Content Analysis for Proactive Intelligence: Marshaling Frame Evidence

AP Sanfilippo[†], AJ Cowell[†], SC Tratz[†], AM Boek[†], AK Cowell[†], C Posse[†], LC Pouchard[‡]

Pacific Northwest National Laboratory, Richland, WA 99352[†]
{antonio.sanfilippo, andrew.cowell, stephen.tratz, annie.boek, amanda.cowell, christian.posse}@pnl.gov

Oak Ridge National Laboratory, Oak Ridge, TN 37831[‡]
pouchardlc@ornl.gov

Abstract

Modeling and simulation have great potential as technologies capable of aiding analysts in making accurate predictions of future situations to help provide competitive advantage and avoid strategic surprise. However, to make modeling and simulation effective, an evidence-marshaling process is needed that addresses the information needs of the modeling task, as detailed by subject matter experts. We suggest that such an evidence-marshaling process can be obtained by combining natural language processing and content analysis techniques to provide quantified qualitative content assessments, and describe a case study on the acquisition and marshaling of frames from unstructured text.

Introduction

The ability to support accurate predictions of future situations and their potential impacts is a crucial task in enabling decision making in areas as diverse as business, politics, security and scientific discovery. Predictive modeling technologies such as evidentiary reasoning (e.g., Bayesian networks, Dempster Shafer theory) or agent-based simulations can greatly facilitate this task by providing an environment in which analysts can model their knowledge of world events to reason about plausible outcomes, to help provide competitive advantage and avoid strategic surprise. For example, suppose an analyst was tasked with modeling the radicalization of a given political activist group to estimate the propensity of the group to engage in violent behavior. Using insights from Social Movement Theory (Wiktorowicz 2004, Johnston and Noakes 2005), the analyst may choose to include in her predictive model messaging strategies that group leaders adopt in an effort to sway target audiences to their interpretation of events as a way to assess the group's radicalization level. These messaging strategies might include contentious rhetoric about unjust repression and political exclusion by social movement entrepreneurs, since unselective and untargeted repression and exclusion from the political process are known to lead to increased radicalization (Hafez 2003). Needless to say, such an approach is predicated on the analyst's ability to collect evidence relevant to the problem under analysis (e.g., group radicalization), as detailed by subject matter experts (e.g., Wiktorowicz, Johnston and Noakes, Hafez), and establish the probative force of the evidence collected to quantify the uncertainty within the model. Due to massive amounts of potentially relevant information available, these evidence-marshaling activities are difficult to perform manually without introducing unwanted biases. The considerable effort needed to distill the relevant evidence and the difficulty in quantifying its probative force render the task too complex for a human to carry out without appropriate computational support.

Information Extraction (IE) (Appelt and Israel 1999) and Content Analysis (CA) (Krippendorff 2004) have separately made substantial progress in providing tools and methods to support the analyst in the evidence-marshaling task. IE technologies enable the automatic recognition of named entities, links and events from unstructured text. CA platforms such as Computer-Assisted Qualitative Data Analysis Software tools (CAQDAS) (www.lboro.ac.uk/research/mmethods/research/software/caqdas.html) support data-theory building through manual annotation, and organization, categorization, quantification and search of evidence in textual and visual data. Yet, IE or CA alone cannot effectively address the analyst's needs in marshaling evidence for predictive modeling. IE tends to be generic, is hard and costly to tailor to a specific domain of application, and is not designed to provide quantitative content assessments. CA relies on word token analytics with untagged text or code-based analytics enabled through manual annotation; so it is either cost-effective but shallow, or enables qualitative analysis but in a laborious fashion.

In this paper, we show that IE and CA can be effectively combined to help the analyst extract evidence signatures from unstructured data that provide qualitative and quantitative content assessments in support of the predictive modeling task. We present a specific case study with reference to the extraction and marshaling of frame information and discuss evaluation techniques to monitor both the soundness of the evidence marshaling process and its results. First, we describe a method for annotating frame content, which takes into account the theories of frames and framing developed in the sociological literature during the last two decades. We validate this frame annotation method and then use IE techniques to apply the frame categories developed automatically to a body of documents. We show how CA methodologies available in CAQDAS tools can be used to deliver frame evidence signatures in a cost-effective manner. We provide an evaluation of the frame evidence signatures obtained through this process, and conclude by adumbrating how the entire process can be made yet more efficient through the adoption of semi-supervised IE techniques.

