query generation, matching, retrieval and presentation.

Narrative Structures

We describe our computer interpretable story encodings as *narrative structures*. They build on a long line of common sense knowledge representation design work stretching back to the early days of Artificial Intelligence (e.g., Schank

& Abelson, 1977; Lehnert, 1981; Domeshek, 1992). When the actions of people are the central concern, representation of intentional causality is essential—that is, systems need to capture behavior explanations by appeal to constructs such as goals and plans. There is a more-or-less standard repertoire of relevant representational components:

1. **Agents:** There are entities that we think of as having intentions—mostly *people*, but sometimes *groups*, *organizations*, *animals*, or even non-sentient entities being treated as having intentions.
2. **Themes:** Agents have *intrinsic* and *relational attributes*, such as being "childish" or being a "parent of" a particular child.
3. **Goals:** From time to time—often depending on context—agent's themes lead to predictable *wants*, *needs*, or *fears*, such as needing to "be fed" or wanting to "provide an education".
4. **Plans:** In response to those goals, the agents decide on *courses of action* that are often also semi-standardized and predictable, such as "showing up for family meals," or "moving to a neighborhood with good schools."
5. **Actions:** Plans may be complex sequences of abstractly specified activities, but eventually they bottom out in some set of more concrete activities, such as "going," "eating," or "buying."
6. **States:** Actions can be carried out only when conditions support them (e.g., to eat you must be near something that counts as food). Actions may also create new conditions (e.g., if you buy something you now own it).
7. **Links:** Instances of many of these categories relate to instance of other categories in meaningful ways: themes generate goals, goals generate plans, plans include actions, states enable actions, actions result in new states (including new relationship states that may change networks of interpersonal themes, and hence future goals, plans, actions, etc.).
8. **Volitions, Intentions, Impacts…:** Various *assessments* of the basic constructs above can be added to the narrative representation scheme. Actions are not always undertaken knowingly or willingly. Results of actions are not always the main (or even foreseeable) effects of acting. States that result from actions may not be intrinsically good or bad, but are perceived to be so by certain agents with respect to certain goals.

These constructs fit together into an abstract standardized kind of explanation, sometimes called an *intentional chain*: some *agent*, due to some *theme*, came to have some *goal*, for which they adopted a *plan*, and took *action* resulting in a new *state*. With the various kinds of assessments added on, various non-standard sequences can be described as well (e.g., someone being forced to carry out an action, or doing something that has an unknown bad effect). Things get even more interesting when intentional chains start to interact—sometimes within the motivational and behavioral structure of an individual, but more often across agents. Narrative structures aim to capture the constructs, linkages, and interpretations of multiple interacting intentional chains. Doing so allows recognition of interesting relationships among situations, including opportunities to influence or change behavior and outcomes.

Narrative Structure Editor

Our work fits into a larger research program concerned with automated situation understanding, including complex textual Information Extraction (IE). By fleshing out case content and representation, and by demonstrating utility in doing so, our work helps to define plausible targets for automated IE and make the case for further investment in such capabilities. At present, robust versions of such high-end IE capabilities do not exist.

The lack of suitable IE capabilities meant that we required an alternate surrogate method for stocking a test library. Our solution to this problem was to design a narrative structure visualization, and an interactive graphical editor based on that visualization. As with the rest of our tools, this was designed for use over the network through a web browser and integrated with the wiki. In later stages—after candidate story retrieval—this visualization was also used as one means to communicate story gist to end users.

A Representative Story Corpus

Our team of Subject Matter Experts (SMEs) built a corpus of two-hundred advice stories for our trial system. We used only open sources, so the corpus is unclassified. We started from news accounts and moved on to consider a broader range of historical sources and analyses, including interview transcripts and items from official lessons learned repositories of various services. We focused on stories useful in supporting planning for four kinds of operations: (1) Influence missions, such as conducting key-leader engagements; (2) SSTR missions, such as developing civilian infrastructure; (3) logistics missions, such as resupplying or expanding remote posts; and (4) kinetic missions, such as engaging high-value targets. The following is an excerpt from a typical story:

> *The increasing involvement of military forces on humanitarian aid and development not only blurs the lines between military and humanitarian aid, it also places the safety of* ***non-governmental organizations*** *(NGOs) aid workers in jeopardy. NATO has actively participated in Afghan development and humanitarian aid through its "civilian-military provincial reconstruction teams" which consist of military staff, reconstruction experts and diplomats. Such efforts could be seen as a bid to foster friendly relations with the local community, and to benefit NATO's military strategy. This raises serious concerns among NGOs, such as French aid group Solidarités, who argue that humanitarian aid should be "independent, neutral and impartial."* [Mojumdar, 2010]

