MITA: An Information Extraction Approach to Analysis of Free-form Text in Life Insurance Applications

Barry Glasgow and Alan Mandell
Metropolitan Life Insurance Co., Individual Business
1 Madison Avenue
New York, NY 10010
Phone: 212-578-1058, Fax: 212-578-5801 bglasgow@metlife.com

Dan Binney, Lila Ghemri, and David Fisher[1]
Brightware, Inc.
2010 Corporate Ridge, Suite 700
McLean, VA 22102
Phone: 703-749-1462, Fax: 703-749-0554
{binney, ghemri, fisher}@brightware.com

Abstract

MetLife processes over 300,000 life insurance applications a year. Underwriting of these applications is labor intensive. Automation is difficult since they include many free-form text fields. MITA, MetLife's Intelligent Text Analyzer, uses the Information Extraction --IE-- technique of Natural Language Processing to structure the extensive text fields on a life insurance application. Knowledge engineering, with the help of underwriters as domain experts, was performed to elicit significant concepts for both medical and occupational text fields. A corpus of 20,000 life insurance applications provided the syntactical and semantic patterns in which these underwriting concepts occur. The extracted information can then be analyzed by conventional knowledge based systems. We project that MITA and knowledge based analyzers will increase underwriting productivity by 20 to 30%.

Problem Description

MetLife's insurance application is designed to elicit the maximum amount of information relating to the client so that a fair contract can be reached between the client and MetLife. The application contains questions that can be answered by structured-data-fields (yes/no or pick-lists) as well as questions that require free-form text answers.

Currently, MetLife's Individual Business Personal Insurance unit employs over 120 underwriters and processes in excess of 300,000 life insurance applications per year. MetLife's goal is to become more efficient and effective by allowing the underwriters to concentrate on the unusual and difficult aspects of a case while automating the more mundane and mechanical aspects. A ten percent improvement in productivity, while maintaining the already existing high quality of the underwriting processing, and/or an increase in the consistency of the process will have sizable effects.

Use of expert systems to improve the insurance underwriting process has been the holy grail of the insurance industry and many insurance companies have developed expert systems for this purpose with moderate success. A daunting problem has been the presence of free-form text fields.

MITA -- MetLife's Intelligent Text Analyzer -- is an attempt to solve this problem using Information Extraction. Applying this technique to the textual portion of an application allows the automation of underwriting review to a greater extent than previously possible.

MITA Overview

The MITA free form text analyzers read in unstructured text fields, identify any information that might have underwriting significance, and return a categorization of the information for interpretation and analysis for risk assessment by subsequent domain-specific analyzers. By localizing the natural language processing of the input text in MITA, other domain specific analyzers can focus on codifying underwriting domain knowledge.

The text fields analyzed by MITA include:

[1] affiliated with the Natural Language Processing Laboratory, Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts, Amherst, MA 01003-4610, dfisher@cs.umass.edu

- a *Physician Reason* field that describes the reason a proposed-insured last visited a personal physician
- a *Family History* field that describes a proposed-insured's family medical history
- a *Major Treatments and Exams* field that describes any major medical event within the last five years
- a *Not Revealed* field that includes any important medical information not provided in other questions
- an *Occupation Title and Duty* field that describes the proposed-insured's employment

AI in MITA

The Information Extraction(IE) approach of Natural Language processing was chosen for use in MITA. The system was engineered based on a corpus of actual application texts. This approach was intended to provide a IE system optimized for MetLife's insurance application text processing needs.

Information Extraction

Analysis of free-form text has been pursued mainly from three viewpoints:

- a keyword approach
- an in-depth natural language analysis approach
- an Information Extraction approach

The keyword approach, whereby the input text is scanned for words that are deemed highly relevant to the application at hand, has the advantage of being relatively easy to implement. However, it is of limited usefulness for accurate data extraction, which is a requirement of MITA.

An in-depth natural language analysis approach in which the input text is fully and completely analyzed, is usually highly complex, costly and relatively brittle. Furthermore it does not answer the need of an application in which the interest is focused on some parts of the text, but not necessarily all.

Information Extraction [Cowie and Lehnert, 1996; Lehnert and Sundheim, 1991] is an ongoing area of natural language research, where the focus is on real working systems with periodic performance evaluations. IE makes use of Natural Language Processing (NLP) tools, (parsing, tagging, dictionaries), but also uses knowledge engineering or machine learning to identify the concepts contained in the texts and to form a framework of these concepts. In this approach, the input text is skimmed for specific information relevant to the particular application.

