Proceedings of the 2006

Intelligence Tools Workshop

23rd of August, 2006
Esbjerg, Denmark

Organized by
Department of Computer Science & Electronics
Aalborg University Esbjerg
www.cs.aaue.dk

Editor
Fredrik H. Madsen

ISBN# 978-87-7606-019-0

Contents

Preface

The Intelligence Tools Workshop (ITW'06) was held on the campus of Aalborg University Esbjerg in Esbjerg, Denmark on 23rd of August 2006. The intention of the workshop was to promote cross fertilization between the academic and industrial areas surrounding intelligence analysis. Initially the workshop was affiliated with the 2006 ACM Hypertext Conference. However, due to a low overlap between the workshop and conference participants, it was eventually split off from the conference, though the interdisciplinary focus was maintained. The organizers still believe there is much potential in the overlap between hypertext and intelligence analysis, but that remains to be seen.

The select group of workshop participants were invited by members of the program committee, and were from both academia and industry. The workshop began with an invited talk, by David Robson from Scottish Enterprise, on cognitive rigidity. This was followed by a session with three presentations on knowledge management tools for intelligence analysis. After having lunch together on campus, an afternoon session was held with two presentations of investigative data mining tools for intelligence analysis. The day concluded with a workshop dinner at Caf Frederik in the heart of Esbjerg. The overall structure of the workshop day was informal and flexible, providing plenty of time for discussion among the participants.

After the workshop there was a post-workshop review cycle, during which one paper was, unfortunately, withdrawn from the proceedings. On a personal note, I would like to thank Sarah Taylor, Lockheed Martin and Nasrullah Memon, Aalborg University Esbjerg, for serving on the workshop program committee. Your help has been invaluable to me and to the workshop. I would also like to thank the workshop participants for their informative presentations and active participation in the workshop.

Investigating the Cognitive Effects of Externalization Tools

Fredrik H. Madsen
Department of Software and Electronics
Aalborg University Esbjerg
Niels Bohrs Vej 8
Esbjerg, Denmark
fhmadsen@cs.aaue.dk

David L. Hicks
Department of Software and Electronics
Aalborg University Esbjerg
Niels Bohrs Vej 8
Esbjerg, Denmark
hicks@cs.aaue.dk

ABSTRACT

In the present paper we present a novel approach to improving national security, by improving the quality of intelligence analysis. We seek to reduce confirmation bias in intelligence analysis by improving the nature of the externalization tools. This is an extensive project and in the present paper we present its context and the first experiment.

1. INTRODUCTION

The basic motivation behind our work is the fact that free societies are unable to fully protect themselves against terrorist attacks. Due to the fact that we have only found sparse information about intelligence agencies other than that of the U.S., we focus our investigation on that. We have no reason to believe that our conclusions should not be valid in other intelligence services as well, however.

There are many things that could be done to increase the protection of said countries. One direction is to invade the countries that harbor terrorists, in order to hamper their ability to carry out successful attacks. This has been tried with the invasions of Afghanistan and to some extent Iraq. Another possibility is to increase the level of surveillance inside the countries in order to increase the intelligence services' ability to pre-empt the terrorist plots. This has been a hot topic in several countries around the globe, mainly due to the linkage between government surveillance and civil rights. A third possible direction is allowing torture of possible informants, in order to increase the level of information they contribute. We focus on increasing the efficiency of the intelligence services.

There are many ways to increase the efficiency of the intelligence services. One way that has been proposed is to restructure the intelligence services so that they become better able to fight loosely coupled networks, rather than a fixed set of countries [7, 55]. Another way that has been proposed is to improve the communication between the intelligence analyst and the intelligence consumer [29, 27]. A third way that has been proposed is that intelligence analysis should develop a meta science of its own, just like medicine [31, 38]. We focus on increasing the ability of the individual analyst to answer the problems he is confronted with. In order to address this, we look at problems in general.

Karl Popper divides problems into two categories: *clocks* and *clouds* [5, 56]. *Clocks* may be very complex, but they function in a predictable and reliable manner. A mechanical system is a good metaphor for this kind of questions, and they may typically be answered more or less algorithmically [5]. The main mental processes that are used in analyzing this kind of systems are *induction*, and *deduction*; the first generalizing rules from particular cases, the second to use the generalized rules in particular cases. The overall process is called *deductive reasoning*, and is related to the inductivist view, introduced by Francis Bacon and commonly associated with physics and mathematics. *Data driven analysis* aims at understanding this kind of systems. The defining feature is that there is a commonly accepted positivist knowledge base behind the reasoning and hence any piece of information has exactly one interpretation. This effectively means that there is no interpretation and that the evidence "speaks for itself". An example of this is a question like "is the sandy beach of region x in country y strong enough to support an invasion with z material?". There is little doubt that with solid knowledge of the geological conditions of region x, and the nature of invasions, this question may be answered with a fairly high degree of certainty. The other kind of problems, *clouds*, are like gasses; highly irregular and not predictable in the same sense [56]. An example of this kind of question is "will Iraq develop into a healthy democracy within the next 10 years?". While both questions have simple yes or no answers, it is plain to see the difference; in analyzing this question there is little commonly accepted positivist knowledge to support the reasoning.

While it is plain to see that intelligence analysis contains clocklike problems, the following section describes how it contains cloudlike problems as well. The third section describes the heuristics that are normally used when working with cloudlike problems. The following section describes a kind of tools that often accompany people in working with cloudlike problems, externalization tools. In the fifth section we describe our hypothesis, followed by a section on how we intend to test it. We finish the paper with a few concluding remarks and a description of future work.

2. THE INTELLIGENCE CYCLE

The intelligence analysis process is often described as a cycle, "the intelligence cycle" [14]. This cycle is often subdivided into 5 phases: *planning and direction*, *collection*, *processing*, *analysis*, and *dissemination*. The intelligence cycle is not a precise description of the process the analyst goes through[32, 61], but it provides a convenient framework for describing its contents.

The main purpose of the *planning and direction* phase is to establish the needs of the customer, and to figure out how best to fulfill them [37, 32]. One source of cloudiness in this phase is the fact that the consumer and the analyst exist in different contexts [29]. The customer may also not know exactly what he needs, or simply not have the time to explain it in sufficient detail, which leads to vague and even slightly wrong requirements. There may also exist a pressure for the analyst to return with a certain answer [28, 25]. The reality of this pressure has been contested, but it remains open nonetheless [25, 57, 58, 7].

In the *collection* phase, information from available sources is collected, based on the requirements defined in previous phase. The output of this phase is normally dubbed "raw intelligence", as opposed to processed and finished intelligence that is output from the following phases. When fighting terrorism, many of the traditional sources are of less use than in conventional warfare. After all, there are no tank divisions to observe on satellite images, there are no known leaders sending out detailed orders to troops. There are also no clever measurements of seismic, nuclear, radiological, etc. activity that can unveil the terrorist plots. The primary way to get solid information is HUMINT, which can be sought in three overall ways; *infiltrate an existing agent*, *recruit an agent*, and *rely on walk-ins*. The first option is hardly even an option; there are few that could do that and even fewer that would. Recruiting an agent is normally a good option, but when dealing with terrorist cells there are additional problems: the cells are small, few people have the complete picture, they are usually very committed to their cause, etc. This leaves relying on walkins as the only real option. The only way to really compensate for these problems is to broaden the collection efforts, so that even though the precision is low, the recall may still be high. This leads to massive data collection, necessarily degrading its value. A former NSA director describes the situation in the following way:

> I felt as though I had a fire hose nozzle held to my mouth, with an endless stream of raw intelligence gushing out [28].

The raw intelligence that has been collected, needs to be *processed* before it is of any use in analysis phase. Many of the activities in this phase may be described as translation efforts; from one language to another, from electrical readouts to something more readable, etc. There are also efforts that have a more additive nature, such as circling areas of interests on large satellite images, collating sources that seem to belong together, annotating and evaluating the evidence. The high output of the collection phase puts stress on this phase, as well as the number of different languages the data exists in.

The processed intelligence serves as input to the *analysis* process. This process has been described as: " . . . the process of evaluating and transforming raw data into descriptions, explanations, and conclusions for intelligence consumers." [30]. This has often been compared to laying a puzzle, and the comparison is important for the similarities as well as the differences [16]. Due to the problems described in the collection phase, the pieces of the puzzle are mixed up with massive amounts of pieces from all sorts of other puzzles. Due to the problems described in the planning phase, there is only a vague description of a box cover to guide the process. Due to the problems described in the processing phase, some of the pieces may seem to fit even though they do not. As an additional requirement, the analyst should be aware of other puzzles that he could assemble, should they prove significant [29, 37]. In relation to fighting terrorism, the information flowing into the process is of much more intangible character, there is no "smoking gun" to be found, there is an even higher risk of deception, etc. This increases the amount of evidence that needs to be considered and the space of hypotheses, leading to an increased complexity.

An intelligence product that is merely produced serves very little purpose, it needs to be *disseminated* to the consumers in order to have its effect. Intelligence consumers want to be informed by good, broad, well sourced, intelligence [7, 41, 57, 34, 58]. The customers may, however, neither have time to read and comprehend 20 pages of highly complex material, nor to wait for its production [29, 27]. This makes the intelligence consumer dependent on, to a certain extent, believing the analyst at his word. This is not a problem, and is in fact necessary, as the analyst has more specific knowledge about the situation and is thus in a better position to know. This does however put much focus on the analysts *credibility*. This opens the doors to a world of trouble, because in some ways the intelligence analyst is like a football referee; the better a job he is doing, the less people will notice him. This is also true for intelligence analysts, because when he delivers good results on time, the dangerous situations are preempted and nothing happens. In other words; the credibility of the analyst does not benefit much from being right, but is punished severely when he is wrong. Another difficult situation is the fact that the analyst needs to produce a correct answer to the consumer. This may seem a simplistic matter to bring up, but the problem is that due to the problems described in the analysis phase, in most non-trivial cases no one knows for certain. On the one hand side, the analyst is obligated to stay close to this truth, but if he stays too close to it, the consumers will dismiss his analyzes as pure hedging [9]. Straying too far to either side of the balance will seriously strain the analysts credibility [41]. The analysts will also be fighting the fact that, when faced with an ambiguous situation, people, and hence also consumers, are prone to believe the positive interpretation [9]. In relation to fighting terrorism, an additional problem of this phase is that the kind of evidence that is found is often more similar to a prolonged buzz, than to a smoking gun [7], [34]. This makes intelligence estimates inherently weaker, reducing the chance that the decision makers will take action. This, in turn, increases the risk of tactical sur-

prise, also referred to as *intelligence failure*. In the following section we take a look at the consequences of this cloudlike nature, for the intelligence analyst.

3. HEURISTICS

Following Poppers theory of falsification, the idealized analyst must hypothesize and test until he understands the situation. Any hypothesis may be proposed and held, as long as it is falsifiable, and has not itself been falsified. The fact that any falsification may itself be falsified increases the complexity of this kind of analysis.

There are few real questions, however, that are patient enough to wait for this kind of answers. This is a common problem in our everyday lives, and a common solution is to use heuristics to arrive at satisfactory answer in the time available. When a person goes shopping for groceries, for example, few make a complete survey of all possible stores and weigh price and quality for each possible scenario. Most make a choice between a couple of stores where they have experienced to get the quality they want, at the price they are willing to pay. Heuristics are good, almost per definition, because in the end we seldom need to make an optimal decision. But heuristics also leave us unprotected against biases.

Biases are systematic deviations from rational decision making [67, 66, 2]. An example could be a person walking down an alley hearing steps behind him. If the person feels safe he will interpret the steps as being, at worst, irrelevant. If the person does not feel safe, the steps may very well be interpreted as a sign of danger. This way the initial hypothesis, safety or danger, has undue impact on the sense that is made from the data. This particular deviation is often labeled *confirmation bias* [67, 57, 2]. Confirmation bias is a term with many definitions, we lean on a fairly open one: "Confirmation bias, ... connotes the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand" [51].

A historical example of this bias lead up to the 1973 Arab-Israel war. Egypt was mobilizing near the Israeli border under the pretense that it was an exercise. This was in fact the prevalent hypothesis in Israel, and it was backed up by the common belief that Egypt would not attack without long-range air-strike capability against Israeli airfields. Most evidence was consistent with this hypothesis. There was another possibility however; that Egypt was secretly preparing for an invasion. As it turned out, all evidence that was not the result of deception was also consistent with this hypothesis. Israel prepared according to the first hypothesis, while the second one turned out to be true, causing tactical surprise or intelligence failure [5]. This was in part due to the fact that what was seen as overwhelming support for the "exercise" hypothesis was in fact also consistent with the "war" hypothesis. In this situation, and others like it, deception was a major part of the opponents success in cloaking his attack. It is easy to produce lots of evidence of a false hypothesis, if you know that the opponent is looking and especially what he is looking for and at [5].

Many ways to reduce the analysts level of bias have been proposed. Devils advocate games have been proposed to keep analyses broad [57, 37]. Conversion to more iterative/agile analysis processes have also been proposed [21]. A third proposition has been to simply force intelligence analysts to be more rigorous in their analyses [41]. We focus on the tools that are available to the intelligence analyst. The reason why we focus on tools is that we believe that they have great influence on our cognitive processes [69].

Many tools have been proposed and developed to support intelligence analysts in their work. It has been proposed establishing a computer based information network between intelligence analysts would improve information sharing, and hence improve the general knowledge level [8, 58]. Due to the massive collection efforts, more data is collected than will ever be translated. Hence automated translation tools have been proposed and developed to help intelligence analysts. Social networks were originally described by Stanley Milgram [65], but Malcolm Sparrow has described how they could be important aids in the fight against organized crime [60]. This has been described by several others, both with respect to organized crime [40, 15, 71], and to terrorist networks [36, 43]. Many of the efforts, within the area of computer science, have centered around devising engines to automatically extract these networks, and to find weak spots. In this relation, their main importance is that they encode knowledge about people and their relationships, that there are advantages from encoding it in computers, and that this encoding is beneficial. This has also been described in some length by Alden Klovdahl [35] in his 1983 article "A Note on Images of Social Networks". We focus on another kind of software tools, namely knowledge externalization tools.

4. EXTERNALIZATION TOOLS

In 1956 George Miller posited that at any one point in time, Man can only have seven things in his mind, give or take two [46]. This is relevant when dealing with most cloudlike problems, because they typically require us to exceed this limitation. A step that is often taken to do so is *externalizing* some of the things [57]. This means creating an external representation, so that it may later be internalized. One example of such an externalization is speaking, where ideas are rendered into external form, sound, by using the vocal cord. Another example is writing on a napkin, where ideas are rendered into external form, lines, using a pen.

An example of a tool that is specialized for attacking a complicated problem by externalizing ones thinking, is a *highlighting* pen. While a very simple tool, anyone who has ever read a complicated piece of text knows that highlighting certain key phrases vastly enhances the process of grasping its contents. The knowledge that is externalized is the knowledge of where the key phrases of the text are located. Another example of this kind of tool is a *list* [57]. A list places important items in close proximity, often with a headline guiding their interpretation. A slightly evolved version of the list is the ranked list, encoding additional information, for example importance, into the positioning of each item. Another example of a slightly evolved version is lists of pro's and con's. Here, the position of an item in either list encodes its relationship to a problem. A slightly different example is the omnipresent "post-it". This tiny piece of paper obviously allows for any combination of text and graphics, but its real genius is the glue. The glue enables the user to anno-

tate any plain surface with text and graphics. This gives the annotation a very precise point of reference. As a main part of his "analysis of competing hypotheses", Richards Heuer proposed a matrix, to externalize the relationship between the different pieces of evidence and the hypotheses [57]. For each entry in the evidence x hypothesis matrix, it is noted the piece of evidence supports or disconfirms the hypothesis. The main benefit of this tool is that it emphasizes the *diagnosticity* of evidence. A piece of evidence has little value in deciding between hypotheses if it is consistent with all possible hypotheses, and hence it is said to have a low diagnosticity. We are, however, not concerned with physical tools, but with computer based ones.

One such tool,although not specifically developed for intelligence analysts, is the "memory extender", or simply *Memex*. In 1949, Vannevar Bush proposed a system devised to externalize idiosyncratic relationships between scientific papers. This was based on the observation that the rate at which the number of scientific papers was growing prevented, especially inexperienced, scientists from reading the important ones. A system such as described would allow experienced researchers to externalize their knowledge of the more important papers within their field, and allow other researchers to benefit from it.

> This is a much larger matter than merely the extraction of data for the purposes of scientific research; it involves the entire process by which man profits by his inheritance of acquired knowledge. The prime action of use is selection, and here we are halting indeed. There may be millions of fine thoughts, ...; but if the scholar can get at only one a week by diligent search, his syntheses are not likely to keep up with the current scene. [13]

Innumerable tools have spawned from the Memex: Augment [20], Notecards [23], KMS [1], Aquanet [39], and Intermedia [44] were the pioneers of a new research area dubbed *hypertext*.

There are plenty of externalization tools that have been developed for the intelligence domain. Analysts Notebook [26] is probably the most common; the "Sandbox" [70], the "Sensemaking Support Environment" [18], the "Entity Workspace" [10], and the "Centrifuge" [17], are other examples. These are examples of different externalization tools with different properties. We are interested in the the influence of these different properties on cognitive processes of the user.

On the very large scale, there have been investigations on how certain key externalization tools have influenced the thinking of Man. Eric A. Havelock has investigated how using the phonetic alphabet changed the thinking in ancient Greece [24]. Walter Ong and Marshall McLuhan have investigated the consequences of using the printing press [52, 42]. These investigations have found major changes in our philosophy, caused by the use of new externalization tools.