Campaign Wiki and Planning Tools

We designed a simplified version of the campaign support wiki, including a simplified version of an LOE editor. The primary constraint was to make the system usable by experienced officers who would receive only a small amount of remote training in how to use the system—limited to a short introductory document and video. Within that framework, we supplied much of the *Caspian Challenge* campaign information, supplemented by more detailed information related to the specific planning problems to be set for our users. We developed four planning problems, reflecting the four main areas covered by the corpus: Stability, Transition, Logistics, and Security. The following is an excerpt from one of the problem statements:

> *You are a member of the Astara Rayon Reconstruction/Development Team (RRT/RDT) planning team. There have been problems in coordination and deconfliction with NGO activities, resulting in duplication of effort, undermining NGO efforts, creating an adversarial relationship, and reduced efficiency for both actors. The existing LOE does not appear to be working, because the situation above is growing worse.*
> ***…*** *deconflict and coordinate operations with NGOs to significantly reduce duplication of effort, both geographically and functionally, along with dramatically reducing related inefficiencies and animosity. … this LOE directly supports RRT/RDT Mission Objective 6.g: "Partner with NGO and international organizations who have pledged support at the rayon level. Ensure the RRT/RDT planning team includes functional, regional, and planning experts representing all the agencies active in the RRT/RDT."*

Query and Matching Algorithms

The project proposal adopted the MAC/FAC framework for retrieval (Forbus, Gentner, & Law, 1995), as described above. From one perspective, MAC/FAC is simply a 2-tier version of the recently popular IR approach of using multi-tier systems that successively winnow down (and re-rank) a retrieval set at each tier (Matveeva et al., 2006). However, the design of MAC/FAC was driven by studies in cognitive science and aimed to duplicate key features of human memory performance—in particular, efficient access to memories of past situations that bear useful analogical relationships to current circumstances. This made MAC/FAC an attractive starting point for our work. In our design, parameters `nMAC` and `nFAC` determine the sizes of the MAC-produced candidate set and the final FAC-produced retrieval set.

SYSTEM IMPLEMENTATION

This section presents the trial system as actually built and evaluated. We built a variant of the structured wiki system described above, simplifying some aspects and adding new case representation, visualization, matching, presentation, and feedback capabilities. The server was built using Java. The client consisted of web pages making light use of JavaScript and heavy use of Asynchronous JavaScript and XML (AJAX) to invoke server-side capabilities. For evaluation purposes, the server was hosted in the Amazon cloud and made available on the open Internet.

Tooling

The structured wiki provides a subset of the capabilities originally developed to support campaign design. It includes (1) pages for the main campaign design products (a *Campaign*, with associated *Guidance*, *Environment*, *Problem*, and *Solution*); (2) pages for basic environmental modeling (*Agents*, *Relationships*, *Goals*, *Actions*, and *States*); and (3) pages for high level plan development (*LOE Diagrams*, with associated *Queries* and *Results*). It also includes (4) additional page types for *Stories* and for the conceptual structures used to index those stories.

Figure 1 shows the encoding of a representative story, illustrating the final form of our narrative structures. In this story, a U.S. Army unit (518th Combat Gun Truck CO), escorting a convoy, avoids an improvised explosive device (IED) ambush because their Division Explosive Ordinance Disposal (EOD) unit (1st Cav Division EOD) spotted the IED and set up a roadblock. Surprised by this turn of events, the insurgents who had planned to ambush the convoy nonetheless try to follow through using their normal ambush tactics. This fails because the forewarned 518th correctly execute appropriate tactical tasks. The episode reveals a kind of inflexibility on the part of the insurgents.

