Intelligence Analysis Ontology for Cognitive Assistants

Mihai Boicu, Gheorghe Tecuci and David Schum

***Abstract*—This paper presents results on developing a general intelligence analysis ontology which is part of the knowledge base of Disciple-LTA, a unique and complex cognitive assistant for evidence-based hypothesis analysis that helps an intelligence analyst cope with many of the complexities of intelligence analysis. It introduces the cognitive assistant and overviews the various roles and the main components of the ontology: an ontology of "substance-blind" classes of items of evidence, an ontology of believability analysis credentials, and an ontology of actions involved in the chains of custody of the items of evidence.**

***Index Terms*—cognitive assistant, ontology, evidence-based hypothesis analysis, types of items of evidence, chains of custody**

I. The Complexity of Intelligence Analysis

Intelligence analysts face the difficult task of analyzing masses of information of different forms and from a variety of sources. Arguments, often stunningly complex, are necessary in order to link evidence to the hypotheses being considered. These arguments have to establish the three major credentials of evidence: its *relevance*, *credibility*, and *inferential force or weight*. *Relevance* considerations answer the question: *So what?* How does this item of information bear on any hypothesis being considered? *Credibility* considerations answer the question: *Can we believe what this item of information is telling us? Inferential force or weight* considerations answer the question: *How strongly does this item of evidence favor or disfavor alternative hypotheses we are considering?* Establishing these three evidence credentials always involves mixtures of *imaginative and critical reasoning*. Indeed, as work on an analytic problem proceeds, we commonly have evidence in search of hypotheses at the same time with hypotheses in search of evidence. First, various hypotheses and lines of inquiry must be generated by analysts who imagine possible explanations for the continuous occurrence of events in our non-stationary world. Second, considerable imagination is required in decisions about what items of information should be considered in the analytic problem at hand. But critical reasoning in intelligence analysis is equally important. No item of evidence comes with its relevance, credibility, and inferential force or weight credentials already established. These credentials must be established by defensible and persuasive arguments which have to take into account that our evidence is always *incomplete*, usually *inconclusive*, frequently *ambiguous*, commonly *dissonant*, and it comes to us from sources having any gradation of *credibility* shy of perfection [1].

But the inherent complexity of the analysts' tasks are only part of their problems. In many cases, analysts are not given unlimited time to generate hypotheses and evidence and to construct elaborated and careful arguments on all elements of the analysis at hand. One way of describing this problem is to say that analysts will neither have the time, or the necessary evidential basis, for *drilling down* or decomposing all elements of the problem being considered. In many instances, analysts are faced with the necessity of having to make various assumptions in which certain events are believed "as if" they actually occurred. And always, the world is evolving and the yesterday's analysis needs to be updated with new items of evidence discovered today.

II. Disciple-LTA: Analyst's Cognitive Assistant

Disciple-LTA is a unique and complex analytic tool that can help an intelligence analyst cope with many of the complexities of intelligence analysis [2], [3]. The name *Disciple*, by itself, suggests that it learns about intelligence analysis through its interaction with experienced intelligence analysts. The word "disciple" has synonyms including: learner, advocate, supporter, and proponent. The addition "*LTA*", refers to the fact that Disciple learns analysis [L], it can serve as a tutor [T] for novice and experienced analysts, and it can assist [A] in the performance of analytic tasks, e.g. in current or in finished intelligence analyses. Disciple-LTA has two very

Manuscript received October 31, 2008. This work was supported in part by several U.S. Government organizations, including the Air Force Office of Scientific Research (FA9550-07-1-0268), the Air Force Research Laboratory (FA8750-04-1-0257), and the National Science Foundation (0750461). The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official polices or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the Air Force Research Laboratory, the National Science Foundation or the U.S. Government.

Dr. Mihai Boicu is Assistant Professor of Applied Information Technology and Associate Director of the Learning Agents Center in the Volgenau School of Information Technology and Engineering, George Mason University, 4400 University Dr., Fairfax VA 22030, USA (phone: 703-993-1591; fax: 703-993-9275; e-mail: mboicu@gmu.edu).

