Chapter 2
Reexamining Computational Support for Intelligence Analysis: A Functional Design for a Future Capability

James Llinas, Galina Rogova, Kevin Barry, Rachel Hingst, Peter Gerken, and Alicia Ruvinsky

2.1 Motivation

Analysis Tool Suites (ATS's) such as Analyst's Notebook[1] Analyst's Workspace (Andrews and North 2012), Sentinel Visualizer,[2] and Palantir Government,[3] Entity Workspace (Bier et al. 2006), and Jigsaw (Stasko et al. 2013), among others are examples of modern intelligence analysis frameworks. A major point for sensibly all these tool-suites is that they start by focusing on the entity level within the environments of interest. None overtly discuss computational support to inter-entity association and attribute/relation fusion. That is, most if not all are single-source-based as regards entity streams, with the tools doing varying degrees of automated link analysis among bounded entity-pairs toward realization of "data fusion" albeit with rather limited rigor. Further, most also assume that any preprocessing that provides entity extraction yields correct results. This framework of tool products provides the basis for identifying and visualizing relational connections between entities, but these connections are largely if not exclusively performed in the mind of the analyst. In most cases, nothing is done in the way of computational support to dealing with entity or relational uncertainties. The primary function of most of these ATS's is relational link discovery to discern inter-entity relations of bounded extent (in graph science terminology, usually single-hop or limited-hop relations),

[1] http://www03.ibm.com/software/products/en/analysts-notebook.

[2] http://www.fmsasg.com/.

[3] http://www.palantir.com/.

J. Llinas • G. Rogova
Center for Multisource Information Fusion (CMIF), State University of New York at Buffalo, Buffalo, NY, USA
e-mail: llinas@buffalo.edu; rogovagl@gmail.com

K. Barry (✉) • R. Hingst • P. Gerken • A. Ruvinsky
Lockheed Advanced Technology Laboratories (ATL), Cherry Hill, NJ, USA
e-mail: kevin.barry@lmco.com; rachel.hingst@lmco.com; alicia.ruvinsly@lmco.com

W.F. Lawless et al. (eds.), *Autonomy and Artificial Intelligence: A Threat or Savior?*, DOI 10.1007/978-3-319-59719-5_2

achieved with quite limited analytical formality regarding issues of uncertainty, inter-data and/or inter-entity associability, and of relational complexities. Thus, deeper and broader analysis of entity and relational connectedness is left for the human analyst. This is especially true in regard to the assembly of typical final desired analysis products in the form of stories or narratives; said otherwise, there is very limited technical support for synthesis or fusion of hypotheses into the larger context of situational understanding. By and large, these tools try to support the Sensemaking (SM) or schema-development loop of SM (Pirolli and Card 2005; Klein et al. 2006), but either have no algorithmic or technological-process support or provide quite-limited automated support to these higher goals; these assessments are summarized in the review paper of Llinas (2014a, b).

Thus we perceive a need first for a processing/reasoning paradigm that can provide the framework for a more holistic, systemic based approach to intelligence analysis. As sensibly all critiques about intelligence analysis as well as the analysis requirements stated in field manuals describe that the main product that an analyst is driving toward is a narrative type description of some world condition/situation, we set this goal for our research presented in this chapter as well. So, primarily we are seeking to study ways that discrete, single-theme hypotheses can be synthesized or fused into a more holistic and semantic construct in the form of a story or narrative. Our approach incorporates methods of associating and fusing so-called hard (sensor) and soft (textual, semantic) information, as many intelligence analysis environments have such disparate data streams as input. (We note that virtually all the work in the areas we studied here only involve soft or textual type inputs.) We believe that the functional design produced here provides a basis for a next step involving research prototype development, and because of this we have also studied ways to test and evaluate such a prototype.

2.2 Goals and Requirements

In this research program, we sought to explore a number of possible computationally-aided enhancements in the ways that technologies can better support and improve the rigor and efficiency of intelligence analysis through the integration of new computationally-based methods and algorithms but also by exploring and nominating new ways in which improved human-machine symbiosis can be realized. Also, we were trying to strike the best balance between technologies and methods that are of the basic research variety while having plausibility in terms of potential for mid-term type operational deployment. Another main goal was toward providing support that can yield the type of "story" or narrative type product that many intelligence analysis environments require. These are those environments that allow for more contemplative methods, accommodating the formulation and evaluation of optional interpretations that have to be weighed and evaluated or argued for. This goal imputes a requirement for capabilities that support what we are calling "hypothesis synthesis" or "hypothesis fusion" as mentioned previously, where competing hypotheses that evolve either: (a) directly from evidence or (b) developed from evidence or assumptions by disparate individual tools are traded off and synthesized into a defendable, integrated hypothesis at the narrative or situational level. In today's analysis environments, these synthesizing operations constitute and demand a high cognitive workload. A major goal is to develop a design whose overall rationale is traceable to and consistent with joint service and Intelligence Community future directions in methodological development balancing effectiveness, efficiency, and rigor; as a result, we have made efforts to garner real-world viewpoints on these directions.

2.3 Future Directions in Intelligence Analysis

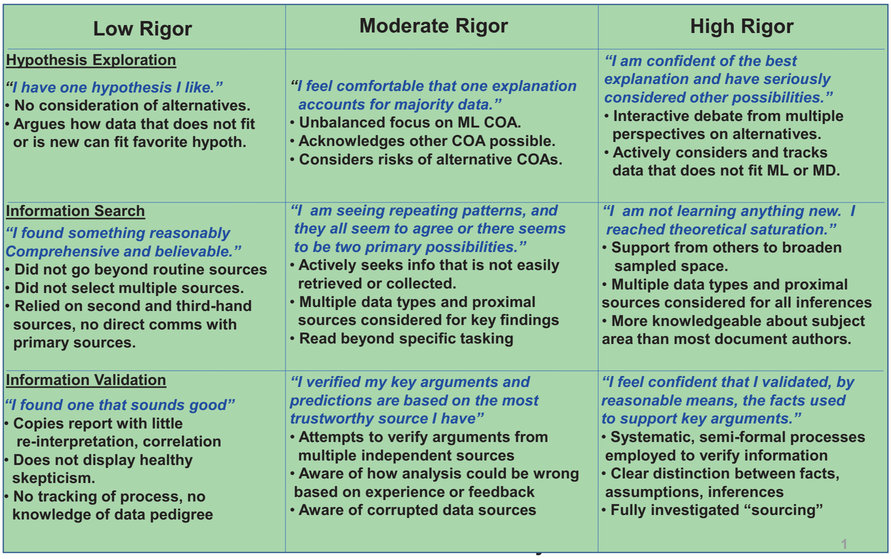
2.3.1 Reviews of Open Literature and Operational Environments

The research described here was in fact partially inspired by our prior exploration of the nature of modern-day computational support for intelligence analysis in the open literature as summarized in Llinas (2014a, b). That work extensively examined much of the literature on such techniques with a focus on technology strategies and interfacing strategies in regard to methods to achieve some level of symbiosis. It should be noted that this survey also collected works from the field of criminal analysis and the related area of Artificial Intelligence and the Law. Our research team at the Center for Multisource Information Fusion has also addressed these topics under a large Army Research Office grant for Unified Research on Network-based Hard and Soft Information Fusion, see e.g., Llinas et al. (2010) and Date et al. (2013a, b) for the Counterinsurgency domain. In both of these surveys, what we primarily saw was a strategy for analytical tool suite design that resulted in collages of disparate tools of various descriptions. Each of these tools can be argued to be individually helpful, producing what we called "situational fragments", i.e. hypotheses, each of which are hypotheses about a particular slice of a situational condition. These problems, and the employment of modern technologies that allow evermore data and information to be available, are extraordinarily complex and it is natural to see "divide and conquer" solution, tool, and visualization strategies being applied. But the latent challenge for sensibly all human analysts involved in these situations is to connect the dots, evolve the most plausible story/narrative, or the most plausible argument in the face of inherent complexity and "big data" quantities and varieties of information. For that type of capability, we saw nothing at all in this survey, leading to our conclusion that there is a significant need for development of both a paradigm and associated technological support for hypothesis synthesis or fusion, aiding human analysts to assemble a more holistic picture (a narrative or story) much more efficiently.

In the Fall of 2015, a team visit to the Air Force National Air and Space Intelligence Center (NASIC) was carried out in order to assess our evolving perspectives regarding future analysis requirements. Because of our future-oriented perspective, our visit focused on the Advanced Analytics Cell (AAC) team, that similarly is studying such future requirements. In summary, this visit revealed that there was considerable commonality in the respective lines of thought across the activities of the AAC and our approach. It also broadly provided a level of confidence that the approach described here was sound and that it resonated with current advanced thinking at least in the Air Force as regards methods and needs of modern intelligence analysis.