Another strain of research that speaks to us on these matters is the Sapir-Whorf hypothesis [12, 33]. It states that: "Structural differences between language systems will, in general, be paralleled by non-linguistic cognitive differences, of an unspecified sort, in the native speaker of the two languages.", and that "The structure of anyone's native language strongly influences of fully determines the world view he will acquire as he learns the language.". In order to relate this to our research, we may perceive the interaction with an externalization tool as a special kind of language, and the externalized thoughts as documents written in that language. In most cases this special language contains a language in the traditional meaning of the word, be it English, Danish, or Chinese. Most tools do however make their own subtle changes.

When we draw with a traditional pen on a piece of paper, straight lines are difficult to make and "organic" lines are easy. When using a computer based drawing application, straight lines are easy and "organic" lines are more difficult. When we write on paper napkins there is an abundance of arrows and circles, while complete sentences more scarce. Most word processing applications allow us to express both arrows and circles as well as full sentences. If we observe the products that are produced in word processing applications, however, there are significantly more full sentences and subsequently less arrows and circles.

There seems to be no difference in the completeness of the two vocabularies, but the frictions are different; some things are more easily expressed in one medium than in another. The Sapir-Whorf hypothesis states that any difference will be paralleled by non-linguistic cognitive differences in native speakers. Another reason that makes even the smallest difference interesting, is that when we have externalized our thoughts, we forget. This way the externalization becomes the only access to the original thought, and in fact, in all relevant aspects, the externalized document replaces the original.

> He who controls the present, controls the past.
> He who controls the past, controls the future [53].

This leads us to believe that even subtle differences between externalization tools could have significant cognitive effect. Hence our project could be described as investigating the side effects of externalization tools. There are many areas in which such side-effects could be found, we have chosen to focus on bias. As previously described, it seems certain that some of our biases are caused our heuristics. Others may, however, be caused by our tools. This is described in Figure 1.

Peter Todd and Izak Benbasat have studied the influence of effort on decision strategy selection [63, 64]. They found that the decision applied by the user was strongly dependent on the effort required to perform it. Peter Pirolli et al. have investigated the influence of a computer based hypothesis vs. evidence matrix (ACH_0) versus a paper based one [54]. Their focus was on the level of confirmation bias, but their results were inconclusive. In another experiment, ACH_0 was extended to allow a group to share their matrices.

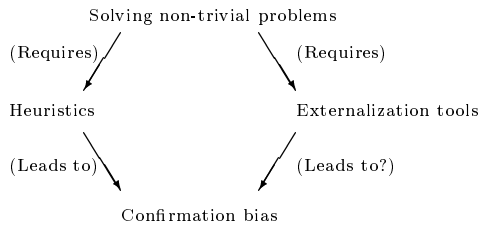


Figure 1: We propose that the nature of our externalization tools has impact on our mental processes, and as such could lead to biases.

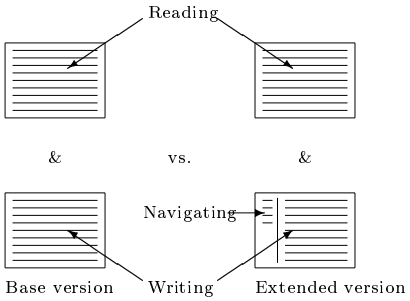


Figure 2: We propose a prototype in two different versions, each consisting of a reading screen and a writing screen. In the first version, the writing tool supports a single document. In the second version the writing tool allows the user to divide the document up into "chunks".

The new collaborative system was dubbed CACHE [11]. In this experiment , the results confirmed that when the group consisted of people with different biases, the bias of the decision process was reduced. In both cases, our research is very related, but the feature under investigation is different.

5. HYPOTHESIS

In order to show that externalization tool influences the level of confirmation bias, I intend to develop a such a tool, in two different versions: a base version, and the base version plus one feature. I am going to have two groups of users working on the same problems, one with each version of the developed piece of software. While they do so, I intend to measure their level of confirmation bias, and in the end compare the results for the two groups.

In order to actually do this, we need to determine a few things. First we need to determine the base version of the software and then the feature we intend to add. Then we need to determine which users will work with the problems, and what these problems might be. In the end we need to determine how to measure the level of bias of the users in the two groups. It is important to note that any configuration of base version, additional feature, users, problems, measures could be used to test the proposed connection between externalization tool and confirmation bias. The exercise here, is to determine a configuration that seems particularly likely to generate a difference in the two groups of users.

The choice of base version is guided mainly by the desire to have the user feel comfortable with the tool in hand. This is important in order not to introduce confusion. This is particularly important as we expect a short time available for each test. The premise is that any confusion on the users part will affect the test results in an unpredictable manner.

The requirement is two "document like" workspaces: one for reading about the problem and eventual source data and another for externalizing the users thinking. This functionality is easily implemented and understood as a surface with two "tabs", one for reading and one for writing. The surface for writing is a standard text editor.

The choice of the feature that is to be added in the extended version of the software, is heavily influenced by the choice of base version. The main importance in this choice is that the feature is likely to affect the level of confirmation bias. We have chosen to introduce the idea that the writing window does not contain one document, but a set of "document chunks" available through a list. The two different versions of the prototype are described in Figure 2.

An externalization tool holds external representations of ideas or more generally, *knowledge*. We are interested in a feature that will reduce the friction put on the externalization process. Hence in order to justify our choice of feature, we need to investigate certain key areas that may supply ideas about the nature of knowledge.

One such key area is the idea of the "struggle over meaning" essential to post-modernist thinking in general, and to Michel Foucault in particular, [59, 6, 62, 4]. The basic postmodern idea, for our purposes, is that there is no singular stable truth, but rather a multitude of different and possibly conflicting truths. A related idea, building on the post-modern project, is that of *intertextuality* [3]. Intertextuality is the idea that all texts rely on other texts for their meaning, creating a massive web that is often referred to as the intertext. It is a key idea of literary hypertext theorists that traditional linear text marginalizes this intertext [49, 50, 45, 47].

A second key area, or set of ideas, is presented by Karl Weick in his book on organizational sense-making [68]. Weick defines sense-making as a process that is similar to interpretation, but different in that it does not presume a stable preexisting truth. Weick describes seven properties, of which we focus on two. The first is that sense-making is "retrospective", meaning that sense is always made of something *past*. This means that there is always a present that impacts the sense that is made. As this present shifts, the sense that is made of past situations shifts as well. Another property described by Weick is that sense-making is focused on and by extracted cues. This means that when we shift our focus from one configuration of cues to another, the sense we make of a situation changes as well.

A third set of ideas is presented by Kathleen Eisenhardt, in a survey on how strategic decision makers manage to make good decisions and still move fast [19]. One of the main points, with respect to our work, is that investigating multi-

ple alternatives is important. It is important for several reasons: decision makers increase confidence in their decisions reducing procrastination, decision makers avoid escalation of commitment to one alternative, and decision makers gain a fall back position.

These three key areas promote an idea of knowledge as a dynamic set of interrelated conflicting *micro narratives*. It seems that replacing a document with a set of chunks may reduce the friction put on the externalization process. Hence, we hypothesize that users of the extended version will exhibit significantly less confirmation bias than the users of the base version.

5.1 Testing

In order to test this hypothesis we need two groups of users, a set of tasks, and a set of measures. The two versions of the prototype are meant to be simple. This means that there are not many requirements when selecting test users. Only very basic computer skills, and probably a certain age. We set the age limit at 20 years, as we want to make sure that the brains of the participants have matured. Otherwise this could have significant impact on our test results. The set of tasks should be non-trivial problems and should be the same for both groups of users. As confirmation bias is not directly available, we must resort to indirect measures.

One measure that could indicate the level of confirmation bias, would be the number of different solutions were explicitly generated by the user [48]. It will require a certain degree of interpretation, depending on the writing style of the user. A sub measure could be to evaluate the relative depth to which the user considers each solution. One could also simply rate the level of confirmation bias in the user, judged on the session. This approach has also been used before [22, 48].

Another measure is the final distribution of belief [11, 48]. Given a data set where there is no obvious conclusion, participants that still reach strong conclusions are quite possibly influenced by bias. A related measure would be to compare the solution judged most likely by the user to the first documented solution. This would establish a measure for the degree of mobility of belief.

A third measure could be the focus of the user over time, measured in the events generated. This data could indicate that the user spends a disproportionate amount of time or characters on a sub set of the solutions. This measure is very time intensive, and should only be pursued if time permits it or the quality of the other measures require it.

A final measure is the NASA TLX [22, 48] score. The NASA TLX is really a fixed questionnaire that presents a final index for the "task load". One part of this index is the mental load, and while not directly signifying a reduction in confirmation bias, it may show something about how well fitted the tool is to the task.

6. CONCLUSION

On one level we are investigating the influence of going from one document to document chunks in an externalization tool on the level of confirmation bias, while dealing with non-trivial problems. It has been shown that confirmation bias haunts intelligence agencies, and lives are lost because of it. This means that, at the very concrete level, it makes sense to investigate the roots of confirmation bias and try to reduce it.

There is a broader angle on it as well though. Throughout the history of tools they have been shaped by what we wanted them to do, but also by what was feasible and possible to construct. As these tools are increasingly being designed by software engineers, this is still true, but there is a difference of scale. While there are limits to what software engineers may construct, they are not many and their numbers are decreasing. This gives increased power to the engineer, but with it comes increased responsibility.

7. FUTURE WORK

If it turns out that there is a significant impact on the level of confirmation bias, it would be very interesting to see if different modes of interacting with the set of chunks would change that impact. In this experiment we have a simple list, but it would be interesting to investigate if a spatial metaphor is a better mode. It would also be interesting to see if the provision of navigational links between points in the chunks would have impact. As a final point, it would be interesting to see if there are other definitions of "impact", than the rather restricted one we have applied in the current work.

8. REFERENCES

[1] Robert Akscyn, Donald McCracken, and Elise Yoder. KMS: a distributed hypermedia system for managing knowledge in organizations. In *Proceeding of the ACM conference on Hypertext*, pages 1–20. ACM Press, 1987.

[2] D. Arnott. A taxonomy of decision biases. Technical Report 2002/01, Melbourne, Australia: Decision Support Systems Laboratory, Monash University, 2002.

[3] Roland Barthes. *S/Z*. Hill and Wang, 1974.

[4] Roland Barthes. *Image Music Text*, chapter 7, pages 142–149. Hill and Wang, 1978.

[5] Isaac Ben-Israel. Philosophy and methodology of intelligence: The logic of estimate process. *Intelligence and National Security*, 4(4):660–718, 1989.

[6] Jeremy Bentham. *The Panopticon and Other Prison Writings (Wo Es War)*, pages 29–95. Verso Books, 1995. http://cartome.org/panopticon2.htm.

[7] Bruce Berkowitz. Better ways to fix u.s. intelligence. *Orbis*, 04:609–619, 2001.

[8] Bruce Berkowitz. Failing to keep up with the information revolution. *Studies in Intelligence*, 47(1), 2003.

[9] Richard K. Betts. Analysis, war, and decision: Why intelligence failures are inevitable. *World Politics*, 31(1):61–89, 1978.

[10] Eric A. Bier, Edward W. Ishak, and Ed Chi. Entity workspace: An evidence file that aids memory, inference, and reading. *Lecture Notes in Computer Science*, Volume 3975:466 – 472, 2006.

[11] D. Billman, G. Convertino, P. Piroll, J. P. Massar, and J Shrager. Collaborative intelligence analysis with cache: bias reduction and information coverage. WWW, 2006.

[12] Roger Brown. Reference: In memorial tribute to eric lenneberg. *Cognition*, 4:125–153, 1976.

[13] Vannevar Bush. As we may think. *The Atlantic Monthly*, 176(1):101–108, 1945.

[14] CIA. Central intelligence agency. www.cia.gov.

[15] Nigel Coles. It's not what you knowit's who you know that counts. analysing serious crime groups as social networks. *British Journal of Criminology*, 41(4):580–594, 2001.

[16] Jeffrey R. Cooper. Curing analytic pathologies: Pathways to improved intelligence analysis. Technical report, Center for the Study of Intelligence, 2005.

[17] Tildenwoods Corporation. Centrifuge. www.tildenwoods.com.

[18] Robert G. Eggleston, Ratna Bearavolu, and Ali Mostashfi. Sensemaking support environment: A thinking aid for all-source intelligence analysis work. In *Proceedings of the First International Conference on Intelligence Analysis*, 2005.

[19] Kathleen M. Eisenhardt. Making fast strategic decisions in high-velocity environments. *The Academy of Management Journal*, 32(3):543–576, 1989.

[20] Douglas C. Engelbart. *AUGMENTING HUMAN INTELLECT: A Conceptual Framework*. STANFORD RESEARCH INSTITUTE, 1962.

[21] Warren Fishbein and Gregory Treverton. Making sense of transnational threats. *The Sherman Kent Center for Intelligence Analysis Occasional Papers*, 3(1), 2004.

[22] Frank L. Greitzer. Methodology, metrics and measures for testing and evaluation of intelligence analysis tools. Technical report, Pacific Northwest Division , Battelle Memorial Institute, 2005.

[23] Frank G. Halasz, Thomas P. Moran, and Randall H. Trigg. NoteCards in a Nutshell. In *Proceedings of the SIGCHI/GI conference on Human factors in computing systemsand graphics interface*, pages 45–52. ACM Press, 1987.

[24] Eric A. Havelock. *Preface to Plato*. Belknap Press, 1963.

[25] Arthur S. Hulnick. Managing intelligence analysis: Strategies for playing the end game. *International Journal of Intelligence and Counterintelligence*, 2:321–343, 1988.

[26] i2 Ltd. Analysts notebook. www.i2.co.uk.

[27] Robert Jervis. What's wrong with the intelligence process? *International Journal of Intelligence and Counterintelligence*, 1(1):28–41, 1986.

[28] Loch Johnson. Seven sins of strategic intelligence. *World Affairs*, 146(2):176–204, 1983.

[29] Loch K. Johnson. Analysis for a new age. *Intelligence and National Security*, 11:657–671, 1996.

[30] Rob Johnston. Developing a taxonomy of intelligence analysis variables. *Studies in Intelligence*, 47(3), 2003.

[31] Rob Johnston. Integrating methodologists into teams of substantive experts. *Studies in Intelligence*, 47(1), 2003.

[32] Rob Johnston. *Analytic Culture in the U.S. Intelligence Community - An Ethnographic Study*. Center for the Study of Intelligence, 2005.

[33] Paul Kay and Willett Kempton. What is the sapir-whorf hypothesis? *American Anthropologist*, 86:65–79, 1984.

[34] Thomas H. Kean and Lee H. Hamliton. *The 9/11 Commission Report*. 2002.

[35] Alden S. Klovdahl. A note on images of networks. *Social Networks*, 3(3):197–214, 1983.

[36] Valdis E. Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.

[37] Lisa Krizan. Intelligence essentials for everyone. *Joint Military Intelligence College Occasional Papers*, 3, 1999.

[38] Stephen Marrin. Preventing intelligence failures by learning from the past. *Journal of Intelligence and Counterintelligence*, 17:655–672, 2004.

[39] Catherine C. Marshall, Frank G. Halasz, Russell A. Rogers, and Jr. William C. Janssen. Aquanet: a hypertext tool to hold your knowledge in place. In *HYPERTEXT '91: Proceedings of the third annual ACM conference on Hypertext*, pages 261–275, New York, NY, USA, 1991. ACM Press.

[40] Duncan McAndrew. The structural analysis of criminal networks. In *The Social Psychology of Crime*, volume 3 of *Offender Profiling Series*, pages 51–94. Ashgate Publishing Company, 2000.

[41] Mary O. McCarthy. The mission to warn: Disaster looms. *Defence Intelligence Journal*, 7(2):19–31, 1998.

[42] Marshall McLuhan. *The Gutenberg Galaxy*. University of Toronto Press, 1962.

[43] Nasrullah Memon and Henrik Legind Larsen. Practical algorithms for destabilizing terrorist networks. In *ISI*, pages 389–400, 2006.

[44] Norman Meyrowitz. Intermedia: The architecture and construction of an object-oriented hypemediasystem and applications framework. In *Conference proceedings on Object-oriented programming systems, languagesand applications*, pages 186–201. ACM Press, 1986.

[45] Susan Michalak and Mary Coney. Hypertext and the author/reader dialogue. In *HYPERTEXT '93: Proceedings of the fifth ACM conference on Hypertext*, pages 174–182. ACM Press, 1993.

[46] George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 56.

[47] Wendy Morgan. Electronic tools for dismantling the master's house: poststructuralist feminist research and hypertext poetics. In *HYPERTEXT '99: Proceedings of the tenth ACM Conference on Hypertext and hypermedia : returning to our diverse roots*, pages 207–216. ACM Press, 1999.

[48] Emile Morse, Michelle Potts Steves, and Jean Scholtz. Metrics and methodologies for evaluating technologies for intelligence analysts. In *Proceedings of the First International Conference on Intelligence Analysis*, 2005.

[49] Stuart Moulthrop. Hypertext and the "the hyperreal". In *HYPERTEXT '89: Proceedings of the second annual ACM conference on Hypertext*, pages 259–267. ACM Press, 1989.

[50] Stuart Moulthrop. Beyond the electronic book: a critique of hypertext rhetoric. In *HYPERTEXT '91: Proceedings of the third annual ACM conference on Hypertext*, pages 291–298, New York, NY, USA, 1991. ACM Press.

[51] Raymond S. Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2:175–220, 1998.

[52] Walter J. Ong. *Orality and Literacy*. Routledge, 2005.

[53] George Orwell. *Nineteen Eighty-Four*. Secker & Warburg, 1949.

[54] Peter Pirolli. Assisting people to become independent learners in the analysis of intelligence. Technical report, Palo Alto Research Center, Inc., 2005.

[55] Robert Popp, Thomas Armour, Ted Senator, and Kristen Numrych. Countering terrorism through information technology. *Commun. ACM*, 47(3):36–43, 2004.

[56] Karl Raimund Popper. *Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man*. Washington University, 1965.