Annotating Frame Content

The objective of Frame Analysis is to understand the process that groups, individuals and the media use to frame specific issues in order to influence or mobilize public opinion. Social movement entrepreneurs and activist groups use frames to process ideas socially (e.g., advocating Shari'ah law against secular government policies) through grammatical constructs (e.g., *Islam is the solution!*) that serve as interpretive lenses creating inter-subjective meaning with the intent of facilitating movement goals (e.g., establish Shari'ah law).

The literature on frames and framing has grown significantly since Goffman's initial exploration of frame analysis (Goffman 1974). Such a surge of interest reflects the increasingly stronger role that frame analysis has come to play across the social sciences within the last two decades (Fisher 1997).

While significant steps have been made in providing a theory of frames and framing (Gamson 1992; Benford and Snow 2000; Entman 2004; Johnston and Noakes 2005), a formal characterization is still largely lacking and there is no recognized set of criteria that can be used to marshal frame evidence reliably (Fisher 1997). Existing approaches to frame annotation and extraction, while producing very interesting results (Miller 1997; Koenig 2004), clearly suffer from the lack of a more systematic approach to frame identification (see section on Related Work below for further comments).

Define and Evaluate Frame Annotation Guidelines

Frames can be seen as cognitive schematas (Rumelhart 1980) that guide social interaction through communication. Following this insight, we propose that a frame be linguistically analyzed as a speech act, i.e., an act of making an utterance (Austin 1962; Searle 1969, 1979) that:

- conveys a particular **intention** in making the utterance–the illocutionary force (Austin 1962, Searle 1969, 1979) of the speech act
- identifies a frame **promoter**, i.e., the person or organization that functions as the speaker
- may identify a frame **target**, e.g., the person or organization to blame for grievances
- specifies one or more **issues**, i.e., referents and predications that form the propositional content of the speech act.

These four categories capture the way frames have been characterized in sociology and political science. The notion of a frame **promoter** is used by Benford and Snow, corresponds to the result of Gamson's *identity* frame function, and overlaps with Entman's notion of *actor*. Frame **intention** is implicit in the framing task classification provided by Gamson (injustice, identity, agency) and Benford and Snow (diagnostic, prognostic, motivational). Frame **target** partially overlaps with the result of Gamson's *injustice* frame function. The category **issues** is used by Entman.

In order to reduce subjective uncertainty in the assignment of frame annotations to text segments, we provide detailed annotation guidelines for each of the four frame categories selected. For example, the code **promoter** is characterized as a word or phrase denoting an individual, a group of individuals, or an organization, that occurs as the subject of the speech act conveying the intent of a frame. The code **intention** is broken down into 13 sub-codes encoding speech act classes. For each intention sub-code, we provide definitions, examples, and a list of distinct lexical realizations derived from WordNet (wordnet.princeton.edu), as shown in Table 1.

Code	criticize
Definition and examples	accuse, blame for: make a claim of wrongdoing or misbehavior against; "he charged the director with indifference". ...
Lexical realizations	accuse, accurse, blame, calumniate, charge, condemn, criticize, defame, denigrate, denounce, ...

Table 1: Example of an **intention** sub-code definition.

We distinguish nine types of **issues–economy, politics, social, law, military, administration, environment, security, and religion–**using WordNet Domains (Bagnini and Cavaglià 2000) as the reference lexical resource.

We validate the reliability of our annotation scheme through inter-rater agreement, utilizing Fleiss' kappa test (Fleiss et al. 2003) to measure the degree of overlap across subjects assigning frame codes to text segments. To obtain inter-rater data, we trained four subjects in the use of the frame annotation guidelines developed, and asked each subject to correct frame annotations which had previously been assigned to the same thirty documents by an automated annotator. The automatic assignment of frame annotation is described in the next section. At present, it will suffice to say that the choice of using automatically annotated texts was motivated by the need to make manual annotation more cost-effective and less prone to individual biases. We used the CAQDAS tool Qualrus (www. ideaworks.com) to facilitate the manual correction process.