2.3.2 Analytical Rigor in Intelligence Analysis/Argument Mapping

Another touchstone for the project as regards vetting our thinking and approach involved discussions with staff from the Army Intelligence Center at Ft. Huachuca, NM. Messrs Robert Sensenig and William Hedges (of Chenega Corp., advisors to the Army on intelligence matters) were our key points of contact. Two main topics were discussed: rigor in analysis, and the use of argument-based techniques of analysis. The Army is quite keen on the entire issue of improving rigor in analysis; this viewpoint certainly is consistent with our own thoughts regarding improvements in the intellectual aspects of analysis. Mr. Sensenig provided the charts of Figs. 2.1 and 2.2 below that depict the mapping/cross-correlation of analysis functions and levels



Low Rigor	Moderate Rigor	High Rigor
Hypothesis Exploration *"I have one hypothesis I like."* • No consideration of alternatives. • Argues how data that does not fit or is new can fit favorite hypoth.	*"I feel comfortable that one explanation accounts for majority data."* • Unbalanced focus on ML COA. • Acknowledges other COA possible. • Considers risks of alternative COAs.	*"I am confident of the best explanation and have seriously considered other possibilities."* • Interactive debate from multiple perspectives on alternatives. • Actively considers and tracks data that does not fit ML or MD.
Information Search *"I found something reasonably Comprehensive and believable."* • Did not go beyond routine sources • Did not select multiple sources. • Relied on second and third-hand sources, no direct comms with primary sources.	*"I am seeing repeating patterns, and they all seem to agree or there seems to be two primary possibilities."* • Actively seeks info that is not easily retrieved or collected. • Multiple data types and proximal sources considered for key findings • Read beyond specific tasking	*"I am not learning anything new. I reached theoretical saturation."* • Support from others to broaden sampled space. • Multiple data types and proximal sources considered for all inferences • More knowledgeable about subject area than most document authors.
Information Validation *"I found one that sounds good"* • Copies report with little re-interpretation, correlation • Does not display healthy skepticism. • No tracking of process, no knowledge of data pedigree	*"I verified my key arguments and predictions are based on the most trustworthy source I have"* • Attempts to verify arguments from multiple independent sources • Aware of how analysis could be wrong based on experience or feedback • Aware of corrupted data sources	*"I feel confident that I validated, by reasonable means, the facts used to support key arguments."* • Systematic, semi-formal processes employed to verify information • Clear distinction between facts, assumptions, inferences • Fully investigated "sourcing"

Fig. 2.1 Mapping of analysis functions vs levels of rigor (Part 1) (Courtesy of Mr. Robert Sensenig, Chenega Corp.)

Low Rigor	Moderate Rigor	High Rigor
Inference Resilience *"My story/explanation/argument seems reasonable to me, independent of available supporting evidence."*	*"I feel that the evidence is reasonably solid for my primary explanation."* • Considers whether being wrong about some inferences would influence or negate the best explanation. • Beware false precision!!	*"I feel comfortable that the key Inferences are resilient to inaccurate Information."* • Uses strategy to systematically consider strength of evidence if individual interpretations debunked. • Actively looked for reasons why a source might misinterpret or manipulate data/information.
SME Collaboration *"I trust my supervisor to cover specialist content area or to be the SME."*	*"I have talked to SMEs, as time allowed, within my personal network."* • Attempts to consult some of the right people.	*" Leading expert in the key content area." (Beware Group Think!!)* • Capital expended to gain access to leading experts in multiple fields related to the analysis.
Information Synthesis *"I compiled the relevant info."* • Numerical values or graphs disconnected from key arguments.	*"I provided insight that goes beyond the source reporting & key documents.* • Validation of events in context. • Understanding depicted as an integrated view including tradeoff dimensions. (Frameworks, models).	*"I considered diverse interpretations trying to identify new concepts* • Sensemaking metrics are high. • Collaborative cross checks applied to data synthesis processes • Collaborative use of diagrams to show relationships between evidence and hypothesis. 2

Fig. 2.2 Mapping of analysis functions vs levels of rigor (Part 2) (Courtesy of Mr. Robert Sensenig, Chenega Corp.)

of rigor, notionally showing an analyst's mind-set across these functions and levels, as well as thumbnails of analysis activities across the matrices. These charts are among the resources we used to direct our efforts.

Mr. Hedges recounted his experience in learning of argument-based methods of analysis and also shared segments of the Army's training activities in the teaching of argument mapping for intelligence analysts. Figure 2.3 shows an excerpt of one of the training segments directed to teaching of argument mapping.

Overall, we believe it is quite clear that the thinking and approaches of this research program are very consistent with modern thoughts in both the Air Force and Army in regard to:

- The use of improved intellectual strategies and methods
- The need for an movements to improve analytical rigor
- The employment of argumentation-based methods and technologies as one framework to achieve these goals

2.4 Approaches to Computational Support

2.4.1 Paradigms and Methods

In today's open-world environment, historical paradigms and methods that rely on deep analysis of an adversary's Tactics, Techniques, and Procedures (TTP's) as a basis for paradigms that can be broadly labeled as of a template-matching type are

U.S. ARMY INTELLIGENCE CENTER AND FORT HUACHUCA
Fort Huachuca, Arizona 85613-7002

LP Narrative & Teaching Plan: Argument Mapping
24 April 2013
PFN:xxxxxxxx

Enabling Learning SLIDE 2: Objective

ACTION:	Create an Argument Map to make analytic assumptions, intelligence gaps, or arguments more transparent.
CONDITIONS:	Given all class handouts to date, appropriate references, an operational framework scenario, and in-class discussion.
STANDARDS:	Create an argument map that incorporates critical and creative thinking and basic and diagnostic structured analytic techniques in order to provide clearer ACH understanding and validate the ACH.

Fig. 2.3 Sample of curriculum at Army Intelligence School Training in argument mapping (Courtesy of Mr. William Hedges of Chenega Corp.)

considered unworkable. Modern-day adversaries and problem conditions demand more flexibility and accommodation of imperfections in analysis techniques. These environments, that we call "weak knowledge" problems, require a more flexible approach and one that allows for unknown states of affairs and degrees of ignorance while carrying out the best analysis possible. Such methods are usually labeled as defeasible and abductive[4] and are directed to the most rational hypotheses that can be defended in some way as "best". In our exploration of alternatives, we narrowed our choices based on two factors: one was the commentaries on intelligence analysis and associated assertions about methodological requirements that balance evidence, arguments, and stories (i.e., nominated hypotheticals), and the other was a body of work we discovered that was centered in Europe that focused on methods of this type, with a deep basis on argumentation-based principles. One clear example of these remarks is shown in the writings of Schum (2005) who suggests that:

[4] We like Stanford's definition here (http://plato.stanford.edu/entries/reasoning-defeasible/): "Reasoning is *defeasible* when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provides support for the conclusion, even though it is possible for the premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information."

- "**Careful construction of arguments** in defense of the credibility and relevance of evidence **goes hand-in-hand with the construction of defensible and persuasive narratives.**"
- "In constructing a narrative account of a situation of interest we must be able to **anchor our story appropriately on the evidence** we have that is relevant to the conclusion we have reached. Careful argument construction provides the necessary anchors."

These remarks, and the results of our surveys, suggest an exploration of methods that jointly exploit the union of evidence, arguments, and stories, in a synergistic dynamic that leads to "best" narratives that holistically convey the most rational explanation of the evidences and sub-stories. These source materials were the foundation of the evolution of our thinking to explore a paradigm of this nature.

2.4.2 Argumentation Methods

As we contend above, one main technological/theoretical theme that we pursue here is the examination of argumentation-based concepts, methods, and computationally-supported tools as one candidate paradigm supportive of intelligence analysis. Argumentation-based methods have a long history in the law and in the teaching of critical thinking, and in the last decade or so have found their way into supporting criminal and intelligence analysis. These extended applications have largely been a result of research and development in the construction of computational tools for "diagramming" or "mapping" arguments that enable and streamline the examination of the veracity of pro and contra arguments in various situations.[5] Before reviewing the state of the art in computational methods for argumentation based reasoning, we briefly review the different paradigms for argumentation itself; that is, there are different flavors or variations of methods that have the core notion of an argument as their foundation. This summary review is shown in Table 2.1 below.

The majority of argumentation-based methods utilize a deterministic formal logic and theorem proof approaches, and the notion of argument acceptance and attack, see, e.g., Simari and Rahwan (2009). There has been multiple argumentation schemes developed with each of them having advantages and drawbacks as methods useful for supporting decisions based on a highly uncertain environment. Most of them represent abstract argumentation, which determines an argument's acceptability

[5] By the way, we see the (necessary) balancing of Pro and Contra arguments as another good feature of these argumentation methods; to some degree this is a built-in preventative to the human foible of confirmation bias.

Table 2.1 Types of Argumentation-based Paradigms

Argumentation types	Methods	Prototypes[a]
Abstract argumentation	Involving formal logic, theorem proof, and based on the notion of argument acceptance and attack	CISpaces, Carneades Araucaria and various others
Story-based argumentation	Abduction-based reasoning about hypothetical stories explaining the evidence	Bex's research on design; AVERS
Hybrid methods	Combination of logic and probability or belief	
Assumption based probability/ belief based argumentation. (A probabilistic extension of abstract argumentation.)	Conjunction of uncertain assumptions to define arguments and disjunction of arguments Assigning probabilities/beliefs to assumptions	ABEL
Belief-story based argumentation	Observations are explained by hypothetical stories Uncertain arguments based on evidence are combined to support alternative stories and select the most credible one (abductions)	This was the goal or the research described in this chapter

[a]See later discussion on Prototypes for citations

on the basis of its ability to counterattack all arguments attacking it. A more promising approach introduced in e.g. Bex (2013) is an abstract story-based argumentation, in which hypothetical "causal stories are hypothesized to explain the evidence, after which these stories can be supported and attacked using evidential arguments." A combination of logic and belief theories for argumentation under uncertainty has been considered for assumption based argumentation, see e.g. Haenni (2001), but these models require a known and complete knowledge base, which does not exist in the context, which we are addressing here. We seek abductive reasoning methods that combine certain desirable capabilities:

- Allowance for open-world reasoning
- Allowance for assigning and combining beliefs in arguments and reliability of the source (i.e., a basis for assigning and combining/propagating uncertainty)
- Integration of human intelligence that enables hypothetical stories to be combined with hypotheses resulting from evidence-based arguments
- Method of evaluating and selecting the most credible stories

Abductive reasoning is often labeled as "backward reasoning" in that it explores/nominates plausible conclusions or assertions that can "explain" or rationalize the evidence available; the notion is that a rearward look is taken from the conclusion toward the available evidence. Abductive reasoning is also often described as reasoning to the best explanation. Our approach is also hybrid in bringing together the abductive reasoning over both the uncertain arguments and human-nominated storylines and rationalizing both lines with the also-uncertain evidence. To deal with these uncertainties, we propose to incorporate the Transferable Belief Model (TBM) see, e.g. Smets (1994). Briefly the TBM is a two-level model, in which quantified beliefs in hypotheses about an object or state of the environment are represented and combined at the *credal* level[6] while decisions are made based on so-called *pignistic* probabilities obtained from the combined belief by the *pignistic*[7] transformation at the *pignistic level*. So taken together, our approach can be summarized as involving the explicit incorporation of uncertainty into hybrid story-based argumentation, depicted in Fig. 2.4.