[57] Jr Richards J. Heuer. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency, 1999.

[58] Charles S. Robb and Laurence H. Silberman. *Report to the President, March 31, 2005*. 2005. Also known as the "WMD" report.

[59] Madan Sarup. *An Introductory Guide to Post-Structuralism and Postmodernism*. University of Georgia Press, second edition, 1993.

[60] Malcolm K. Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks*, 13(3):251–274, September 1991.

[61] Sarah M. Taylor. The several worlds of the intelligence analyst. In *Proceedings of the First International Conference on Intelligence Analysis*, 2005.

[62] Peter Thielst. *Man bør tvivle om alt - og tro på meget*. DET lille FORLAG, second edition, 1996.

[63] Peter Todd and Izak Benbasat. An experimental investigation of the impact of computer based decision aids on decision making strategies. *Information Systems Research*, 2:87–115, 1991.

[64] Peter Todd and Izak Benbasat. The influence of decision aids on choice strategies under conditions of high cognitive load. *IEEE Transactions on Systems, Man, and Cybernetics*, 24:537–547, 1994.

[65] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[66] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

[67] Peter Cathcart Wason. On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, 12:129–140, 1960.

[68] Karl E. Weick. *Sensemaking in Organizations*. SAGE Publications, 1995.

[69] Joseph Weitzenbaum. *Computer Power and Human Reason - Fron Judgement to Calculation*. Penguin Books, 1984.

[70] William Wright, David Schroh, Pascale Proulx, Alex Skaburskis, and Brian Cort. The sandbox for analysis - concepts and methods. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006.

[71] Jennifer Xu and Hsinchun Chen. Criminal network analysis and visualization. *Commun. ACM*, 48(6):100–107, 2005.

Automated Capture and Representation of Date/Time to Support Intelligence Analysis

David M. Cassel
Lockheed Martin
P.O. Box 8048
Philadelphia, PA 19101
610-354-4909
david.cassel@lmco.com

Sarah M. Taylor
Lockheed Martin
4350 N. Fairfax Drive
Arlington, VA 22203
703-351-4440x135
sarah.m.taylor@lmco.com

Gary J. Katz
Lockheed Martin
P.O. Box 8048
Philadelphia, PA 19101
610-354-5880
gary.j.katz@lmco.com

Lois C. Childs
Lockheed Martin
P.O. Box 8048
Philadelphia, PA 19101
610-354-5816
lois.childs@lmco.com

Raymond D. Rimey
Lockheed Martin
P.O. Box 277004
Mail Stop DC3535
Littleton, CO 80127
303-977-4811
raymond.d.rimey@lmco.com

ABSTRACT
Most intelligence analysis tasks require the analyst to have a mastery of a sequence of events. Tools supporting intelligence analysis must aid the analyst in untangling a variety of time related problems, including overlapping durations for events, and imprecise, incomplete, or conflicting information. Tools require both strong date and time identification/extraction and a flexible presentation of information on a time scale. We propose a method of presenting events, linked to a timeline, which addresses the problems of overlapping events, vague date/time references, and the need for the analyst to move easily between different time scales.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *graphical user interfaces, natural language.*

General Terms
Design, Human Factors, Languages

Keywords
Event analysis, intelligence analysis, time expressions, timeline

1. INTRODUCTION
Intelligence problems come in great variety, ranging from the development of situation awareness in preparation for a battle, to trying to understand the effects of a natural disaster on an economy, to following the progress and associations of a target person or object of interest. In almost all cases, an understanding of past events, and an ability to track the progress of events as they unfold, is a necessary part of the analysis applied to the problem. Additionally, fusion of information from multiple sources requires the best possible information about the time, as well as location, of the information – both the time and place of collection, and the time and place of the recorded events. Some sources of information may come to the analyst already supplied with precise date and time information of this type. However, in text sources, with which we primarily are concerned for this discussion, much information about the timing of events is vague or otherwise incomplete. Yet, the analyst must make the best of the information he has. His tools should support him.

A tried and true method for fusing and understanding events has been the timeline, a linear scale to which the analyst can link events. These were hand drawn for many years, of course, and now can be provided, in a variety of flavors, by commercially available software tools, or constructed manually (and thus tailored to a particular problem and data) using drawing software. However, there remain some challenges for automated timelines which have not been overcome to our satisfaction: automated input of information to the timeline; the effective presentation of the time scale, with events attached, at different levels of granularity; the effective representation of events with overlapping durations, as well as the ability to include events with vague date/time references and differing degrees of uncertainty. These representation issues need to be addressed with methods that are reasonably intuitive and provide clarity, not visual clutter.

We describe here our understanding of the needs that a timeline tool must meet, our approach to the automated extraction of time information from text for ingest into a tool, and our concepts for representing time information on a zoomable timeline. This project is on-going. The required information extraction capability, from text, is available from Lockheed Martin's AeroText™ software and the zoomable timeline has been implemented, with many of the features described here, in a research prototype.

2. TASKS REQUIRING TIMELINES

Our timeline tool targets event information included within text. The types of text for which it may be useful are news reports, government message traffic, and internet sources. Because of the volumes of such materials and the large numbers of events involved, particularly at the single event level, tools to automatically assist intelligence analysts in discovering and understanding events of interest are becoming increasingly important. We start with some examples of the types of analysis for which timelines can be particularly helpful. These are provided to clarify the need for a tool for those readers not familiar with the methods of intelligence analysis.

Recurring patterns of events can suggest methods of operation for planning and executing terrorist acts, for money laundering, drug smuggling, and any number of criminal activities. They can be helpful in understanding negotiating behaviors and diplomatic maneuvering as well as military operations. What series of observable behaviors typically appear before elections in this country? What are the typical preparatory steps for a certain organization to carry out a terrorist attack? Timelines make such patterns easier to see and to compare.

Exact sequences of events can be highly important. Did a certain meeting take place before or after the decision was made to carry out a particular policy? Are policy statements regularly softened or hardened after input by any one official? Did a certain piece of information become available only after the press conference by a particular agency? This kind of sequence information can be critical to building an analyst's understanding of the viewpoints and information that influenced policy determinations or decisions to act. Using a timeline can help the analyst determine and explain such sequences. [4]

Timelines are particularly useful in deconflicting different versions of events. If incompatible time expressions are used, it can be difficult to see the conflict between different versions of an event until the time expressions have been normalized – one of the effects of putting events onto the timeline. For example, the conflict between the following two statements is more easily spotted in Figure 1 than in the text versions. "On March 15th (2001 understood from previous context), Mr. Jones left London for Kabul, beginning a trip of several months to countries in South Asia and the Arabian Peninsula." "Mr. Smith arrived in London in 2000, establishing a close working relationship with Mr. Jones there during the late spring of the following year." Mapping the events of both statements to a single timeline shows immediately that the statements cannot both be true.

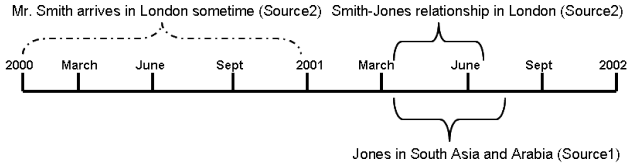


Figure 1. Depicting events on a timeline aids deconfliction.

Plaisant et al. showed in [5] the flexible utility of timelines by applying very similar systems to the juvenile justice and medical domains. For juvenile justice, the timelines illustrated events in from the records of troubled youths, enabling an easier method to search for patterns of behavior. In the medical domain, the LifeLines tool showed problems, allergies, diagnoses, complaints, lab work, imaging, medications, immunizations, and patient communication on a single view, while allowing the viewer to specify the interval and level of detail to be observed.

3. REQUIREMENTS FOR TIMELINES

Guidelines for annotation of time expressions in text [6], [3] make clear the wide range of time expressions which must be captured and represented by timelines. These include not only the obvious, such as "yesterday", "10:50am", "following", and the information conveyed by verb tenses, but also the less overt, such as the existence of states, e.g. "married" which in general usage imply a previous event, e.g. "wedding", or states of planning or expectation, which mean that something has not yet occurred, but may do so in the future. To truly capture the useful information from the text, any one of these kinds of time expressions must be able to be represented on a timeline.

There is a legitimate question, however, whether the timeline should in fact be made to represent every type of time expression; is it useful to try to recreate in the graphic form all the nuances of the linguistic form of the information? Without considerable experimentation and testing, we do not believe that question is definitively answerable. However, the approach we take is that there is a set of features that must be represented on a timeline, to which most time expressions can be mapped, and that these cover most of the cases for which timelines are typically used. We begin by addressing these clearly necessary features with our timeline tool.

Timelines are used to represent the ordering and spacing of events, that is, something, signaled by a verb in the language, that happened, a change from one state to another. States of being can also be represented, e.g. "Mr. Smith was in London from May until September", but these may be considered in terms of their beginning and ending events, e.g. in this case, possibly an arrival in May and a departure in September, depending upon the context. But every event, or state, that must be represented is either a point or an interval in time. The differences in time expressions reflect not a wide range of different kinds of time, but differing degrees of knowledge and precision of expression concerning the point or interval of time of the event. If we wish to be extremely accurate, of course, all events are intervals of time, but many of them are short enough, in whatever time scale is being used, to be usefully represented as points.

Vague or missing time information is extremely common in text sources. Information may simply not be known, or even not knowable, in any greater detail or with any greater level of confidence than is expressed in the source. Where an event has not been observed directly, but is being inferred from the existence of other events, or is known from reporting by a third party, the language may be a faithful expression of the level of detail available to the speaker. In other cases, the level of detail available in the text is related to the importance accorded to the event in that particular discussion, with the key events of a news report, for example, more likely to be specified closely and the background information referred to with more general expressions of time. Thus, the same event may be referenced in one source with a specific time expression and in another with something much vaguer. These imprecise expressions – "sometime last month", "several times a year" – may be all that an analyst has available, and must be represented as faithfully as possible on the timeline in conjunction with the more specific information.

Reliability of the source information is also important for the timeline, as in any other analysis tool. Combined into a calculation of reliability are estimates of the worth of the original report and of the chain of automated and human processing the information underwent before it reached the timeline display. Methods for automatically deriving this kind of provenance are not yet mature enough to include in our current version of the tool. Eventually, source reliability will necessitate a second layer of uncertainty information being incorporated into the timeline, over and above the information about the imprecision of the language content which we are incorporating in this version.

Aside from imprecision of expression and estimates of reliability/uncertainty, three major challenges for representation on timelines remain: a wide range of scales, large numbers of events, and characterizing event content.

Time scales must be able to represent scales from the level of seconds to the level of centuries. Similar to the changes in scale required for geographic analysis, it may be necessary to view events at any number of levels – at a higher level to see the evolution of a situation over a ten year period, yet also being able to focus on the detailed happenings of any particular week or day within that period. This requires not only an ability to grow or shrink the scale of the timeline, while retaining the user's focus, and orientation, but also the ability to meaningfully summarize events for display on the higher scales, and to decompose those summaries into their constituent pieces for more detailed analysis. Otherwise the more general scale – the decade view for example – simply becomes indecipherable for all the detailed events that have been attached to it. So, concomitant with our work on the multiple scales of the timeline itself, is work, not reported in detail here, on the association of component events into larger and more abstract constructs.

Large numbers of events on a timeline present similar problems to other analysis tools. But the possible solutions are more constrained. Unlike a link chart, for example, the time scale, which must remain easily understood, can be less easily stretched or rearranged to accommodate more material.

Finally, for the timeline to be most useful, the events displayed on it must be represented in such a way that their most important characteristics are readily recalled and understood by the viewer. Lines, bars and dots are insufficient without additional information. It is necessary for these to be as efficient in their use of space, as intuitive, and as distinctive as possible. We are experimenting with different representations and expect to report that work elsewhere.

4. EXTRACTION OF DATES/TIMES FROM TEXT

Any system extracting date and time information in text must accomplish three tasks: first, it must locate time expressions in the text; second, it must normalize the expressions into some standard and structured format; and finally, it must associate the date/time structures with the events or relationships they describe. The system we use for date/time extraction is Lockheed Martin's Information Extraction software tool, called AeroText™. We describe our approach to the location and normalization of date/time expressions below.

Two widely recognized schema for structuring date/time information are Timex2 and TimeML. Timex2 is an SGML-based format used to represent a broad array of times, including specific dates, durations, sets, fuzzy dates, and non-specific dates [3]. TimeML includes Timex3 tags, which supplement the attributes of Timex2, as well as other tags for marking events [6].

These standards have the flexibility to represent the breadth of temporal expressions necessary for dealing with natural language. Many programming languages provide some sort of date object, but these can typically only express specific dates. In contrast, consider the Timex2 specification for three expressions: "8:15 tonight", "Thursdays in May", and "the past three summers" (these examples each assume a reference date of June 1st, 2006).

<timex2 val="2006-06-01T20:15">8:15 tonight </timex2>

<timex2 set="YES" val="2006-05-WXX-4>Thursdays in

<timex2 val=2006-05">May</timex2></timex2>

<timex2 val="P3SU" anchor_dir="BEFORE" anchor_val="2006-06-01">the past three summers</timex2>

The first example shows a specific date and time, which Timex2 represents precisely using the same level of granularity provided by the expression. The second shows a set of dates. This set is precise, but would typically require one date object per Thursday to represent it. The third example shows a case that is vague – it does not specify when the summers began or ended.

Note the difference between the "val" attributes for "8:15 tonight" and "May". In the first case, the val attribute is specified down to the minute, whereas the second drops everything after the month. This approach does not assign arbitrary values for information that is not available; rather, the representation reflects all the information given and no other.

In 2004, a Time Expression Recognition and Normalization (TERN) [7] evaluation was first conducted by a team consisting of MITRE Corporation, SPAWAR Systems Center and the National Institute of Standards (NIST). Then, in 2005, the Automated Content Extraction (ACE) [1] evaluation included TERN as a separate task as well as requiring that extracted events, in other tasks, have a time associated with them. ACE and TERN rely on the Timex2 standard.

Lockheed Martin used the AeroText tool to address tasks in both the 2004 TERN and the 2005 ACE evaluations. AeroText software recognizes temporal expressions with a hand-crafted set of rules, normalizing them for internal storage to an interval form using Coordinated Universal Time (UTC). Approximately 250 rules were required for the TERN evaluations, including adaptations of the AeroText native interval forms to the output normalized forms required by TERN.

A general issue for time handling is the interpretation of expressions with respect to a reference time, which must be identified. The reference time may be the date of the document, as in "yesterday" used in a news story, and referring to the day before the date of the by-line. However, other expressions can require more complicated reasoning. For example, the interpretation of "in July", from a document written in August, might depend upon whether an associated verb was in the past or present tense to distinguish between the preceding or the following July. Embedded time expressions, such as "on Friday night in the Fall of 1998", are another complication addressed by the system. To handle these more difficult cases, our system looks for combinations of time expressions that were meaningful as a whole and applies the rules after the initial recognition phase has identified the single time expressions.

For TERN, it was necessary to find and output times in eight categories of normalized value: calendar point (a specific date or time, such as 1999), week point (includes a week indicator such as week 20 in the year 1999), non-specific (expressions underspecified in relation to the standard), duration (fully specified time periods), token (a reference to a previously specified value); prefixed (a year plus a two character prefix), no val (the actual date value is not expressed, e.g. "the anniversary"), and other (other partially specified expressions, e.g. "tonight"). It is instructive that this kind of normalization is by itself not sufficient for representation on a timeline, but requires further interpretation. The AeroText system handled calendar point, week point, token, and prefix expressions with fairly straight-forward translations from the original system. The remaining categories, dealing with less specific expressions required more significant addition of rules.

In 2004, the TERN evaluation used two scoring programs, one from MITRE and one from NIST. The MITRE scores are similar to those used in the Message Understanding Conferences (MUC) of the 1990s. The 2005 evaluation dropped the MITRE scoring; however, we provide those unofficial scores for comparison. The official NIST scores from the 2005 evaluation have been posted on the Internet [1]. There were only four participants in 2005 TERN. This is the first year such scores have been posted publicly.

The AeroText system for TERN and ACE, called HyperTERN, achieved excellent performance in the 2004 evaluation and saw only a slight decline in 2005 due mainly to two easily fixed glitches. The VAL attribute score fell somewhat because the system misidentified the document date in two document types, causing a miscalculation of normalization values. Secondly, the function to use the document date as the default Anchor_Val failed, resulting in a drop in recall from 75% to 6% in that slot. The MITRE score shows a 15% drop in F-Measure from 2004 to 2005. The drop in NIST value score was more dramatic, going from 78.1 to 56.2. This may be due to weighting factors applied by the NIST scoring program. A full description of the NIST metric can be found with the official posting of the 2005 results. Fixing the one document date glitch allowed our NIST score to rise to 83.4, an unofficial score, but one that clearly shows a continued excellence in extracting and normalizing temporal expressions.

Despite the extra work of normalizing the extracted expressions, HyperTERN achieves processing speeds greater than 200 Mbytes per hour. This system provides the basis of the date/time identification for our timeline solution.

5. A SOLUTION FOR AN ANALYST TIMELINE

Effective analysis requires that time expressions be normalized and attached to events. Furthermore, effective visualization techniques are required for analysts to be able to use the information. One solution is a timeline so that the analyst can see the ordering of events and look for patterns over time. We are implementing a new timeline view that we believe will improve on previous event visualizations for analysis. Our timeline provides smooth zooming, displays fuzzy date/time expressions, provides for abstract representations, shows a context bar, and allows for stacked visualizations, as explained below. We believe these meet the requirements already detailed, for representing imprecise times, expanding and compacting time scales, displaying large numbers of events, summarizing and decomposing events, and intuitively representing the key elements of events. We will continue to test and evolve these elements of the solution.