Dr. Gheorghe Tecuci is Professor of Computer Science in the Volgenau School of Information Technology and Engineering and Director of the Learning Agents Center at George Mason University, Fairfax VA 22030, USA, and Visiting Professor and former Chair of Artificial Intelligence at the US Army War College (e-mail: tecuci@gmu.edu).

Dr. David Schum holds the rank of Professor in the Systems Engineering and Operations Research Department in the Volgenau School of Information Technology and Engineering, and in the School of Law, at George Mason University, Fairfax VA 22030, USA. He is also Honorary Professor of Evidence Science, University College London, UK (e-mail: dschum@gmu.edu).

distinct differences from other knowledge-based or rule-based "expert systems" developed in the field of artificial intelligence over the years. Such systems are developed by knowledge engineers who attempt to capture and represent the heuristics or rules of the experienced expert users so that they could be preserved and utilized in new situations. This is a very long and difficult process that results in systems that are even more difficult to maintain. But Disciple-LTA is qualitatively different from these earlier expert systems.

Instead of being programmed by a knowledge engineer, Disciple-LTA learns its expertise directly from expert analysts who can teach it in a way that is similar to how they would teach a person. However, when it is first used by an expert analyst, Disciple-LTA does not engage in this interaction with a blank mental tablet. Disciple-LTA already has a stock of established knowledge about evidence, its properties, uses, and discovery. Some of this knowledge may not be already resident in the minds of its expert users, who apply their experience with certain analytic contexts that *Disciple* will learn. So, Disciple does learn about specific intelligence problems from its users, but it can combine this knowledge with what it already knows about various elements of evidential reasoning. Conventional expert systems can be no better than the expertise of the persons whose heuristics are trapped; this represents a "ceiling" on the suitability of these earlier systems. But this ceiling is actually the "floor" for Disciple-LTA, since this system incorporates basic knowledge of the evidential reasoning tasks analysts face in addition to the substantive expertise of the analysts who interact with it.

One basic feature of Disciple-LTA is that it provides the analyst the opportunity to decompose a complex problem into finer levels; i.e. it rests upon a "divide and conquer" strategy for dealing with the analytic complexity of hypothesis in search of evidence. In particular, it allows "top-down" decompositions to deduce from a stated hypothesis what needs to be proven in order to sustain this hypothesis. This decomposition eventually results in the identification of possible sources of evidence relevant to this hypothesis. Consider, for example, the problem of assessing whether Al Qaeda has nuclear weapons. This problem can be reduced to three simpler problems of assessing whether Al Qaeda has reasons, has desires, and has ability to obtain nuclear weapons. Each of these simpler problems is further reduced to even simpler ones (e.g. by considering specific reasons, such as deterrence, self-defense, or spectacular operation) that could be solved either based on the available knowledge or by analyzing relevant items of evidence. An abstraction of these decompositions is presented in the left-hand side of Fig. 1. Let us consider "Spectacular operation as reason" which is a short name for "Assess whether Al Qaeda considers the use of nuclear weapons in spectacular operations as a reason to obtain nuclear weapons." As indicated in the left-hand side of Fig. 1, to solve this hypothesis analysis problem Disciple-LTA considered both favoring evidence and disfavoring evidence. Disciple-LTA has found two items of favoring evidence, EVD-FP-Glazov01-01c and EVD-WP-Allison01-01, and it has analyzed to what extend each of them favors the hypothesis that Al Qaeda considers the use of nuclear weapons in spectacular operations as a reason to obtain nuclear weapons. EVD-FP-Glazov01-01c is shown in the bottom right of Fig. 1.