To illustrate why IE provides a good match to the problem, consider the text field that describes why the client has last seen a physician. Frequently this field describes a visit for a 'checkup' with additional modifiers for a specific disease, a chronic condition, or to fulfill an occupation or athletic need. A simple keyword search finding 'checkup' would not provide sufficient information. Alternately, full natural language processing with deep semantics is not required because most *Physician Reason* texts could be analyzed in terms of four concepts:

- *Reason*, which usually describes the type of visit such as "regular visit", "school checkup", or "post-partum checkup"
- *Procedure/Treatment*, which describes a procedure or a treatment such as "prescribed antibiotics", "ear wax removed", or "HIV test"
- *Result*, which describes the outcome of the exam or the procedure such as "nothing found", "all ok", or "no treatment"
- *Condition* which describes a medical condition that the applicant has such as "high blood pressure", "ear infection", or "broken leg"

In building MITA, knowledge engineering was used to create a representation of the domain to be analyzed in terms of important concepts. NLP techniques were used to process text and to extract these concepts when they are present in the text.

Recent advances in Information Extraction from text have been well-documented in a series of rigorous performance evaluations sponsored by DARPA called the Message Understanding Conferences (MUC) [MUC-3,4,5,6].

A variety of techniques have been tested at MUCs, ranging from full-sentence syntactic parsers to machine learning algorithms. The majority of the systems tested such as CIRCUS (Lehnert 1990), BADGER (Soderland et. al. 1995) use robust partial parsing followed by pattern matching to achieve information extraction. MITA's design was based primarily on the approach of BADGER.

Description of the Corpus

The corpus used for this application contains some 20,000 applications, each with at least one free form text field. A basic analysis of the sample was performed to determine the range of syntactic forms commonly used. It was found that most sentences were cryptic, and major syntactic constituents such as subjects and verbs were freely omitted. Punctuation was terse and often incorrect. The text was basically a series of noun-phrases clumped together as in: *"c-section childbirth 1979 no complications"*.

Project History

The MITA system was built in three phases, an analysis phase, a design and prototyping phase, and an implementation phase.

Initial Analysis

During the Spring of 1996, a team composed of MetLife underwriters and business analysts and Brightware consultants conducted an analysis of the free-form text fields in MetLife's New Business life insurance applications. The goal was to determine the best way to handle these fields in automating the underwriting initial review process. Factors considered in this analysis of each field included

- the nature and strategy of the question
- the potential underwriting significance and use of the information
- the syntactic and semantic structure of the text.

The resulting recommendations for each field fell into two categories:

- conversion of the free-form questions into structured questions(such as check-boxes and drop-down lists)
- the development of intelligent text analyzers to process the free-form text and convert it into a form that can be used by downstream analyzers

The results also included a list of the categories of potentially significant underwriting information provided by each field. These categories were called *concepts*. For example, for the Occupation analyzer three concepts were found to be of significance : the job *title* as in "waiter", the description of the *duties* as in "serves food and beverages" and the work *environment* as in "restaurant".

Design and Prototype

During the second phase of the project, a general system design was developed, along with a prototype system.

The prototype was based on the Physician Reason field. Physician Reason was chosen because it is a field that is populated on 75% of insurance applications. It also presented a challenge since the medical terminology used was fairly complex.

The prototype and design were developed by three Brightware consultants over a one-month period. Small lexical and medical dictionaries were built as part of the prototype.

Implementation

Based on the success of the prototype, an implementation phase was begun. Text analyzers were to be built for four Medical fields and one Occupational field.

The core project team consisted of seven Brightware consultants working closely with four underwriters. Each underwriter represented expertise from a different area of life insurance underwriting. A number of MetLife system engineers participated in the project as well.

Implementation began in late August 1996 and the completed system was released to system test in January 1997. At the time of this writing MITA is scheduled to be deployed into production in the second quarter of 1997.

MITA Design

The diagram below shows the major components of MITA.