We worked with the vendor to develop xml-based capabilities which made it possible to import automatically annotated texts into the Qualrus environment.[1] Annotation is always manually managed in CAQDAS tools, and this is the first time that a CAQDAS tool has been enabled with automatically annotated functionality, to the best of our

[1] We would like to thank Ed Brent and Ted Carnahan at *Idea Works Inc.* for working with us on adapting Qualrus to the purposes of this research.

knowledge. We also added to Qualrus facilities to enable the calculation of kappa scores.

Table 2 shows the kappa scores resulting from comparing frame annotations across two sets of raters: four humans, and the same four humans plus the automated annotator described in the next section. Given the total number of ratings (1433), codes (27), and raters (4 and 5), the kappa scores obtained for all codes (0.499 and 0.422) strongly support the reliability of the annotation system developed. The z-scores (46.2 and 50.5) and p-values (0) for the kappa scores indicate the high statistical significance of these results.

A few caveats are in order to help the reader better understand the context of this evaluation. First, annotations were assigned to sentences, rather than to sentence components. While the automatic annotator has the ability to annotate sentence components as shown in Figure 1, we simplified the annotation procedure in the evaluation phase to ease the rating process. Secondly, we only took into account sentences that were annotated by all raters. This leaves out some 300 code rating events, for which groups of less than four raters made a code selection. Still, even with these limitations, the results in Table 2 prove that both manual and automated frame annotation is feasible.

	4 human raters			4 human raters + computer		
Codes	kappa	z	p-value	kappa	z	p-value
ACCEPT (6)	0.259	1.56	0.12	0.233	1.8	0.0712
ADMIN. (72)	0.536	11.1	0	0.467	12.5	0
ASSERT (103)	0.528	13.1	0	0.509	16.3	0
BELIEVE (18)	0.41	4.26	2.07E-05	0.317	4.25	2.14E-05
CRITICIZE (35)	0.433	6.28	3.44E-10	0.475	8.9	0
ECONOMY (19)	0.029	0.314	0.754	-0.085	-1.17	0.244
EMPHASIZE (7)	0.417	2.7	0.00693	0.536	4.48	7.31E-06
ENVIRONMENT (1)	none	none	none	none	none	none
EXPLAIN (8)	0.451	3.13	0.00176	0.57	5.1	3.45E-07
FRAME (231)	0.506	18.8	0	0.474	22.8	0
IMPUTE (3)	none	none	none	none	none	none
INTENTION_ NEGATION (2)	0.467	1.62	0.106	0.2	0.894	0.371
JUDGE (1)	-0.33	0.816	0.414	-0.25	-0.79	0.429
LAW (71)	0.564	11.6	0	0.446	11.9	0
XXX_ PROMOTER (99)	0.438	10.7	0	0.483	15.2	0
XXX_TARGET (4)	-0.33	-1.63	0.102	-0.25	-1.58	0.114
MILITARY (45)	0.555	9.12	0	0.235	4.98	6.30E-07
POLITICS (188)	0.484	16.2	0	0.311	13.5	0
PROMOTER (231)	0.505	18.8	0	0.472	22.7	0
REJECT (19)	0.384	4.1	4.12E-05	0.359	4.94	7.63E-07
RELIGION (51)	0.562	9.84	0	0.439	9.92	0
REQUEST (27)	0.626	7.96	1.78E-15	0.493	8.1	4.44E-16
SECURITY (54)	0.508	9.14	0	0.176	4.09	4.25E-05
SOCIAL (59)	0.458	8.62	0	0.177	4.3	1.69E-05
SUPPORT (15)	0.2	1.9	0.0578	0.241	2.95	0.00316
TARGET (51)	0.405	7.09	1.32E-12	0.529	11.9	0
URGE (13)	0.35	3.09	0.00197	0.462	5.27	1.37E-07
All Codes (1433)	0.499	46.2	0	0.422	50.5	0

Table 2: Inter-rater agreement results. The numbers in parentheses indicate the count of ratings per code.

