

Integrated Knowledge Management for Early Warning and Risk Assessment of Complex Humanitarian Emergencies

by Krassimir Krastev

Information about world events resulting from *intense social interactions* on the political and economic arenas is now abundantly available due to technology and involvement providing extensive media coverage.

Instantaneous visibility of disasters and conflicts on television has made it less and less acceptable either to do nothing or to offer only a confused effort in emergency assistance (Tanguy).

In this aspect the Information Age has brought about greater chances for survival to people caught in the midst of crisis or disaster but also time of ever increasing responsibility for governments and institutions that influence or neglect the policies inflicting conflict and economic collapse.

Research in the Social Sciences is long set to study the processes and forces motivating collective and individual behaviors. Statistics has introduced powerful tools for assessment of ongoing situations, learning from experience and forecasting of trends. Decision analysis, the "game played by the experts against nature" (Korsan) has now become a popular profession engaging human intellect in search for models and solutions to problems from the real world. Numerous methods have been proposed, many of them "tested" with dubious success. The purpose of this paper is to review the method of Integrated Knowledge Management for Early Warning and to introduce the reader to an existing implementation of this approach by the VRA Knowledge ManagerTM.¹

¹ The flagship software system developed by the Virtual Research Associates.

Our discussion will begin with a short introduction to the idea of Knowledge Management with relevance to the domain of Complex Humanitarian Emergencies. Next we shall explain coding for political event data and summarize the problems that occurred with this approach prior to the introduction of automated tools. The most essential part of the text will focus sequentially on the different stages of the KM method and will give examples of how these have been implemented by an existing Early Warning system. The VRA Knowledge Manager™ will be discussed, the technologies involved as well as some of the design decisions that have been considered. Throughout analysis and suggestions for improvement will be offered as appropriate. Ultimately the author shall seek to provide enough detail so that the reader can appreciate the power and universality of Knowledge Management as a general approach to problem solving..

Belligerent behavior among social clusters is suspected to arise due to power transition cycles that occur throughout the course of time (Doran). What is the precise dynamics of these cycles, why do they occur and how we can control them to our own advantage are all questions, answers to which we are yet to find. Integrated Knowledge Management for Early Warning is a method for gathering analysis and presentation of information that might provide critical insight into a social process preceding conflict or economic collapse. It relies on tools for automated development of data, customizable manipulation and forecasting technology as well as human expertise for risk assessment, strategic planning and policy prescriptions. The short-term objective of the method is to supply real-time operational decision support data for risk assessment and logistic

guidance in CHEs. The long-term goal is to engineer datasets for scientific modeling and studying of the power transition cycle dynamics.

The method is currently being operationalized through the development of automated software tools by the Virtual Research Associates, Inc. The leadership team of VRA thrives on 10-years research activity in the area of automated coding for political event data headed by Dr. Douglas George Bond at the Weatherhead Center for International Affairs, Harvard University. *and sponsored by the COE.*

In "Timely Conflict Risk Assessments and the PANDA Project" Doug Bond provides an overview of the Conflict mapping process as performed by the PANDA protocol. In particular he discusses the initial implementation of the protocol by the Kansas Event Data System. Robin Hayden in vol. I, no.3 of the Liaison magazine authors an article "Asia Specific Information System." This text discusses the Knowledge Manager™- the flagship software product developed by VRA that improves on the idea of streamlined data development for Early Warning and assessment. The article presents also a few screen shots from the KM visualization module. The PANDA P24 Code Book available at <http://data.fas.harvard.edu/cfia/pnscs/DOCS/contents.htm> is an excellent resource covering the conceptual framework as well as technicalities of PANDA. Again discussion of the protocol is with reference to the implementation offered by the KED System. The "VRA Knowledge Manager Developer's Manual" is the first and first-hand source on protocol development issues for the VRA Early Warning System.

Automated coding for political event data has been around for several decades now. Human coders for the most part, consulting a codebook, examined the texts of journal and newspaper articles in order to distill statistically relevant information with reference to a predefined modeling paradigm² into code sheets according to the rules of a transformation protocol. This process, noun as event mapping, has been in one form or another attempt to systematically represent the course of history into event records of the form Who did What to/with Whom When Where How and Why? Historically this formula has been derived from the so prescribed “good” practices of journalism that investigation and report of an event should seek to provide explicit answers to these seven questions. Because of their completeness and inherent orthogonality six of them have conveniently outlined the set of explanatory variables for the KM method which ultimately seeks the Why variable value for an event of interest, for example war. The Why variable value for each event is thus the causal chain of implications that could be traced through the entire knowledge base provided values for all other independent variables have been filled in. In an ideal world that may be the case but gleaning data from news text in practice results in sparse event matrices. Human coders resort to inference in order to fill the missing values and that eventually leads to inconsistencies of data in the sets due to coders’ diverse backgrounds and consequently varied modes of interpretation. Human-based event coding was too expensive and too slow. Development of data never occurred in real time and consequently rapid assessment has not been feasible.

why they
are orthogonal

yes,
but the
“why” also
be natural
of the
as
topic
or
issue in
contention

² Implemented by the rules of the codebook.