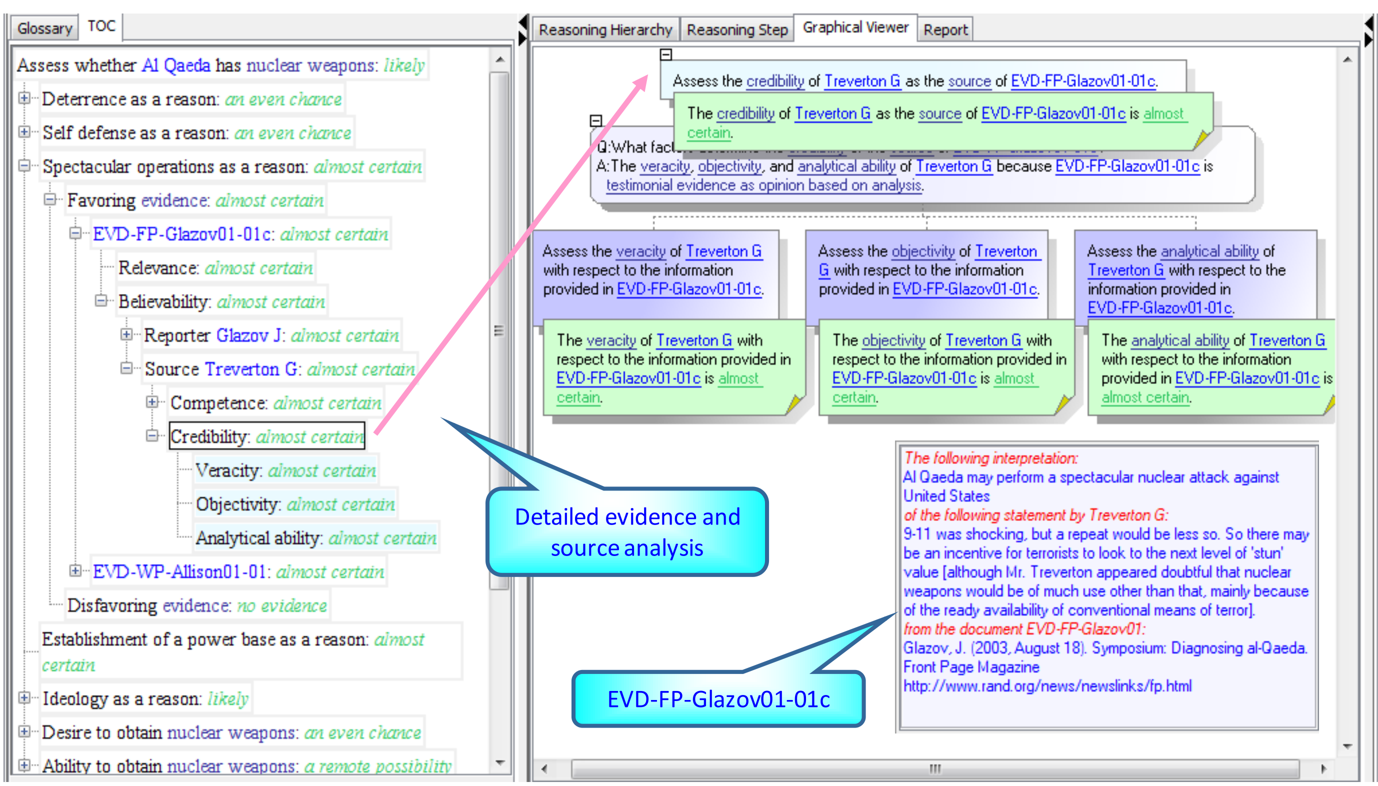


Fig. 1. Hypothesis analysis through problem reduction and solution synthesis.

It is a fragment from a magazine article published in the Front Page Magazine by Glazov J. where he cites Treverton G. who stated that Al Qaeda may perform a spectacular nuclear attack against United States [4]. To analyze EVD-FP-Glazov01-01c, Disciple-LTA considered both its relevance and its believability [1], [5]. The believability of EVD-FP-Glazov01-01c depends both on the believability of Glazov J. (the reporter of this piece of information) and the believability of Treverton G. (the source). The believability of the source depends on his competence and his credibility. The credibility of Treverton G. depends on his veracity, objectivity, and analytical ability. When the analyst clicks on a problem, such as “Credibility” from the left-hand side of Fig. 1, Disciple-LTA displays the details on how it solved that problem, as shown in the right-hand side of Fig. 1. For example, to “Assess the credibility of Treverton G as the source of EVD-FP-Glazov01-01c” Disciple-LTA has assessed his veracity, objectivity, and analytical ability. Then the results of these assessments (almost certain, almost certain and almost certain) have been combined into an assessment of the credibility (almost certain). Disciple-LTA may use different synthesis functions for the solutions (such as, minimum, maximum, average, etc.), depending on the types of the problems. A abstraction of the synthesis process is displayed in the left hand side of Fig. 1, where the solutions appear in green, attached to the corresponding problems. Notice that this problem-reduction/solution-synthesis approach enables a natural integration of logic and probability.