The basic ideas of the story-based approach are presented in Fig. 2.5 that shows that:

- Arguments are derived from evidential foundations
- Stories are analyst-nominated (with computational support, e.g., prior case libraries) hypotheticals
- Together these lead to the assembly of sub-stories and, again with computational support (see Sect. 2.7 on our ideas), to the development of an integrated Narrative/Story

In the following, we provide our view of the state of the art in each of several functional areas necessary toward realization of a desired level of automated capability for a future semi-automated, computationally supported analysis prototype that realizes the hybrid capability described. We note, from the literature, a set of particular argumentation-related functional categories: Argument Detection-Construction-Invention-Mining-Accrual and, importantly (as it dominates the literature) Visualization that will serve as the basis for our review.

2.4.3 Computational Support to Argumentation: The State of the Art

It is realized that the input to any modern intelligence analysis system could be in a wide variety of formats and types in terms of media and modalities. As regards the role of these varying inputs toward supporting argument formation, however, it is considered that textual inputs provide the most likely format for somewhat-direct input-to-argument formulation. Most other input types would more likely represent evidential data (such as sensor data) and require a more complex structuring process to frame the data into argument forms. (Later it will be seen that we address sensor

[6] Credal will be seen to mean belief but in regard to conducting analysis this term is taken to mean a (human's) conviction of the truth of some statement or the reality of some being or phenomenon especially when based on examination of evidence.

[7] Pignistic is a term coined by Smets and is drawn from the Latin pignus for "bet", and can be taken to imply or relate to a probability that a rational person would assign to an option when required to make a decision.

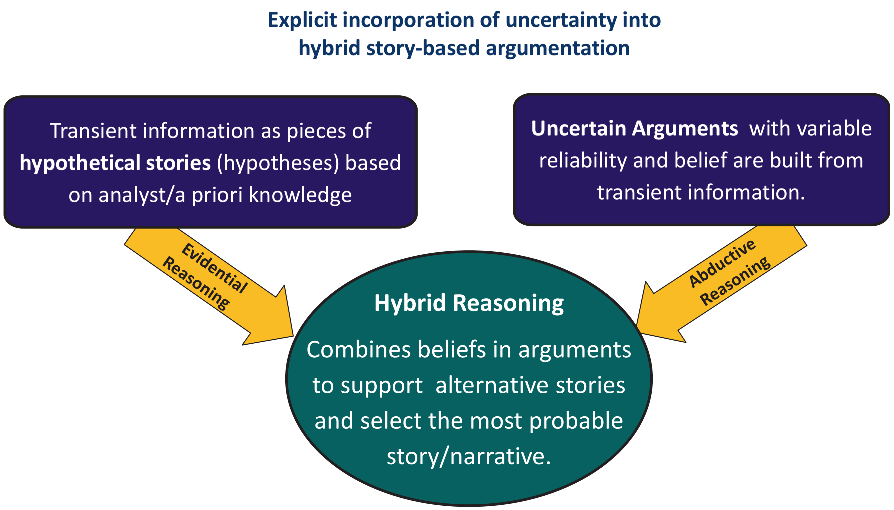


Fig. 2.4 Depiction of the proposed hybrid approach

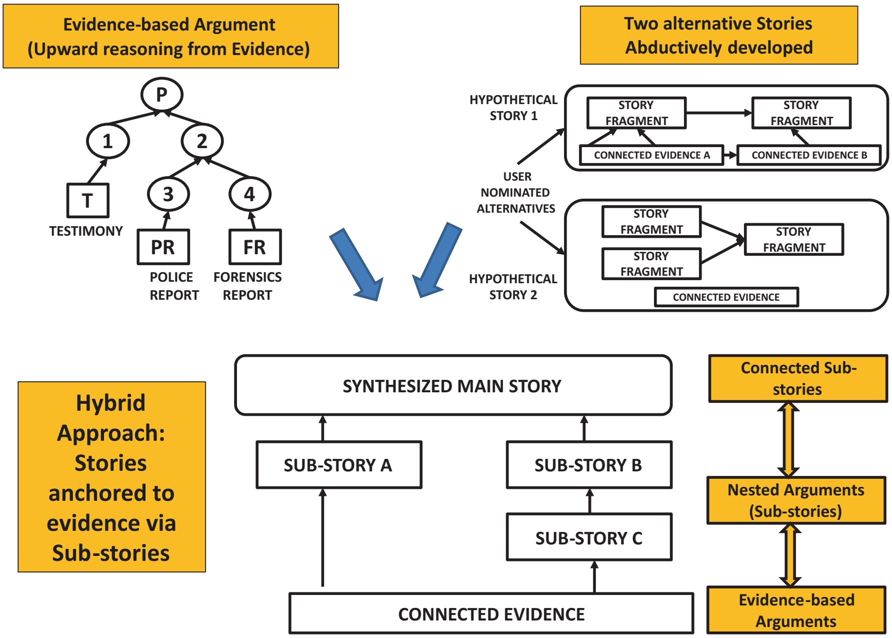


Fig. 2.5 Overview of Bex's scheme for joint argument-story exploitation

data as an input stream of interest in proposing our design but we note that sensibly all current systems do not include such "hard" data as input.).

Our review of current prototype argument systems shows that the front-ends of these prototypes do not currently provide any automated support to the identification of either the basic linguistic form of an argument (based on lexical content and other factors) or types of argument structures based on argument taxonomies (usually called "schemes" in the argument literature) from textual report, prose-type input, whether structured or not. Thus, a significant human cognitive operation is needed in these prototypes for the formulation of the very basic constructs (arguments) upon which next analysis steps, some computationally-aided, depend. Moen et al. (2007) in discussing the Araucaria prototype[8] designed for argument visualization, say that "The manual structuring of an argumentative text into a graph visualization as is done in the Araucaria research is a very costly job."

However, we will see that approaches to computational support for extracting parts of or entire argument schemes from text has been addressed but has not, for whatever reasons, been integrated into modern prototype systems. As noted above, this functional activity comes under different names, such as argument detection, argument construction, and argument mining—we simply use the term detection here but draw on works having these other labels to describe what is happening in

[8] An argument mapping tool developed at the University of Dundee; see http://www.arg-tech.org/index.php/projects/.

the research community. We will review some sample works in this area and also provide a broader summary view of the state of the art in the next section.[9]

2.4.3.1 Argument Detection

Moen et al. (2007) Automatic Detection of Arguments in Legal Texts

This paper describes the results of experiments on the detection of arguments in texts with a focus on legal texts. As will be seen in related works on detection, the detection operation is seen as a classification problem based on defined features of a postulated argument scheme. A classifier is developed in the paper and trained on a set of annotated arguments. Different feature sets are evaluated involving lexical, syntactic, semantic, and discourse properties of the texts, and each of their contributions to classifier accuracy is examined.

Strategies for detecting argument constructs clearly require some defining process for the nature of argument forms or schemes in a linguistic sense; said otherwise, an ontology of argument forms is required. Moen et al state that "The most prominent indicators of rhetorical structure are lexical cues (Allen 1995), most typically expressed by conjunctions and by certain kinds of adverbial groups." Humans can do this well but one important factor exploited by humans to do so is the context of the textual phrases, and this is very hard to do automatically. The approach in Moens et al. (2007) is admitted to be a bounded first step toward automating this process, and they take an approach built on isolated sentences. They represent sentences as a vector of features and use annotated training data to train a classifier. (It will be seen that this problem is broadly treated as a classification problem in the literature.) We will not review the details of the features and methods but they use a multinomial Bayes classifier and a Maximum Entropy based classifier in this work. It is interesting to see that even simple feature sets yield reasonable (~70+% accuracy) results. The paper also reviews related works and remarks that this type of research on detection is very limited in the legal domain at least (as of the date of this publication, 2007).

Mochales-Palau and Moens (2007)

In a later work, these authors develop an approach to detect sentences that contain argument structures (i.e., apart from not discerning the existence of Walton-type schema; in Walton et al., 2008). A maximum-entropy-based classification is used to determine if input sentences are argumentative or not, and more specifically if they contain a premise, a conclusion or a non-argumentative sentence. These same

[9] For the Reader: our reviews in the next section are running commentaries about selected papers from the literature that address each reviewed topic; in various places any emphasis provided is our own. Some excerptions from the original papers are included without quotation.

authors also study and develop a context-free grammar for argument detection in Mochales and Moens (2008), but this was a very limited study across a ten document training set.

Feng and Hirst (2011), Classifying Arguments by Scheme

This work is oriented to a subtle issue in argumentation, the issue of enthymemes; as part of an approach to argument detection, in reasonably-frequent cases, there are implicit premises that are never present in the prose text, and these are called enthymemes. To address this issue however, the authors argue that by first identifying the particular argumentation scheme that an argument is using will help to bridge the gap between stated and unstated propositions in the argument, because each argumentation scheme is a relatively fixed "template" for arguing. The idea here is that the argument scheme classification system is a stage following argument detection and proposition classification; that is, a two-stage system involving two different classification systems.

