

1-1-2008

Enabling Synergy between Psychology and Natural Language Processing for e-Government: Crime Reporting and Investigative Interview System

Alicia Iriberry '06

Claremont Graduate University

Chih Hao Ku '12

Claremont Graduate University

Gondy Leroy

Claremont Graduate University

Recommended Citation

A. Iriberry, C.H. Ku , and G.Leroy, Enabling Synergy between Psychology and Natural Language Processing for e-Government: Crime Reporting and Investigative Interview System, Ninth International Conference on Digital Government Research (DG.O 2008), May 18-21, 2008, Montreal, Canada.

This Poster is brought to you for free and open access by the CGU Faculty Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in CGU Faculty Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

Enabling Synergy between Psychology and Natural Language Processing for e-Government: Crime Reporting and Investigative Interview System

Alicia Iriberry

Claremont Graduate University
130 9th Street, Claremont, CA 91711

alicia.iriberri@cgu.edu

Chih Hao Ku

Claremont Graduate University
130 9th Street, Claremont, CA 91711

ku.justin@gmail.com

Gondy Leroy

Claremont Graduate University
130 9th Street, Claremont, CA 91711

gondy.leroy@cgu.edu

ABSTRACT

We are developing an automated crime reporting and investigative interview system. The system incorporates cognitive interview techniques to maximize witness memory recall, and information extraction technology to extract and annotate crime entities from witness narratives and interview responses. Evaluations of the IE components of the system show that it captures 70 to 77% of information from witness narratives with 93 to 100% precision. Our development goal is for the system to approximate progressively the performance effectiveness of a human investigative interviewer and to generate graphical visualizations of crime report information.

Categories and Subject Descriptors

C.4 [Performance of Systems]: Design studies, measurement techniques, performance attributes. H.5.2 [User Interfaces]: Natural Language. I.2.7 [Natural Language Processing]: Text analysis. J.1 [Administrative Data Processing]: Government. J.4 [Social and Behavioral Sciences]: Psychology.

General Terms

Algorithms, Design, Human Factors, Performance

Keywords

Information Extraction, Crime, Cognitive Interview, e-Government

1. INTRODUCTION

Every year millions of crimes are committed in the US. According to the FBI, every 3.2 seconds a property crime occurs, and every 22.2 seconds a violent crime is committed [4]. Research indicates that the principal determinant to solving crimes is the completeness and accuracy of eyewitness reports [10]. However, 50% of crimes go unreported due to fear and privacy concerns [2]. In addition, police resource shortages (e.g., investigators, training, time to respond and transcribe reports) often lead to incomplete or inaccurate information.

We are developing and evaluating a crime reporting and investigative interview system that can extract information from witness crime narratives and simulate an investigative interview. This system will provide a tool police can use to collect accurate and complete information when it is not feasible or possible to conduct face-to-face interviews. In addition, witnesses can maintain their privacy by reporting information anonymously or using secured IDs. Specifically, the system asks witnesses to

provide written narratives of a crime they witnessed and extracts relevant facts from the narrative using natural language information extraction techniques. The system then generates questions and an interview strategy to help witnesses recall missing facts about the crime they witnessed conforming to cognitive interview principles. Finally, our system produces a written standard police report.

The importance of this contribution lies in the possibility of using natural language information extraction technology and Internet-based interviewing systems to gather information from the public. Such a system may help alleviate the shortage of police resources while maximizing the quality of information collected from witnesses. Lessons learned will be useful in the design of usable e-Government applications and services.

2. COGNITIVE INTERVIEW

The system incorporates interviewing techniques from the Cognitive Interview (CI) [5,6]. The CI is based on theories and principles from memory and eyewitness research. Evidence from CI research indicates that when investigative interviewers use CI, as opposed to standard, questionnaire-based interviews, the accuracy and richness of the information obtained from witnesses is significantly higher.

Even though the CI is a very effective investigative interviewing technique, it places high demands on the interviewer in terms of training, time, ability, and availability [6]. Due to limited police resources, investigators might focus only on interviewing key witnesses at the time of a criminal incident, and important pieces of information from other witnesses may be lost. Findings also indicate that if witnesses are not interviewed shortly after the incident, their memory of it can suffer from cross-contamination and decay [7].

3. INFORMATION EXTRACTION

Information extraction (IE) uses a range of natural language processing (NLP) techniques to produce fixed-format data about domain-specific entities found in written narratives (i.e., texts, documents, articles, webpages). The resulting data then may be used for database queries or further analysis.

We use a knowledge engineering, rule-based approach to IE. In this approach, knowledge engineers with the help of domain experts analyze example narratives to identify such text snippets as named entities, grammars and text patterns. Then, they create rules that model the grammar and text patterns they identified.