In some situations the analysts will not have the time to deal with all of the complexities their own experience and Disciple-LTA makes evident. In other situations, analysts will not have access to the kinds of information necessary to answer all questions regarding elements of an analysis that seem necessary. In such situations Disciple-LTA allows the user to decompose (“to drill down”) an analysis to different levels of refinement in order to reach conclusions about necessary analytic ingredients, by providing mechanisms necessary to identify assumptions that are being made and by showing the extent to which conclusions rest upon these assumptions [3].

For evidence in search of hypotheses, Disciple-LTA allows the construction of "bottom-up" structures in which possible alternative hypotheses are generated. No computer system, even Disciple-LTA, is capable of the imaginative thought required to generate hypotheses and new line of inquiry. But Disciple-LTA can assist in this process by prompting the analyst to consider the inferential consequences of chains of thought that occur in the process of generating hypotheses and new lines of inquiry and evidence.

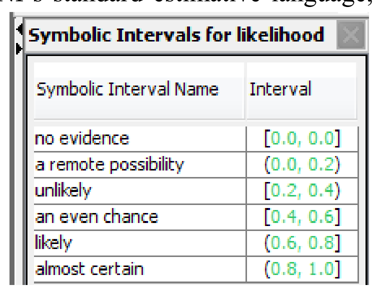
The following sections will discuss the general features of the intelligence analysis ontology of Disciple-LTA.

III. Knowledge Base Structure for Sharing and Reuse

In addition to the separation of knowledge and control (which is a characteristic of all the knowledge-based systems), Disciple-LTA is characterized by an additional architectural separation at the level of the knowledge base. Its knowledge base is structured into an object ontology that defines the concepts of the application domain, and a set of problem solving rules expressed in terms of these concepts. While an ontology is characteristic to an entire domain (such as intelligence analysis), the rules are much more specific, corresponding to a certain type of applications in that domain, and even to specific subject matter experts. This separation allows one to easily share and reuse the ontology developed for a given intelligence analysis application, when developing a new one. Additionally, the ontology in Disciple-LTA is organized as a distributed hierarchy of several ontologies, which further facilitate its sharing and reuse, as well as its development and maintenance.

IV. Multiple Roles for Ontology

The object ontology plays a crucial role in Disciple-LTA and in cognitive assistants, in general, being at the basis of knowledge representation, user-agent communication, problem solving, knowledge acquisition and learning [6]. First, the object ontology provides the basic representational constituents for all the elements of the knowledge base, including the problems, the problem reduction rules, and the solution synthesis rules. The ontology language of Disciple-LTA is an extension of OWL-light [7] that allows the representation of partially learned concepts and features. A partially learned feature may have both it domain and its range represented as plausible version space concepts [6]. One may also define different symbolic probability scales, such as Kent, DNI, IPCC or legal [8], and automatically convert from one to another and into the Bayesian probabilities. For example, the left hand side of Fig. 2 shows the symbolic probabilities for likelihood, based on the DNI’s standard estimative language, while the right hand side shows the corresponding Bayesian probability intervals. The ontology also allows the representation of items of evidence that may contain different or even contradictory views on some entities.


Symbolic Intervals for likelihood

Symbolic Interval Name	Interval
no evidence	[0.0, 0.0]
a remote possibility	(0.0, 0.2)
unlikely	[0.2, 0.4)
an even chance	[0.4, 0.6]
likely	(0.6, 0.8]
almost certain	(0.8, 1.0]

Fig. 2. Symbolic probabilities for likelihood.