This paper (and some others) relies on the notion of argument schemes or schemata; such schemes are structures or templates for forms of arguments Walton's set of 65 argumentation schemes is one of the most-cited scheme-sets in the argumentation literature. According to Feng and Hirst (2011), the five schemes defined in Table 2.2 copied below are the most commonly used ones, and they are the focus of the scheme classification system that is described in this paper. The functional

Table 2.2 Five top argument schemata from Walton et al. (2008)

Argument from example
Premise: In this particular case, the individual *a* has property *F* and also property *G*
Conclusion: Therefore, generally, if *x* has property *F*, then it also has property *G*
Argument from cause to effect
Major premise: Generally, if *A* occurs, then *B* will (might) occur
Minor premise: In this case, *A* occurs (might occur)
Conclusion: Therefore, in this case, *B* will (might) occur
Practical reasoning
Major premise: I have a goal *G*
Minor premise: Carrying out action *A* is a means to realize *G*
Conclusion: Therefore, I ought (practically speaking) to carry out this action *A*
Argument from consequences
Premise: If *A* is (is not) brought about, good (bad) consequences will (will not) plausibly occur
Conclusion: Therefore, *A* should (should not) be brought about
Argument from verbal classification
Individual premise: *a* has a particular property *F*
Classification premise: For all *x*, if *x* has property *F*, then *x* can be classified as having property *G*
Conclusion: Therefore, *a* has property *G*

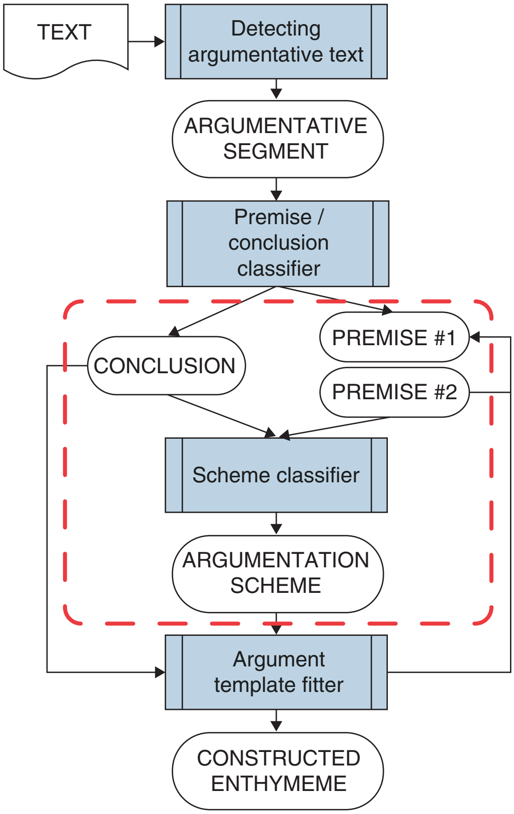


Fig. 2.6 Functional flow of argument scheme detection

approach is shown in Fig. 2.6, where it can be seen that argument detection from text precedes the argument *scheme* classification step.

The classifier approach is essentially entropy based. Performance is quite variable, since the various argument schemata vary significantly in the specificity of cue phrases; this is an issue to be dealt with in classifying argument schemata. Note that a training data set for either argument detection or scheme detection requires that the textual corpus be labeled with the "true" argument constructs. This study used the Araucaria data set available at the Araucaria research project website, http://www.arg-tech.org/index.php/projects/.

2.4.3.2 Argument Mining

Moens (2013), State of the Art in Argument Mining

Argumentation mining is defined by Moens as the (automated/automatic) detection of the argumentative discourse structure in text or speech and the recognition or functional classification of the components of the argumentation. It is clear from this definition that various functional capabilities are required in mining to include detection of lexical units, identification of sentences containing arguments, and the fit of an argument sample to a predefined argument schema. This type of functionality falls into the domain of Information Retrieval systems, to provide the end user with instructive visualizations and summaries of an argumentative structure. Moens dates argument mining as having started in 2007. The notion of argument "zoning" is mentioned as an area of some study, where a document or corpus is examined to localize sections possibly containing argument-based content. Moens reviews some works that perform these types of functions as typical of the current state of the art; typical Precision/Recall/F measures are in the high 60 to low/mid 70% range, which is just fair performance.[10]

This paper also describes some capability goals for argument mining systems. While discussing the use of machine learning methods, the goal of detecting or recognizing a "full argumentation tree" is mentioned. Cited papers use either a set of piecewise classifiers or a single set-wise or tree-wise classifier, but these are cited only as methodological examples, i.e., these works do not apply such methods to the argument mining problem. Another important argumentation mining issue stated by Moens is the correct identification of the relationships between text segments (e.g., the relationship of being a premise for a certain conclusion) and defining appropriate features that indicate this relationship. Moens suggests that textual entailment in natural language processing, which focuses on detecting directional relations between text fragments may be useful.

2.4.3.3 Argument Invention

Walton and Gordon (2012), the Carneades Model of Argument Invention

This paper seems a bit off-topic for our purposes but one aspect that may be of interest is that the mechanics involved in argument invention may hint at how stories (in a knowledge base) and arguments achieve some symbiosis. Argument invention is a method used by ancient Greek philosophers and rhetoricians that can be used to help an arguer find arguments that could be used to prove a claim he needs to defend. The Carneades Argumentation System (named after the Greek skeptical philosopher Carneades) is said by Walton and Gordon to be the first argument mapping tool with an integrated inference engine for constructing arguments from knowledge-bases, designed to support argument invention. It can be said that the notion of invention revolves around the notion of how arguments are evaluated or defended; the idea is to provide automated support to improve the acceptability of an argument. This tool is intended for rhetorical-type applications but conceptually could have applicability in analysis frameworks.

[10] The F measure is the harmonic mean of precision and recall, and can be viewed as a compromise between recall and precision. It is high only when both recall and precision are high.

We offer an aside regarding argument evaluation, drawn from Walton and Gordon (2012), as follows: one approach to argument evaluation revolves around the idea of "critical questions" to evaluate an argument. Walton and Gordon (2012, p. 1) suggest: "Critical questions were first introduced by Arthur Hastings (1963) as part of his analysis of presumptive argumentation schemes. The critical questions attached to an argumentation scheme enumerate ways of challenging arguments created using the scheme. The current method of evaluating an argument that fits a scheme, like that for an argument from expert opinion, is by a shifting of the burden of proof from one side to the other in a dialog. When the respondent asks one of the critical questions matching the scheme, the burden of proof shifts back to the proponent's side, defeating or undercutting the argument until the critical question has been answered successfully. At least this has been the general approach of argumentation theory." Thus, the presence of critical questions could serve as a mechanism to assure that pro and contra sides of an argument receive attention.

The Carneades design approach provides a number of "assistants" for helping users with various argumentation tasks, including a "find arguments" assistant for inventing arguments from argumentation schemes and facts in a knowledge base, an "instantiate scheme" assistant for constructing or reconstructing arguments by using argumentation schemes, and a "find positions" assistant for helping users to find minimal, consistent sets of statements which would make a goal statement acceptable. The schemes representing knowledge of the domain in the knowledge base must be programmed manually by an expert. A distinctive contribution of the Carneades system is the integration of an inference engine in an argument mapping tool. Although the paper does not emphasize application in the legal domain, it seems clear that this system is oriented to either legal applications or in rhetorical applications as mentioned previously.

2.4.3.4 Argument Visualization (a.k.a. Mapping, Diagramming)

Argument visualization is often claimed to be a powerful method to analyze and evaluate arguments by providing a capability to perceive dependencies among argument components of evidential components, premises, and conclusions, focusing on the logical, evidential or inferential relationships among propositions. Argument visualization and theoretical modeling play important roles to cope with working memory limitations for problem solving, providing some relief to the cognitive workload that these analyses impute. Since the task of constructing such visualizations (also described in the literature as argument mapping or diagramming) is laborious, researchers have turned to the development of software tools that support the construction and visualization of arguments in various representation formats that have included graphs and matrices among other forms. To say that there have been a number of prototype systems developed that support argument diagramming is rather an understatement—a website provided by Carnegie-Mellon University (http://www.phil.cmu.edu/projects/argument_mapping/) shows, just on the first page, the following subset of tools shown in Table 2.3; the complete table goes on

Table 2.3 Sampling of computer-supported argument diagramming tools (see http://www.phil.cmu.edu/projects/argument_mapping/)

Tool	Description	Representation	Audience
Athena	Argument mapper from Blekinge Institute of Technology and CERTEC, Sweden	Simplified Toulmin	Education
ArgMAP	Argument mapper	Simplified Toulmin	Research
ArguMed	Argument mapper based on DEFLog	DEFLog (Toulmin extension)	Research
Argutect	Argument mapping-like "thought-processor" from Knosis, Pittsburgh	Thought tree (tree of questions and answers, can be used as simplified Toulmin)	Productivity, education
Araucaria	Argument mapper from University of Dundee, UK	Simplified Toulmin	Education
Belvedere	Collaborative concept mapper and evidence matrix originally developed by D. Suthers at LRDC, Pittsburgh, now at LILT, University of Hawai'i at Manoa	Inquiry/evidence maps and matrices (links between claims and supporting data)	Education
Causality Lab	Allows students to solve social science problems by building hypotheses, collecting data and making causal inferences	Causal diagram and data charts	Education
Carneades (.pdf)	Toulmin based mathematical model for legal argumentation	Toulmin	Law
ClaimMaker/ ClaimFinder/ ClaimMapper	Concept mapping of knowledge claims from S. Buckingham Shum's Scholarly Ontologies Project, KMI, Open University, UK	Concept map with semiformal ontology for argumentation	Research

(continued)

Table 2.3 (continued)

Tool	Description	Representation	Audience
Compendium	IBIS mapping tool originally developed by Verizon Research Labs and associated with CogNexus Institute and KMI, Open University	Dialogue map (concept map with ontology: nodes can represent issues, ideas, pro, con, and notes)	Ill-structured problems
Convince Me	Creates diagrammatic representations of hypothesis and evidence	Evidence map	Education
Debatabase	Debatabase is the world's most useful resource for student debaters. Inside you will find arguments for and against hundreds of debating topics, written by expert debaters, judges and coaches	Communal, simplified Toulmin	Education

for 2–1/2 pages. Note also the range of representational forms, in part dependent on the argument-model used in the application.