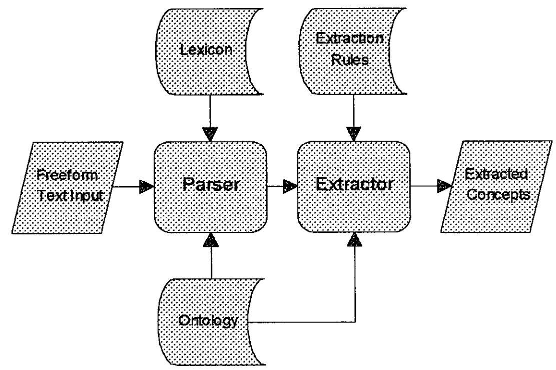

Figure 1. Overview of MITA's Process

A freeform text field is first processed by the Parser which tags the text syntactically and semantically. Next, the Extractor compares the parsed and tagged text to known extraction patterns. When a pattern applies, the extractor *extracts* important words and their classifications and categorizes them into what are called the text's *concepts.*

System Inputs and Outputs

The primary input to MITA is freeform text. An ungrammatical but typical example from a Medical field is the string *"after childbirth applicant suffers from iron deficiency medication"*.

The exact output of MITA is specific to the particular field, but all analyzer outputs follow the same general format:

- **Concept** - a particular type of information that can be found in this field
- **Value** - the actual word(s) in the text which are associated with this particular instance of the concept
- **Class** - the semantic class that the value denotes. Semantic classes are inferred from the ontologies

The output for this example would be:

Concept	**Value**	Class
Condition	*"childbirth"*	PARTURITION-NOS[2]
Condition	*"iron deficiency"*	IRON-DEFICIENCY-NOS

[2] The NOS suffix means "not otherwise specified"

Procedure/ Treatment	*"medication"*	DRUG-NOS

Figure 2. A sample output frame

Major MITA Modules

MITA accomplishes its work through three major modules, including a parser, a set of extraction rules, and a set of dictionaries.

The Parsing Module. The Parser first tags each incoming word with a semantic classification, and with a part-of-speech. It then uses syntactic knowledge to segment the text into noun phrases, prepositional phrases, and verb phrases. These constituents are then grouped together into larger text fragments and assigned the clausal roles of subject, verb, direct object, and prepositional phrase.

The Extraction Module. Extraction rules are then applied to

1. determine the presence of a given concept in the input text being processed, and
2. fill out slots for the value and the class of concept found.

Each extraction rule is an if-then rule which uses the pattern matching mechanism of ART**Enterprise*®. The left hand side of the rule describes a set of constraints on words and word senses found in the syntactic buffers of the parser output. If these constraints are met, specified roles are filled by asserting the facts from the right hand side (the -then part-). A single sentence can generate an arbitrary number of concepts depending on the complexity of the sentence being processed

Dictionaries. MITA relies on several dictionaries. It uses a lexical dictionary containing words in their various inflected forms along with their parts of speech. It also uses a description(*ontology*) of the medical and occupational domain in terms of a hierarchy of classes.

The lexical dictionary has connections to the classifications in the two other ontology dictionaries so that words can be associated with semantic classes.

MITA's dictionary tables contain about 75MB of data. There are approximately 268,000 lexical entries and approximately 135,000 classes, including all domains. There are approximately 4,000 composite classes(these are described later in the paper).

Supportive Modules

An additional set of modules provide critical support to the major modules.

Recognizers. The Recognizers find and normalize nonstandard text contained in the input to make it easier to handle during formal parsing. Some examples of nonstandard text normalized are abbreviations, contractions, and time periods.

Spell Correction Module. The Spell Correction Module uses a dictionary of common misspellings to correct misspelled words in the input. This dictionary was custom-built for MITA and is composed of the words that were commonly misspelled in the corpus.

Multi-word Lookup Module. The Multi-word Lookup Module consists of a set of rules that look for typical word patterns that should be handled as a single term. An example in the medical domain is the phrase "high blood pressure".

Composite Classifier. While Multi-word Lookup handles phrases at the word level, the composite classifier handles them at the class level. This is useful for phrases that hold their useful meaning at a more abstract level.

For example, "tonsillectomy" is defined as a *composite class* in MITA's ontology. It has *participant classes* of "tonsils-nos" and "excision-nos". When the text being analyzed is "had tonsils removed", an extraction rule produces the sequence {"tonsils-nos" "removal—nos"} as a classification. The composite classifier recognizes "removal-nos" as a child of the participant class "excision-nos", and thus *composes* this sequence into the class "tonsillectomy".

Confidence Factor. MITA achieves a high level of performance because it does its best work on the most common texts. However, in the life insurance business, a misinterpretation of text, although rare, can be very costly to the insurer.

MITA achieves its reliability, not because it analyzes text perfectly, but because it can determine when it has done well, and when it has not done well. This prediction of extraction quality is accomplished through Confidence Factors.