These rules are fed to an IE system that will use them to process new domain-specific narratives. When the system detects instances of the modeled snippets, it automatically extracts and annotates them. To produce annotations, IE systems preprocess narratives with tokenizing, sentence-splitting, and part-of-speech tagging tools. IE systems use the output of these tools and process it further using lexicon lookups to identify named entities. Finally, they use the grammar and text pattern rules to produce the required output.

In the crime domain, named entities include people, locations, personal physical attributes, weapons, vehicles, acts, and personal property. A text pattern in this domain may be "blue eyes." A crime rule-based IE system would annotate this phrase as *eye:body part* and *blue:eye color*.

The IE rule-based approach is labor-intensive, but it is useful when the number of example narratives available to creating and testing rules is limited and the level of precision required is high. Rule-based IE systems often achieve these levels [1].

4. THE SYSTEM

Our crime reporting and investigative interview system incorporates CI techniques to maximize witness memory recall and IE capabilities to extract crime information from interviews. We expect incremental developments will approximate gradually the performance effectiveness of a human interviewer.

The system comprises Internet, database and Java technology, and leverages open-source IE tools from the General Architecture for Text Engineering (GATE) [3]. Using these technologies, the system simulates the tasks that a CI interviewer normally performs. First, it asks witnesses to provide general information and then a written narrative of the incident witnessed. Second, it uses IE tools to extract named entities and relevant facts from this narrative. Third, using the output of the initial extraction, the system assembles questions and designs an interviewing strategy, just like a CI interviewer does, and presents these questions to help witnesses recall facts that are missing in their report. Last, the system produces a written standard police report for confirmation.

5. FINDINGS

Evaluations of the IE components of the system show promising results (see [8] and [9]). Evaluations of the interviewing component will follow. The first completed component is the suspect-description IE module. This module was able to capture 70% (recall) of the information with 100% precision using narrative descriptions of individuals. These narratives were collected from participants who looked at suspect mugshots (initial narratives only without answering probing questions) [8]. Recall improvements were achieved after several rounds of fine-tuning and testing of lexical lookups and semantic-tagging rules.

The second increment of our system included modules to extract information about weapons, vehicles, time, people, articles of clothing, and locations. These modules were able to capture 77% (recall) of the information with 93% precision using police and witness narrative crime reports collected from various websites (e.g., Unsolved Crime International, Chat Law Info forum, True Crime blog and Expert Law) [9]. Improvements were achieved after using spellchecking capabilities before extraction. We tested

"first alternative" and "best alternative" spellchecking approaches. Results were 79% (recall) with 94% precision and 78% (recall) with 93% precision respectively [9].

Real, first-hand narratives are not readily available to use as test data, since they are treated as confidential by police. We are collecting realistic report examples from subjects acting as witnesses to videotaped crimes (real and staged) to conduct more "real life" evaluations of the IE and interviewing modules.

6. FUTURE PROJECTS

The next steps in our system development are user evaluation of the IE and interviewing components as a unit to compare the effectiveness of our interviewing system to the effectiveness of a human CI interviewer. We will then add graphical visualizations and data mining capabilities to allow discovery of crime entities and behavior patterns and of conflicting and complementary information in the text-based standard police reports resulting from the interviewing components.

7. REFERENCES

- [1] Appelt, D. & Israel, D. (1999). An introduction to information extraction technology. Tutorial prepared for the IJCAI Conference. <http://www.ai.sri.com/~appelt/ietutorial>.
- [2] BJS. 2005. Reporting crime to police, 1992-2000. Bureau of Justice Statistics Special Report.
- [3] Cunningham, H., et al. 2002. GATE: A framework and graphical development environment for robust NLP Tools and Applications. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia.
- [4] FBI. 2006. Crime clock. Federal Bureau of Investigation. http://www.fbi.gov/ucr/cius2006/about/crime_clock.html.
- [5] Geiselman, R.E., Fisher, R.P., MacKinnon, D.P. and Holland, H.L. 1985. Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, 70, 401-412.
- [6] Geiselman, R. E. and Fisher, R. P. 1997. Ten years of cognitive interviewing. In *Intersections in Basic and Applied Memory Research*, D. G. Payne and F.G. Conrad, Eds. Lawrence Erlbaum Associates.
- [7] Hope, L., Gabbert, F. and Fisher, R. 2006. Developing a scene of crime recall tool: Investigating the impact of immediate recall on later crime event memory. *American Psychology and Law Conference*, St. Petersburg, Florida.
- [8] Iriberry, A. and Leroy, G. 2007. Natural Language Processing and E-Government: Extracting Reusable Crime Report Information, IEEE International Conference on Information Reuse and Integration. Las Vegas, Nevada, USA.
- [9] Ku, C. H., Iriberry, A and Leroy, G. 2008. Crime Information Extraction from Police and Witness Narrative Reports, 2008 IEEE International Conference on Technologies for Homeland Security, Forthcoming.
- [10] Rand. 1975. The criminal investigation process 1-3. Rand Corporation Technical Report. Santa Monica.