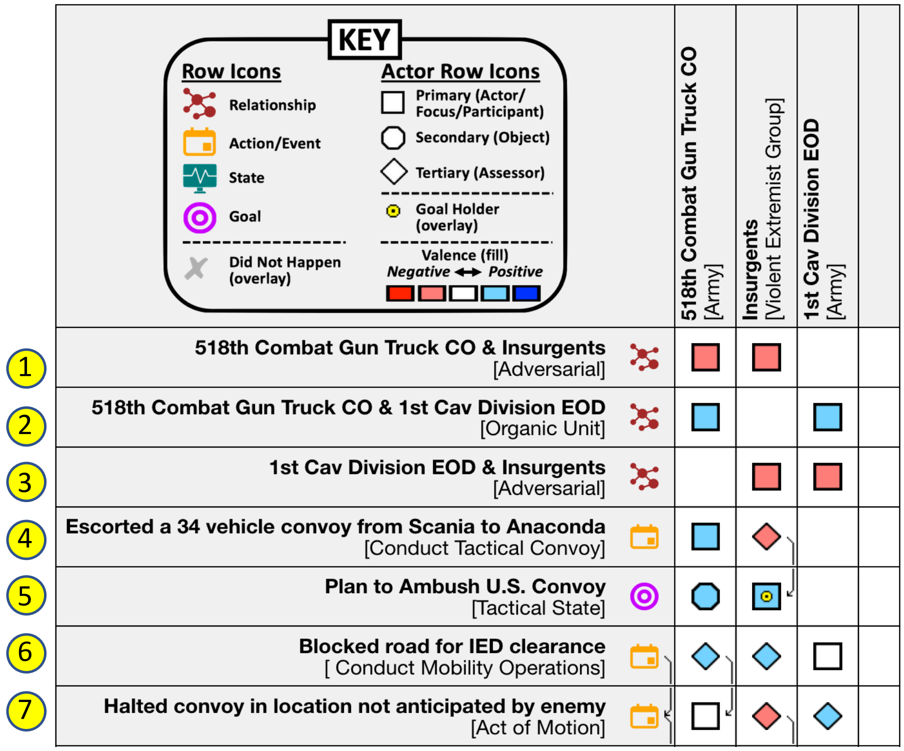

Figure 1. Sample Story Encoding

The narrative grid in Figure 1 has one column for each involved Agent (the characters in the story), and one row for each "happening" discussed in the story—a covering category for relationships, goals, states, and actions. Happenings typically occur in time, and the order of the rows reflects a rough chronology. The grid cells contain icons that represent the ways in which a given agent participates in or assesses a given happening (their role in the happening). Roles for actions include 'actor' and 'object'; in the fourth row, the 518th is the actor of the "escort" action. Roles for states include 'focus' and 'object'; in the eighth row, the Insurgents are the focus (here, the experiencer) of the "Surprise." Roles for relationships include 'primary' and 'secondary' for non-symmetric relationships, such as owner or supervisor; in the first row the symmetric Adversarial relationship has two primary role-fillers: the 518th and the Insurgents. Goals extend states by allowing an additional role annotation—'goal-holder'—indicating which agents desire the state in question; thus in row five, the focus agent (the Insurgents) is also noted as the goal-holder.

An additional role for 'observer/assessor' is available for all happenings; it is used to capture important sentiment reactions from agents who might not otherwise be involved. In row seven, for example, the Insurgents and the 1st Cav EOD are both noted as having opinions about the 518th halting short of the IED: the Insurgents are unhappy (red) while the EOD unit is happy (blue). So while role icon shape encodes these role types, icon color encodes the agent's sentiment. All these graphical conventions are summarized in the "key" included in the upper left corner of each narrative visualization. Finally, arrows depict links between happenings, or between roles associated with happenings. Paying attention to causal/motivational (and plan/impact) links is important because they capture in some detail the mechanisms underlying story sequences. They thereby help identify means to disrupt or facilitate those sequences—e.g., to plan or counterplan.

Figure 2 shows an example of the system's Lines-of-Effort (LOE) diagram. It largely follows conventions suggested in doctrinal and training publications (JP 5-0, 2017; DoD Joint Staff, 2011). Initial conditions appear on the left and desired conditions on the right. Actions intended to create the desired conditions are arrayed across columns representing campaign phases and segmented into horizontal bands representing distinct LOEs (here we show only one phase and one LOE). As with the Narrative Grid visualization, details can be accessed by hyperlinking to pictured States and Actions, and an editable version of the visualization is provided to create or update LOE diagrams.