Second, the agent’s ontology enables the agent to communicate with the user and with other agents by declaring the terms that the agent understands. As illustrated in the upper-right part of Fig. 1, the agent uses natural language phrases where the terms from the ontology appear in blue. Consequently, the ontology enables knowledge sharing and reuse among agents that share a common vocabulary which they understand. Third, the problem solving rules of the agent are applied by matching them against the current state of the agent’s world which is represented in the ontology. The use of partially learned knowledge (with plausible version spaces) in reasoning, allows solving of problems with different degrees of confidence [2]. Fourth, the object ontology represents the generalization hierarchy for learning, general rules being learned from specific problem solving examples by traversing this hierarchy [2], [3], [6].

V. Ontology of "Substance-blind" Classes of Items of Evidence

Being able to categorize evidence is vitally necessary for many reasons, one of the most important being that we must ask different questions of and about our evidence in the process of intelligence analysis in which we encounter different recurrent forms and combinations of evidence. If we were not able to categorize evidence in useful ways we might not be aware of many different questions we should be asking of our evidence. However, asked to say how many kinds of evidence there are, we could easily say that there is near infinite amount, if we considered its substance or content. This presents a significant problem: how can we ever say anything general about evidence if every item of it is different from every other item? Fortunately there is a "substance-blind" way of categorizing evidence that does not rely at all on its substance or content, but on its inferential properties: its relevance and believability.

Disciple-LTA includes an ontology of "substance-blind" classes of items of evidence. Some of the classes based on their believability attributes are shown in Fig. 3 [1].

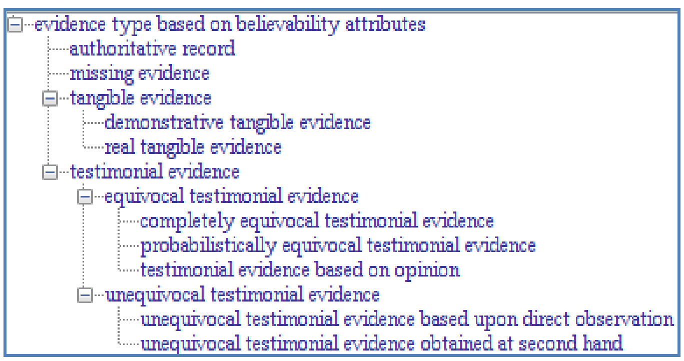


Fig. 3. "Substance-blind" classes of items of evidence.

If you can pick up the evidence yourself and examine it to see what events it might reveal, we say the evidence is *tangible* in nature such as objects, documents, images, and tables of measurements. We distinguish between *real tangible evidence* which is an actual thing itself (such as a captured weapon component), and *demonstrative tangible evidence*, which is a representation or illustration of this thing (such as a diagram of that component). Now suppose you have nothing you can examine for yourself and must rely on someone else who has made some observation and who will tell you about the occurrence or nonoccurrence of some event. This is called *testimonial evidence*, as in a HUMINT report from an asset. This person may state unequivocally that some event has occurred or has not occurred. Of great concern is how the person providing testimonial evidence obtained the information reported. Did this person make a *direct observaton* or did he/she learn about the occurrence or nonoccurrence of the reported event from another person, in which case we have *secondhand or hearsay evidence*. Moreover, there are classes of evidence mixtures, such as *testimonial evidence about tangible evidence*. It would not be uncommon in intelligence analysis to encounter evidence obtained through a chain of sources (see section VII).