5.1 Smooth Zooming

Time sequencing can occur at many levels of detail. Our timeline provides a smooth zoom, whereby the amount of time displayed expands and contracts in response to mouse actions. As the timeline zooms in, additional detail becomes visible – first years, then months, then days; zooming as far as seconds is possible. Events on the timeline spread out and expand their time markings in real time as the zooming continues. Zooming out brings items together and hides time markings as they become too small to be useful. This allows analysts to work at the level of detail appropriate to their tasks.

As events are drawn together, it becomes necessary to use some of the display's vertical space to show events. Boxes pile in order to maintain their correct horizontal placement. Animation is used when a box is repositioned vertically; it slides into its new location in order to help the analyst maintain orientation.

5.2 Fuzzy Dates and Times

In order for analysts to work at different levels of detail they must understand the level of precision of their data. Improper analysis could occur if an analyst were sequencing events that occurred within minutes and the precision of the data was not easily recognizable. For example an event might start on a timeline at 1/5/06 at 12am. The display must clearly indicate whether the event began at exactly at 12:00:00 am, during the hour of 12 am, or sometime during the course of that day. This careful reflection in the timeline display of the imprecise language is necessary to facilitate correct analysis.

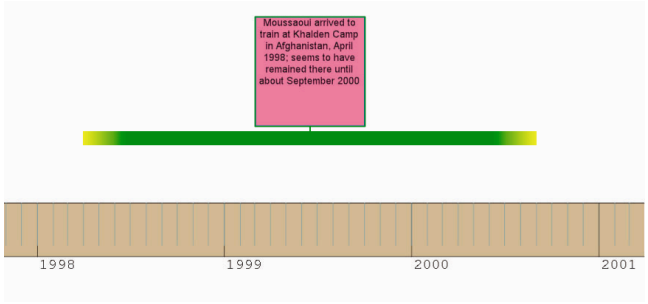


Figure 2. Gradient shading illustrates the likely range of an event.

Fuzzy times, along with more complex time expressions (examples above), can be represented internally by the Timex2 representation. Our timeline provides a visual representation of a variety of time expressions using the granularity that is supported by the available data. We are experimenting with various gradient representations for imprecise expressions as shown in Figure 2. The precision of an event is shown by shading its timeline to describe the probability of the event occurring during that slice of the time-span. In this case, the text says that "Moussaoui arrived to train at Khalden Camp in Afghanistan, April 1998; seems to have remained there until about September 2000." The bar under the box containing this text is a dark green for most of the period from April 1998 to September 2000. However, at the beginning and end of the period, the color fades and turns yellow at the outer

points. This shows that the information about Moussaoui's arrival and departure are approximate. Traditional representations would simply show Moussaoui in Khalden Camp from April 1st, 1998 to September 30th, 2000, creating the illusion that he could not have been somewhere else in the beginning or ending period. With the gradient representation, an analyst can reconcile another report that showed the subject in another location on April 3rd, 1998, for instance.

5.3 Abstraction

Each event shown on the timeline may be a single event or a collection of events. It is represented by a box depicting its contents either pictorially or in text. Double-clicking on a collection event causes it to explode to its constituent parts, as shown in the transition from part A of Figure 3 to part B. Similarly, a user may collapse the constituent pieces back into their container, cleaning up the display. A sub-item may be locked so that it remains visible when its siblings are collapsed. These transitions are animated in order to make the connection between parent and child nodes clear, and to help the viewer track the new items as the contents of the display are rearranged.

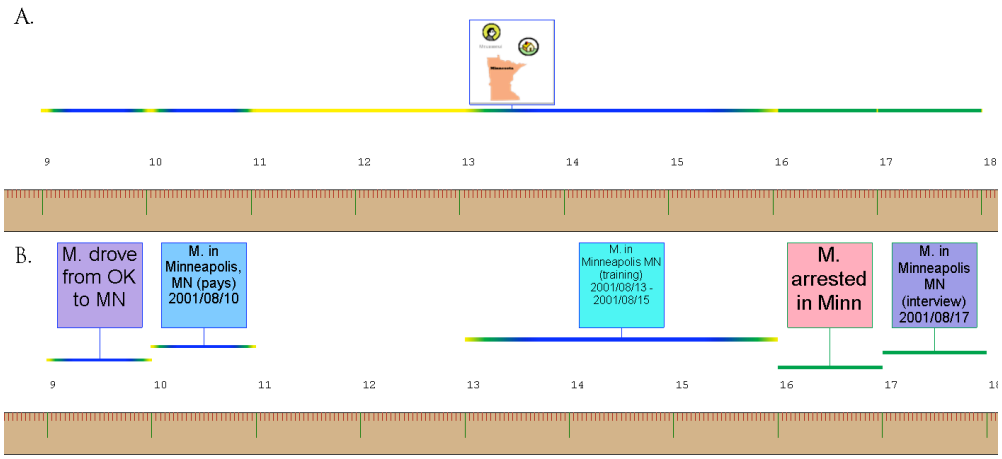


Figure 3. Abstract events contain more detailed sub-events.

As the parts of a collected event separate, each becomes an item with its own pictorial or text representation and its own time indicator. Collection items may in turn hold other collection items. In this way, an analyst may get an overview of a set of events and then drill down to view more detail. To reduce clutter, the analyst can hide items that are not of interest. The collection items form a hierarchy, the creation and maintenance of which are beyond the scope of this paper.

As events are combined in a hierarchy, more complex shaded timelines must be created from sub-event timelines being added together. When the collection item is broken down, the shaded time display itself breaks into constituent pieces, with each box showing its own time independently. This effect can be seen in parts A and B of Figure 3. Note that in part A, a section of the line is yellow, stretching from the 11th to the 13th. In part B, we see that this corresponds to a period not covered by any of the sub-events. The color coding provides insight into the abstraction presented in part A. Similarly, the lines under the boxes in part B use gradient shading to signify the vagueness of the endpoints. This shows the uncertainty as to when events actually happened.

5.4 Context Bar

Our timeline also employs a Context Bar, as shown in Figure 4. This bar appears just below the timeline and acts as a master timeline. It spans the range of time over which items of interest occur. Lines within the bar represent times when items of interest appear, giving the analyst context at a glance. The bar functions as a scroll bar across time: the span being displayed corresponds to the scroll bar thumb. Moving the thumb translates the view right (ahead in time) or left (backward in time). An analyst can also resize the thumb by positioning the mouse over the left or right edge and dragging. Resizing the thumb is one means of zooming the view. An approach like this was also used in the University of Maryland's LifeLines system. [5], [8]

The end points of the Context Bar are selected based on the overall span of events in which the analyst has indicated interest. In some cases, this selection can be problematic, such as "Up until 2002 …." but we use the earliest and latest events that have at least approximate start and end points, respectively.

Within the Context Bar itself, the lines indicating the presence of viewable events use the same level of abstraction as the view above on the main timeline; thus opening a container or collapsing children back into one changes the lines held within the Context Bar.

5.5 Stacked Timelines

The Context Bar provides another benefit: the system can present a stacked set of context bar timelines, with each showing a fixed period of time: a week, a month, or a year. In this way, the analyst can look for events that appear to occur at somewhat regular intervals. While data mining techniques can identify an event that occurs every Friday at 3pm, this view enables the analyst to find events that happen at less precise intervals, for instance, within the last few days of each month.

5.6 Traditional Timelines

[2] lists six shortcomings of traditional timeline visualizations.

1. *Views are discrete and fixed, thus they are bound to one specific [level of detail] and have a static size.*

In a display with a fixed level of detail, the user lacks the flexibility to determine how much information to view. Some kinds of analysis track events that happen over decades; for others, weeks or days may cover all the relevant events. Even within one of these types, a user may wish to shift back and forth between the big picture and other views of varying levels of detail.

Our timeline provides a smooth zooming capability, in which the timeline expands and contracts in response to user commands. The events on the timeline spread out or move together as the scope of the timeline changes. The context bar indicates how the period in the current view relates to the task as a whole.

2. *High-level views hide too much data, whereas detailed views suffer from lack of context and orientation.*

This is a continuation of the problem mentioned above. Neither a high-level nor a low-level view is correct all the time, so zooming plays a role in the solution. However, zooming by itself does not provide context or orientation. Our timeline provides two mechanisms: abstraction and the context bar.

The abstraction concept provides the analyst with control over the level of detail. By providing the option to limit the display to high-level concepts and then to drill down, the timeline offers conceptual context. Temporal context is provided by the context bar feature, which illustrates what portion of the overall period of interest is being shown. Smooth transitions, including animation as boxes move in the display, show the analyst what is changing as it happens. This helps the analyst maintain orientation.

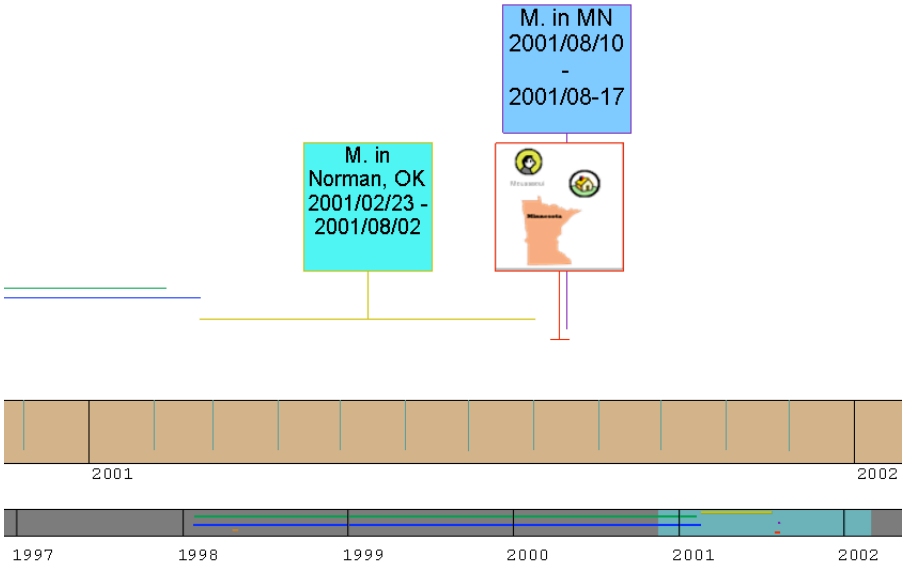


Figure 4. The Context Bar provides a broader view and zooming control.

3. *Scalability, e.g. for mobile devices is not supported.*

Some tools are built with the goal of scalability in mind, to allow for the same tools to be used regardless of the system on which the work is being done.

In our tool, the amount of the timeline shown depends on the width of the display window and the degree of zooming selected by the user. Given the nature of the analysis task, we believe that scalability up (for instance, to multiple screens) is more important for our work than scalability down (to mobile devices). The windows of the timeline can be expanded to take advantage of large or multiple screens.

4. *Missing support for zooming lenses, i.e. high-detail views or focus areas within a coarser time view.*

This weakness addresses the situation where groups of events are clustered at different timeline points, not distributed evenly along the line. For instance, consider the case where several events happen in January and several more happen in August. If the timeline scale is set broad enough to include all the events, very little detail will be discernible. If the scale shows enough detail, not all the events will fit on the display at once.

To address this problem, we began with the notion of a curved timeline, in which concave areas (bulging down) would bring events closer together and convex areas (bulging up) would spread them out. However, we now prefer [2]'s solution to this problem and plan to evaluate it more thoroughly. It changes the scale of some portion of the timeline to provide more detail than is shown in neighboring sections, while maintaining a straight timeline. We believe this approach provides the same ability to collapse gaps in time in a manner that is easier to understand than the curved approach we originally considered.

5. *Requirement of additional cognitive efforts for reinterpretation and orientation due to missing smooth transition between views.*

In some timelines, changing the zoom is a step function, causing a sudden change in the period of time displayed. When this happens, the user must pause for reorientation to the new display. This problem is solved in our timeline through the use of smooth zooming, in which the display gradually changes in real-time in response to the user's actions. In this way, the user is able to watch the changes as they occur and thus maintain orientation.

6. *Display of absolute time is missing, i.e. scroll-bar based views (e.g. in e-mail applications) only show relative position within the collection.*

When a timeline shows only relative position, the user can easily lose track of when the observed events actually took place. Alternatively, the user may have to refer to other data points to determine temporal position, creating a distraction and breaking the user's focus.

Our timeline uses labeled hash marks to indicate the absolute time. These hash marks spread out and new ones appear as the user zooms in. Figure 3 shows hash marks indicating days of a month, whereas the marks in Figure 4 show years. Note that in Figure 4, hash marks for the months are also visible. These marks are unlabeled due to the level of detail selected. If zooming continued, the months would be labeled once they had spread out and marks for the days would appear.

In addition to the labeled hash marks, the context bar keeps the display grounded in the overall time span of the analysis.

6. CONCLUSIONS

We have indicated some of the intelligence problems for which timelines are useful to the analyst and the requirements for timelines that derive from these. We have outlined the approach we take to automated extraction of dates/times for the attachment of events to timelines. We have described the features of our timeline tool which we are implementing to address the requirements and compared our solutions to the critique of timelines contained in [2]. We believe we have addressed the major issues with workable solutions and will continue to test and

refine our approach through work with real data and with the help of feedback from working analysts.

7. ACKNOWLEDGMENTS

This research has been conducted using Lockheed Martin Internal Research and Development funding.

8. REFERENCES

[1] ACE2005: http://www.nist.gov/speech/tests/ace/ace05/doc/ace05eval_official_results_20060110.htm

[2] Dachselt, R. and Weiland, M. 2006. TimeZoom: a flexible detail and context timeline. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM Press, New York, NY, 682-687. DOI= http://doi.acm.org/10.1145/1125451.1125590

[3] Ferro, Lisa. 2004. TIDES: 2003 Standard for the Annotation of Temporal Expressions. http://timex2.mitre.org/annotation_guidelines/2003_timex2_standard_v1_3.pdf

[4] Moore, David T. and Krizan, Lisa 2003. Core Competencies for Intelligence Analysis at the National Security Agency. In Swenson, ed., *Bringing Intelligence About*. Joint Military Intelligence College, May 2003.

[5] Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., Shneiderman, B. (1998) **LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records** Revised version in *1998 American Medical Informatic Association Annual Fall Symposium (Orlando, Nov. 9-11, 1998)*, p. 76-80, AMIA, Bethesda MD. HCIL-98-08, CS-TR-3943, UMIACS-TR-98-56

[6] Saurí, Roser, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky, 2004. TimeML Annotation Guidelines, Version 1.1. April 2, 2004.

[7] TERN 2004: http://timex2.mitre.org/tern.html

[8] University of Maryland's LifeLines on-line demonstration http://www.cs.umd.edu/hcil/lifelines/latestdemo/chi.html

Re-Using Millions of Visualizations

Raymond D. Rimey, David S. Bolme
Lockheed Martin Corporation, Denver, CO
raymond.d.rimey@lmco.com

ABSTRACT

Our goal is to enable an individual analyst to utilize and benefit from millions of visualization instances created by a community of analysts. A visualization instance is the combination of a specific set of data and a specific configuration of a visualization providing a specific visual depiction of that data. As the variety and number of visualization techniques and tools continues to increase, and as users increasingly adopt these tools, more and more visualization instances will be created (today, perhaps only viewed for a moment and thrown away) during the solution of analysis tasks. This paper discusses what fraction of these visualization instances are worth keeping and why, and argues that keeping more (or even all) visualization instances has high value and very low cost. Even if a small fraction is retained the result over time is still a large number of visualization instances and the issue remains, how can users utilize them? This paper describes what new functionality users need to utilize all those visualization instances, illustrated by examples using an information workspace tool that is like a multimedia spatial hypertext system. The paper concludes with a concise set of principles for future analysis tools that utilize spatial organization of large numbers of visualization instances.

Categories and Subject Descriptors

I.6.9 [Simulation, Modeling, and Visualization]: Visualization – information visualization, knowledge visualization. H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – visualization retrieval. H.5.4 [Information Interfaces and Representation (HCI)]: Hypertext/Hypermedia.

General Terms

Algorithms, Design, Experimentation, Theory.

Keywords

Intelligence Analysis, Visualization Retrieval, Sensemaking, Information Workspace, Knowledge Visualization, Information Visualization, Spatial Hypertext, Zoomable User Interface.

1. INTRODUCTION

This paper addresses user communities that analyze massive and complex data to discover patterns of interest. Our motivation is to address problems faced by the user community of intelligence analysts, but many commercial applications face similar problems. A massive flow of raw data is currently accessible by an intelligence analyst, and the volume and variety of that raw data will continue to increase. An ever growing variety of algorithms operate on the raw data flow to filter, alert, reduce, transform, and otherwise process that data to extract intelligence value from the raw data. Intelligence value means that the data contributes to the solution of an identified analysis task. Information visualization techniques and tools also help the intelligence analyst extract intelligence value from the raw and transformed data.

The combination of increasing data flows and increasing use of visualization tools will lead to an increasing flow of visualization instances created by the intelligence community (Figure 1). A visualization instance is the combination of a specific set of data and a specific configuration of a visualization providing a specific visual depiction of that data. If a few thousand analysts each generate a few tens of visualization instances each day that are worth keeping, then each year they produce on the order of 10 million visualization instances. Thus, we refer to the general problem as re-using "millions" of visualization instances. Should all those visualizations be viewed momentarily and then destroyed? Or is there value in saving them? Should they be saved for a few hours, or days, or years? How can they be used analytically in the future and what new tool functionality is needed to enable that? This paper addresses these questions.

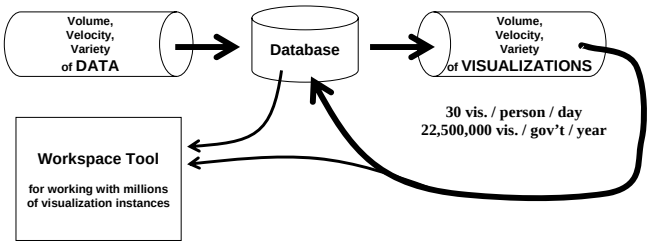


Figure 1. The massive number of visualization instances soon to be created across the intelligence community contain analytic value unrealized today.