The Confidence Factor is derived primarily from the types of words that were *not* extracted from a text. Unextracted words are given penalties based on part-of-speech, and semantic class. Penalties also occur when a spelling correction was performed or when a composite classification was attempted but was unsuccessful.

The higher the confidence factor, the more trustworthy the output. Downstream analyzers decide whether a particular output is reliable to process automatically based on the Confidence Factor. Each field has its own unique confidence factor threshold and its own set of penalty values.

Lexical Coder. Each class within MITA's ontology has a code which is used in subsequent automated underwriting steps. This module lists out the hierarchical path of standard codes starting from the code of the extracted concept itself.

Input and Output Tables

Input Tables. The input to MITA is provided through two tables:

- The Work Queue table which contains a queue of requests waiting to be processed by MITA.
- The Raw Text Table which contains the actual text and field information.

Output Tables. The output from MITA is written to five separate tables all linked together by a common key. These are:

- The Results table which contains the extracted concepts.
- The Confidence Factor table which contains the confidence factor for the concept extraction at a field level.
- The Lex code table which lists all the medical, occupational, or Met codes associated with a particular concept.
- Two Audit tables which contain useful statistics for continued maintenance of MITA.

Hardware and Software

MITA Platform

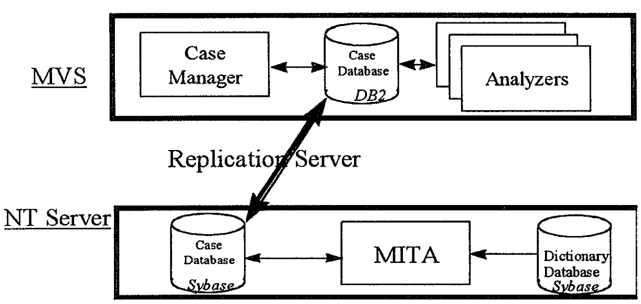

Figure 3. Deployment Environment

MITA runs on a Microsoft Windows NT^{TM} server platform. The dictionaries as well as the input and output tables are all implemented in SYBASE$^{®}$.

MITA communicates with a "Case Manager" which resides on an MVS^{TM} mainframe. The Case Manager coordinates all analyzers and uses DB2^{TM} tables. MITA's SYBASE input and output tables are replicated in DB2 using SYBASE RepServer. This approach makes all inter-platform communication transparent to all processes.

ART*Enterprise

MITA is implemented in Brightware's ART**Enterprise*. ART**Enterprise* is a set of programming paradigms and tools that are focused on the development of efficient, flexible, and commercially deployable knowledge-based systems.

ART**Enterprise* rules provide an expressive pattern-matching language and an efficient inference engine. Rules are extensively used in MITA processing. They allow MITA to be data-driven, efficient, flexible, and easy to maintain.

Procedures are implemented as ART**Enterprise* functions and methods. They are coded in ART**Enterprise's* ARTScript language.

Data structures within MITA are represented as *facts* (which are essentially lists of values and symbols) and as ART**Enterprise objects*. The object system allows full object-oriented programming within MITA, and is fully integrated with rule and procedural components.

The ART**Enterprise* Data Integrator allows easy generation of an object-oriented interface between MITA and the database tables.

MITA uses memory-resident hash-tables, which are another ART* *Enterprise* data structure, to make its frequent navigation among semantic class relationships fast and efficient.

Expected Application Use and Payoff

By applying MITA, in conjunction with the downstream analyzers, to this initial review process MetLife hopes to save up to a third of total underwriting time. In addition, it is expected that MetLife will achieve greater underwriting consistency.

Validation

There were two phases of MITA validation. The first was ongoing during the development phase, and its purpose was to focus development on the areas that would maximize MITA's performance. The second was begun in the System Test phase, and its purpose is to measure MITA outputs in the context in which they will be used to make underwriting decisions.

An important metric for MITA is its *recall,* which is the percentage of concepts that are successfully extracted from the text[Lehnert and Sundheim, 1991]. MITA was built to extract all concepts in the text field but *not* to evaluate the underwriting significance of these concepts. Therefore MITA must have a high recall to guarantee that the significant concepts *are* extracted.

Confidence Factors must have a high correlation with recall to be a reliable predictor of extraction quality. A combination of high recall with a good confidence factor will allow a low Confidence Factor *threshold* to be set. This threshold is the point above which MITA extractions are sufficiently reliable. The lower the threshold can be set, the more cases can be automated.