The effectiveness of such diagramming or mapping tools is reviewed in (van den Braack et al. 2006). Among the tools that were experimentally tested for their effectiveness were Belvedere, Convince Me, Questmap, and Reason!Able, which are a sampling of tools from Table 2.3.[11] While there are many issues regarding such evaluations discussed by van den Braack including criticisms about statistical testing methodology, the paper concludes that (p. 7) "most results indicated that the tools have a positive effect on argumentation skills and make the users better reasoners. However, most experiments did not yield (statistically) significant effects." Another study (Twardy 2004) showed that (manual) argument mapping generally helped in understanding arguments and also enhanced critical thinking; the study also showed that the benefits were greater with computer-based argument mapping.

In Mani and Klein (2005), they review structured argumentation as an analysis framework for "open-ended" (i.e., in operational cases where absolute truth is unknown) intelligence analysis. The paper is a short, opinion-type paper and asserts that structured arguments are a means not just of representing and reusing reasoning (one useful benefit), but also a means of communicating and sharing the argument, as analysis is often collaborative. They suggest that one way of assessing the quality of the associated reasoning is to determine how easy the argument is to follow and

[11] See the website listed at Table 2.3 for further details on these systems.

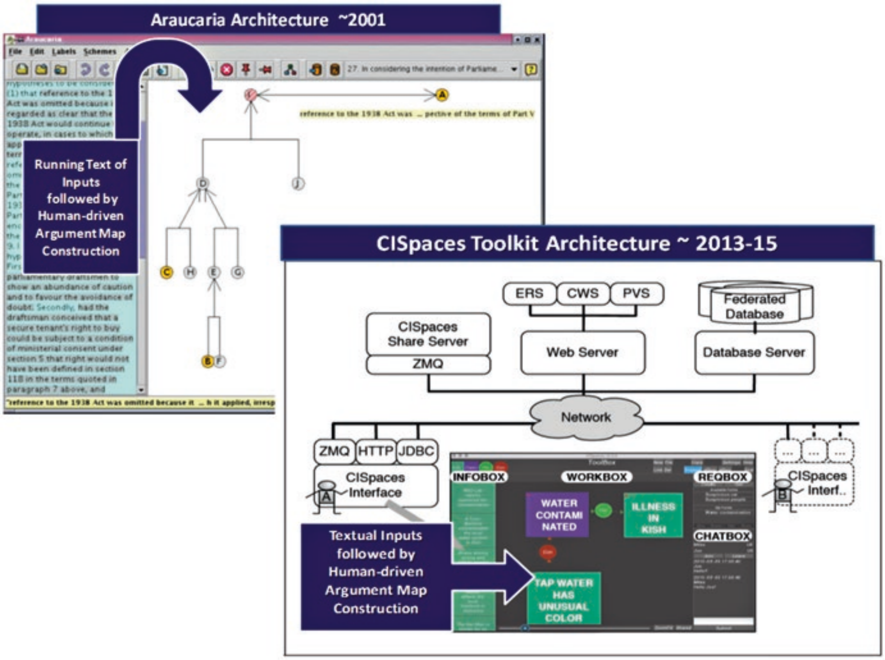


Fig. 2.7 Sampling of argument visualization prototypes

understand. If arguments are constructed in agreeable ways (e.g., based on argument models/schema) and correspondingly visualized, presumably they can be more easily communicated and discussed with others.

To allow an appreciation for what such visualizations look like, we show some examples in Fig. 2.7; these are drawn from Gordon's presentation in (1996); we use his format as it typically provides a screenshot with some remarks on associated features. Araucaria is very frequently cited as an exemplar of relatively recent prototypes for argument visualization (see for example Suthers et al. 1995; Reed and Rowe 2004). The most recent prototype we are aware of is CISpaces, developed under joint US-UK efforts and led by Norman at the University of Dundee. It can be seen that Araucaria, while having many attractive features, still imputes a high cognitive load onto human analysts is working with streaming text and manually developing the diagrammatic argument constructs. CISpaces incorporates various additional features such as chat for collaborative analysis but still imputes similarly high cognitive workloads for argument mapping; see additional comments below.

2.5 Current-Day Computational Support to Argumentation

One other remark that we will offer here is that the greater proportion of research along the lines of computational support schemes for analysis has been carried out in Europe or at least outside of the USA. Among the leading centers of such research are:

- ARG-Tech, at the University of Dundee in Scotland (http://www.arg.dundee.ac.uk/)
- Centre for Research in Reasoning, Argumentation and Rhetoric, University of Windsor, Canada (http://www1.uwindsor.ca/crrar/)
- Intelligent Systems Group, University of Utrecht, Holland (http://www.cs.uu.nl/groups/IS/)
- Intelligent Systems Group, University College London (http://is.cs.ucl.ac.uk/introduction/)

To the extent that there is belief that computationally-supported argumentation methods can be helpful to intelligence analysis, this situation should be of concern to the US academic and industrial research communities.

2.5.1 AVERS and CISpaces as Leading Relevant Prototypes

The research program described in this paper was largely initiated by an early review of a dissertation in Holland having to do with "Sensemaking software for crime analysis" (van den Braack et al. 2007) by Susan van den Braack. That dissertation provided the spark of thinking, as was first explored in that work, for a hybrid, story and argumentation based approach to intelligence analysis since intelligence and criminal analysis requirements have quite similar requirements. This dissertation described AVERS as a prototype developed within the dissertation effort that was designed to explore alternative "scenarios" (stories in effect) based on evidentially-supported arguments. A prototype was developed in the university framework but unfortunately the code for that prototype was not subsequently maintained (we had contacted Dr. van den Braak to explore this). Nevertheless, as described in van den Braack (2010), it is clear that the thinking related to the design and realization of AVERS was very synergistic to our line of research. Formalisms for combining stories and arguments in this hybrid environment were put forward in Bex et al. (2007a, b).

During our program, largely because of our close relations to researchers at the Army Research Laboratory, we learned that, under the "International Technology Alliance (ITA)" program (a US-UK cooperative research program) that a team at the University of Aberdeen (at ARG-Tech as noted above) was carrying out the development of a prototype called "CISpaces", with goals also similar to ours.

CISpaces was conceptualized as an initial set of tools for collaborative analysis of arguments and debate, providing a uniform way of constructing and exchanging

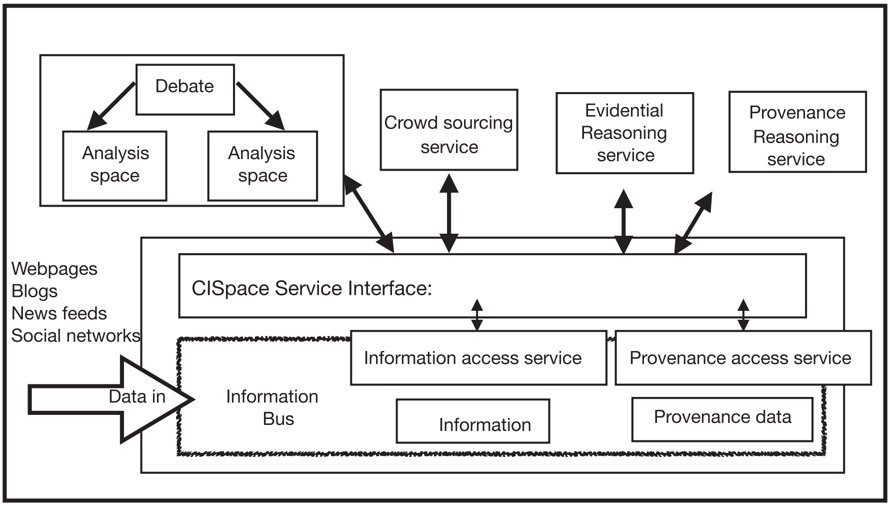


Fig. 2.8 CISpaces functional architecture

arguments based upon argumentation schemes. The top-level functional design is shown in Fig. 2.8 below (Toniolo et al. 2014) and comprises three main services in a service-based architecture:

- the evidential reasoning service, supporting collaboration between users in drawing inferences and forming opinions structured by argumentation schemes;
- the crowd-sourcing service, enabling users to post requests for aggregated opinions from samples of a population;
- the provenance reasoning service, facilitating the storage and retrieval of provenance data including provenance of information and analysis.

The core components of CISpaces, as it is highly oriented to a collaborative, multi-analysts environment, are the WorkBox, the ChatBox and the ReqBox. As described by Toniolo, the WorkBox permits users to elaborate information by adding new claims or by manually importing information and conclusions from different locations; e.g., social networks, blogs. Different forms of argumentation-based dialogue are supported through the ChatBox: collaborative debate, information retrieval through crowd-sourcing, and reasoning about provenance. The list of active debates is intended to be maintained in the ReqBox.

While the development of a real software prototype of this type should be applauded for its forward-thinking approach and for moving the bar of computational support to argumentation to a new level, our thoughts on prototype design addressed other, additional issues:

- Inclusion of both Hard/sensor data as well as Soft/textual/linguistic data as input
  - This is a major change as sensibly all existing argumentation support prototypes are strictly text-input-based

- Major reduction in analyst cognitive workload
  - We see this as involving an aggressive inclusion of front-end, automated processing to aid in argument detection and construction, a major cognitive workload factor of all current prototypes, to include CISpaces.
  - Another aspect is in automated support to final analysis product development, seen as a narrative or story descriptive of a situational estimate of interest (none of the computational systems described here address this at all)
- Major concern for managing information quality along various lines, including automated support for relevance-checking and tracking and assessing provenance of input sources.

