1-1-2008

# Crime Information Extraction from Police and Witness Narrative Reports

Chih Hao Ku '12
*Claremont Graduate University*

Alicia Iriberri '06
*Claremont Graduate University*

Gondy A. Leroy
*Claremont Graduate University*

## Recommended Citation

C. H. Ku, A. Iriberri, and G.Leroy, "Crime Information Extraction from Police and Witness Narrative Reports," 2008 IEEE International Conference on Technologies for Homeland Security, May 12-13, 2008, Boston.

Chih Hao Ku, A. Iriberri, and G.Leroy, "Crime Information Extraction from Police and Witness Narrative Reports," 2008 IEEE International Conference on Technologies for Homeland Security, May 12-13, 2008, Boston

# Crime Information Extraction from Police and Witness Narrative Reports

Chih Hao Ku, Alicia Iriberri, Gondy Leroy
School of Information Systems and Technology
Claremont Graduate University
ku.justin@gmail.com, Alicia.IriberriAjuria@cgu.edu, gondy.leroy@cgu.edu

**Abstract**–*To solve crimes, investigators often rely on interviews with witnesses, victims, or criminals themselves. The interviews are transcribed and the pertinent data is contained in narrative form. To solve one crime, investigators may need to interview multiple people and then analyze the narrative reports. There are several difficulties with this process: interviewing people is time consuming, the interviews – sometimes conducted by multiple officers – need to be combined, and the resulting information may still be incomplete. For example, victims or witnesses are often too scared or embarrassed to report or prefer to remain anonymous. We are developing an online reporting system that combines natural language processing with insights from the cognitive interview approach to obtain more information from witnesses and victims. We report here on information extraction from police and witness narratives. We achieved high precision, 94% and 96%, and recall, 85% and 90%, for both narrative types.*

## 1. Introduction

Homeland security focuses on the protection of populations and essential infrastructure [1]. Information technology can contribute by helping solve and prevent crimes more efficiently. In the United States, millions of crimes go unreported [2]. Many victims and witnesses are too scared or embarrassed to report crime incidents. In some cases, interviewers may record answers inaccurately or illegibly, or may fail to record them [3]. In addition, report data such as Uniform Crime Reports (UCR) collected by the FBI often contains missing, incomplete, or incorrect data [4].

We are developing an online reporting system that addresses these problems. It will allow those too scared or embarrassed to report anonymously and, because it is based on principles of the cognitive interview [5], it may also help them remember more information correctly. Our approach will also enable us to combine the reported information in one report for law enforcement personnel.

Our system combines information extraction (IE) and principles of the cognitive interview [6]. The cognitive interview is a psychological technique that helps people remember more information about an incident. In order to leverage its principles, we encourage the use of natural language so that people do not need to fill out numerous structured reports. Such forms may be complex or difficult to understand leading fields left blank and incomplete information. By using natural language, people can report a crime more easily and thus more information can be collected. To enable such reporting, we need to extract crime-relevant information to ask follow up questions and compile a final report. Our goal is to obtain as much information as possible. To this end, we developed a large lexicon that combined with rule-based system can extract crime-related entities. Those extracted entities are triggers for our system to ask questions according to the principles of the cognitive interview.

To evaluate the information extraction component, we collected police and witness narratives from web sites, blogs, and forums. We test these by calculating precision and recall of information in comparison with a separately developed gold standard. The diversity of data sources shows the dependability.

## 2. Information Extraction

Information extraction aims to extract pre-specified elements. For example, the names of people, places, or organizations can be extracted from documents without "deep understanding" [7] of the text. IE techniques have been used for many different purposes such as to extract auction prices from eBay and Yahoo web pages [8], to extract text information from PDF files [9], , or in bioinformatics, to extract named entities and relationships of genes, proteins, and RNA from scientific publications [10].

Commonly extracted crime-related entities are race, gender, age, weapons [11], addresses, narcotic drugs, vehicles, and personal properties [12]. However, such annotated information is often short [13] and may be difficult for investigators to correctly interpret. 'Victoria', for example, can refer to the name of a person, a location, or clothing.

The techniques used for IE range from lexical lookup and rule-based approaches [12] to fully automated machine learning. Lexical lookup uses hand-crafted lexicons, such as people's names, and compares these with target texts. Most entity IE systems employ some type of lexical lookup to match named entities. Rule-based systems use hand-crafted rules to recognize entities. However, in some hybrid approaches, these rules are learned automatically. The pure statistics-based systems require statistical models trained on large data sets to identify entities in documents. Another common IE techniques is the use of a hidden Markov model, e.g., to extract headers of computer science research papers [14].

Most IE systems combine multiple techniques and algorithms. For example, Chau et al. [12, 15] proposed a system that used noun phrasing, lexical lookup, and neural network to extract entities: person, address, narcotic drug, and personal property. Feldman et al. [16] proposed the TEG (Trainable Extraction Grammar) system that integrated a statistical and knowledge-based approach to extract named entities and relations. Maynard et al. [17] used a rule-based approach and lexical lookup to extract entities such as person, organization, location, and date from news texts. Srihari and Lei [18] used machine learning with a FST (finite state transducer) rule-based approach and lexical lookup to carry out IE for question answering.