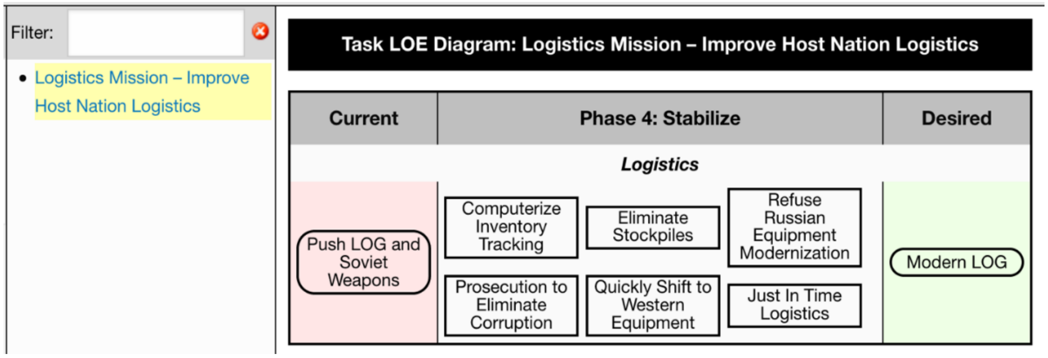

Figure 2. Sample Story Encoding

From LOEs to Queries

The primary link between the wiki as a notional planning environment and the retrieval algorithms required for an advisory system is made through the automated construction of queries based on the users' work on LOE diagrams. Development of LOEs is an appropriate time at which to consider the kinds of planning inputs offered by the system. To create queries from LOE diagrams, we adopt the following conventions:

- Each LOE is mapped to a query;
- Each current-condition is mapped to a query State;
- Each desired-condition is mapped to a query Goal;
- Each task is mapped to a query Action for blue force;
- All Agents mentioned in the above states, goals, and actions are added to the query;
- Relationships linking the included Agents are also added to the query;
- Negative impact links are added between current-condition States and desired-condition Goals;
- Causal sustainment links are made between current-condition States and the first Action;
- Successive Actions are given plan-sequence links;
- Final Actions are given motivation links from desired-condition Goals.

Matching Algorithm Variants

The MAC component uses a relatively simple feature-based approach to identifying relevant items. In our system, those features were originally based on hand-engineered concept tags attached to narrative elements. The algorithms were later generalized so they could alternately rely on the phrase embedding vectors also associated with narrative elements. Thus we effectively had two main MAC implementations: *MAC-Tags* and *MAC-Vecs*. Both of these MACs, involve a linear scan through the feature-encoding of all stories, which is easily done for our 200-story test corpus. Scalable algorithms for high-dimensional approximate k-Nearest Neighbor (kNN) are an active area of research, which is relevant because, for normalized vectors, distance in Euclidian space is equivalent to the standard IR cosine distance metric, which is what our MACs use to build up their match scores. Many competing approximate kNN algorithms offer different trades on issues such as indexing speed and size, retrieval speed and quality, scalability with increasing corpus size and dimensionality, empirical and theoretical analysis, etc. (Li et al., 2016). It appears that larger (billion-scale) corpora with thousands of feature dimensions can be handled by proven sub-linear algorithms (Sun et al., 2014; Arora et al., 2018).

The FAC component pays attention to the structural arrangement of narrative elements. For instance, say an agent-1 holds a goal with respect to another agent-2, which leads agent-1 to take an action that produces an effect that bears on agent-2. A structurally aware matcher will prefer matches where there is a goal that leads to the action that produces an effect, and the same agent appears as the holder of the goal and actor of the action, while a different actor appears as both the subject of the goal and the subject of the effect. Our FAC implementations further aim to enforces constraints that studies of analogy suggest are important for human judgements (Falkenhainer, Forbus, & Gentner, 1989). Our initial FAC implementation was based on recent generalizations of the Structure Mapping Engine (Forbus et al., 2017). The final implementation was substantially simplified by taking advantage of the specific narrative structure developed over the course of the project. As with our final MAC, this FAC implementation was generalized to allow the use of phrase vectors as an alternative to concept tags when scoring the lowest level element matches. Thus we effectively had two main FAC implementations: *FAC-Tags* and *FAC-Vecs*.

To round out our experimentation with matching algorithms, we also implemented a MAC matcher (designated *MAC-Text*) based on standard text-based IR tools. We used the open source Apache Lucene framework. This was intended to provide a plausible baseline against which to judge performance of our custom algorithms. When using MAC-Text, we treated our SME's original writeups of our sample queries as the text queries to be evaluated. Finally, we developed a dummy implementation of FAC (designated *FAC-Null*) that does nothing but truncate the previous phase's MAC results list of length `nMAC` to one of length `nFAC`. This allowed us to easily compare raw MAC performance (for any of our three MAC implementations) with combined MAC/FAC performance. With three MACs and three FACs, we were able to compare performance across nine configurations.