Because of our concern for these information quality factors, we established a research thrust along these lines. A later section also addresses our ideas, largely from our Lockheed teammates, on computational support to narrative development.

2.6 Computational Support for Narrative Development

As described earlier, for a broad range of intelligence analysis requirements, the desired final output of analysis is a situational picture of some type. In most cases these situations are best communicated as a story or narrative description. However, none of the system concepts and prototypes reviewed here addresses the issue of providing computational support to narrative development. In this next section, we describe our team's approach and some actual prototyping (done by Lockheed in conjunction with Virginia Tech in a separate effort).

2.6.1 Using Topic Modeling to Assess Story Relevance and Narrative Formation

As was remarked in particular for Sect. 2.4.3, here too we note that some elements of this section were extracted closely from the conference paper that reported the original work on Topic Modeling carried out in part by Lockheed ATL[12]; see Schlacter et al. (2015) for the original paper.

Storytelling as a data-mining concept was introduced by Kumar et al. (2008). Storytelling (or "connecting the dots") aims to relate seemingly disjoint objects by uncovering hidden or latent connections and finding a coherent intermediate chain of objects. This problem has been studied in a variety of contexts, such as entity networks (Hossain et al. 2012a, b, c), social networks (Faloutsos et al. 2004),

[12] Lockeed's Advanced Technology Laboratories; see http://www.lockheedmartin.com/us/atl.html.

cellular networks (Hossain et al. 2012a), and document collections (Hossain et al. 2012b; Shahaf and Guestrin 2010; Shahaf et al. 2012, 2013). The unsupervised learning technique for storytelling called Story Chaining links related documents in a corpus to build a story or narrative arc. The story chaining approach uses a real-time, flexible storytelling approach that can be used for streaming (online) data as well as for offline data. Because it is fully unsupervised, this approach does not carry the costs of competing approaches such as the need for configuration with domain knowledge or labeling of training data. As such, Story Chaining is ideal for new and frequently evolving domains. Figure 2.9 presents an example of a story chain generated from a corpus of news stories published in Brazil in 2013. The story chains generated from this approach can potentially tell a story about what is happening over time and across news articles by focusing on how the same people, organizations, and locations occur between documents. For this reason, story chains may be considered to be a narrative structure.

Because story chaining is an unsupervised, automated process that generates many results, there is a need to identify the story chains that contain the clearest narratives. One technique uses context overlap as a measure to produce stories that stick to one context by extracting context sentences from a document using a Naive Bayes classifier. Others, for assessing quality, also use dispersion plots and dispersion coefficient to evaluate the overlap of contents of the documents in a chain and thereby quality. Shahaf et al. (2013), as referenced above, define concepts of chain coherence, coverage, and connectivity that offer more insights into the storytelling process. Our approach differs in that it learns a topic model over the corpus and tries to associate certain types of topic change across a story chain as an indicator of how clear of a narrative structure is contained within a story chain.

Topic models are probabilistic models for uncovering the underlying semantic structure of a document collection based on a hierarchical Bayesian analysis of the original texts (Blei et al. 2003). They have been applied to a wide range of text to discover patterns of word use, or topics, across a corpus and to connect documents that share similar structure. In this way, topic models provide a way to create a structure from unstructured text in an unsupervised manner. We leverage them in our work primarily for this reason.

In our research, we have investigated the use of a topic model based analytics to evaluate the clarity of the story chain narrative structure. This work proposes two different kinds of measures of assessment, representativeness and quality.

Firstly, we considered a measure of representativeness that captures how well a story chain represents the corpus from which it was generated. For example, the story chain in Fig. 2.9 was generated from a corpus of thousands of documents published in Brazil in 2013 and it tells a clear story about the Pope visiting Brazil. The stories in the chain take place over a period of 11 days and fit well with the dominant theme of the corpus during that time period which focuses on social issues and protests. Our measure of representativeness is assessed by comparing the similarity of topics found over time in a story chain against those expressed in the corpus during the same time period. This measure assumes the corpus contains dominant topics that are desirable to understand. Our hypothesis for investigating

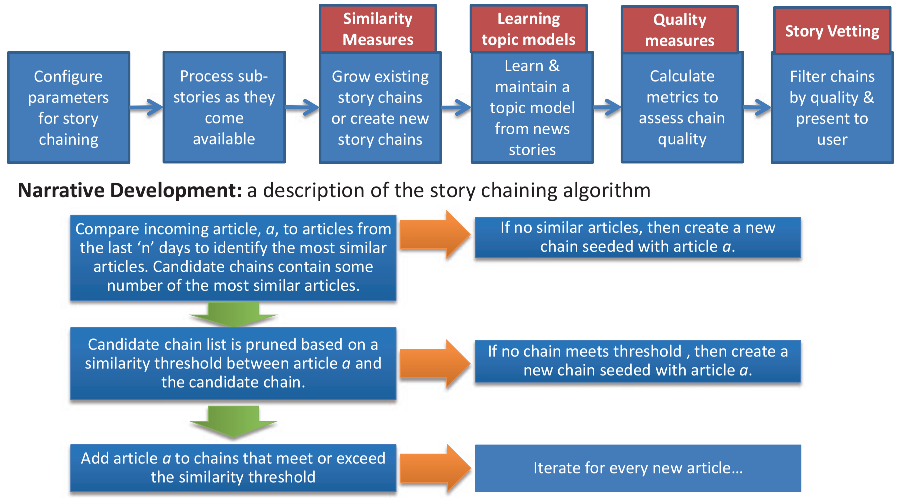


Fig. 2.9 Overview of topic modeling strategy for narrative development

representativeness was the idea that story chains with similar topic expressions to the corpus will convey narratives that are central to the corpus.

Secondly, we considered a measure of quality in which higher quality story chains exhibit a characteristic of focusing on a small number of stable topics, rather than many interleaved or shifting topics. To evaluate this form of quality, we decomposed the measure into two contributing measures, topic persistence and topic consistency.

Topic persistence was designed to capture volatility in topic focus within a story chain. In other words, how often does the topic of a chain shift across each link in the story chain? For example, consider a story chain that has 11 articles such that there are 10 transitions in the story chain connecting one article to the next article in the chain. Topic Persistence (TP) will indicate how well topics persist between links. If most of those ten transitions represent a change in the main topic of the article, then that story chain would have a lower TP score than a chain in which most of those ten transitions represented no change in the main topic. In this way, if a story chain has a high TP score, then most of the links in the chain represent connections between two articles that are discussing the same main topic, and hence, the narrative structure is exhibiting more stable structure for a, hypothetically, better quality chain.

Topic consistency (TC) is a relative assessment of the stability of the main topic of the story chain. More specifically, TC assesses how regularly the main topic of the story chain appears as a main topic of an article within the story chain. For example, if a story chain is made up of ten articles and has a main topic of political unrest, TC will indicate how stable that main topic of political unrest is by looking at each of the ten contributing articles and seeing if political unrest appears as the primary topic within those ten articles. If only three of those ten articles are focused on political unrest for a TC = 3/10 or 30%, that means that most of the articles in the chain are focused on (1) different topics, and (2) a variety of different topics such that consensus did not exceed three. Compare this to a scenario in which the story chain had seven articles focusing on political unrest where TC = 7/10 or 70%. In this case, the topic is much more consistent throughout the chain (not necessarily consecutively) and hence, the narrative structure more centered on political unrest and, hypothetically, of better quality.

Our results indicate that using topic model based analytics to predict the quality of a narrative structure is a promising avenue of research. We found correlations between all of our analytics and the human scoring of our story chains, with particularly strong correlation to the relevance metric.

The need to build situational awareness from increasingly large sets of textual data requires automatic methods to construct narrative structures from text without regard to domain factors such as actors, event types, etc. The metrics presented in this paper provide a means to assess these narrative structures so that only the most useful narrative structures are transformed into narratives. In this work, we define three metrics of relevance, topic persistence and topic consistency to assess narrative structure. We specify and implement these measures with respect to a narrative structure of story chains generated by an unsupervised narrative generation technique presented in Hossain et al. (2012b). This data is processed to provide analytical evidence for the usefulness of these metrics for identifying high quality story chains.

2.7 Developing a Functional Design for an Advanced-Capability Prototype

An effective approach to architecting our proposed decision-support concept requires that we assert our views of the overall reasoning process from evidence to decision-making and decision enablement. Most traditional characterizations describe decision-making (DM) as contemplative, analytic, involving nomination and evaluation of options that are weighed in some context, eventually leading to a choice of a "course of action (COA)". This model, often labeled as the "System 2" model, can be seen in most descriptions of the "Military Decision-Making Process" or MDMP as for example in published military Field Manuals such as in HQ, Dept of Army (2010). The literature also identifies a "System 1" or largely intuitive decision-making paradigm (IDM) that operates in conjunction with System 2 processes in what is argued to be an improved DM process model, often called the "Dual-Process Model". Most research in decision support however has focused on System 2 DM ideas since this model is quantitative and can be mathematically studied using notions of utility theory and other frameworks for mensuration. We intend however to factor the Dual-Process Model concept into our systemic design approach; the basis of this rationale cannot be elaborated here but we offer our references for the interested reader, e.g., Croskerry (2009) and Djulbegovic et al. (2012).

Furthermore, in our view of the System Support context for DM, we see what today are called Sensemaking processes, as lying between automated System Support capabilities, such as Data Fusion processes and DM processes, in a stage wherein "final" situation assessments and understandings (in the human mind) are developed. Thus, our view of this meta-process is as a three-stage operation: System Support (SS) as an automated process that nominates algorithmically-formed situational hypotheses (such as from the combined operations of data fusion and argumentation), followed by human-computer, mixed-initiative processes for Sensemaking and symbiosis, whose narrative-type products provide the vetted situational assessments needed for decision-making. There is a substantive literature on Sensemaking, such as those previously cited (Llinas 2014a, b; Gross et al. 2014). Our key thoughts on and rational for the meta-architecture for System Support described briefly here have been summarized in (Llinas 2014a, b). Finally, in the face of significant production pressures and rapidly proliferating data availability—and the resulting data overload deluging the professional analyst—it is increasingly easy for analysts and decision-makers to be trapped by shallow, low-rigor analysis; improvements in rigor have been previously discussed and are part of our proposed design. At the highest level, and consistent with the System Support/Fusion—

Sensemaking–Decision-making interdependent processes concept, we see our initial prototype as embedded in the Sensemaking dynamic (note that this is an initial, design-in-process), as shown in Fig. 2.10.