Corpus Segmentation

The corpus was segmented into a "training" set of cases -- cases that knowledge engineers could analyze and use to develop extraction rules and dictionary entries; and a "performance" set – cases which were set aside to test system performance. Several hundred cases from each of these segments were annotated by the underwriters with the expected extractions and classifications.

Development Phase Validation

During the development phase, periodic evaluations of MITA performance on small sample sets were conducted. Performance results rose steadily, and the last measurements taken indicated overall recall around 85%, and confidence factors with the ability to reliably predict at least 70% of the correct cases.

System Test Phase Validation

A stricter Validation of MITA is in progress at this writing -- one that measures MITA quality against the decision-making challenges that will be required of its output. One approach that is being used is "blind testing" to determine whether the output of MITA is sufficient to make underwriting decisions equivalent to those produced by an underwriter with access to full text. This test is being conducted in conjunction with MetLife's actuarial staff. The goal is to determine the reliability and appropriate thresholds of Confidence Factors.

The blind test set of examples was selected from 3,800 sample cases that the MITA development staff and the underwriters had not seen. From this set, approximately 180 were selected representing various confidence factor levels.

For each of these cases, underwriters were presented first with only the MITA extracted concepts and their classifications. They were asked to draw and record underwriting conclusions from these. Next they were given the actual text from which the extractions had been made, and were asked again to draw and record their underwriting conclusions. The recorded results were then reviewed by actuaries, to identify and characterize any differences in underwriting decisions.

Preliminary results show that *none* of the extractions resulted in underwriting errors. However, several were over-general and led the underwriters to take extra review steps. Since the validation sample was not random, it is not known yet what percentage of over-general extractions occur in the general case population and how well they correlate with low Confidence Factors.

Another test now under way involves sorting the extractions from a randomly selected set of cases in descending order of Confidence Factor by field. The underwriters review each extraction in the sorted order and identify where they see the first incorrect extractions. The threshold will be set at a level somewhat higher than the level where those inaccuracies were first encountered.

Hopefully these two tests will yield comparable results. These tests, along with a period of parallel operation prior to production should provide a reliable validation and calibration of MITA.

Knowledge Engineering

The Process

The system was engineered to imitate the ability of underwriters to read text and recognize the potentially important pieces of information. The focus of this engineering was to achieve optimal performance for the five designated text fields. 20,000 sample cases were made available to the team for testing and engineering the system.

The knowledge engineering process therefore, was focused both on underwriter knowledge as well as text characteristics. These objectives required a knowledge acquisition tool that would allow underwriters to annotate cases with the correct results. This was coupled with a validation tool that could compare MITA's output with that specified by the underwriters.

The underwriters' ability to recognize significant concepts was essential to the project. During every phase, the knowledge engineers consulted with them to determine what combinations of syntax and semantics indicated the presence of a potentially significant concept. The underwriters also provided a strong direction on how to classify words and phrases, and what the relationships should be among classes.

Custom Development Tools

Several tools were developed to assist in MITA development:

Knowledge Acquisition Tool. The Knowledge Acquisition(KA) Tool allows the underwriter to create annotated cases to be used in system development and testing. The KA tool provides the underwriter with an interface showing a text, and some preliminary extractions(concept, value, and class). The underwriter can change or add concept names, concept values, and classes through a link with the ontology tool.

MITA Validator. The Validator is a utility created in ART**Enterprise* that compares the outputs of MITA for a particular text, to the annotated expected results. Reports show where and how the two sets of results differed.

Ontology Tool. The ontology tool assists the knowledge engineer in browsing and modifying the dictionary. It allows the user to navigate between words and their classes, and even among the participants of a composite class. Entries and relationships can be added or changed at any point.

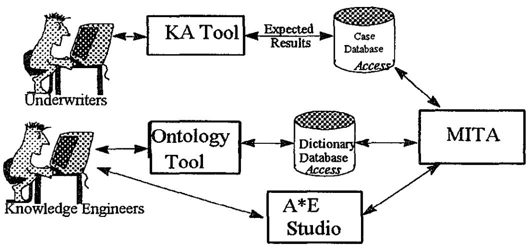

Figure 4. The MITA Development Environment and Tools

System Maintenance

It is expected that system maintenance will consist of monitoring the cases that have a low confidence value for commonalties and frequencies. Most often the solution will involve using the ontology tool to add or adjust relationships among words and classes. . Less often it will involve adding or modifying extraction rules. Occasionally it may involve changing entries in the lists of abbreviations, etc. Changes may involve adjustments to the confidence factor penalties themselves to correct an errant confidence factor. It will be rare that adjustments will have to be made to the parser or other major infrastructure components after deployment.

