A Vision of a Toolbox for Intelligence Production

Joel Brynielsson, Andreas Horndahl, Lisa Kaati, Christian Mårtenson, Pontus Svenson
Swedish Defence Research Agency
SE-164 90 Stockholm, Sweden
Email: firstname.lastname@foi.se

***Abstract*—In this paper, we describe preliminary work on a toolbox aiming to help analysts involved in the intelligence production process. Intelligence analysts are overwhelmed by information, both in the form of sensory data, text stemming from human observations and other sources. In order to make sense of this information and to produce the intelligence reports needed by decision-makers, assisting computer tools are needed. We briefly describe parts of the intelligence process and touch upon the subject of what parts can and cannot be automated. A tool for tagging information semantically that we are currently working on is described, and ideas for two other tools are briefly outlined.**

I. Introduction

Intelligence analysts of today are overwhelmed by information that they must take into account when producing their analyses and assessments. They must also produce far more content today than previously while at the same time taking into account material from a wide variety of sources, ranging from IMINT (image intelligence) to text articles in newspapers. Both the pace of the information push to the analysts and the information pull by the commanders and decision-makers who need output from the analysts has also increased: the time-scales involved at operative and tactic levels are shorter compared to the days of the cold war. The types of conflicts that we are involved in have also changed: in peace-keeping and peace-enforcing missions, we are faced with a multitude of actors, who are of different types and it is not always clear whether they should be regarded as friends, foes or neutrals in a given situation. When analyzing, for instance, a report on a confrontation between two clan leaders, it is important not only to know the interests, capabilities and motivations of these two, but also those of the other actors in the area who are connected to the involved parties. The goals of network-based defense and network-based intelligence will further exacerbate these problems: more sensors mean more data for the analysts to consider, and also more decisions to be made regarding where to put the sensors. The need for continuous action can make the traditional intelligence cycle obsolete. Instead, it is necessary to make all stakeholders, including customers, involved in the intelligence process, enabling a creative collaboration process where several analysts may contribute in parallel to a continuously refined shared picture of the target, see, e.g., [1] who proposes a "target-centric" intelligence cycle.

In order to meet these requirements, intelligence analysts need better information handling tools and concepts. Fusion tools can be an important help, enabling automatic clustering of similar reports and using aggregation methods to produce meaningful labels on the information displayed to them. Techniques from natural language processing and text mining can be used to fuse information in different languages and to produce summaries of vast amounts of textual data.

Complete automation of the intelligence production process, however, is neither possible nor desirable. Hence, the task for fusion researchers is both to build automatic tools that process information and to build tools that help humans to do further processing: ultimately, fusion is a process involving humans. Situation awareness is created in human minds, not in machines. In this paper, we present some ideas on how computer tools for intelligence analysis can be created, i.e., tools that can be used at the discretion of a human analyst to produce intelligence products in alternative ways. What is needed, rather than constructing one single tool, is a "toolbox" containing several different tools, each helping the analyst with one specific task.

The work described in this paper emphasizes intelligence analysis, i.e., the work of intelligence analysts which can best be described as the art of creating useful intelligence products. At the same time, however, the work described herein is strongly related to work within other areas currently performed at the Swedish Defence Research Agency (FOI) in various ways. All the ideas described in this paper will eventually be included in the Impactorium toolset which is described further in [2]. A future version of the semantic tagging tool described in Section III will be used to help users input information into the Reportorium tool of the Impactorium suite as well as to input information into the Semantic MilWiki [3]. The threat model construction tool described in Section IV-A will be used to construct the threat models used in Impactorium, and one of the uses of the need-based situation picture tool described in Section IV-B is to determine which results from the Impactorium and what information from the Semantic MilWiki should be displayed to what users at what time.

This paper describes ongoing research and does not present finished research or even finished thoughts about further research. Section II contains a description of different parts of intelligence analysis work where it could be possible to use computer-aided tools. Section III contains some details about a tool for semantically tagging information that we are currently working on, while Section IV gives a brief overview of envisioned tools for constructing threat models and need-based situation pictures. Finally, Section V concludes and discusses future work.