The Integrated Knowledge Management approach proposed by VRA seeks to overcome the deficiencies of human coding by streamlining the stages of data mining, relegation and representation by means of a single production-scale data development system. The data acquisition facility of this production system has been conveniently set similar to the now increasingly popular survey tools in the form of a web-based application providing the interface for rapid structured data entry to field reporters from around the world. This approach is by far the fastest way available to bring data into operational state so that statistical research can be applied to it. In the future this application may be ported to the PalmTop platform so that accessibility to terminals in disaster situations be not an issue.

Alternatively VRA have developed a full syntax natural language parser to take advantage of the vast availability of news reports from online services such as the Reuters Business Briefing. The later has been selected as the preferred source because of reports format consistency and extensive topical coverage.

Conceptual navigation though the lexicon space has only now become possible due to availability of AI technology implementing dictionaries as networks of semantic relationships. Such tools forge "the passage from computing with numbers to computing with words." (Kohl et al.) The VRA parser is capable of "calculating" semantic categorizations at the clausal level. This approach results in data sets that register the greatest number of events per volume of text examined. Relegation of the semantic information is handled by a customizable protocol. Reprogramability of the

transformation phase is an extremely useful and powerful facility which has been provided to the analysts so that she can tailor the data development process to the requirements of specific research exercises. In this way scientific theories can be extensively tested against curves from the real world. *both historically and predictively*

The function of the front end of the Knowledge Management System is to process information into an intermediate structured representation so that data becomes manipulable by statistical and machine learning algorithms. One interesting line of investigation would be to explore possibilities for mapping report sentences directly into a logic knowledge base so that inference queries could be executed against the predicates by means of an automated back-chaining theorem prover such as the Prolog interpreter.

The analysis and presentation phases constitute the back end of the Knowledge Management process. Output from this side is what constitutes the input to the analyst. Currently the VRA Knowledge Manager is equipped with a visualization module that supports diverse but limited number of views of the underlining data sets. A truly customizable research environment such as the Mathematica Frontend by Wolfram Research, Inc. would better meet the need for user programmable simulations and experimentations to be carried out. Multimedia capabilities can add high-impact to presentations and help the presenters communicate their findings in a seamless and impactful manner. Not in the last place a high-level programming language such as the Mathematica Programming Language can boost creativity and productivity by allowing

for ever faster prototyping of the novelties and effective capturing of the researcher's know-how that is being developed along the way.

For the past several years of computational linguistics research, the VRA have become increasingly good at the data acquisition stage of the Knowledge Management process. In authors view, the VRA NL parser has no competing analog on the market of linguistic tools. Improving the backend of the research environment has now become a hot area of concentration. Along this line of development, options should be considered for integration with existing statistical packages such as Statistica and SPSS to leverage on analysts existing skills and wide availability of statistical tools based on these environments. The author himself is presently engaged in this line of activities investigating integration options with the R-language environment for statistical research and the BUGS compiler due to the increasing interest in Bayesian Inference as one alternative and powerful decision-support methodology.

The VRA Knowledge Management System is currently undergoing field-testing as an integral part of the ASIA project at the COE in Disaster Management and Humanitarian Assistance at Tripler Army Medical Center, Hawaii. This project aims at monitoring the political stability of regimes in the Asia-Pacific region. The cost of the method essentially amounts to the price of the software tools and support as well as the subscription fees for newswire services like the Reuters Business Briefing which costs approximately half a million dollars³; also hardware and not in the last place domain experts and statisticians to operate the tools and verbalize the policy prescriptions.

³Information has been supplied by the VRA staff.

Installations are presumably affordable to government-related or academic organizations. Given the cause and the potential benefits of saving human life or alleviating human suffering, reducing the costs of intervention operations through well advised allocation of resources, the practical utility of IKM is by all means confirmed, although achievement so far can be only claimed "in the anecdotal sense."⁴ The technology and the experience are in their infancy, yet for the first time now they have advanced the productivity factor of the event coding process to effective rates such that statistical inference begins to work. Presumably, modeling in the area of Social Sciences has never been an easy job given the typical length of the time intervals until result from an experiment can be observed.