Utilizing massive numbers of visualization instances is a new research area spanning the two research communities for information visualization and human-machine interfaces. Consider the three axes in Figure 2(a). The information visualization research community is focused on the far end of the x-axis, developing novel and highly sophisticated visual depictions of data. Classic, simple depictions of data (charts, timelines, maps, node-link diagrams) fall closer to the near end of the x-axis. Dynamic query techniques also fall roughly here. The simpler visualization techniques are well understood, have a proven track record of usefulness, and are more accepted by analyst end-users. Visualization tools that are more integrated with analytic tradecraft today are depicted by the shape in the x-y plane. The z-axis shows the number of visualization instances used by one analyst to complete one analytic task.

Today, analysts use (point A in Figure 2) small numbers of simple visualizations, which are well integrated with analytic

tradecraft. More sophisticated visualizations (point B) are used by fewer analysts, who use fewer instances of these visualizations, which are poorly integrated with analytic tradecraft. The frontier surrounding point B is slowly but surely being settled. This paper goes in a different direction, up from Point A, to a new part of the problem space denoted by point C. We address use of very large numbers of simple visualizations. The same issues and answers apply to large numbers of more sophisticated visualizations.

Solution processes that utilize many visualization instances can further be characterized by the axes in Figure 2(b). The first axis is the number of visualization instances *created* in the course of solving one analytic task. The second axis is the number of previously created visualization instances (created earlier in a complex task, or during previous tasks, or by other users) that are accessed again and *re-used* in the solution of the current analytic task. Today, visualizations are typically viewed for a relatively short period of time, and most or all are then destroyed. There is analytic value in saving more visualizations.

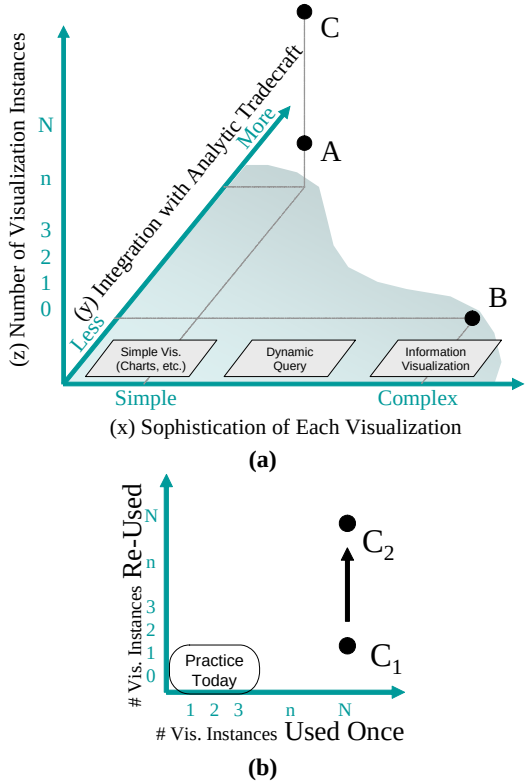


Figure 2. The new capabilities described in this paper enable analysts to access and utilize massive numbers of visualization instances. Point C in the problem space depicted here is elaborated as points C_1 and C_2.

This paper considers the future analyst work environment, where every analyst regularly utilizes a number of varied visualization tools. The visualization tools support simple visualizations (chart, scatterplot, etc.), dynamic query capabilities, and information visualizations for macro views and detailed study of massive datasets. The proportion of visualizations viewed to raw data viewed will be very high, perhaps 10 to 1. The process of completing one complex analytic task will involve creating or touching large numbers of visualization instances, perhaps hundreds.

1.1. Related Work

Storing and enabling re-use of large numbers of documents has a long history of research. Lifestreams [6] organized every document created or received by a user into a time-ordered stream of documents to which filters can be applied. Time Machine Computing [11] persists the small spatial area of a user's conventional computer desktop and allows the user to view the desktop as it existed at any point in the past. Emerging internet and desktop search capabilities can automatically organize text documents, for example by keyword or topic, creating document organizations on the fly to suite the immediate need of the user. The Keeping Things Found project [9] is focusing on how to organize things (e.g., retrieved web pages) so they can be easily re-accessed and re-used in the future. We are not aware of any work on re-use of *large* numbers of *visualization* instances.

VIKI [10] is a spatial hypertext system that supports information seeking with more emphasis on sorting through and (spatially) organizing the items as part of the information seeking process. NaviQue [7] was a similar but more powerful system, employing a zoomable canvas with scale (based on Zoomable User Interface or ZUI principles [1]) and allowing a wider variety of object types. DLITE [4] presents small persistent spatial work areas that contain all the data and tool objects needed for one task, and information seeking activities are mapped to spatial arrangements and manipulations of those objects. Our work maps information seeking and sensemaking activities to spatial organizations and manipulations of massive numbers of visual work materials on a ZUI canvas. Our work addresses analysis tasks that involve complex solution processes and extensive study of work materials.

Starlight is a 3D information workspace that has a small number of 3D information visualizations embedded within a 3D environment [12]. Those sophisticated information visualizations empower Starlight for intelligence analysis tasks. Data Mountain utilizes 3D spatial arrangements for managing a few hundred documents [13]. Our work addresses information workspaces that contain large numbers of visualization instances, many types of work materials, and large complex sets of work materials.

CoMotion and its ancestor Visage [14] treat data and visualizations on equal terms and allow them to be drag-and-dropped to create new visualizations during an analysis task. Visage can also directly map the analyst's exploration process to tree structured layouts of visualization instances [5]. CoMotion builds on the Visage ideas to create a collaborative information visualization environment [3]. Sandbox [17] is a 2D canvas that implements a few of our spatial organization ideas.

This paper is organized as follows. Section 2 discusses why an analyst should store and re-use visualization instances at potentially all points in their analytic solution processes. Section 3 introduces the new functions needed to enable re-use of large numbers of visualization instances, and Sections 4-6 discuss them in detail. Section 7 presents a set of principles for using large numbers of visualization instances.

2. WHY STORE AND RE-USE VISUALIZATIONS?

It is important to understand the work materials used by an analyst while performing an analytic task. We use the term *work materials* for all the things an analyst creates, touches or views while performing an analytic task. The work materials include raw data (e.g., short cables, long finished reports, tables, events, images, signals), a variety of simple visual objects (e.g., graphic

annotations, colored notecards, item glyphs), and, of course, visualization instances. While this paper focuses on large numbers of visualization instances, it is important to keep in mind that there are large numbers of these other work materials too. Some concepts in this paper apply to large numbers of work materials, not just visualization instances.

One of our colleagues, a highly experienced analyst, once commented about analysis that "you certainly don't want to keep everything in some kind of elaborate permanent structure; but some things it pays to keep up and available all the time." The crux of this comment is a value-cost tradeoff. Ignoring the cost, what is the value of keeping literally everything? What is the value of keeping 25% more than one analyst normally keeps today? The tools available to analysts today impose a high cost for keeping and organizing "everything", but they were never intended to do that, which must be done manually with little automation support, so of course the cost is high. Our proposed new functionality for future tools significantly lowers the cost.

An analyst needs only a limited amount of information to solve a task, and more experienced analysts often need less information [15]. They need the right pieces of information, not large amounts of information. Our concept of visualization re-use requires all visualization instances to be persistent; They will be invisible most of the time but be available for recall on demand. It is difficult to estimate the future "re-use value" of a visualization instance and decide whether to persist it at the time it is created, which suggests a strategy of simply saving everything. The point of visualization re-use is to provide tools to the analyst that lets her effortlessly retrieve a key previously created visualization instance -- one of those right pieces of information.

One of the most important tools for an analyst is her "shoebox", a highly selective collection of work materials the analyst has touched in the past and felt was worth saving because they may be useful for future related analysis tasks. The widespread use of shoeboxes tells us there *is* value in saving and re-using *some* of the analyst's work materials. The shoebox may contain raw data, annotated report-level items, etc., and in the future will increasingly include visualization instances. The ability to re-use visualization instances eases the process of finding items in a shoebox, enabling a larger and more valuable shoebox.

Where do visualization instances that are worth re-using come from? Visualization instances come from a single analyst's past work materials, and they come from other analysts' work materials. An intelligence analyst typically has one or more topic areas that they specialize in, and that analyst will be assigned many analytic tasks in that area over time. An analyst may find that her work materials from solving task B a month ago contribute to solving task G today. Complex tasks may involve a large amount of work materials, and some of those work materials will fade in the analyst's memory. So, visualization instances worth re-using during task G can also come from earlier in task G. Visualizations worth re-using also come from other analysts' work materials, because analysts often work in small groups and many analysts have overlapping areas of expertise/responsibility.

The ability to re-use visualization instances augments the analyst's short-term memory, long-term memory, and recall. It enables the analyst to better integrate her work materials and to better integrate materials from other analysts. For example, new patterns can be detected by correlating a visualization instance in the analyst's immediate work materials with one or more visualization instances in historical work materials.

Externalization is widely considered to be good analytic methodology. Externalization is "getting the decomposed problem out of one's head and down on paper or a computer screen in some simplified form that shows the main variables, parameters, or elements of the problem and how they relate to each other [8]." We believe that more externalization than can be achieved with today's tools is better. The new functionality presented in our paper here enables more externalization. Today's tools can force the analyst into single-threaded thinking. Our new tool functionality enables the analyst to explore multiple concurrent threads of thought.

The bottom line is that the ability to re-use visualization instances enables an analyst to find more patterns, find the patterns faster, and find new patterns that are not found using today's tools.

3. DESIGN ELEMENTS ENABLING USE OF A MILLION VISUALIZATIONS

Future analyst tools need several new functions in order to enable re-use of visualization instances:

- Methods for transferring visualization instances between tools and for storing them;
- Methods for organizing large numbers of visualization instances;
- Methods for querying for visualization instances and for relating visualization instances.

The next few sections discuss these functions and give examples using our 2D information workspace tool. Similar functionality can be implemented in other types of 2D workspace tools, 3D tools or WIMP tools.

3.1. Experimental Study Environment

Our Analytic Landscape (Anlan) tool, used for our examples, is briefly described here. Anlan is based on zoomable user interface (ZUI) principles [1], and is built on the Piccolo ZUI toolkit [2]. A ZUI presents the user with a single essentially infinite 2D canvas on which objects can be placed at any scale. The use of scale enables large amounts of work materials to be organized on the canvas (Figure 3). Objects must be organized on the canvas using spatial organization techniques, and the interface naturally leverages the excellent spatial abilities that people have.

The Anlan tool's canvas supports a variety of high-level object types such as notecards, images, graphical/text annotations, piles (representing sets of data items), visualization instances, and viewpoint bookmarks (that bring you to another area on the canvas). See Figure 4. Every object can be a container for other objects, done by dragging one object on top of another. Every object has a variety of automatically maintained metadata tags including date-time, location name and latitude-longitude, entities and keywords. Lens tools provide dynamic visualizations (such as maps and timelines) that operate on object metadata, and lens tools provide controls to marshal objects (e.g., sort them in time order). Visualization instances and other object types can be drag-and-dropped from third-party visualization tools. A simple query tool is built into Anlan and query results can also be obtained from third-party query tools.

The Anlan tool provides an intuitive spatial metaphor for organizing a wide variety of work materials. It is well suited for working with the large variety of data types, visual objects and visualization instances that an intelligence analyst encounters. The Anlan tool can be used for both analysis and presentation, allowing drilldown into analysis results from within the presentation. The Anlan tool provides the foundation for us to address issues involving the use of large numbers of visualization instances.

4. TRANSFERRING AND STORING VISUALIZATIONS

The foundation for re-use of large numbers, millions, of visualization instances is a standard XML representation for a visualization instance. The standard representation allows visualization instances to be transferred between tools (from any vendor) and it allows them to be stored in a database. This enables visualizations to be shared between users and between tasks. Communities of analysts will never be able to create and share large numbers of visual depictions of their data without such a standard. Of course, the greater challenge is the adoption not the design of a standard.

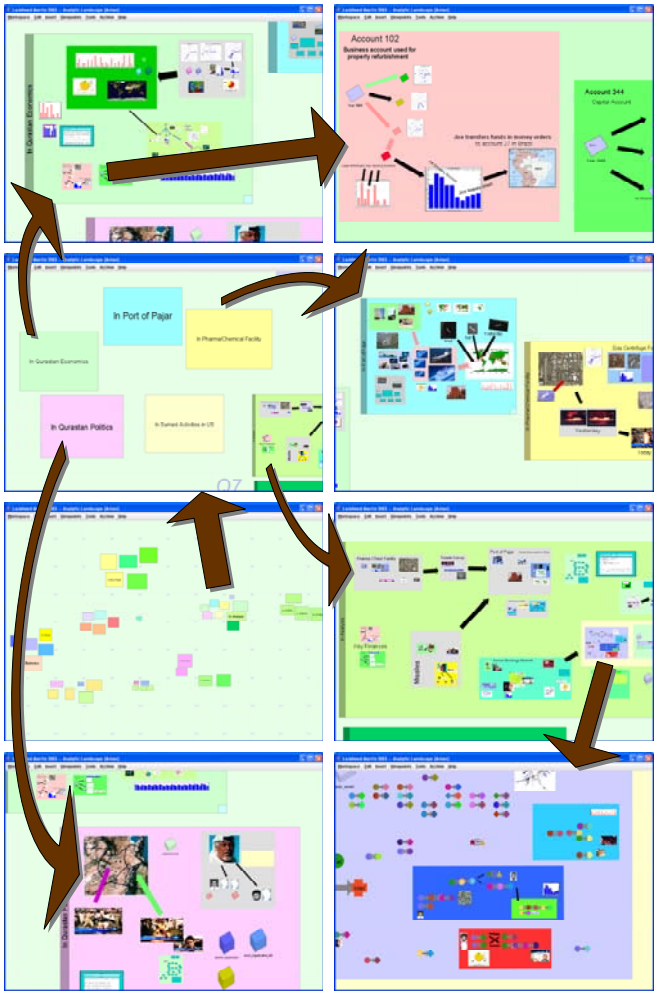


Figure 3. Work materials are organized on the Anlan tool's canvas using spatial organization techniques including the use of scale.

The standard representation must include metadata, for example all metadata associated with the underlying data for a visualization instance. Metadata may also include the task context in which the visualization instance was created. Visualization metadata makes it possible to query for visualization instances. Visualization metadata makes it possible to define relationships between visualization instances, and more generally relationships between collections of work materials. These topics are discussed in a later section.

Figures 5-7 show examples of drag-and-drop transfer of visualization instances and data collections between tools. The first example depicts transfers between our Anlan tool and a map tool. The analyst can use query capabilities in either the Anlan tool or the map tool to create a new data collection, which is depicted as a pile icon in the Anlan tool or a map layer in the map tool. Dragging a pile from Anlan to the map causes a new map layer to be added. For example, Figure 5 shows a pile of ELINT intercepts in Anlan, which is displayed as a set of ellipses in the map tool. Dragging a layer from the map's legend to Anlan transfers that layer's data collection into a pile in Anlan, and dragging a map icon from the map tool to Anlan transfers a copy of the current map view to Anlan. So, for example, as the analyst performs some work using the map tool, the key data collections and map views are dragged into Anlan, where they are persisted. Then, the analyst may do further non-map analysis of that data, for example using some of Anlan's capabilities to create charts and timelines of the data, or dragging the data to a third tool where charts and timelines are created, and eventually dragging refined sets of data back to the map tool. (These drag-and-drop capabilities in the map tool did require modifications to the map tool software. Our implementation of drag-and-drop uses both XML and HTML and is illustrative of the ideas here but it is not yet standardized. Anlan currently stores visualizations in a flat field-value format in a relational database.)

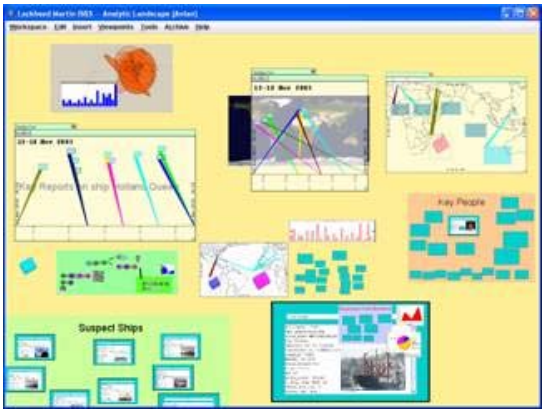


Figure 4. The Anlan tool allows a user to work with a wide variety of data types and a wide variety of visualizations.

Spreadsheet tools are commonly used by analysts. Dragging a pile from Anlan to a spreadsheet tool transfers the pile's data into an area of cells in the spreadsheet (Figure 6). Visualizations created in the spreadsheet tool can be dragged into Anlan. Transfers can also be made directly between the spreadsheet tool and our map tool, rather than going through our Anlan tool. (Currently, the visualization is dropped into Anlan as an image, however the spreadsheet application could be modified to transfer fuller descriptions of the objects, as we did with the map tool. Then, clicking on the visualization in Anlan could re-launch the visualization in a third-party visualization tool.)

The final example shows transfers from a query tool implemented within a web portal (Figure 7). We constructed a web portal that mimics some of the capabilities an analyst might use, such as the capability to query for image, ELINT and COMINT data. The query results pages contain drag-and-droppable icons for the entire query result set, the textual elements of a data item (e.g., image metadata), and an optional graphical element (e.g., the

image). Dropping these icons into Anlan respectively creates a pile, a notecard, or an image from the transferred data. Thus, Anlan is used to save queries (query definition and query result) and multiple visualizations of a query result, and Anlan enables the analyst to organize multiple queries and visualizations, discussed in the next section. Interchange with other tools is also possible, for example dropping a query result set icon or an image icon from the web portal into the map tool creates a new map layer containing that data.