II. Intelligence Analysis

A requirement for high-quality intelligence analysis is that the data to analyze is of sufficient quality. When faced with ever-increasing amounts of information to analyze and shorter and shorter time to do the analysis, it is vital that the quality of the produced intelligence does not decrease. Commanders and decision-makers must still be able to rely on the results of the process. It is therefore important that intelligence products come with a marking that states its level of quality and confidence in the presented results, as discussed in, e.g., [4]. In addition to such metadata, intelligence data also needs to be semantically tagged to enable quicker searching of information. In order to produce a fusion result, the fuser (whether it is a machine or a human) must first find all relevant information. For sensor data that is about a given object, this is the association or clustering problem. For text data, text clustering could be used. When combining sensory and textual data, it is possible to use semantic queries, but this requires that all data is semantically tagged—a process that needs to be automated as much as possible. In Section III of this paper, we describe a prototype for how semantic information could be automatically extracted from text documents.

Another important part of intelligence analysts' work is to construct the models used in automatic fusion tools. Section IV-A describes an idea for how a computer assistant tool that helps in this process could be constructed using case-based reasoning. As mentioned above, one possible application of this is in the model-construction part of the Impactorium tool. Section IV-B, in turn, briefly describes an idea for how a tool helping the user to determine what information should be displayed to them could be constructed.

The ideas for tools described in this paper are envisioned to be used together, and as briefly outlined in Section I they will also be combined with other tools that are currently developed at FOI. It is, however, also important that each tool in the intelligence analysis toolbox is useable (and use worthy) by itself. It will be impossible to postulate beforehand what combinations of tools that are useful for an analyst working on a specific case. Instead, the analysts must be able to choose for themselves what tools to use. This thinking is in line with the emphasis on a service-oriented architecture which is affecting the current development of command and control and intelligence analysis systems in Sweden.

The intelligence analysis process can generally be divided into three phases: search, analyze and present/disseminate. The phases are generally connected to form a loop in the so-called intelligence cycle; it is important to realize that intelligence analysis is not a linear process, but involves jumping back-and-forth between the different phases. During the search phase, information is gathered from different kinds of sources, such as databases, sensors, and newspapers. In order to reduce the time spent in the search phase, a new phase dealing with information input and structuring of information can be introduced.

If information is structured using semantic techniques, it might be possible to turn the solid curve in Figure 1 into the dotted one. Decreasing the amount of time spent on searching for information enables the analyst to spend more time on the later phases which, typically, involve more creative thinking and analysis. Such benefits could prove to be especially useful in time critical situations, when it is of utmost importance to gain situation awareness in a short period of time. It is in this structuring phase of the intelligence process that the semantic tagging tool will fit. We believe that some parts of this process can be automated using a tool such as the one described in Section III. It is, however, likely that it will still be necessary to have humans participating in this process: language technology is not yet mature enough to be able to produce semantic tags with enough certainty. Since the consequences of making wrong decisions based on the processing of the semantic tags could be severe, it is doubtful whether complete automation of this process would even be desirable.

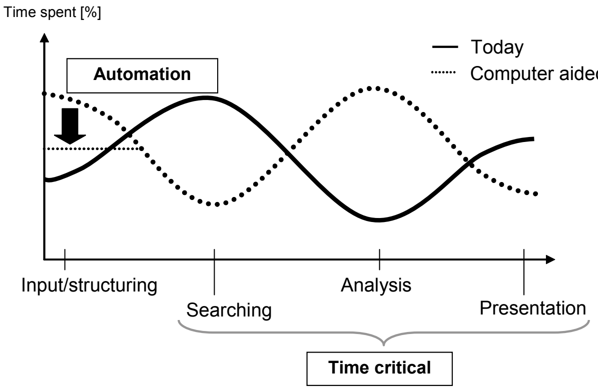

Fig. 1. Theoretical intelligence workflow. The solid line shows the amount of time spent in the different phases in a typical case of today, whereas the dotted line shows how much time may be spent in each phase if semantic techniques could be used.