Reliability in any data development process is oftentimes an issue concerning origin, respectively sources. The sheer fact that in parsing the news feeds, we are largely dealing with just reporter's view of the historical event, not to mention the wide editorial interference aiming to "standardize" the language that is being used, legitimately rises a discussion of whether the data developed from this interpretations and any classifications concluded afterwards, reflect the actual significance of that "historical" event? Although study of the editorial process could theoretically provide guidelines for adjustment of relegation rules, event coding claims no validity of the underlying conceptual approach. Automated event coding though can guarantee consistency, i.e. reproducibility of the

engineered data sets and the scientific experiment. It can also facilitate the triangulation among multiple readings of competing perspectives.

⁴ Quotation is by Dr. Bond

→ is associated with a unique reading/observation.

of that which is being observed
Indeed, the authors of the reports are treated as a variable as each report.

Among the numerous advantages that KM brings in service to the analyst, automated navigation from trend to detail is the essence and notably the one carrying a revolutionary breakthrough in the investigation search. What really happens in the information-processing aftermath is that the software effectively builds a view from where the researcher can visually detect an outlier and bring up the story that is responsible for the underlying statistic. This is in fact the mechanism that "mines" the data and "discovers" the knowledge!

The information age has brought about new and unusual habitat, the impact of which on the evolution of social interactions is tremendous and yet unknown. The VRA Knowledge Manager is one example of a decision-support system for dealing with the emerging search complexity in the information world. The technology supporting the approach will continue to evolve with our understanding of the problems we attempt to solve. Diverse information formats will certainly drive innovations on the information acquisition side. One interesting application would prepend phonetical analysis on sound waves compiling them into text so that it be digestible by the semantic analyzer. This will allow for telephone and radio channels to be monitored similarly to the newswire feeds. Also computer vision will bring ideas for analysis of imagerial content. We should not ignore the availability of neural interfaces (Charles) that could effectively link to minds and monitor for "evil thoughts." In other words, information available in multiple formats could well inform the Knowledge Management approach as long as tools for data acquisition and relegation be available to handle correspondingly content. Extensive use will be made of machine learning algorithms for automated model generation from

examples and classification feedback. Also the presentation stage will increasingly acquire multimedia capabilities. Animation, for example, can be effectively used for visualization of a dynamic process.

efficiency

Reprogramability of the relegation phase, also known as flexibility at the protocol development level, will certainly trigger the attention of specialist from other areas who wish to tailor the approach and make it appropriate for use in their particular domains of application. Development of economical data from news reports to study market dynamics and early-warn daily or recessional financial collapse is a straightforward and useful application that can engineer investment decisions.

Before we proceed to summarize KM in a way that fits the method in a wider context, we propose to the reader the following two quotations by administrators of two leading businesses of our time.

"An organization's ability to learn, and translate that learning into action rapidly, is the ultimate competitive advantage."

Jack Welch

Chairman, General Electric

"The most meaningful way to differentiate your company from your competition is to do an outstanding job with information. How you gather, manage, and use information will determine whether you win or lose." (Gates)

Bill Gates

COE of Microsoft

Control of the business workflow means bringing the right information to the right person at the right time to solicit the “right” decision. The purpose of organizing knowledge is fast search, i.e. efficient navigation. Object-oriented technology provides a powerful approach to structured representation of knowledge. Knowledge Management is the process of capturing the intellectual know-how for the sake of providing partial or complete automation to human inference activities in the form of software components – reusable entities that implement the so called “best practices.”⁵ The availability of these in reusable format has turned the activity of problem solving into an engineering process known as programming.

Krasin's
This is already done.
I would like to use the
some of the cap on the
in writing.
system On

⁵ or “patterns” also known as the “know-how of objects”: well-understood solutions to well-specified problems.

References

- Charles, John. "Neural Interface Link the Mind and the Machine." Computer (1999): 16-18.
- Doran, Charles F. "'Why Forecasts Fail: The Limit and Potential of Forecasting in International Relations and Economics'." International Studies Review 1, No. 1 (1999): 11-41.
- Gates, Bill. Business @ the speed of thought: Using a Digital Nervous System: Warner Books, 1999.
- Kohl, Karen T., et al. "Representing Verb Alternations in WordNet." WrodNet. Ed. Christiane Fellbaum. Cambridge, Massachusetts: The MIT Press, 1998. 153-178.
- Korsan, Robert J. "Decision Analytica: An Example of Bayesian Inference and Decision Theory Using Mathematica." Economic and Financial Modeling with Mathematica. Ed. Hal R. Vrian. Santa Clara, CA: Telos, 1993. 407-458.
- Tanguy, Joelle. Responding to Complex Humanitarian Crises and Massive Abuses of Human Right: Reflections on the legal, political and humanitarian framework, 1998.