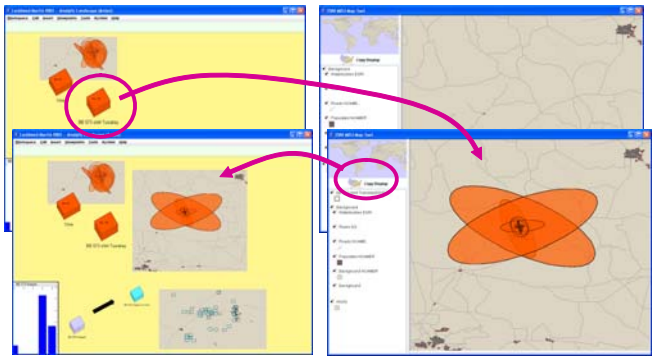


Figure 5. Drag-and-drop transfer of visualization instances and data collections between our Anlan tool and a map tool.

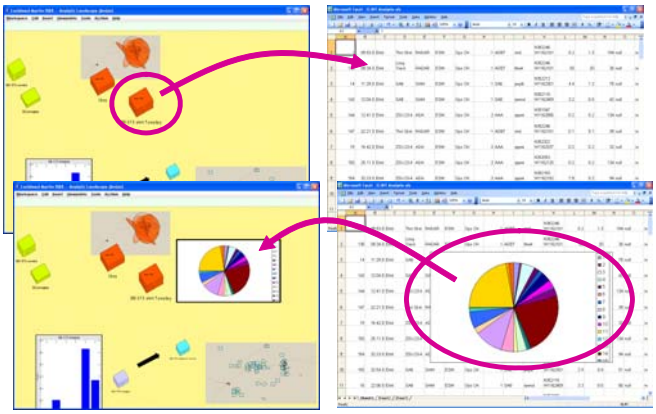


Figure 6. Drag-and-drop transfer of visualization instances and data collections between Anlan and a spreadsheet tool.

5. ORGANIZING VISUALIZATIONS

Once large numbers of visualization instances are being created and stored, methods are needed for organizing them. The classic approach is to assign a textual/iconic handle to each visualization instance and organize those in nested lists, where the user can select items from the list and see those visualization instances in a few separate windows. This approach does not scale well to very large numbers of objects (millions), and it introduces a mis-match between the spatial/visual nature of the visualization objects and the point/linear method of organization.

Our approach is to organize visualization instances (and all other types of work materials) spatially. Visualizations are inherently spatial things (a visual depiction of data, covering a local spatial area) and spatial methods for organizing multiple visualizations is a natural choice. Our spatial approach for organizing visualizations builds on zoomable user interface principles. The spatial organization method scales well to large numbers of (visual) items because it leverages human spatial skills and spatial memory. For example, most people have a good ability to find one of thousands of items spatially organized within their home, and have more difficulty navigating a pure textual list of those thousands of items.

Spatial organization techniques are relatively straightforward, and fall into two categories, area and path organizations. The key design issue is how the various structures within an analyst's task solution processes can be mapped to spatial organization techniques. Task solution processes involve two top-level ways to organize work materials: organization by topic and organization by thought process.

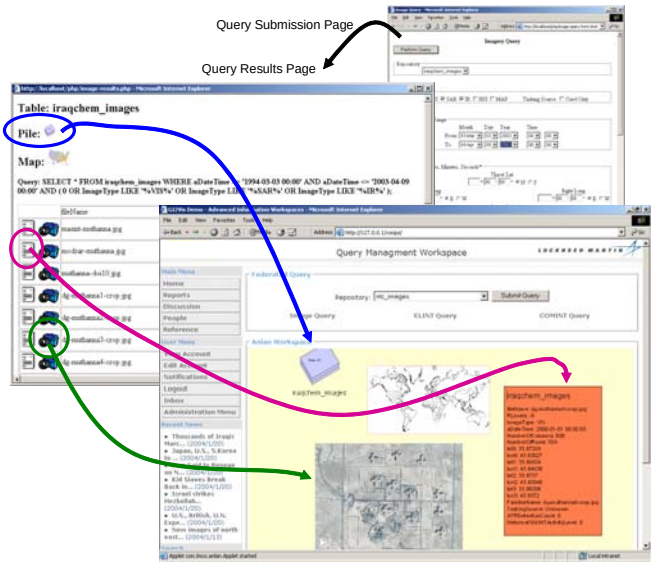


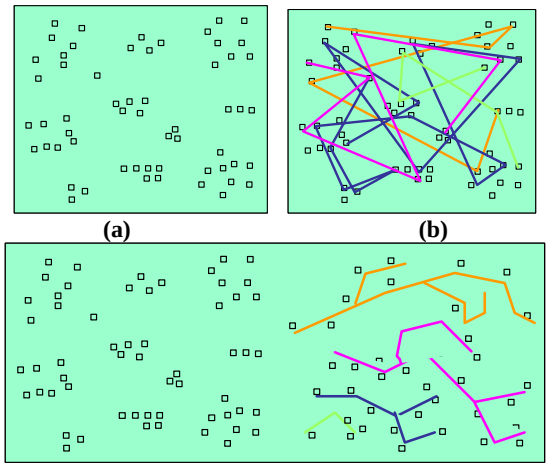
Figure 7. Drag-and-drop transfer between a web-based query tool and our Anlan tool. (The Anlan canvas in this example is within a web portal rather than in an application window.)

Solving a complex task involves a large number of organizations by topic and organizations by thought process, all interconnected (Figure 8). Working through a single thread of thought (externalized as a path organization of work materials on the Anlan canvas) will involve reaching into several previously organized topic areas (items organized using area structures, elsewhere on the Anlan canvas) and linking or cloning objects from those topic areas into the current thought process (the path organization currently under construction). We call this process *lassoing:* The analyst reaches into work materials previously organized by topic on the left side of Figure 8(c), clones a selected visualization instance, and places the clone within the work area for the current thought process on the right side Figure 8(c).

Area organizations are based on the idea of a container. The container can be a simple rectangle, or any object (e.g., a notecard – recall that every object in Anlan is a container), or a multi-compartment object. Scale is used to nest containers. Figure 9 shows some examples. (The Anlan tool includes a top-level container type, called a workcenter, which is currently the atomic unit of work materials that can be saved to and loaded from the database.)

Area organizations are commonly used to group objects by topic or some other shared criteria. Area organizations could be used to organize work materials according to structured analytic methods, for example by creating top-level workcenters named Hypotheses, Gather, Organize, and several Analyze areas to hold

the work materials in the corresponding parts of a structured analysis process. A customizable multi-compartment object can serve as a template, a standard spatial structure for organizing work materials for common analytic tasks. Multi-compartment objects can be extendable, providing controls around all the edges to add new area structure inside or outside the object, similar to building new walls inside an existing house floorplan or building a new addition.



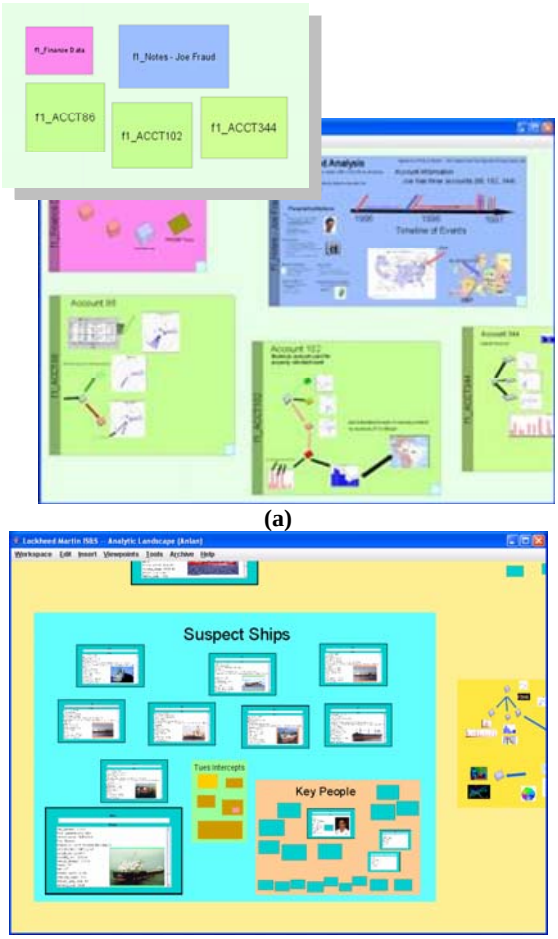
(a) (b) (c)

Figure 8. (a) Work materials spatially organized by topic; (b) The analyst's solution process requires jumping around the topic organization; (c) Work materials in the topic organizations (on left) are cloned and spatially organized according to solution processes (on right). Each tiny box denotes one visualization instance.

Path organizations utilize a single path, a branching path (tree), or a general node-link structure as part of the overall spatial structure. The path or links connect spatial objects, typically areas that contain additional internal structure.

Path organizations are typically used to depict thought processes or ordered analytic structure. An example of externalizing thought processes is that analysts often perform a series of queries with interspersed visualizations of the query results. That thought process is mapped rather directly to a branching path structure, where a series of queries corresponds with a branch in a spatialized tree structure, with piles and visualizations organized at each node along the branch. For example, in Figure 10(a), a financial analyst has identified a specific account for further study, so given an initial collection of bank transactions for that account, the analyst has broken the transactions into three categories, and is analyzing transaction patterns using some spiral visualizations. Each of these category piles is created as a subset of the original pile. Visualizations that spur further questions and thoughts cause the analyst to create additional queries and visualizations down that spatialized path of thought. (These branching structures are currently created manually in our Anlan tool, but a specialized tree container object could be provided with controls for adding a child (itself a container) at any point in the overall spatialized tree structure.)

As an example of an externalized analytic structure, Figure 11(a) depicts the analyst's mental model for a specific problem, and Figure 11(b) shows part of the analyst's workspace with work material structure that matches the mental model structure. Mental models [16] can be externalized as a spatial organization of work materials. This part of the workspace is a "summary" area, an area where the analyst has assembled key pieces of information (raw data, visuals, visualizations) from other areas of the canvas where the detailed analysis was performed. Those areas of detailed analysis could alternatively be located inside the summary area but at smaller scale.



(a) (b)

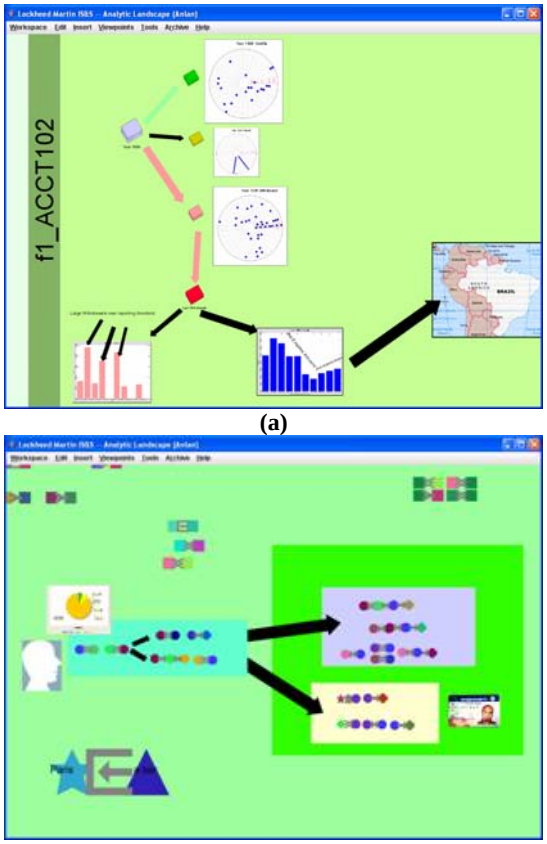
Figure 9. Area organization examples.

Analysts need to share work materials and to work collaboratively using shared work materials. The units of spatial organization (either area or path organizations) are the natural choice for units of work material (containers) to share between users. Each container and its contents are persisted in a database, so multiple clients can easily provide a shared view of a container, and some additional permission and control structure enables collaborative editing of container content. A user needs the ability to make private extensions to a shared container, and sophisticated user groups may need a "version control system" to manage multiple versions of containers.

6. QUERYING FOR AND RELATING VISUALIZATIONS

Re-using visualization instances from the large numbers created in the past, requires capabilities to retrieve visualization instances. Two kinds of information are available to help automate the

retrieval of visualization instances: stored metadata and computed context. The most reliable information is metadata stored with each visualization instance. The metadata for a visualization instance must include the metadata for the source data being visualized. Metadata automatically computed and attached to all source data includes date-time, location name and latitude-longitude, entities, topics and keywords. The metadata for a visualization instance should also include the task-specific context in which the visualization was created, for example the name of the workcenter or other labeled spatial container in which the visualization instance was created. When one visualization instance is derived from another visualization instance, that information should also be stored in the metadata. The term "computed context" refers to the idea that algorithms can operate on the work materials on the canvas nearby a visualization instance and compute contextual attributes for that visualization instance. The labels of the containing and nearby containers is one valuable source of contextual attributes. A visualization instance's current location on the canvas may be different from where it was created.

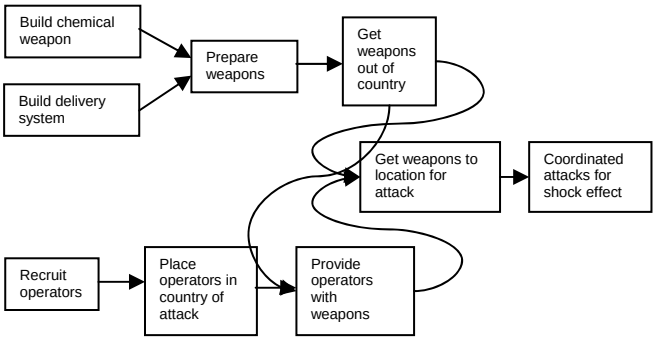


(a)

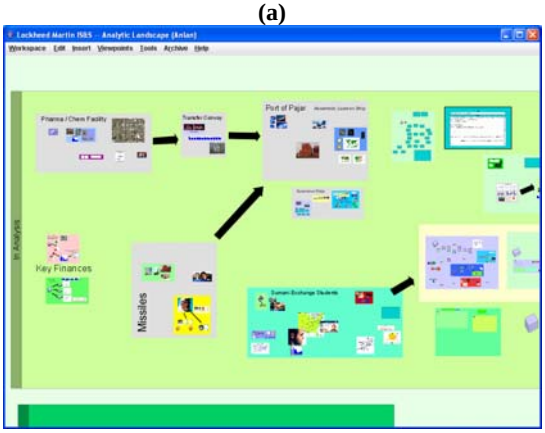
(b)

Figure 10. Path organization examples. (a) Analysis of a bank account. (b) Assembling events into larger activities.

An analyst needs a capability to explicitly query for visualizations (Figure 12). This is analogous to their need to query for raw data. When the analyst works primarily with visualizations rather than raw data, then queries will be performed for visualization instances rather than data. At the lowest level these are equivalent, since the most basic visualization metadata is a copy of the metadata for the data used to create that visualization instance. A second level of metadata can be added to visualization instances as stated in the previous paragraph. Our new capability to query for visualization instances in the analyst's work materials is also analogous to current capabilities to query for images on the internet.



(a)



(b)

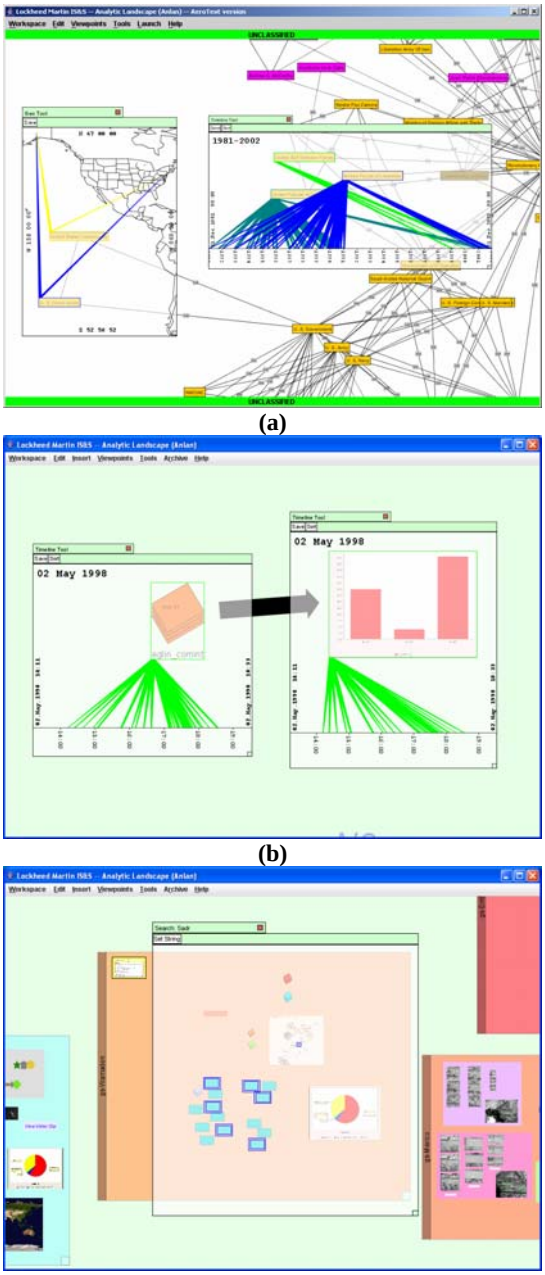
Figure 11. (a) Depiction of analyst's mental model for a specific analytic problem. (b) Analyst's spatial organization of work materials reflects the structure of her mental models.