SYSTEM EVALUATION

Evaluation Design

Our system evaluation was carried out using a cloud-hosted version of the tools available over the Internet. The participants were a panel of forty-nine active duty and retired officers, drawn from all services (and, in one case a GS-15 civilian). They ranged in rank from O-3–O-6 (i.e., Captain to Colonel, or equivalents), with most at the higher end (e.g., only 4 O-3s participated). All had operational planning experience (some for many years). Almost all had COIN/SSTR experience. They were recruited through the personal networks of our three project SMEs.

Each officer was provided an introduction to the project, basic instruction on use of the system, and an introduction to the *Caspian Challenge* campaign planning context. Each officer was then asked to attempt two planning problems chosen from a repertoire of four problems covering the four COIN/SSTR operations planning areas for which we had collected stories. One problem was worked using the wiki tools but without any story retrieval and presentation. The other was carried out with the full system including story retrieval. The selection and order of the exercises was balanced and participants were assigned to a scripted sequence randomly. The result was that each problem was attempted by about 24 different officers—roughly 12 with story support, and 12 without story support; for 12 as a first problem, and for 12 as a second problem. Prompts and input widgets were integrated into the page displays to collect data on: (1) story reading time; (2) story understanding; (3) story relevance; and (4) story influence. Additional questions about the overall system appeared in a final questionnaire. The evaluation was opened on 3 Aug 2019 and the last evaluator completed their work on 13 Nov 2019.

Evaluation Results

Evaluators provided feedback at the level of individual stories (i.e., retrieved stories' perceived value and time spent on reading), planning problems/solutions (i.e., self-ratings of plan quality at completion of each problem), and the system overall. Among the story retrieval ratings we focus on the most stringent criterion: were retrieved stories influential in users' planning. Figure 3 plots the distribution of responses to the prompt *"The story is affecting the plan I am developing."* There are 177 ratings of "Agree" to "Strongly Agree" versus 69 ratings of "Disagree" or "Strongly Disagree." If we count neutral ratings as unhelpful stories, nearly half of the rated stories were not perceived as having direct or immediate planning impact. We found this result neither surprising nor alarming, neither from the perspective of the system's technical performance nor its planning value. Even between people, advice is not always perceived as helpful and is not always followed. The real questions are: *Does the system deliver value in aggregate* and *How much overhead does it impose in doing so*.

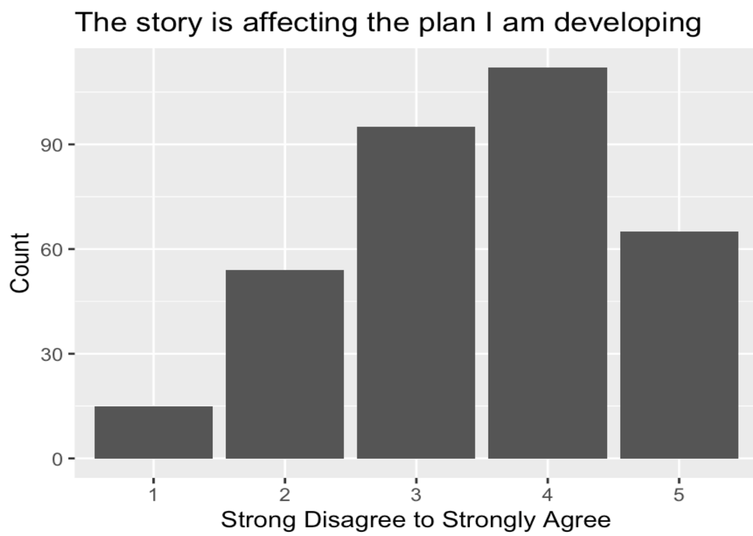

Figure 3. Response Distributions for Story Impact

The above plot suggests that users, on average, must read two stories to find one that truly affects their plan. Figure 4 summarizes distributions of story reading times conditioned on several different story ratings, including the one above (labeled "**Impact**"). The general trend is that users spend more time (and more variable amounts of time) on "better" stories. In particular, looking at the plots for "Story Impact Ratings," users spend consistently less time on low-rated stories—less than half as much as on the best stories, typically ranging up to 5 minutes. This suggests that time wasted on less relevant or less impactful stories is lower than a 50% hit rate might suggest.