Another benefit is that if information is structured, it is easier to see how new information would affect a given situation. If structured information is added, the new information might cause a chain of reactions that may lead to new statements that will drastically change the situation picture. For example, a list of potential threats that is dynamically updated when new input is added may be very useful in time critical situations.

To put our work into perspective, it should be contrasted and compared to the research and development efforts currently undertaken in support of the Swedish military intelligence function [5]. This work is done in close cooperation with intelligence personnel who have been continuously participating using action research methods. That is, analysts and researchers have been working collaboratively in order to improve intelligence work procedures through continuous reflection on, and adjustment of, the actions taking place in the actual intelligence unit. In even more recent work, the same authors elaborate on the multifaceted and somewhat vague concept of an "intelligence architecture" [6]. In this work, the authors' long-term goal is to integrate the human and the technological aspects of intelligence work. Although resisting formal analysis and definition, the ISTAR (intelligence, surveillance, target acquisition, reconnaissance) concept provides precisely this, i.e., a possibility to overcome the problem of uniting widely differing pieces of information by careful integration of human work and technical compositions. A conclusion made is that there is little need for automated quantitative processes in the intelligence domain, largely because knowledge production is politically and command informed rather than being the result of a formalized objective process. We support this viewpoint and interpret it as a recommendation: intelligence technology ought to be looked upon and constructed in the form of a toolbox, i.e., a variety of supporting tools that the analyst can use to easily view and enlighten, possibly large, pieces of information. In the following sections, we will describe some preliminary thoughts on three tools that could be part of such an "analyst toolbox." Other papers at this conference [2], [3], [7] and presented elsewhere [8], [9], [10], [11] describe other tools that could also be part of this toolbox.

III. Semantically Tagging Information by Extracting Knowledge from Data

The semantic web is the term used for the vision of making Internet content interpretable by machines. The semantic web helps computers gain a better understanding of what the information really means. Subsequently, when the information is understandable by computers, the computers would also be able to infer new facts. Hence, the semantic web concept, where information can be interpreted by machines, may not only improve the Internet, it may also be useful in other information sharing domains such as information sharing within the armed forces.

In order to make full use of the semantic techniques, it is necessary for the content of knowledge-bases to be semantically tagged. Although there is still much work to be done, research in this area is active and making good progress [12], and performing semantic tagging will inevitably be a major part of intelligence analysts' work in the future. Hence, automating some parts of this process would be very useful. One approach to doing so is to use text analysis techniques to perform entity extraction and present a list of entities found in the document to the user who is doing the semantic annotation. A more challenging problem is to also find relations in the document, and to add semantically marked links to the intelligence document. Figures 2 and 3 show a simple prototype system.

In addition to the annotation of each document, it is also important for the intelligence analysis assistant program to be able to analyze several documents at the same time. For this, summarization techniques from natural language processing could be used. It should also be possible to obtain lists of the extracted entities that are present in the different documents and fuse these to provide a "situation picture" of what objects are referred to in the document collection. In addition to standard techniques for this, we are also investigating the use

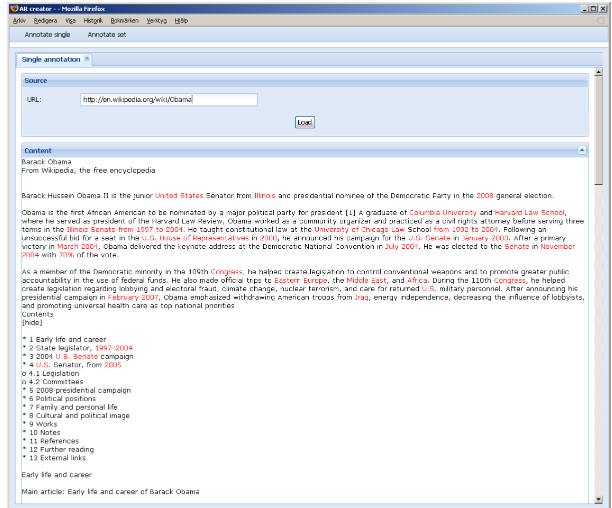

Fig. 2. An intelligence document where entity extraction has been made and entities which are known in the used ontology have been marked in red.