The analyst needs a capability to find a visualization or a unit of work materials that is similar to the unit of work materials currently being studied. This capability is derived from (i) relationships between two visualizations and (ii) similarity criteria that combine several relationship metrics. A relationship between two visualizations is derived from a relationship between the two sets of data used to create the visualizations. For example, two bar charts that depict financial transactions and meetings between individuals are related if the same individual is included in each bar chart. Relationships between two visualization instances are derived from the source data's metadata for date-time, location name, latitude-longitude, entities, events, topics, keywords, etc. Finally, a higher-level similarity measure is constructed from several individual relationship metrics. This similarity measure is used to give the user a list of historical units of work, meaning previously created spatial containers on the canvas, that have strong similarity to the current unit of work, meaning the container currently being viewed on the canvas.

7. PRINCIPLES

This paper has identified the new functionality needed in future tools that will enable analysts to easily utilize and extract value from very large numbers of visualization instances. Many of the concepts apply more broadly to large numbers of work materials, rather than specifically visualization objects. Recall that work

materials include raw data and visual objects that are too simple to be considered visualizations (such as notecards, images, documents, annotations). This section presents a concise set of principles for future analysis tools that utilize spatial organization of large numbers of visually-oriented work materials.



(a)

(b)

(c)

Figure 12. Each visualization instance includes metadata derived from the raw data used to create it. Lens tools depict the metadata in (a) for each node of a node-link diagram and (b) in a chart. (c) This lens tool identifies all visualizations in a large area that meet a query criterion, simply a keyword query in this example.

Spatial Externalization: Important parts of the analysis process should be externalized on a persistent spatially organized workspace. Greater amounts of spatial externalization will improve the quality of work performed by the analyst. There are a few key concepts behind this principle: Externalized information is organized on a large persistent canvas. Analytical work involves modifying the externalization. At any given time the externalizations represent the solution state of the current task.

Integrated Process: The workspace should integrate all steps of the analysis process. An integrated workspace provides a seamless framework to hypothesize; gather, organize, and analyze information; and present results. Because each step in the process depends on the previous steps, it is important that modification from one step persist into the next. A totally integrated workspace has many advantages by providing continuity of the working areas between steps in the analysis process.

Integrated Information: The workspace should allow all types of analytic object to be placed on the work surface. The workspace provides one surface where reference materials, source data, queries, visualizations, report products, and personal notes can be processed simultaneously. This allows all the information to be easily compared and intermingled as the problem demands. The workspace contains both the user's private work and the external information world. The context and meaning of related objects is emphasized by local spatial organizations. A local area can be associated with a unit of work, enabling the relevant objects and tools to reside where that work is performed. Analytic objects should be directly manipulatable, support drilldown, and visually depict relationships with other objects.

Visual Thoughts: Important thoughts and processes should be visually represented on the work surface. The process of externalization allows the analyst to capture key evidence, thoughts, ideas, and processes in a way that is persistent. Those externalizations should exist on the work surface with the information that led to the conclusions about importance. The work surface provides trace of the analysis process. As the analyst works on the problem, she leaves trails of information and ideas that lead to important conclusions. Complex analytic processes can be retroactively studied via these visual traces, trails and any other dynamically created visual thought structures. Similar processes should look similar visually.

Related Object Neighborhoods: Analytic objects in the workspace should be placed near other related objects. Organizing information into neighborhoods on the workspace allows a set of related information objects to be viewed simultaneously on the screen. The analyst can use visual perception to augment working memory by using the work surface in this way. When new analytic objects are first placed on the work surface they will naturally be grouped based on the source of the information. Subsequently created neighborhoods may represent relationships to people, events, equipment, geographic location, hypothesis, etc. Since thought processes can have visual representation, they too can be organized using neighborhoods.

Pattern Discovery: Visuals in the workspace should aid in the discovery of patterns. One goal of the workspace is to aid the analyst in the discovery of interesting patterns. The patterns are found via manipulation of the visuals in the workspace. These manipulations allow the analyst to explore, organize, and visualize the information on the work surface. Creating multiple visualizations of some information may expose patterns. Information will often be projected into different spaces (e.g., plotted on a timeline, map, or organizational diagram) or into different contexts within one space to aid in the discovery of

patterns. Comparing two or more externalized thought processes side-by-side may also expose patterns.

Perceptual Leverage: The objects in the workspace should leverage the perceptual system. The human visual system can process complex information with little conscious effort from the analyst, and a workspace that efficiently uses the human perceptual system can greatly increase the analyst's productivity. It is important that visuals be easily understood and accurately convey the context and meaning of the information. Visual externalizations will increase the analyst's memory capacity by removing some of the load from her short-term memory. The interface should keep the user's attention at the work location.

External Interaction: The workspace should smoothly interact with the user's external work environment. Some aspects of the user's work environment will forever remain external, such as some hardcopy work materials and some third-party software tools, and the workspace should interface as smoothly as possible with them.

9. SUMMARY AND CONCLUSION

This paper presents the idea that analysts will benefit from future information visualization environments that have the new capability to access and re-use large numbers of visualization instances created by a community of analysts. An increasingly higher percentage of analyst work materials in the future will be visualization instances. The paper discusses the value of storing and re-using more (perhaps even all) intermediate work materials than is done today. The new capabilities needed to make this happen are described and illustrated with examples from a prototype workspace tool. Finally, the paper presents a concise set of principles to guide future work involving the storage and re-use of large numbers of intermediate work materials or visualization instances.

10. ACKNOWLEDGEMENTS

This work was supported by Independent Research and Development (IR&D) projects within Lockheed Martin IS&S (Integrated Systems & Solutions).

11. REFERENCES

[1] B.B. Bederson, et al., "Pad++: A Zoomable Graphical Sketchpad for Exploring Alternate Interface Physics," *Journal of Visual Languages and Computing*, 7:3-31, 1996.

[2] B.B. Bederson, J. Grosjean, J. Meyer, "Toolkit Design for Interactive Structured Graphics," *IEEE Transactions on Software Engineering*, 30(8):535-546, 2004.

[3] M.C. Chuah, S.F. Roth, "Visualizing Common Ground," *International Conference on Information Visualization*, 2003, pp. 365-372.

[4] S.B. Cousins, et al., "The Digital Library Integrated Task Environment (DLITE)," *ACM Conference on Digital Libraries*, 1997, pp. 142-151.

[5] M. Derthick, S.F. Roth, "Data Exploration Across Temporal Contexts," *International Conference on Intelligent User Interfaces (IUI)*, 2000.

[6] E. Freeman, D. Gelernter, "Lifestreams: A Storage Model for Personal Data," ACM SIGMOD Bulletin, March, 1996.

[7] G.W. Furnas, S.J. Rauch, "Considerations for Information Environments and the NaviQue Workspace," *ACM Conference on Digital Libraries*, 1998, pp. 79-88.

[8] R.J. Heuer Jr., *Psychology of Intelligence Analysis*, Center for the Study of Intelligence, U.S. Gov't Printing Office, 1999.

[9] W. Jones, H. Bruce, S. Dumais, "Keeping Found Things Found on the Web," *International Conference on Information and Knowledge Management*, 2001, pp. 119-134.

[10] C.C. Marshall, F.M. Shipman III, "Spatial Hypertext and the Practice of Information Triage," *ACM Conference on Hypertext*, 1997, pp. 124-133.

[11] J. Rekimoto, "Time-Machine Computing: A Time-Centric Approach for the Information Environment," *ACM Symposium on User Interface Software and Technology (UIST)*, 1999.

[12] J.S. Risch, et al., "A Virtual Environment for Multimedia Intelligence Data Analysis," *IEEE Computer Graphics and Applications*, 16(6):33-41,1996.

[13] G. Robertson, et al., "Data Mountain: Using Spatial Memory for Document Management," *ACM Symposium on User Interface Software and Technology (UIST)*, 1998.

[14] S.F. Roth, et al., "Visage: A User Interface Environment for Exploring Information," *IEEE Symposium on Information Visualization (InfoVis)*, 1996.

[15] S.M. Taylor, "How Much Information is Enough?", International Command and Control Research and Technology Symposium (ICCRTS), 2005.

[16] S.M. Taylor, "The Several Worlds of the Intelligence Analyst," *Int'l Conference on Intelligence Analysis*, 2005.

[17] W. Wright, et al., "Advances in nSpace - The Sandbox for Analysis," *Int'l Conference on Intelligence Analysis*, 2005.

Investigative Data Mining: Connecting the dots to disconnect them

Shaikh Muhammad Akram
Tsinghua University
Dept. of computer Science & Technology
Beijing, China
+86-10-62777703
alm04@mails.tsinghua.edu.cn

Wang Jiaxin
Tsinghua University
Dept. of computer Science & Technology
Beijing, China
+86-10-62777703
alm04@mails.tsinghua.edu.cn

ABSTRACT
The concern about national security has increased significantly since the 9/11 attacks. However, information overload and lack of advanced, automated techniques hinders the effective analysis of criminal and terrorist activities. Data mining applied in the context of law enforcement and intelligence analysis, called Investigative Data Mining (IDM), holds the promise of alleviating such problems. In this paper, we present an understanding to show how IDM works and the importance of this approach in connecting the dots to disconnect them in the context of terrorist network investigations.

Categories and Subject Descriptors
H.2.8 **[Database Applications]**: *Data mining*

Keywords
Investigative Data Mining, Link Analysis, Social Network Analysis, Visualization, Destabilization

1. INTRODUCTION
Criminals/terrorists seldom operate in a vacuum but interact with one another to carry out various illegal activities. In particular, organized crimes such as terrorism, drug trafficking, gang-related offenses, frauds, and armed robberies require collaboration among criminals/terrorists. Relationships between criminals/terrorists form the basis for organized crimes [1] and are essential for smooth operation of a criminal/terrorist organization, which can be viewed as a network consisting of nodes (terrorists) and links (relationships).

Criminal organizations such as terrorist networks are termed as covert organizations have network structures that are distinct from those in typical hierarchical organizations; e.g., they are cellular and distributed [2].

In criminal networks, there may exist groups or teams, within which members have close relationships. One group also may interact with other groups to obtain or transfer illicit goods. Moreover, individuals play different roles in their groups [3]. For example, some key members may act as leaders to control activities of a group. Some others may serve as gatekeepers to ensure smooth flow of information or illicit goods and some act as outliers in a group.

To analyze such criminal networks, investigators must process large volumes of crime data gathered from multiple sources. This is a non-trivial process that consumes much human time and effort. Current practice of criminal network analysis is primarily a manual process because of the lack of advanced, automated techniques. When there is a pressing need to untangle criminal networks, manual approaches may fail to generate valuable knowledge in a timely manner.

Fighting against criminal networks requires a more nimble intelligence apparatus that operates more actively and makes use of advanced information technology. Investigative Data-mining and automated data analysis techniques are powerful tools for intelligence and law enforcement officials fighting against such networks.

The rest of the paper is organized as follows: Section 2 gives background with specific reference to terrorist networks; Section 3 describes the concept of IDM framework in terrorist networks and discusses how to Destabilize these Networks; section 4 present a case study of a simple embassy bombing terrorist network in order to have clear understanding of IDM framework and section 5 concludes the paper and gives some future directions.

2. BACKGROUND
The information problem facing intelligence and law enforcement in preventing future terrorist acts is, large data volumes and limited analytic resources. However, compounding the problem is the fact that relevant data (that is, information about terrorist organizations and activities) is hidden within vast amounts of irrelevant data and appears innocuous (or at least ambivalent) when viewed in isolation. Individual data items – relating to people, places, and events, even if identified as relevant – are essentially meaningless unless viewed in context of their relation

to other data points. It is the network or pattern itself that must be identified, analyzed, and acted upon [4]. The main goal is the quest for interesting and understandable patterns. This search has always been, and will always be, a critical task in law enforcement, especially for criminal investigation, and more specific for the fight against terrorism. Thus, data mining for domestic security requires development of additional capabilities because the traditional data mining techniques were primarily developed to analyze propositional data – to analyze transactional data from unrelated subjects to make inferences about other unrelated subjects – and may be poorly suited for relational analysis in the context of domestic security. Examples are the discovery of interesting links between people (social networks, see, e.g., [5]) and other entities (means of transport, modus operandi, locations, communication channels like phone numbers, accounts, financial transactions and so on). *Post-hoc* analysis of the September 11 terror network shows that these relational networks exist and can be identified, at least after the fact [6].

The research into data search and pattern recognition technologies is based on the idea that terrorist planning activities or a likely terrorist attack could be uncovered by searching for indications of terrorist activities in vast quantities of transaction data. Terrorists must engage in certain transactions to coordinate and conduct attacks, and these transactions form patterns that may be detectable. Initial thoughts are to connect these transactions (e.g., applications for passports, visas, work permits, and drivers' licenses; automotive rentals; and purchases of airline ticket and chemicals) with events, such as arrests or suspicious activities.

A major challenge to terrorist detection today is the inability to quickly search and correlate data from the many databases maintained legally by our intelligence, counterintelligence, and law enforcement agencies.

Recently, it has become quite fashionable to analyze data by building an abstract graph (network) out of a set of observations and applying tools and techniques of graph theory. The recent emphasis in social network theory is one example of this approach. Usually, the network is created in such a way that nodes correspond to observational units of interest (for instance, people), while links between nodes reflect the strength of some known connection (i.e. the higher the support for the connection, the higher the strength of the link). Thus, even though the graph allows the analyst to focus on localized interactions (between two individuals, or institutions, or places, at a time), it still carries a notion that high-support events are more important than low-support ones.

3. INVESTIGATIVE DATA MINING

The rapid growth of available data in all regions of society requires new computational methods. Besides traditional statistical techniques [7] and standard database approaches, current research known as Investigative Data Mining (IDM) uses modern methods that originate from research in Algorithms and Artificial Intelligence. The main goal is the quest for interesting and understandable patterns. There are many ways in which IDM can be defined, one of its approach is states as: "The technique which is used for determining associations and predicting criminal behavior in criminal/terrorist networks in order to destabilize them".

The ultimate goal of IDM is to investigate terrorist networks in order to find out who the suspicious people are and who is capable of carrying out terrorist activities and how to destabilize them.

Investigative Data Mining (IDM) offers the ability to firstly map a covert cell, and to secondly measure the specific structural and interactional criteria of such a cell. This framework aims to connect the dots between individuals and "map and measure complex, covert, human groups and organisations". The method focuses on uncovering the patterning of people's interaction, and correctly interpreting these networks assists "in predicting behaviour and decision-making within the network". Investigative Data Mining usually uses SNA techniques and graph theory connecting the dots in order to disconnect them.

The main focus of IDM approach is to identify important actors, crucial links, subgroups, roles, network characteristics, and so on, to answer substantive questions about terrorist organizational structures. There are three main levels of interest: the element, group, and network level. On the element level, one is interested in properties (both absolute and relative) of single actors, links, or incidences. Examples for this type of analyses are bottleneck identification and structural ranking of network items. On the group level, one is interested in classifying the elements of a network and properties of sub networks. Examples are actor equivalence classes, cluster identification and associations. Finally, on the network level, one is interested in properties of the overall network such as connectivity or balance. The overall objects of interest are emergent patterns of relationships and their interplay with entity attributes in order to destabilize terrorist networks.

3.1 SOCIAL NETWORK ANALYSIS

Social network analysis (SNA) primarily focuses on applying analytic techniques to the relationships between individuals and groups, and investigating how those relationships can be used to infer additional information about the individuals and groups [8]. There are a number of mathematical and algorithmic approaches that can be used in SNA to infer such information, including connectedness and centrality [9].

Law enforcement personnel have used social networks to analyze terrorist networks [6,10] and criminal networks [11]. The capture of Saddam Hussein was facilitated by social network analysis: military officials constructed a network containing Hussein's tribal and family links, allowing them to focus on individuals who had close ties to Hussein [12].

SNA, originating from social science research, is a set of analytical tools that can be used to map networks of relationships and provides an important means of assessing and promoting collaboration in strategically important groups [13].

SNA has recently been recognized as a promising technology for studying criminal and terrorist networks. Social Network Analysis (SNA) provides a set of measures and approaches for the investigation of terrorist networks. These techniques were originally designed to discover social structures in social networks [14] and are especially appropriate for studying criminal networks [9,15,11].

Social network analysis describes the roles of and interactions among nodes in a conceptual network. Investigators can use this technique to construct a network that illustrates criminals' roles, the flow of tangible and intangible goods and information, and associations among these entities. Further analysis can reveal critical roles and subgroups and vulnerabilities inside the network [11]. The overall contribution of social network analysis to counter-terrorism is the ability to map the invisible dynamics inside a terrorist community.

The structural properties of a social network can be described and analyzed at four levels: node, link, group, and the overall network. SNA provides various measures, indexes, and approaches to capture these structural properties quantitatively.

Specifically, in the literature the use of centrality and structural equivalence measures from SNA are used to measure the importance of each network member. Several centrality measures, such as degree, betweenness, closeness, and eigenvector can suggest the importance of a node in a network [14] and can automatically identify the leaders, gatekeepers, and outliers from a network.

The degree of a particular node is its number of links; its betweenness is the number of geodesics (shortest paths between any two nodes) passing through it; and its closeness is the sum of all the geodesics between the particular node and every other node in the network whereas eigenvector centrality acknowledges that not all connections are equal. An individual's having a high degree, for instance, may imply his leadership; whereas an individual with a high betweenness may be a gatekeeper in the network. Baker and Faulkner [16] employed these four measures, especially degree, to find the central individuals in a price-fixing conspiracy network in the electrical equipment industry.

In general, the network studied in this paper can be represented by an undirected and un-weighted graph G = (V, E), where V is the set of vertices (or nodes) and E is the set of edges (or links). Each edge connects exactly one pair of vertices, and a vertex pair can be connected by (a maximum of) one edge, i.e., multi-connection is not allowed.