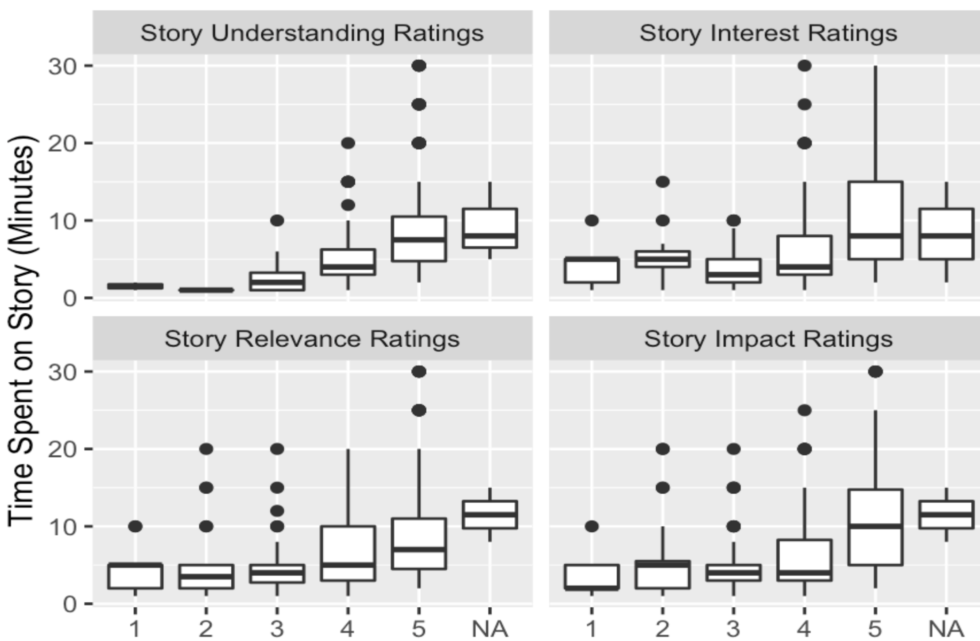

Figure 4. Story Reading Time Versus Story Ratings

Figure 5 plots the difference in time spent on plans with and without the system's story retrieval and advice support. Not surprisingly, the figure shows that it takes more time to plan when stopping to read stories—according to our data, about 1/3 more time. We believe this is a fair tradeoff for gaining better insight into the likely effects of proposed actions when the country is risking blood, treasure, and reputation

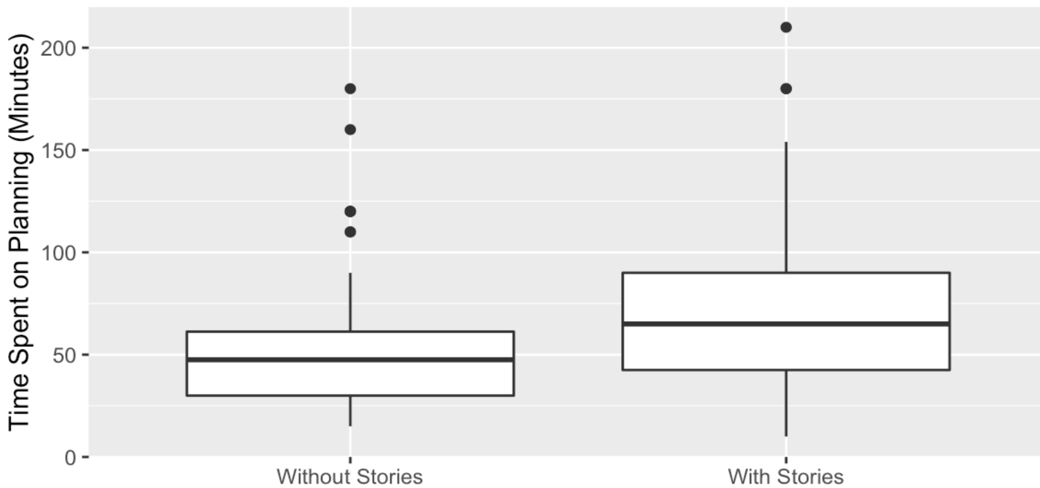

Figure 5. Boxplot of Time Spent Planning With and Without Story Support

Figure 6 shows the distributions of the differences between plan quality ratings with and without stories for each evaluator. Given the coarseness of our five-point rating scale, the majority of evaluators (30) reported no difference (difference value of 0) in the quality of their plans with and without story support. Five evaluators indicated their plan with stories was somewhat worse than their plan without stories (difference value of -1). Fourteen evaluators indicated their plan with stories was better than their plan without stories (positive difference values).