Building on these ideas, we formed our initial functional design as shown in Fig. 2.11. Included in this design are the specifics of the Hard-Soft data association operations that would be part of the Fusion/System Support segment in an eventual final design. The figure can be examined by starting at the bottom where notional Use Cases are also shown—these include current service-specific mission operations, Joint service operations, and a technological type thrust that examines the proposed methods as having disruptive properties:

- **Army: Operations in Megacities, Syrian Civil War**
  - **Megacity operations are an evolving new Army interest**
- **Navy: Piracy (NATO), Autonomous ISR Systems**
  - **Piracy is a continuing NATO interest, ONR has considerable interest in UAV/UXV operations**
- **Joint: Expeditionary Operations (Anti-Access Area Denial, A2AD),**
  - **Joint operations dealing with A2AD issues are an evolving widespread interest**
- **Assess Hybrid Argumentation Technology as Disruptive**
  - **And of course these proposed methods can be studied from the technological point of view as a new and disruptive capability**

For any Use Case, we envision that there would be the opportunity or need to enable both Hard and Soft data stream inputs of various types as peculiar to each of the Use Cases. Based on our own research in computational support techniques for Relevance filtering and Provenance accounting, we show those two functional blocks first, operating on both data streams. (Note that there may be some preprocessing required for the Hard Data stream to frame the results into Entity-Attribute sets.) These filters ideally provide relevant and qualified data to two processes: a Natural Language Processor (NLP) and Argument Detection and Nomination (ADM) process. The functions of each of these operations are:

- NLP: extract Named Entities and associated features and attributes of those Named Entities
- ADM: detect and construct argument phrases with labeled Schemas as possible

Metadata is also considered for both processing operations. The outputs of both NLP and ADM (and possible Hard Data preprocessing) are inputs to the Hard/Soft Data Association process that correlates the Entity-Attribute sets and forms the associated and reconciled fused Entity/Attribute results, i.e., the associated, fused Entity/Enriched Attribute evidential data set as shown on Fig. 2.11. This output provides a feedback to the Argument Detection processing (that contains labeled

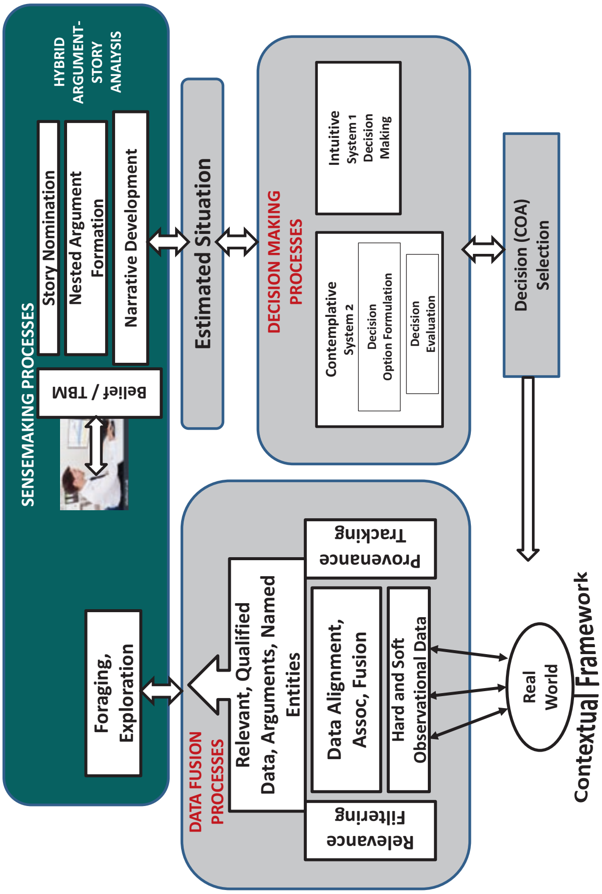


Fig. 2.10 The hybrid scheme in the context of a meta-architecture involving fusion-sensemaking-decision-making (Llinas 2014a, b)

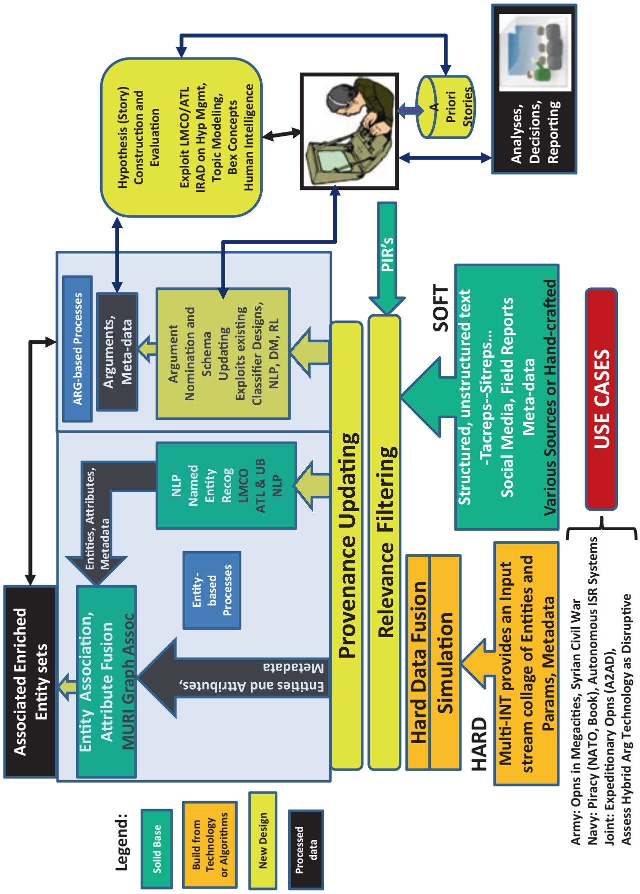


Fig. 2.11 The final functional design

Entities) so that these identified Entities can be enriched with the associated/fused Attributes. Note that there can be possible outlier Entities here, since the ADM process is only Soft-data-based; this is a reconciliation issue yet to be determined. One idea is to engage the human analyst in the process of integrating and managing these outlier Entities. At this point, this front-end processing has *automatically* produced nominated arguments with associated and enriched/fused Entity/Attribute pairs—this capability is a high-priority goal of our approach as this capability has the potential to greatly reduce human cognition workload in terms of argument construction, a major issue even in the most modern prototypes we have reviewed. These nominated arguments then are vetted with analyst intervention and once vetted can provide draft input to our proposed Topic Modeling/Narrative Construction software that aids in a mixed-initiative, human-machine symbiotic process of hybrid argument/story combination. This approach takes into account the uncertainty inherent into the environment as well as the results of argument detection and nomination. These operations will likely involve the management of competing hypotheses for which Lockheed Internal Research and Development (IRAD) software may also provide automated support. These operations would take advantage of Bex's theories and methods for hybrid correlation of the evidentially-grounded arguments and stories emanating both from the analyst and from the Topic Modeling story-nomination process.

This is of course an ambitious vision but is one that sets a new milestone we think for automated support to intelligence analysis. A number of details have to be worked out but the considerably advanced capabilities that a system like this can provide will move the bar forward in terms of revolutionary, disruptive automated support to intelligence analysis.

2.7.1 Looking Ahead: Possible Test and Evaluation Schemes

Given that our end-goal of this project was to develop initial thoughts on a functional design, it was considered necessary to explore possible strategies for Test and Evaluation (T&E) as well as possible metrics for evaluation, since the quality of any possible prototype would be measured by some appropriate T&E approach.

There are various important functions in the proposed top-level design of Fig. 2.11. As the multisource Data Association process is considered key in any Information Fusion process, one critical aspect of a T&E approach would suggest a scheme for evaluating Hard-Soft Data Association. Here, we would suggest the approach of the MURI program that the Center for Multisource Information Fusion at the University at Buffalo developed as at least a starting approach (this is well-documented in Gross et al. 2014; Date et al. 2013a, b); this technique was explored and tested with good success on that program.

Testing of Natural Language Processing (NLP) methods is a very broad topic but one focus for the proposed design in on Named Entity extraction, a key capability for good performance in the proposed scheme. Here too the methods employed on the prior MURI program could be applied to evaluate performance in any Use Case application; these techniques are discussed in Shapiro (2012).

There is not much literature on specific evaluation techniques for the various front-end argument detection/construction methods we would intend to explore, but most of these rely on some type of classification framework, and evaluation of such text extraction methods. The cited literature of Sect. 2.6, along with various survey papers on classifier evaluation form an adequate starting point for developing an evaluation approach.

Evaluating the quality of argument constructs is an area where there is considerable literature. There are various websites on this topic (e.g., http://www.csuchico.edu/~egampel/students/evaluating.html) and a wide variety of papers that address this topic, e.g., Corner and Hahn (2009). Much of the literature discusses notions of argument strength, different for deductive, inductive, and abductive arguments and introduces related ideas on validity of premises and other issues. This literature is helpful toward test planning but we prefer Dahl's ideas on the notion of argument persuasiveness that in turn relates to ideas on "explanatory coherence" as a technique for evaluating the persuasiveness of arguments; see Dahl (2013), Thagard (2000), and Ng and Mooney (1990).