A terrorist network consists of V set of actors (nodes) and E relations (ties or edges) between these actors. The nodes may be individuals, groups (terrorist cells), organizations, or terrorist camps. The ties may fall within a level of analysis (e.g. individual to individual ties) or may cross-levels of analysis (individual-to-group analysis). A terrorist network can change in its nodes, links, groups, and even the overall structure. In this paper, we focus on detection and description of node level dynamics.

Mathematically, a network can be represented by a matrix called the adjacency matrix A, which in the simplest case is an n x n symmetric matrix, where n is the number of vertices in the network. The adjacency matrix has elements.

$$A_{ij} = \begin{cases} 1 & \text{if i and j are connected,} \\ 0 & \text{otherwise.} \end{cases}$$

The matrix is symmetric since if there is an edge between i and j then clearly there is also an edge between j and i. Thus

$$A_{ij} = A_{ji}.$$

Turning to the analysis of network data, we start by looking at centrality measures, which are some of the most fundamental and frequently used measures of network structure.

Centrality measures address the question, "Who is the most important or central person in this network?"

There are many answers to this question, depending on what we mean by important. Perhaps the simplest of centrality measures is degree centrality, also called simply degree. The *degree* of a vertex in a network is the number of edges attached to it. In mathematical terms, the degree D_i of a vertex *i* is [17]:

$$D_i = \sum_{j=1}^{n} A_{ij}$$

A network member with a high degree could be the leader or "hub" in a network.

Betweenness measures the extent to which a particular node lies between other nodes in a network. The betweenness B_a of a node *a* is defined as the number of geodesics (shortest paths between two nodes) passing through it:

$$B_a = \sum_{a}^{n} \sum_{j}^{n} g_{ij}(a)$$

Where $g_{ij}(a)$ indicates whether the shortest path between two other nodes *i* and *j* passes through node a. A member with high betweenness may act as a gatekeeper or "broker" in a network for smooth communication or flow of goods (e.g., money, arms).

Closeness C_a is the sum of the length of geodesics between a particular node *a* and all the other nodes in a network. It actually measures how far away one node is from other nodes and is sometimes called farness [16,20]:

$$C_a = \sum_{i=1}^{n} l(i,a),$$

Where $l(i,a)$ is the length of the shortest path connecting nodes *i* and *a*.

Though simple, degree is often a highly effective measure of the influence or importance of a node: in many social settings people with more connections tend to have more power.

A more sophisticated version of the same idea is the so-called eigenvector centrality. Where degree centrality gives a simple count of the number of connections a vertex has, eigenvector centrality (EC) acknowledges that not all connections are equal. If we denote the centrality of vertex *i* by x_i, then we can allow for this effect by making x_i proportional to the average of the centralities of i's network neighbors [72]:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

Where λ is a constant. Defining the vector of centralities x = (x1; x2; : : :), we can rewrite this equation in matrix form as:

$$\lambda x = A \bullet x$$

Hence we see that x is an eigenvector of the adjacency matrix with eigenvalue λ. Assuming that we wish the centralities to be non-negative, it can be shown that λ must be the largest eigenvalue of the adjacency matrix and *x* the corresponding eigenvector.

The equation lends itself to the interpretation that a node that has a high eigenvector score is one that is adjacent to nodes that are themselves high scorers. The idea is that even if a node influences just one other node, if that node influences many others (who themselves influence still more others), then the first node in that chain is highly influential. Hence, the eigenvector centrality measure is ideally suited for influence type processes.

3.2 DESTABILIZING TERRORIST NETWORKS

Destabilizing techniques traditionally aim at neutralizing members of terrorist networks either through capture or death. The removal of a node from a network can make a cell less able to adapt, reduce its performance, and reduce its ability to communicate. These nodes are known as the 'critical' nodes within a network. The removal or isolation of these nodes ensures maximum damage to the network's ability to adapt, performance, and ability to communicate.

In network analysis, node changes are the standard approach to network destabilization [18]. Using standard social network techniques, individuals who are key in the terrorist networks are identified and then removed. The argument is that their removal serves to weaken or break the network so that messages flow slower and so that the network as a whole is no longer a single entity [19].

The centrality approach, consisting of measuring the centrality [20] of each node in the network, then selecting a small number of most central nodes as targets for further action, is an intuitive approach to finding a core group of leaders within a terrorist network.

4. APPLYING IDM CONCEPT TO USA EMBASSY (TANZANIA) BOMBING NETWORK

To better understand how IDM works, we employed aforementioned IDM methods on Embassy bombing network dataset. For illustrative purposes we will use in this paper a reduced form of the embassy bombing data (EB data set) [21], and its Agent-Agent relationship between 16 terrorists in the form of an adjancey matrix (considering the value "1" for the presence of connection and value "0" for the absence of connection between them) as shown in Table 1 and Table 2 respectively.

Table 1. Terrorists of EB data set

S #.	Terrorists Original ID	Terrorist ID used for illustrations
1	Mohammed Rashed Daoud al-Owhali	MRDalOwhali
2	Khalfan Khamis Mohamed	KKM
3	Mohammed Sadiq Odeh	MSodeh
4	Ahmed the German	ATGerman
5	Fazul Abdullah Mohammed	Fazulam
6	Wadih al Hage	WAHage
7	Usama Bin Ladin	UBL
8	Ali Mohammed	AliM
9	Ahmed Khalfan Ghailani	AKGhailani
10	Mohammed Salim	Msalim
11	al-Fadl	alfadl
12	al-Fawwaz	alfawwaz
13	Jihad Mohammed Ali	JihadMA
14	abouhalima	Ahalima
15	Abdullah Ahmed Abdullah	Aaa
16	Abdal Rahmad	Arahmad

Table 2. Adjancey Matrix Data

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 MRDalOwhali	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1
2 KKM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 Msodeh	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4 ATGerman	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
5 Fazulam	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6 WAHage	0	0	1	0	1	0	1	1	0	0	0	1	0	1	0	0
7 UBL	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0
8 AliM	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
9 AKGhailani	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10 Msalim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11 alfadl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12 alfawwaz	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
13 JihadMA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14 Ahalima	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
15 Aaa	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
16 Arahmad	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Terrorist Link Analysis chart (network) of USA embassy bombing in Tanzania using NetDraw [22] is drawn and shown in figure 1.

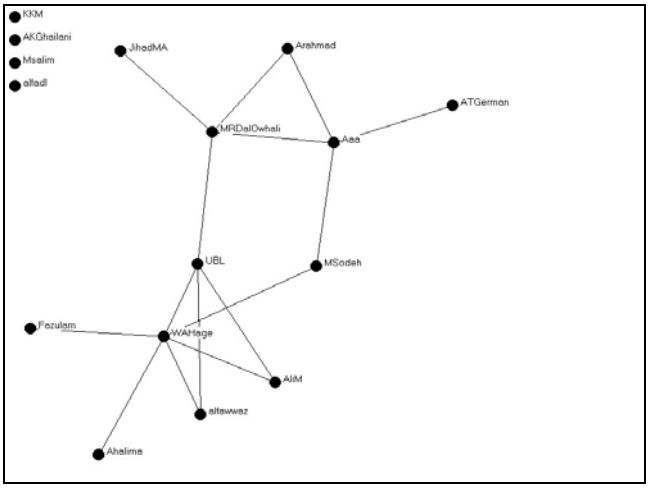


Figure 1. Terrorist Link Analysis chart (network) of USA embassy bombing in Tanzania

Table 3 shows four basic centrality measures of terrorists of USA embassy bombing in Tanzania and are calculated using UCINET [23].

Table 3. Basic Centrality measures of terrorist network

S #	Terrorist ID	Degree	Betweenes	Closeness	EC
1	MRDalOwhali	**4**	**18.5**	**7.2**	**11**
2	KKM	**0**	**0**	**0**	**0**
3	MSodeh	**2**	**9.5**	**6.4**	**10**
4	ATGerman	**1**	**0**	**4.5**	**4**
5	Fazulam	**1**	**0**	**5**	**6**
6	WAHage	**6**	**24.5**	**8**	**12**
7	UBL	**4**	**16**	**7.4**	**14**
8	AliM	**2**	**0**	**5.8**	**10**
9	AKGhailani	**0**	**0**	**0**	**0**
10	Msalim	**0**	**0**	**0**	**0**
11	alfadl	**0**	**0**	**0**	**0**
12	alfawwaz	**2**	**0**	**5.8**	**10**
13	JihadMA	**1**	**0**	**4.7**	**4**
14	Ahalima	**1**	**0**	**5**	**6**
15	Aaa	**4**	**14.5**	**6.9**	**9**
16	Arahmad	**2**	**0**	**5.5**	**8**

By analyzing and using standard social network techniques in the embassy bombing terrorist data the individuals who are key in the network are identified (red nodes) and then removed as shown in figure 2.

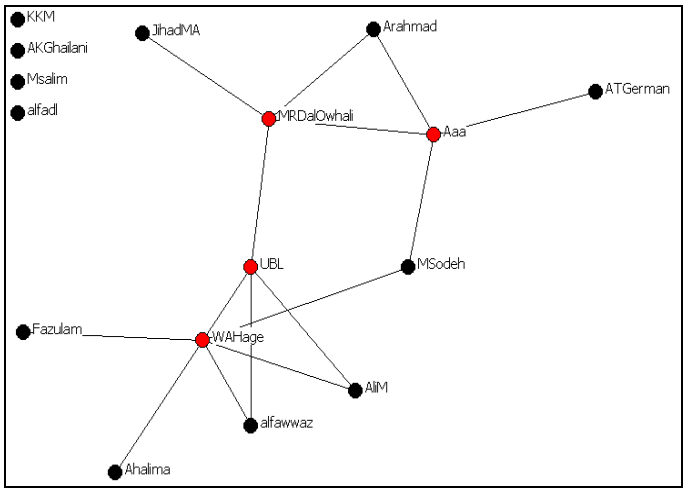


Figure 2. The important nodes marked as red in embassy bombing terrorist network using centrality concept.

After removing the important nodes WAHage, UBL, MRDalOwhali, and Aaa (red) the shape of the disconnected terrorist network is shown in figure 3:

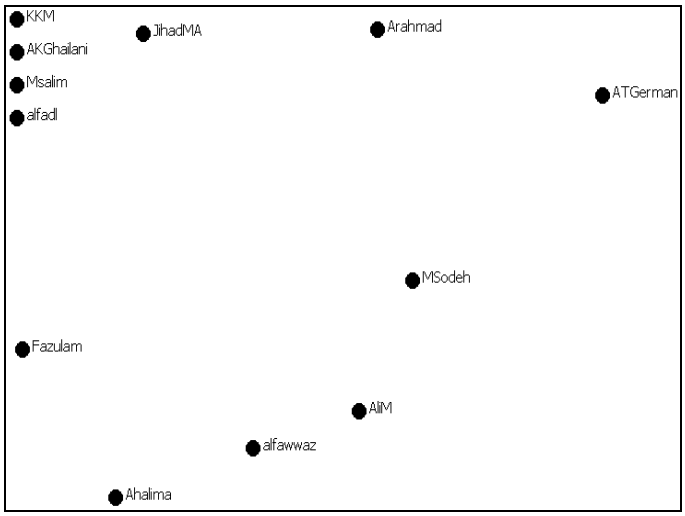


Figure 3. The disconnected embassy bombing terrorist network after the removal of important nodes.

The argument is that the removal of important nodes serves to weaken or break the network as shown in the analysis. However, despite the apparent success of this approach, evidence indicates that this technique is not effective as in the case of dynamic, distributed and decentralized terrorist network such as alqaeda [24]. These type of networks have the ability to heal themselves after the removal of any key member (leader) form the network [25]. In fact, leadership removal may make the network more dense to future analysis given the emergence of new leadership that may not be known. Kathleen Carley et al. proposed three indicators of destabilization [26]:

- The rate of information flow through the network has been minimized (perhaps to zero).
- The network, as a decision making body, cannot reach on a joint consensus.
- The ability of the network to accomplish tasks is totally impaired.

Based on the above discussion, it is very important to know about the structural properties to gain insight to the following questions in order to not only disconnect but also destabilize these terrorist networks:

- What is the overall network structure?
- Who is the key player in the network?
- What is the efficiency of the network?
- Who is highly connected to whom?
- Who shares most resources to whom?
- Which terrorist if removed would highly disrupt the network communication?

These and many more questions could be answer by following the overall structure and dynamics of the terrorist networks.

5. CONCLUSION AND FUTURE DIRECTIONS

It is believed that reliable data and sophisticated analytical techniques are critical for law enforcement and intelligence agencies to understand and possibly disrupt terrorist or criminal networks.

In this paper, we have presented an overview of investigative data mining with its basic framework and tried our best to shed some light on the issues. We believe that investigative data mining has a promising future for increasing the effectiveness and efficiency of counter terrorism and intelligence analysis. In addition to this we have also discussed the few available approaches and their shortcomings for destabilizing terrorist networks. Many future directions can be explored in this still young field. For example, more visual and intuitive criminal and intelligence investigation techniques can be developed for counter-terrorism. Moreover we are also working on some practical approaches and algorithms for destabilizing terrorist networks, which are the integration of different techniques in which the concept is borrowed not only from SNA measures but also from mathematical order theory [27] and web structure analysis [28].

6. ACKNOWLEDGMENTS

The authors gratefully acknowledge the useful suggestions and comments given by Ms. Zhu Hongmei and Mr. Honbo Liu time by time.

7. REFERENCES

[1] McIllwain, J. S. *Organized crime: A social network approach*. Crime, Law & Social Change, Vol. 32. (1999). 301–323.

[2] Carley M. Kathleen. *Estimating Vulnerabilities in Large Covert Networks*. Proceedings from 9th ICCRTS Command and Control Research and Technology Symposium September 14-16, 2004, Copenhagen, Denmark.

[3] Jennifer Xu and Hsinchun Chen. *Criminal Network Analysis and Visualization: A Data Mining Perspective*, Communications of the ACM (CACM), 48(6), pp. 101-107, 2005.

[4] Hazel Muir, Email Traffic Patterns can Reveal Ringleaders, New Scientist, available at http://www.newscientist.com/news/news.jsp?id=ns99993550

[5]Paul R. Pillar. *Counterterrorism after Al Qaeda*. The Washington Quarterly. 27 (3) pp. 101–113 (2004).

[6]Valdis E. Krebs, Uncloaking Terrorist Networks, First Monday, volume 7, number 4 (April 2002),available at http://firstmonday.org/issues/issue7_4/krebs/

[7] Heuer, R.J. *Psychology of Intelligence Analysis*. Center for the study of Intelligence, Central Intelligence Agency, 2001.

[8] Degenne, A. & Forse, M. Introducing Social Networks. London: Sage Publications (1999).

[9] Wasserman, S. & Faust, K. Social Network Analysis: Methods and Applications. Cambridge University Press (1994).

[10] Stewart, T. (2001). Six Degrees of Mohamed Atta. http://money.cnn.com/magazines/business2

[11] Sparrow, M. The application of network analysis to criminal intelligence: An assessment of the prospects. Social Networks 13, 251-274 (1991).

[12] Hougham, V. Sociological Skills Used in the Capture of Saddam Hussein (1991). http://www.asanet.org/footnotes/julyaugust05/fn3.html.

[13] Chan, K. and Liebowitz, J. The synergy of social network analysis and knowledge mapping: a case study, Int. J. Management and Decision Making, Vol. 7, No. 1, pp.19–35 (2006).

[14] McIllwain, J. S. Organized crime: A social network approach. Crime, Law & Social Change, Vol. 32. (1999). 301–323.

[15] McAndrew, D. The structural analysis of criminal networks. In: Canter, D., Alison, L. (eds.): The Social Psychology of Crime: Groups, Teams, and Networks, Offender Profiling Series, III, Aldershot, Dartmouth (1999) 53–94.

[16] Baker, W. E. and Faulkner, R. R. The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. American Sociological Review, 58, No. 12, 837-860 (1993).

[17] Newman, M. E. J. The structure and function of complex networks, SIAM Review 45, 167- 256 (2003).

[18] Borgatti, S.P. *The Key Player Problem*. Proceedings from National Academy of Sciences Workshop on Terrorism, Washington DC (2002).

[19] Jennifer J. Xu and Hsinchun Chen. *CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery*, ACM Transactions on Information Systems, Vol. 23, No. 2, April 2005, Pages 201–226

[20] L. C. Freeman, *Centrality in social networks: Conceptual clarification*, Social Networks 1 (1979) 215-239

[21] Center for Computational Analysis of Social and Organizational Systems (CASOS), Accessed on January 02, 2006. http://casos.isri.cmu.edu/index.html

[22] Netdraw software. http://www.analytictech.com/download_products.htm

[23] UciNet Software. http://www.analytictech.com/download_products.htm

[24] Robb John, *Destabilizing Terrorist Networks*, 2004. Accessed on January 02, 2006. http://globalguerrillas.typepad.com/globalguerrillas/2004/03/destabilizing_t.html

[25] Memon, N.; Larsen, H.L. *Practical Approaches for Analysis, Visualization and Destabilizing Terrorist Networks.* First International Conference on Availability, Reliability and Security (ARES'06), 20-22 April 2006, IEEE Conference Proceedings Page(s): 906 – 913 (2006).

[26][5] Carley M. Kathleen, Lee Ju-Sung, David Krackhardt. *Destabilizing Networks*. 2002, *Connections 24 (3)* 79-92.

[27] Farely David J. *Breaking Al Qaeda Cells: A Mathematical analysis of counterterrorism*. Operations.Studies in conflict terrorism. 26:399–411 (2003).

[28] Scott White, Padhraic Smyth. *Algorithms for Estimating Relative Importance in Networks*, SIGKDD '03 August 24-27, 2003 Washington D.C., USA Copyright 2003 ACM ACM 1-58113-737-0/03/0008