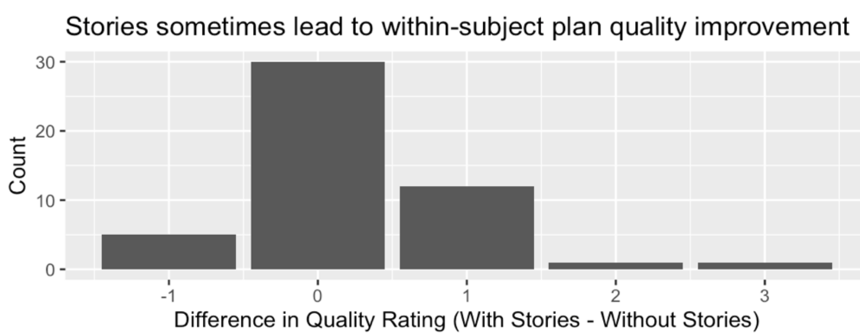

Figure 6. Sample Story Encoding

We ran a paired t-test on the users' self-ratings of plan quality across the with/without story support conditions. The results showed a small but significant increase: difference in means = 0.24 higher with story support on the 5-point ratings scale; paired $t(48)=2.3$ $p=0.027$. Any follow-up study would benefit from a more sensitive measure of plan quality differences.

CONCLUSION

Major Findings

Experience is a critical teacher when it comes to military planning. Our panel of evaluators provided clear, strong support for a capability of the kind envisioned in this project. In the world as it is today, large databases of lessons learned are being built at great expense (in both dollars and blood); there is tremendous value locked up in those experiences, but they are not, for the most part, being used. Despite DoD mandates requiring consultation of lessons learned, making use of those lessons remains too slow and difficult to be practical in most situations.

This work points to a solution: Integrate technology that provides relevant excerpts from and links into lessons learned documents directly as part of the user's work context. We demonstrated the principle by integrating narrative-based representation and retrieval algorithms with a high-level operational planning tool (the LOE diagraming tool). The principle should apply in other work contexts as well.

Key quantitative results from the evaluation include:

- 34 out of 49 (69%) subjects agreed or strongly agreed that the experimental system helped with planning;
- 45 (92%) agreed or strongly agreed that the concept of story-based advice is valid and could be useful to military planners;
- 43 (88%) agreed or strongly agreed that story retrieval should be further developed for military planning; and
- 40 (82%) agreed or strongly agreed that work should be continued on the wiki and integrated planning tools for military planning.

Finally, qualitatively, evaluators provided feedback that strongly supports both project premise and execution:

- *"This is essentially like having contextualized lessons learned available during planning."*
- *"It helps you think of things that you would not have thought of. The stories can give you radically different perspectives to consider."*
- *"What you are providing is the most functional application of a lesson-learned tool I've ever seen."*
- *"I wish this tool had been available for the 2002 Afghan campaign."*

Next Steps

We believe four primary steps should be taken to build on this work:

1. **Work with a Larger Corpus:** The first and most essential step is to experiment with a version of this system working over a much larger corpus. Both for pragmatic and technology advancement reasons, that corpus should be largely (if not exclusively) built using automated IE techniques. Those techniques will have to be chosen and tuned to enable construction of narrative structures. A larger corpus will provide coverage of more planning problems, while also providing greater challenge to the system and its algorithms. A corpus interpreted by automated extraction algorithms will provide more complete and uniform (if perhaps less accurate) encodings that will offer advantages and challenges for retrieval algorithms.

2. **Work with More and Better Planning Tools:** The wiki and embedded LOE editor were viewed positively by most evaluators. However, some comments noted that the prototype did not offer the polished look-and-feel or usability expected of modern software tools (e.g., the LOE editor did not support drag-and-drop rearrangement of elements). In addition, the LOE editor only supported one slice of planning activity. For a fairer and broader evaluation of such a system's potential impact on planning, it would be worth integrating story retrieval with a more complete set of more polished planning tools. An expanded and refined version of the wiki environment should be used to host future studies of story-based planning support.

3. **Study Story Support for Other Tasks in Other Contexts:** Feedback from evaluators suggested a range of DoD contexts other than operational planning where these capabilities could be useful. Evaluators involved in military education have requested that we pilot the system in the schoolhouse. Others have suggested that such tools would be particularly valuable in support of pre-deployment (or early-deployment) briefings and training. These suggestions should be followed up on.

4. **Explore Query-Construction from Less Structured Tools:** So far we have focused on instrumenting task-focused tools that inherently provide a lot of context and structure to inform implicit query construction (e.g., the LOE editor used in the evaluation). It would be worth exploring whether useful queries can be built by instrumenting more open-ended tools such as MS Word or web browsers, perhaps if combined with a richer sense of cross-tool user activity context.