Of course, the best evaluation approach would reveal the impacts of these combined technologies on mission-based analysis effectiveness; however, since the proposed design and suggested methods are, in our opinion, still at the formative stage, much testing and evaluation would have to be done to first establish technological credibility before mission effectiveness assessments could (or should) be carried out.

Acknowledgement This publication results from research supported by the Naval Postgraduate School Assistance Grant No. **N00244-15-1-0051** awarded by the NAVSUP Fleet Logistics Center San Diego (NAVSUP FLC San Diego). The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

References

Schlacter, J., et al. (2015), Leveraging Topic Models to Develop Metrics for Evaluating the Quality of Narrative Threads Extracted from News Stories, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015 Procedia Manufacturing, Volume 3

Allen, J. (1995), *Natural Language Understanding* (2nd ed.). Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA

Andrews, C. and North, C. (2012), "Analyst's Workspace: An Embodied Sensemaking Environment For Large, High-Resolution Displays", *Proc. 2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Seattle, WA.

Bex, F., S. van den Braak, H. van Oostendorp, H. Prakken, B. Verheij, and G. Vreeswijk (2007a), "Sense-making software for crime investigation: how to combine stories and arguments?," *Law, Probability and Risk*, vol. 6, iss. 1-4, pp. 145-168.

Bex, F, et al. (2007b), Sense-making software for crime investigation: how to combine stories and arguments?, *Law, Probability and Risk*.
Bex, F. (2013) Abductive Argumentation with Stories. ICAIL-2013, in: *Workshop on Formal Aspects of Evidential Inference*, 2013
Bier, E.A., Ishak, E.W., and Chi, E. (2006), "Entity Workspace: An Evidence File That Aids Memory, Inference, and Reading", In *ISI*, San Diego, CA, 2006, pp. 466-472.
Blei, D.M., Ng, A. Y., and Jordan, M. I. (2003), "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022.
Corner, A. and Hahn, U. (2009)., Evaluating Science Arguments: Evidence, Uncertainty, and Argument Strength, *Journal of Experimental Psychology Applied* 15(3):199-212.
Croskerry, P. (2009), *A Universal Model of Diagnostic Reasoning*, Academic Medicine, Vol 84, No 8, pp1022–8.
Dahl, E. (2013), *Intelligence and Surprise Attack: Failure and Success from Pearl Harbor to 9/11 and Beyon*d, Georgetown University Press.
Date, K., Gross, G. A., Khopkar, S, Nagi, R. and K. Sambhoos (2013a), "Data association and graph analytical processing of hard and soft intelligence data", *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey
Date, K., GA Gross, and Nagi R. (2013b), "Test and Evaluation of Data Association Algorithms in Hard+Soft Data Fusion," *Proc.of the 17thInternational Conference on Information Fusion,* Salamanca, Spain
Djulbegovic, B., et al. (2012), Dual processing model of medical decision-making, BMC *Medical Informatics and Decision Making*, Vol. 12
Faloutsos, C. KS McCurley, and Tomkins A. (2004), "Fast discovery of connection subgraphs." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* 22: 118-127.
Feng, V.W. and Hirst, G.. (2011), Classifying Arguments by Scheme, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 987–996, Portland, Oregon.
Gordon, T.F. (1996), Computational Dialectics, In Hoschka, P., editor, *Computers as Assistants - A New Generation of Support Systems*, pages 186–203., Lawrence Erlbaum Associates.
Gross, G., et al. (2014), Systemic Test and Evaluation of a Hard+Soft Information Fusion Framework; Challenges and Current Approaches, in: *"Fusion2014," International conference on Information Fusion*,
Haenni, R. (2001) Cost-bounded argumentation, International Journal of Approximate Reasoning, 26(2):101–127.
Hastings, A.C. (1963), *A Reformulation of the Modes of Reasoning in Argumentation*, Ph.D. dissertation, Northwestern University, Evanston, Illinois.
Headquarters, Dept of Army (2010), Army Field Manual 5-0, The Operations Process
Hossain, MS, M Akbar, and Nicholas F Polys (2012a). "Narratives in the network: interactive methods for mining cell signaling networks." *Journal of Computational Biology* 19.9:1043-1059.
Hossain, MS., et al, (2012b), Connecting the dots between PubMed abstracts PloS one 7.1
Hossain, M. S., Butler, P., Boedihardjo, A. P., and Ramakrishnan, N. (2012c). Storytelling in entity networks to support intelligence analysts*, Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*.
Klein, G., et al. (2006), Making Sense of Sensemaking 2: A Macrocognitive Model, *IEEE Intelligent Systems*, Volume:21, Issue: 5.
Kumar, D., Ramakrishnan, N., Helm, R. F., and Potts, M. (2008), Algorithms for storytelling, *IEEE Transactions on Knowledge and Data Engineering*, 20(6), 736-751.
Llinas, J. (2014a), Reexamining Information Fusion--Decision Making Inter-dependencies, in *Proc. of the IEEE Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, San Antonio, TX.
Llinas, J, Nagi, R., Hall, D.L., and Lavery, J. (2010), "A Multi-Disciplinary University Research Initiative in Hard and Soft Information Fusion: Overview, Research Strategies and Initial Results", *Proc. of the International Conference on Information Fusion*, Edinburgh, UK.

Llinas, J. (2014b), A Survey of Automated Methods for Sensemaking Support, *Proc of the SPIE Conf on Next-Generation Analyst*, Baltimore, MD

Mani, I. and Klein, G.L. (2005), Evaluating Intelligence Analysis Arguments in Open-ended Situations, *Proc of the Intl Conf on Intelligence Analysis*, McLean Va.

Mochales, R. and Moens, M. (2008), Study on the Structure of Argumentation in Case Law, *Proceedings of the 2008 Twenty-First Annual Conference on Legal Knowledge and Information Systems: JURIX 2008*

Mochales-Palau, R. and Moens, M. (2007), Study on Sentence Relations in the Automatic Detection of Argumentation in Legal Cases, *Proceedings of the 2007 Twentieth Annual Conference on Legal Knowledge and Information Systems: JURIX 2007*

Moens, M., et al (2007), Automatic Detection of Arguments in Legal Texts, *Proceedings of the 11th international conference on Artificial intelligence and Law*.

Moens, M. (2013), Argumentation Mining: Where are we now, where do we want to be and how do we get there?, *FIRE '13 Proceedings of the 5th 2013 Forum on Information Retrieval Evaluation*.

Ng, H.T., and Mooney, R.J. (1990), On the Role of Coherence in Abductive Explanation, in *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI-90)*

Pirolli, P. and Card, S. (2005), The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, *In Proceedings of 2005 International Conference on Intelligence Analysis* (McLean, VA, USA, May, 2005). pp.337-342, Boston, MA.

Reed, C. and Rowe, G. (2004), ARAUCARIA: Software for Argument Analysis, Diagramming and Representation, *International Journal on AI Tools* 13 (4) 961–980.

Schum, D. (2005), Narratives in Intelligence Analysis: Necessary but Often Dangerous, *University College London Studies in Evidence Science*.

Shahaf, D., and Guestrin, C., (2010), Connecting the dots between news articles, *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*: 623-632.

Shahaf, D, Guestrin, C, and Horvitz, E. (2012) Trains of thought: Generating information maps." *Proceedings of the 21st international conference on World Wide Web*: 899-908.

Shahaf, D. et al (2013), Information cartography: creating zoomable, large-scale maps of information. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*: 1097-1105.

Shapiro, S. (2012), Tractor: Toward Deep Understanding of Short Intelligence Messages, University Seminar, available at: http://studylib.net/doc/10515245/tractor-toward-deep-understanding-of--short-intelligence-m., 2012

Simari, G and Rahwan, I (2009) *Argumentation in artificial intelligence*, Springer.

Smets, P. (1994), The transferable belief model, *Artificial Intelligence*, Volume 66, Issue 2, Pages 191-23.

Stasko, J., Gorg, C., Liu, Z., and Singhal, K. (2013), "Jigsaw: Supporting Investigative Analysis through Interactive Visualization", *Proc. 2007 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Sacramento, CA.

Suthers, D. et al, (1995),, 'Belvedere: Engaging students in critical discussion of science and public policy issues', in *AI-Ed 95, the 7th World Conference on Artificial Intelligence in Education,* pp. 266–273, (1995).

Thagard, P. (2000), *Probabilistic Networks and Explanatory Coherence*, Cognitive Science Quarterly 1, 93-116

Toniolo, A., Ouyang RW, Dropps T, Allen JA, Johnson DP, de Mel G, Norman TJ, (2014), Argumentation-based collaborative intelligence analysis in CISpaces, in Frontiers in Artificial Intelligence and Applications; Vol. 266, IOS Press

Twardy, C. (2004): Argument maps improve critical thinking. *Teaching Philosophy* 27 (2):95--116

van den Braack, S. W. et al, (2007), AVERs: an argument visualization tool for representing stories about evidence, *Proceedings of the 11th international conference on Artificial intelligence and law*, Stanford, CA.

van den Braack, S. W. (2010), *Sensemaking software for crime analysis*, Dissertation, Univ of Utrecht, Holland.

van den Braack, S.W., et al (2006), A critical review of argument visualization tools: do users become better reasoners?, *ECAI-06 CMNA Workshop*, 2006

Walton, D., Reed, C., and Macagno.F. (2008), *Argumentation Schemes*, Cambridge University Press.

Walton, D. and Gordon, T.F. (2012), The Carneades Model of Argument Invention, Pragmatics & Cognition, 20(1), Web 2.0, Amsterdam, The Netherlands

http://www.springer.com/978-3-319-59718-8

Autonomy and Artificial Intelligence: A Threat or Savior?
Lawless, W.F.; Mittu, R.; Sofge, D.A.; Russell, S. (Eds.)
2017, XIV, 318 p. 102 illus., 86 illus. in color., Hardcover
ISBN: 978-3-319-59718-8