Automating Frame Annotation

One of the main objectives of the approach pioneered in this paper is to augment Content Analysis (e.g. CAQDAS tools) with Information Extraction techniques. As mentioned in the previous section, this augmentation in our case involved developing xml-based capabilities that made it possible to load the results of automated annotation in the Qualrus CAQDAS tool. These capabilities consisted of an xml schema which mirrored the internal data representation in Qualrus and a mapping function from the xml format into the internal Qualrus data format. The automated frame annotation processes we implemented produced output which conformed to the Qualrus xml-schema. The integration of IE into CA was therefore realized by mapping IE output into the CA tool.

We collected 619 documents from the website of a militant group, which has been outlawed in its own country as a political organization. Our objective was to examine the messaging strategies adopted by the group to gain insights as to the radicalization potential of the group. Consequently, the documents were chosen so as to privilege texts which reported on interviews, comments, issues and opinions. The 619 documents cover a 2-year period, 2005-2006.

We designed and implemented a fully automated frame extraction algorithm and used it to assign the frame annotations described above in the 619 documents harvested from the group's web site. More specifically, for each segment which was recognized as a frame, the IE algorithm establish the presence of a **promoter,** an **intention**, a **target** (when present) and one or more **issues** with probabilities indicating relative prominence, as shown in Figure 1, where the sequence "XXX" replaces sensitive words or phrases.[2]

XXX demands the government recognizes it as a political entity. ⇨

XXX = promoter
urges = intention:request
government = TARGET
government = issues:[admin, 0.25]
legal = ISSUES:[law, 0.25]
political, party = issues:[politics, 0.50]

Figure 1: Sample results of automatic frame annotation.

[2] US Government sponsored the work described in this paper. Therefore, certain data had to be omitted due to the terms of the sponsorship. These omissions do not affect the technical content and expository clarity of the paper.

We used four separate language processing components to analyze the 619 documents: the OpenNLP sentence detection Java package (opennlp.sourceforge.net), the Connexor syntactic dependency parser (www.connexor.com), the LingPipe named entity recognition (www.alias-i.com), a gazetteer-based named entity recognizer developed in-house, and a word domain recognizer based on the approach described by Sanfilippo et al. (2006). These four tools were integrated into a single process using IBM's Unstructured Information Management Architecture (www.research. ibm.com/UIMA/). Table 3 provides a sample output of such an integrated process.

		Parsing	NER	Word Domains
1	XXX	subj:>2	ORG	
2	demands	main:>0		
3	the	det:>4		
4	government	subj:>5	ORG	ADMIN
5	recognizes	obj:>2		
6	it	obj:>5		
7	as	copred:>5		
8	a	det:>11		
9	legal	attr:>10		LAW
10	political	attr:>11		POLITICS
11	party	pcomp:>7		POLITICS

Table 3: Sample output of integrated language processing facilities for frame annotation.

The frame extraction algorithm leverages the tags assigned by the UIMA-based syntactic and semantic analysis process to detect frame categories in sentences, yielding frame annotations such as those shown in Figure 1 through an intermediate syntactic and semantic tagging stage, as shown in Table 3. The following sequence of steps summarizes the main generalities of this algorithm, with exemplifications of specific annotation heuristics.

1. Split each text in the document collection into sentences and process sentences one by one
2. If the sentence's main verb is recognized as the lexical realization of an **intention** sub-code, assign the **intention** sub-code to the verb and proceed to step 3, or else process the next sentence in the queue
3. Find the frame **promoter**, e.g., if the intention verb is in the active voice and its subject is a person, a group or an organization, mark the subject as **promoter**
4. Find frame **target**, e.g., if the **intention** verb takes a sentential complement where the verb is in the active voice and the subject is a person, a group or an organization, mark the complement subject as **target**
5. Find frame **issues** sub-codes and assign a probability to each sub-code.

Using this approach we extracted 4288 frames from the 619 documents selected. Each frame consisted of a sentence structure with annotations for **promoter**, **intention**, **target** (if detected), and **issues,** as shown in Figure 1.

We have evaluated the results of the automatic frame annotation by correlating automatically assigned annotations with manually corrected annotation, using the Fleiss kappa test. The results of this evaluation are shown in the right half of Table 2, where the results of automated annotation are compared with the annotation choices of the four human subjects who participated in the inter-rater agreement study. As discussed earlier, these results indicate both annotation reliability and high statistical significance.