The above actions will help make *task-embedded advice-bearing story retrieval* a reality. We have taken a significant step towards demonstrating both feasibility and utility. Now we need to demonstrate operations at larger scale, enabled by more thorough use of automation on the corpus construction side. DoD needs this capability. And once proven there, it will find application in a much wider range of civilian and commercial settings where those who cannot remember the past fear they are condemned to repeat it.

ACKNOWLEDGEMENTS

Stottler Henke gratefully acknowledges the sponsorship of this work by the Office of Naval Research (ONR), and Dr. Martin Kruger. However, the content of this document does not necessarily reflect the position or policy of the Government and no official endorsement should be inferred. We also appreciate the contributions of all the current and former officers who provided evaluation input on the prototype BEACON system.

REFERENCES

Arora, A., Sinha, S., Kumar, P., & Bhattacharya, A. (2018). HD-index: Pushing the scalability-accuracy boundary for approximate kNN search in high-dimensional spaces. In *Proceedings of the VLDB Endowment*, *11*(8), 906-919.

Bach, K., Szczepanski, T., Aamodt, A., Gundersen, O. E., & Mork, P. J. (2016). Case representation and similarity assessment in the self BACK decision support system. In *International Conference on Case-Based Reasoning* (pp. 32-46). Springer, Cham.

Domeshek, E., (1992). *Do the right thing: A component theory for indexing stories as social advice*. PhD Dissertation, Yale University.

Domeshek, E., & Kolodner, J. (1993). Using the points of large cases. *AI EDAM*, *7*(2), 87-96.

Domeshek, E., Tuohy, D., Spangler, D. (2015). "Concepts for Collaboration in Campaign Design". *Proceedings of the 20th International Command and Control Research and Technology Symposium (ICCRTS-2015)*.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, *41*(1), 1-63.

FM 3-24 (2006). *Counterinsurgency*. Headquarters, Department of the Army, Washington, DC, 15 Dec 2006.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*(2), 141-205.

Forbus, K. D., Ferguson, R. W., Lovett, A., & Gentner, D. (2017). Extending SME to handle large-scale cognitive modeling. *Cognitive Science*, *41*(5), 1152-1201.

Gentner, D. & Maravilla, F. (2018). Analogical reasoning. L. J. Ball & V. A. Thompson (eds.) *International Handbook of Thinking & Reasoning* (pp. 186-203). NY, NY: Psychology Press.

Hammond, K. J. (1989). On functionally motivated vocabularies: An apologia. In *Proceedings of the International Joint Conference on Artificial Intelligence*.

Holyoak, K. J., Gentner, D., & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. In *The analogical mind: Perspectives from cognitive science* (pp. 1-19).

JP 5-0 (2017). *Joint Planning.* Department of Defense, Washington, DC, 16 June 2017.

Kolodner, J. (1993) *Case-Based Reasoning*. San Mateo, CA: Morgan Kaufmann.

Li, W., Zhang, Y., Sun, Y., Wang, W., Zhang, W., & Lin, X. (2016). Approximate nearest neighbor search on high dimensional data—Experiments, analyses, and improvement. arXiv. Retrieved from https://arxiv.org/abs/1610.02455

Matveeva, I., Burges, C., Burkard, T., Laucius, A., & Wong, L. (2006). High accuracy retrieval with multiple nested ranker. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 437-444). ACM.

McLaren, B. M., & Ashley, K. D. (2001, July). Helping a CBR program know what it knows. In *International Conference on Case-Based Reasoning* (pp. 377-391). Springer, Berlin, Heidelberg.

Mojumdar, A. (2010). Afghanistan Aid Groups Say NATO Threatens Their Neutrality, PANOS London. On-line at https://www.globalpolicy.org/ngos/role-of-ngos-in-the-international-arena/49066-afghanistan-aid-groups-say-nato-threatens-their-neutrality.html

SAMS (2008). *Art of Design: Student Text, Version 1.0*. 24 September 2008.

Sun, Y., Wang, W., Qin, J., Zhang, Y., & Lin, X. (2014). SRS: Solving c-approximate nearest neighbor queries in high dimensional Euclidean space with a tiny index. *Proceedings of the VLDB Endowment*, *8*(1), 1-12.

TRADOC (2008). Pamphlet 525-5-500: *Commander's Appreciation and Campaign Design, V 1.0*. 28 January 2008.