## 3. Cognitive Interview

Accurate recall of a lot of information is difficult to achieve. Most investigators are trained to rely on 'who, what, where, when, and why' questions [19] when they interview people. Unfortunately, in many cases this results in only a subset of the relevant information being gathered. It has been shown that a technique such as the cognitive interview (CI) [19] can help people remember more information with high accuracy.

The cognitive interview technique is based on psychological principles of memory storage and retrieval of information [20]. The first step of the cognitive interview includes mental reinstatement, a way to help witnesses mentally reconstruct the context of an incident, and encouraging witnesses to recall as much detailed information as possible. Next, interviewers ask witnesses to recall the event in different temporal orders. Finally, witnesses are encouraged to remember an incident from different perspectives, i.e., from different physical locations [20].

The cognitive interview has many advantages. Unfortunately, there are also several disadvantages. For example, it is time and labor intensive. Detectives have to be trained in advance to use this technique and it takes a long time to conduct a thorough interview. An online interviewing system may have the same benefits as the cognitive interview while avoiding some of the problems. The system can interview people without time constraint and people can use this system at any place where computers and Internet are available.

## 4. System Development

Our information extraction system combines a large crime-specific lexicon, several GATE (General Architecture for Text Engineering) modules, and an algorithm to recognize the relevant entities among the phrases generated by the system.

### LEXICON DEVELOPMENT

Our lexicon contains several subsets that help recognize entities, such as weapons, vehicles, scenes, clothes, shoes, and physical features. To build this lexicon we used several different resources.

We collected crime types and crime definitions from the Uniform Crime Reports (UCR). To build the vehicle and weapon lexicons, we used encyclopedia data sources such as Wikipeida[1] and MSN Encarta[2]. To collect abstract lexicons such as scene and physical features, we used FrameNet[4] . To build specific lexicons such as brands of cars, web sites such as Serious Wheels[3] were used. To further expand and complete our lexicons, we used thesauri dictionaries such as Collins Cobuild[6], MSN Encarta, Thesaurus.com[5], and Microsoft Word.

Each category includes several sub-lexicons. For example, 'Personal Property' includes 'Bag', 'Jewelry', 'Money', 'Computer', and 'Phone'. 'Act/Event' includes 'Cheat', 'Flee', 'Harm', 'Kidnap', 'Steal', 'Threat', and 'Trespass'. This led to 126 sub-lexicons. The sub-lexicons were combined into 15 categories (also used for detailed evaluations): 'Act/Event', 'Scene', 'People', 'Personal Property', 'Vehicle', 'Weapon', 'Body Part', 'Time', 'Drug', 'Shoes', 'Electronic', 'Physical Feature', 'Physical Condition', 'Hair', and 'Clothing'. Instead of using the entire lexicon, using the 126 sub-lexicons separately makes our rule development more efficient when testing and debugging our rules.

We ensured that there were no overlapping terms between the categories. We selected the most frequently used definition based on Collins Cobuild[6] dictionary to retain or remove a term when overlapping terms occurred.

### GATE MODULES

We used GATE and leveraged several of its modules and plug-ins. We adopted, without adjustment, the tokenizer, sentence splitter, part-of-speech (POS) tagger, noun chunks, and ortho-matcher. We developed our own JAPE rules and

[1] Wikipedia, http://wikipedia.org/
[2] MSN Encarta, http://encarta.msn.com/
[3] Serious Wheels, http://www.seriouswheels.com/cars.htm
[4] FrameNet, http://framenet.icsi.berkeley.edu/
[5] Thesaurus.com, http://thesaurus.reference.com/
[6] Collins Cobuild, http://www.elearnaid.com/basiccobuildcd.html