Marshaling Evidence Signatures

Using the results of automated annotation for the 619 documents selected, we created xml project files with automatically coded documents, and loaded these into Qualrus. In addition to code building and annotation capabilities, most CAQDAS tools provides a variety of search facilities that leverage code annotation to make quantified qualitative content assessments. For example, Qualrus allows users to formulate code-based searches in the form of abductive statements and return statistics about the distribution of the codes used to formulate the searches with reference to all the codes within the annotated documents consulted, in addition to provenance information, as shown in Figure 2.

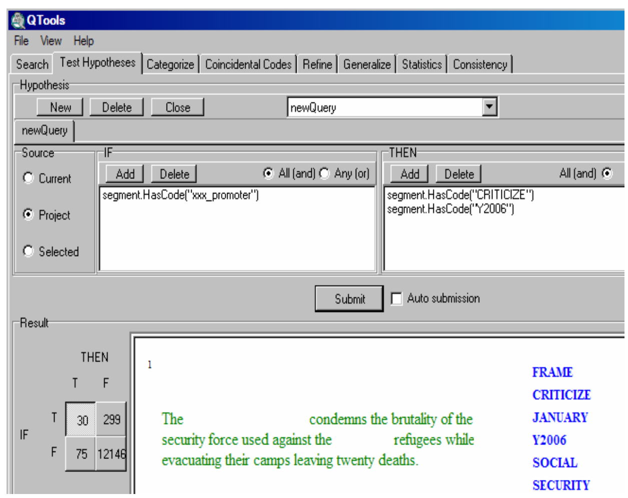

Figure 2: Abductive search in Qualrus.

Using the code-based search facilities in *Qualrus*, we charted the distribution of messaging strategies conveyed by all frame promoters and the frame promoters belonging to the group under analysis, for the two years covered by the data collected (2005 and 2006). Even with our relatively small data set, the results of this analysis yield some very interesting patterns. For example, Table 5 shows that while the overall distribution of frames where the group under analysis acted as a promoter was fairly balanced for the period considered (153 in 2005 vs. 176 in 2006), there is a significant preponderance of **assert**, **criticize** and **request** frames in 2006 vs. 2005, while we find more **support** frames in 2005 than we find in 2006. This distribution suggests that the group in question may have assumed more challenging messaging strategies in 2006 as compared to 2005. This surmise is supported by the distribution of issues relative to **criticize** messaging strategies in 2006. As shown in Figure 3, **criticize** messaging strategies by the group for 2006 are primarily focused on security, social, political and military factors. This suggests that the group is placing a strong emphasis on framing issues about unjust repression and denied political access. This hypothesis is confirmed by manual inspection of the relevant document sources. As we remarked earlier citing work by Hafez (2003), unjust repression and denied political access often occur as precursors of increased radicalization. Whether deliberate or not, these messaging strategies may indeed have the effect of swaying their intended constituencies toward more radical anti-government attitudes.

	ALL PROMOTERS		GROUP PROMOTER	
	2005	2006	2005	2006
ALL FRAMES	1464	2824	153	176
accept	77	210	5	12
assert	167	363	19	35
believe	186	404	11	14
criticize	44	105	4	30
emphasize	29	43	2	3
explain	54	107	4	8
impute	4	10	1	0
judge	31	41	2	2
reject	109	167	8	6
request	136	344	9	26
support	77	173	17	6
urge	24	48	2	5

Table 4: Distribution of messaging strategies.

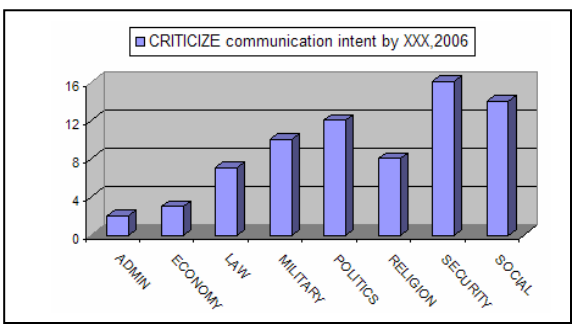

Figure 3: Distribution of issues associated with **criticize** messaging strategies by a specific group **promoter**.