Since it is expected that enhancements to MITA over time will be common, MITA has been designed with a great deal of flexibility. However, any careless change has the potential to degrade MITA performance considerably.

To assist with the smooth maintenance of MITA, the MITA team created a regression test base. After a system change, these cases are run through MITA and their results compared to baseline results. Any differences found are investigated, resolved, and incorporated into a new baseline.

Technology Transfer

As the MITA implementation moves toward completion, a technology transfer effort was begun in conjunction with testing and deployment. The goal of technology transfer is to transfer the everyday operations of MITA over to MetLife staff, and to prepare MetLife staff as much as possible to support and enhance MITA in the future.

Initial technology transfer tasks include involving MetLife personnel in the analysis of potential dictionary and ontology changes to improve MITA's performance. Subsequent steps involve the participation of MetLife people in performing tests and analyzing MITA test results. Next they will become familiar with the ART**Enterprise* toolset and begin participating in making code changes, particularly to the extraction rules. Ultimate full support will require an understanding of the functioning of all MITA modules.

Summary and Further Work

Initial results indicate that MITA will enable a high degree of automation of the labor-intensive tasks of initial underwriting review, and will pave the way for more extensive automation in the future. The same techniques can be extended to other text fields in life insurance, to other types of underwriting (property and casualty) and to other insurance systems i.e. claims processing. Immediate extensions of MITA to handle underwriting correspondence and to process electronic transmissions of reports and test results are contemplated.

The large number of applications which will become candidates for automation as a result of MITA is indicative of the high percentage of business knowledge that appears as text, relative to structured information. Use of Natural Language Processing to manipulate this body of information is now becoming possible due to improved algorithms and increased computational power available for dedicated tasks.

Acknowledgments

We would like to acknowledge the contribution of the MetLife underwriters Eileen Kosiner, Bill Rowe, Mike Breen, and Joe Grecco, without whose enthusiastic participation MITA would not exist. We would also like to thank Linda Mattos for making the underwriters available to us, Joe Casalaina for his assistance in numerous technical tasks, and Bob Riggio for providing guidance and support throughout the project. Thanks also to Ellen Riloff, who provided the benefit of her extensive knowledge of NLP and Information Extraction, from both a theoretical and practical point of view. Finally, thanks to the members of the Brightware team, Dan Grudus, Fred Simkin, Michael Belofsky, David Reel, and Jay Runkel.

Inc. Windows NT is a trademark of Microsoft Corporation. MVS and DB2 are trademarks of International Business Machines Corporation.

References

Cowie, J., and Lehnert, W.G. (1996). Information Extraction. In Communications of the ACM vol. 39 (1), pp. 80-91.

Lehnert, W. (1990). Symbolic/Subsymbolic Sentence Analysis: Exploiting the Best of Two Worlds. Advances in Connectionist and Neural Computation Theory. vol. 1. Norwood, NJ: Ablex Publishing, 151-158.

Lehnert, W.G., and Sundheim, B. (1991). A Performance Evaluation of Text Analysis Technologies. AI Magazine: 81-94.

(MUC3) Proceedings of the Third Message Understanding Conference, Morgan Kaufmann Publishers, 1991.

(MUC4) Proceedings of the Fourth Message Understanding Conference, Morgan Kaufmann Publishers, 1992.

(MUC5) Proceedings of the Fifth Message Understanding Conference, Morgan Kaufmann Publishers, 1993.

(MUC6) Proceedings of the Sixth Message Understanding Conference, Morgan Kaufmann Publishers, 1995.

Riloff, E. Automatically Constructing a Dictionary for Information Extraction Tasks. In Proceedings of the Eleventh National Conference on Artificial Intelligence, 811-816, 1993.

Riloff, E. "An Empirical Study of Automated Dictionary Construction for Information Extraction in Three Domains", AI Journal , Vol. 85, August 1996.

Soderland, S., Fisher, D., Aseltine, J., Lehnert, W. CRYSTAL: Inducing a Conceptual Dictionary. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1314-1321, 1995.

Soderland, S., Fisher, D., Aseltine, J. and Lehnert, W. "Issues in Inductive Learning of Domain-specific Text Extraction Rules." In Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing. Wermter, S., Riloff, E., and Scheler, G. (Eds). Springer-Verlag 1996.