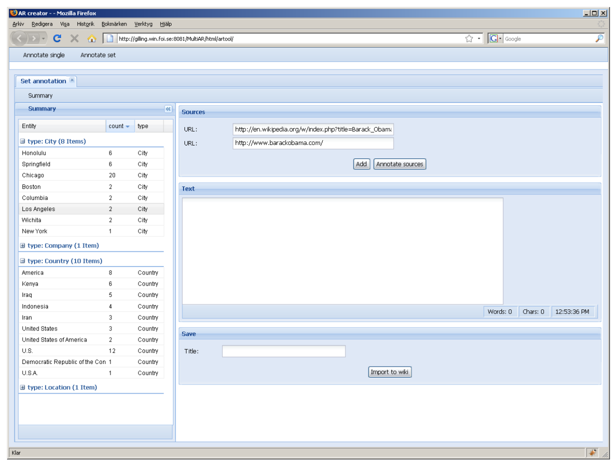

Fig. 3. The figure shows how an interface to be used by an intelligence analyst looking at a number of intelligence documents could look like. The panel on the left shows aggregate statistics about entities found in documents—information that could be of use when writing an intelligence report.

of topic models to fuse text documents with sensor data and semantically annotating the result.

As outlined in Section II, we do not believe it will be possible (or desirable) to automate the semantic tagging process. However, computer tools that act as assistants to the human, suggesting tags for documents, would provide a valuable enhancement of the intelligence input/structuring process.

IV. Tools Based on Case-based Reasoning

A. Threat Model Construction Assistant

As outlined in Section II, fusion tools need models in order to work. Constructing these models is an important, and difficult, part of intelligence analysts' work. We think that it is possible to use case-based reasoning to help the humans in this process. Case-based reasoning relies on a case database consisting of previously seen solutions to problems. For a recent survey of case-based reasoning and, in particular, its relationship to analogical human reasoning, see [13].

In our application, a case would correspond to a "situation," as described, for instance, by the set of active indicators. A solution would correspond to a particular model to use in the fusion tools, for instance, the Bayesian networks to be included in the Impactorium tool. The current situation is compared to those in the case database, and the best matches are extracted and presented to the user who can choose which models to include. In order to make the match, situations must be compared with each other and a similarity measure calculated. For attributes that adopt numerical values, it is easy to calculate the similarity distance. For other kinds of attributes, other models need to be used.

To determine the extent this process can be automated to, it is necessary to first make an adequate description of a situation. For this, it will be necessary to define an ontology or information model that is rich enough to be able to distinguish between different situations. Construction of a relevant similarity measure between situations is also a challenge.

B. Situation Picture Construction Assistant

Network-based defence decision-makers and analysts are given the opportunity to make use of a wide range of information services. For a specific mission in a specific situation, however, it is only a small part of these services that are useful. A choice needs to be made regarding what services to use. Since each service will provide a "building-block" of a situation picture, it is possible to create different views on the situation by making use of different services. In order to support an officer in the selection of information, one could make use of techniques from case-based reasoning. Good solutions for what set of services to use are stored in a case database along with a description of the overall situation when the set of services were selected. Factors such as the role of the decision-maker and the status of own operations should also be included in the case description, as well as external factors such as mood and cultural status in the region.