With these content analytics in mind, we proceeded to consolidate the frame evidence marshaled into signature constructs that can be used as input to the predictive modeling task, with specific reference to the radicalization problem discussed in the introductory section. Figure 4 offers a graphic example of frame signatures produced, relative to **criticize** messaging strategies. The signature construct as a whole is identified as a "Diagnostic Anti-System Master Frame" for the group, country and historical context under consideration, following insights from Benford and Snow (2000). It encodes information about the **promoter**, the communicative **intent** conveyed, the **targets** to blame for grievances, the **issues** which constitute the focus of the framing activities with an indication of relative prominence, an estimation of probative force obtained as the proportion of **criticize** frames out of all frames where the group in question is the promoter, plus time, source and provenance information. The same approach can be used to break down a framing strategy into more specific ones (e.g. issue-driven diagnostic anti-systems master frames), represent other framing strategies, or deal with other evidence constructs.

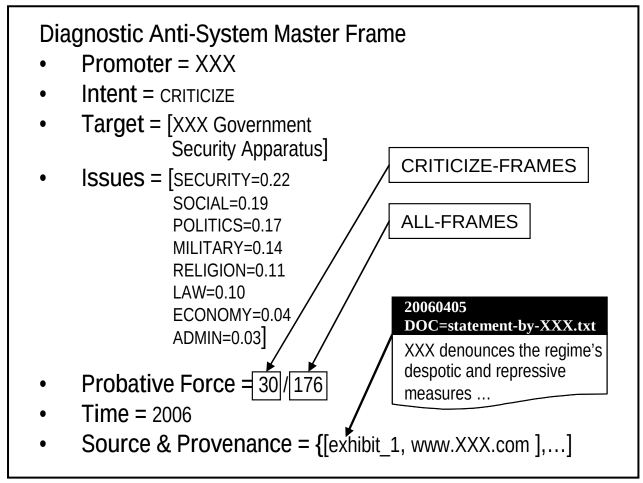

Figure 4: Frame signature.

Related Work

The literature on frame analysis has grown at a very substantial pace, but relatively little effort has been devoted to using computer-aided analysis methods to carry out frame analysis with greater empirical rigor. Miller (1997) and more recently Koenig (2004) are some of the exceptions to this trend and represent the studies most closely related to the work presented in this paper. While differing in scope and focus, the main approach adopted in these studies consists in using *in vivo* (automatic) coding facilities available in CADQAS tools to recognize frames through keyword association. For example, Koenig gathered textual evidence supporting master frames (e.g., "Liberal Individualist Citizenship Rights") from postings in a German internet forum using key terms (e.g. *Freiheit, Sanktionen, Zensor, Grundrechte*) instantiating configurations of concepts (e.g. "freedom of speech, repression, censorship, constitution") forming frames (e.g., "freedom of speech").

These studies have contributed to a more systematic approach to frame extraction, and Koenig's work makes significant advances through the use of techniques such as term lemmatization with thesaurus-assisted term expansion, and the use of exclusion lists to block the inclusion of false positive terms. The approach presented in this paper takes this research a step further by providing a framework which enables: (1) validation of the coding scheme used for frame evidence marshaling; (2) integration of IE techniques to extract more detailed information about frame components and functions; (3) evaluation of frame extraction results, and (4) creation of frame signature for predictive modeling purposes.

Conclusions and Further Work

If frame analysis is to inform predictive modeling, a way to extract frame signatures from relevant document collections must be developed which enables the analysis of messaging strategies and their probative force. We have shown that such a goal can be achieved by using IE to inform CA processes such as those implemented by CAQDAS tools. The approach developed represents a significant evolution of frame extraction work, and can be applied to any evidence-marshaling problem which requires a strong theoretical underpinning in support of the modeling effort for specific domains of application (e.g., group radicalization as a precursor of violent behavior, in our case). The initial evaluation results for the annotation scheme and automated annotation are very promising, and we are currently working to extend the existing testing suites to increase the statistical significance of these results.