VI. Ontology of Believability Analysis Credentials

As discussed above, the "substance-blind" ontology of classes of evidence is based on their *believability* and *relevance* credentials. That is, there are specific credentials for each such class. For example, the believability of a source of *direct testimonial evidence* depends on the source's *competence* and *credibility* [1], [5]. Assessments of the competence of a source require answers to two important questions. First, did this source have *access* to, or did actually observe, the events being reported? If it is believed that a source did not have access to, or did not actually observe the events being reported, we have very strong grounds for suspecting that this source fabricated this report or was instructed what to tell us. Second, we must have assurance that the source *understood* the events being observed well enough to provide us with an intelligible account of these events. So, access and understanding are the two major attributes of a human source's competence. Assessments of human source credibility require consideration of entirely different attributes: *veracity* (or *truthfulness*), *objectivity*, and *observational sensitivity under the conditions of observation*. Here is an account of why these are the major attributes of testimonial credibility. First, is this source telling us about an event he/she believes to have occurred? This source would be untruthful if he/she did not believe the reported event actually occurred. So, this question involves the source's *veracity*. The second question involves the source's *objectivity*. The question is: did this source base a belief on sensory evidence received during an observation, or did this source believe the reported event occurred either because this source expected or wished it to occur? An objective observer is one who bases a belief on the basis of sensory evidence instead of desires or expectations. Finally, if the source did base a belief on sensory evidence, how good was this evidence? This involves information about the source's relevant *sensory capabilities and the conditions under which a relevant observation was made*.

Answers to these competence and credibility questions require information about our human sources. But one thing is abundantly clear: *the competence and credibility of HUMINT sources are entirely distinct. Competence does not entail credibility, nor does credibility entail competence.* Confusing these two characteristics invites inferential disaster **Error! Reference source not found.**. Disciple-LTA includes an ontology of these credentials and Fig. 1 shows an example of using such credentials in analyzing the believability of an item of evidence.

VII. Ontology of Actions from Chains of Custody

A crucial step in answering questions on the believability of the items of evidence involves having knowledge about the chain of custody through which the testimonial or tangible item has passed en route to the analyst who is charged with assessing it. Basically, establishing a chain of custody involves identifying the persons and devices involved in the acquisition, processing, examination, interpretation, and transfer of evidence between the time the evidence is acquired and the time it is provided to intelligence analysts. Lots of things may have been done to evidence in a chain of custody that may have altered the original item of evidence, or have provided an inaccurate or incomplete account of it. In some cases original evidence may have been tampered with in various ways, the analysts risking of drawing quite erroneous conclusions from the evidence they receive. Suppose we have an analyst who is provided with an item of testimonial evidence by an informant who speaks only in a foreign language. We assume that this informant's original testimony is first *recorded* by one of our intelligence professionals; it is then *translated* into English by a paid translator. This translation is then *edited* by another intelligence professional; and then the edited version of this translation is *transmitted* to an intelligence analyst. So, there are four links in this conjectural chain of custody of this original testimonial item: recording, translation, editing, and transmission. Various things can happen at each one of these links that can prevent the analyst from having an authentic account of what our source originally provided. Fig. 4 shows how the believability of the testimonial evidence provided to the analyst (EVD-Wallflower-5) depends on the believability of the testimony of the informant (i.e. EVD-Wallflower-1), but also on the believability of the *Recording*, *Translation*, *Editing*, and *Transmission* actions. Disciple-LTA has an ontology of actions that may be involved in a wide variety of chains of custody for different types of evidence, such as HUMINT, IMINT, SIGINT or TECHINT. For example, Fig. 5 shows the representation of a translation action. The believability of this translation depends both on the translator's competence (in the two languages, as well as the subject matter being translated) and on his/her credibility.

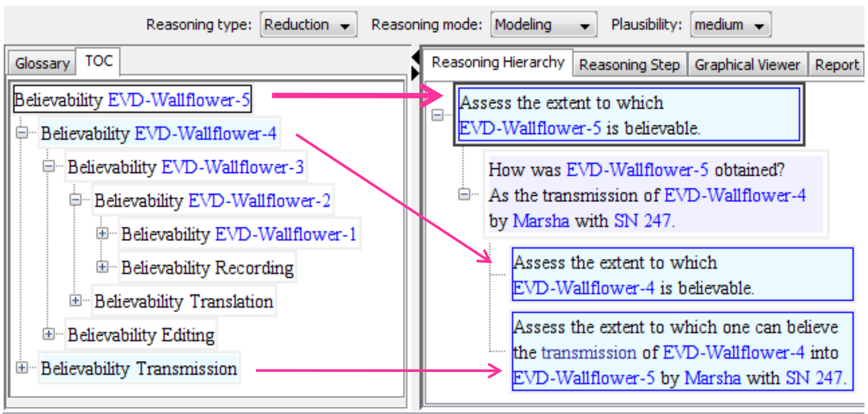


Fig. 4. Evaluating the believability of an item of evidence obtained through a chain of custody.