User-tailored situation-views could also be constructed using other methods. An example might look something like this. An intelligence analyst is searching for information to respond to an RFI (request for information). In a given moment, the analyst is looking at the information provided by the knowledge base about the warlord X. The system should then show related information that it believes the user could be interested in, at the side of the computer screen. For instance, if the context within which the analyst is working relates to smuggling, information about boats owned by X or an associate, and that have been seen near a border could be shown. A more sophisticated example would be to show information about a boat that moved anomalously and which is linked to a subordinate of X. The idea is similar to the recommendation system used by many online stores, see, e.g., [14], [15], but would require more advanced methods.

Another example is if the analyst is interested in criminal activities in the theatre of operations. The side-screen will then display an aggregated statistical view of the entire area showing crime statistics. The user can choose to zoom in on parts of the data, for example, look at a map where crimes committed by a member of a certain ethnical group and that have a member of another ethnical group as victims have been committed.

V. Summary and Continued Work

This paper has described some research directions being pursued at FOI related to constructing a toolbox aimed at helping intelligence analysts to fuse information more efficiently. Work will continue on developing concept prototypes for the ideas presented, and then testing them in a relevant setting using experienced intelligence analysts.

References

[1] R. M. Clark, *Intelligence Analysis: A Target-Centric Approach*. Washington, D.C.: CQ Press, 2004.

[2] R. Forsgren, L. Kaati, C. Mårtenson, P. Svenson, and E. Tjörnhammar, "An overview of the impactorium tools 2008," in *Skövde Workshop on Information Fusion Topics (SWIFT 2008)*, Skövde, Sweden, Nov. 2008.

[3] C. Mårtenson and A. Horndahl, "Using semantic technology in intelligence analysis," in *Skövde Workshop on Information Fusion Topics (SWIFT 2008)*, Skövde, Sweden, Nov. 2008.

[4] S. Arnborg, H. Artman, J. Brynielsson, and K. Wallenius, "Information awareness in command and control: Precision, quality, utility," in *Proceedings of the Third International Conference on Information Fusion (FUSION 2000)*, Paris, France, July 2000, pp. ThB1/25–32.

[5] P.-A. Persson and J. M. Nyce, "Intuitive tools? design lessons from the military intelligence community," *American Intelligence Journal*, pp. 38–50, Summer 2007.

[6] ——, "Integrating human effort and technology in the ISTAR model: An ethnographic perspective," *Submitted for publication*, 2008.

[7] P. Svenson, "Social network analysis of uncertain networks," in *Skövde Workshop on Information Fusion Topics (SWIFT 2008)*, Skövde, Sweden, Nov. 2008.

[8] J. Brynielsson and S. Arnborg, "An information fusion game component," *Journal of Advances in Information Fusion*, vol. 1, no. 2, pp. 108–121, Dec. 2006.

[9] L. Ferrara, C. Mårtenson, P. Svenson, P. Svensson, J. Hidalgo, A. Molano, and A. L. Madsen, "Integrating data sources and network analysis tools to support the fight against organized crime," in *Proceedings of the IEEE ISI 2008 International Workshops: PAISI, PACCF, and SOCO 2008*, Taipei, Taiwan, June 2008, pp. 171–182.

[10] R. Suzić and P. Svenson, "Capabilities-based plan recognition," in *Proceedings of the Ninth International Conference on Information Fusion (FUSION 2006)*, Florence, Italy, July 2006.

[11] P. Svenson and C. Mårtenson, "SB-Plan: Simulation-based support for resource allocation and mission planning," in *Proceedings of the Conference on Civil and Military Readiness 2006 (CIMI 2006)*, Enköping, Sweden, May 2006.

[12] V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, and F. Ciravegna, "Semantic annotation for knowledge management: Requirements and a survey of the state of the art," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4, no. 1, pp. 14–28, Jan. 2006.

[13] R. L. D. Mantaras, D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T. Cox, K. Forbus, M. Keane, A. Aamodt, and I. Watson, "Retrieval, reuse, revision and retention in case-based reasoning," *The Knowledge Engineering Review*, vol. 20, no. 3, pp. 215–240, Sept. 2005.

[14] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, June 2005.

[15] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, Jan./Feb. 2003.