Moving forward, one important area of development is the use of semi-supervised learning techniques to facilitate automatic annotation. The approach we are investigating is modeled on the semi-supervised IE approach pioneered by Stevenson & Greenwood (2005), and consists in using iterative expansion of an initial small set of seed patterns to detect relevant patterns from a document collection. Our algorithm differs from previous approaches in that it combines the resource effectiveness of a semi-supervised method with the reliability of a fully supervised approach. More specifically, we utilize the final results of the iterative semi-supervised learning process to build a classification model with both positive and negative training samples (Tratz and Sanfilippo 2007). We have evaluated this approach with the MUC-6 data set (www.ldc.upenn.edu) and obtained results which are 10% better than the current state of the arts: 0.68 F-measure vs. the 0.58 reported by Stevenson & Greenwood (2005). We hope that these results, relative to succession patterns in management change, will extend to the extraction of frame structures.

Another area of enhancement concerns the construction of an ontology-driven knowledge base using the frame signatures produced. We have developed an OWL-DL (www.ksl.stanford.edu/people/dlm/webont/OWLFeatureSynopsisJan22003.htm) frame ontology based on the annotation guidelines and the frame literature consulted. This ontology is being used to organize the frame evidence marshaled in a Knowledge Base (KB). The creation of this KB will enable the user to perform semantic-based searches and inferences which will provide added functionality for the use of frame evidence in the predictive modeling task.

References

Appelt, D. and D. Israel (1999) Introduction to Information Extraction Technology. IJCAI-99 Tutorial. August 2, 1999, Stockholm, Sweden.

Austin, J. L. (1962) *How To Do Things with Words*. Oxford: Oxford University Press.

Benford, D. and R. Snow (2000) Framing Processes and Social Movements: An Overview and Assessment. *Annual Review of Sociology*, Vol. 26, pp. 611-639.

Entman, R. M. (2004) *Projections of power: Framing news, public opinion, and U.S. foreign policy*. Chicago: University of Chicago Press.

Fisher, K. (1997) Locating Frames in the Discursive Universe. *Sociological Research Online*, vol. 2, no. 3.

Fleiss, J.L., Levin, B., & Paik, M.C. (2003) *Statistical Methods for Rates and Proportions*. 3rd Edition. New York: John Wiley & Sons.

Gamson, William A. (1992) *Talking Politics*. Boston: Cambridge University Press.

Goffman, Erving (1974) *Frame Analysis: An Essay on the Organization of Experience*. London: Harper and Row.

Hafez, M. (2003) *Why Muslims Rebel: Repression and Resistance in the Islamic World*. Lynne Rienner Pub.

Johnston, H., and Noakes, J. (2005) *Frames of Protest: Social Movements and the Framing Perspective*. Rowman & Littlefield, Lanham, MD.

Koenig, T. (2004) Reframing Frame Analysis: Systematizing the empirical identification of frames using qualitative data analysis software. Presented at the ASA Annual Meeting, San Francisco, CA, August 14-17, 2004.

Krippendorff, K. (2004) *Content Analysis: An Introduction to Its Methodology*. 2nd ed. Thousand Oaks, CA: Sage.

Miller, M. (1997) Frame Mapping and Analysis of News Coverage of Contentious Issues. *Social Science Computer Review*, 15 (4): 367-78.

Rumelhart, D. (1980) Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, and W. Brewer eds. *Theoretical Issues in Reading Comprehension*, pp. 33-58. Hillsdale, New Jersey: Lawrence Erlbaum.

Sanfilippo, A., S. Tratz, M. Gregory (2006) Word Domain Disambiguation via Word Sense Disambiguation. *Proceedings of HLT/NAACL*. New York, June 5-7, 2006.

Searle, J. R. (1969) Speech Acts. Cambridge: Cambridge University Press.

Searle, J. R. (1979) Expression and Meaning. Cambridge: Cambridge University Press.

Stevenson, M. and M. A. Greenwood. (2005) A Semantic Approach to IE Pattern Induction. *Proceedings of the 43rd Annual Meeting of the ACL*, Ann Arbor, Michigan.

Tratz, S. and A. Sanfilippo (2007) A High Accuracy Method for Semi-supervised Information Extraction. *Proceedings of HLT/NAACL*. Rochester, NY, April 22-27.