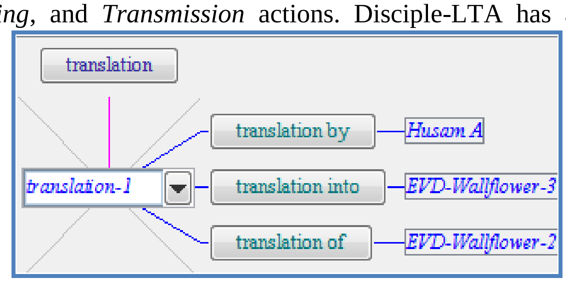


Fig. 5. Action involved in a chain of custody for an item of evidence.

VIII. Lessons and Stories about Intelligence Analysis Concepts

Disciple-LTA can be used to helps new intelligence analysts learn the reasoning processes involved in making intelligence judgments and solving intelligence analysis problems. In particular, its ontology includes lessons and stories about a wide range of intelligence analysis concepts, such as the lesson on veracity illustrated in Fig. 6 [5]. Moreover, its stock of established knowledge about evidence, its properties, uses, and discovery, makes it a suitable educational tool even for expert analysts.

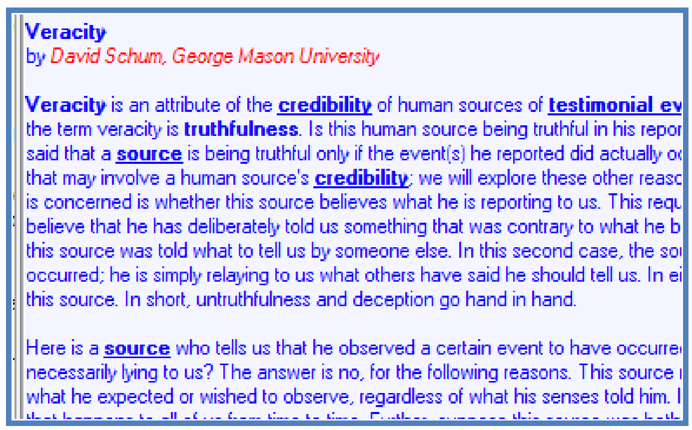


Fig. 6. Fragment from the lesson on veracity.

References

[1] Schum D.A., The Evidential Foundations of Probabilistic Reasoning, Northwestern University Press, 2001.

[2] Tecuci G., Boicu M., Marcu D., Boicu C., Barbulescu M., Ayers C., Cammons D., Cognitive Assistants for Analysts, Journal of Intelligence Community Research and Development, 2007.

[3] Tecuci G., Boicu M., Marcu D., Boicu C., Barbulescu M., Disciple-LTA: Learning, Tutoring and Analytic Assistance, Journal of Intelligence Community Research and Development, 2008.

[4] Glazov J., Symposium: Diagnosing Al Qaeda, *FrontPageMagazine.com*, 18 August 2003.

[5] Schum D.A., Lessons and Stories about Concepts Encountered in Disciple-LTA, Research Report 2, Learning Agents Center, 2007.

[6] Tecuci G., Boicu M., A Guide for Ontology Development with Disciple, Research Report 3, Learning Agents Center, 2008.

[7] W3C, W3C Recommendation, *OWL Web Ontology Language Overview*, http://www.w3.org/TR/owl-features/, February 2004.

[8] Weiss, C., Communicating Uncertainty in Intelligence and Other Professions, *International Journal of Intelligence and CounterIntelligence*, 21:1, pp 57-85, 2008.