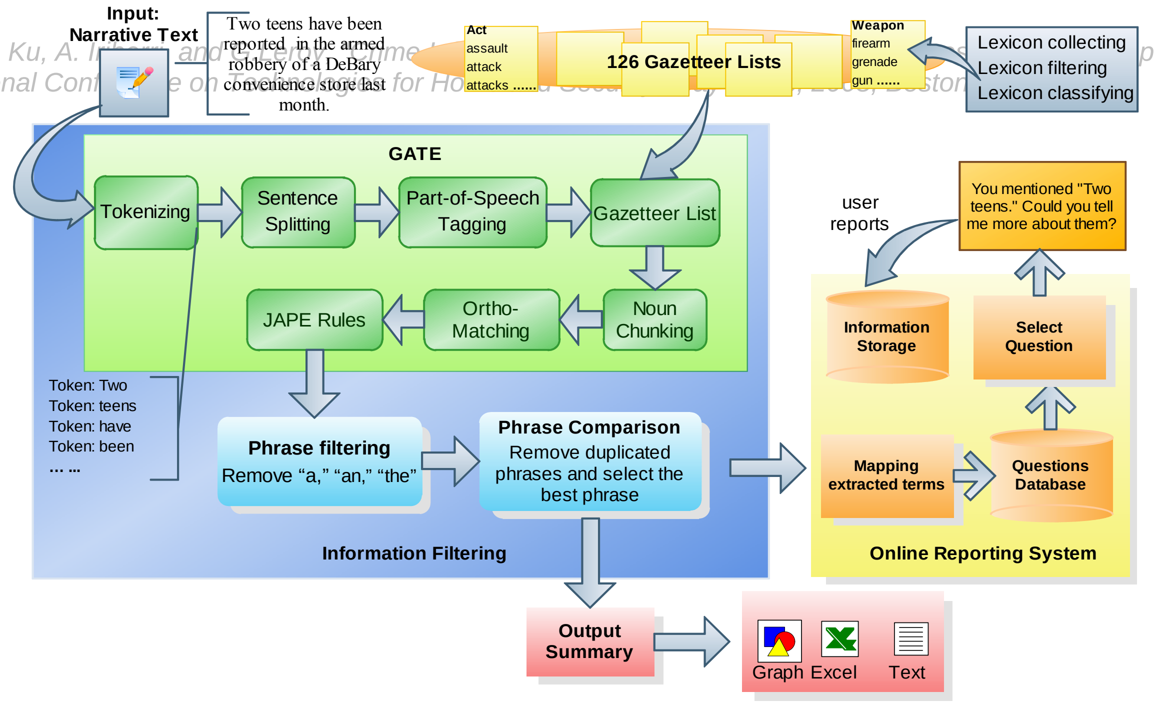

**Figure 1 –** Crime Information Extraction Module

gazetteer lists. Figure 1 provides an overview. We illustrate how each module works with the following example "Two teens have been reported in the armed robbery of a DeBary convenience store last month."

*Tokenizer*−The tokenizer splits text into tokens such as punctuation, words, numbers, and symbols. The example is tokenized as:

*Two / teens / have / been / reported / in / the / armed / robbery / of / a / DeBary / convenience / store / last / month*

*Sentence Splitter*−The text is split into several sentences. Since the example only contains one sentence the result is the same as the example.

*POS Tagger*−Each token is annotated with its part-of-speech (POS) tag. This is a grammatical tag, e.g., verb, noun, or adjective. There are 42 different possible tags as default in GATE. For example, in our sentence DT refers to determiner, NNS refers to plural nouns, VBN refers to past participle verbs, and IN refers to prepositions. The result for the example is:

*Two***[CD]** */ teens***[NNS]** */ have***[VBP]***/ been***[VBN]***/ reported* **[VBN]** */ in***[IN]***/ the***[DT]** */ armed***[VBN]** */ robbery***[NN]** */ of***[IN]** */ a***[DT]** */ DeBary***[NNP]** */ convenience***[NN]** */ store***[NN]** */ last***[JJ]** */ month***[NN]**

*Noun Phrase Chunker*−The chunker uses the tags from the previous components to mark noun phrases. The noun phrases in our example are:

*Two teens / the armed robbery / a DeBary convenience store / last month*

*Gazetteer List*−Our lexicons are used as gazetteer lists. We have divided our lexicon into 126 gazetteers. Each rule only uses related gazetteers rather than the entire gazetteer. For example, a street rule only uses the 'Street, 'Direction', and 'Street Abbreviation' lexicons. An example of the 'Act/Event' gazetteer is: *Kill / murder / rape / slay / rob / massacre / carjacking / assassinate / wallop /smack /garrote / bludgeon / …*

These gazetteer lists were used by our JAPE rules and algorithms. An index file lists majorType and minorType for each gazetteer file so the JAPE rules can use minorType to include different gazetteer lists. For example, the "*Jewelry.lst:property:jewelry*" section refers to the file "*Jwerlry.lst*", majorType "*property*", and minorType "*jewelry*", which are split by colons. The words in the text were annotated using this gazetteer information.

*Ortho-Matcher*−This component can recognize specific names such as people's names and cities based on their orthographical information, such as the presence of uppercase letters. The *{Token.orth == upperInitial}* section recognizes the word "Arizona," for example, as an entity. The ortho-matcher does not require specific lexicons such as gazetteer lists to match target names. Sometimes it is difficult to collect complete lexicons for people names and city names. The ortho-matcher can help our system match those names that were not included in our lexicon lists.

*JAPE Rule*−JAPE (Java Annotations Pattern Engine) [21] is a GATE-specific format to define regular expressions over annotations needed for pattern matching. We created 14 JAPE files. An example rule for the sub-lexicon 'harm' is:

```
Rule: harm
   (
      {Lookup.minorType == harm}
   )
   :harm -->
  :harm.Rule = {majorType= "act", minorType = "harm",
   Rule = "harm Rule"}
```

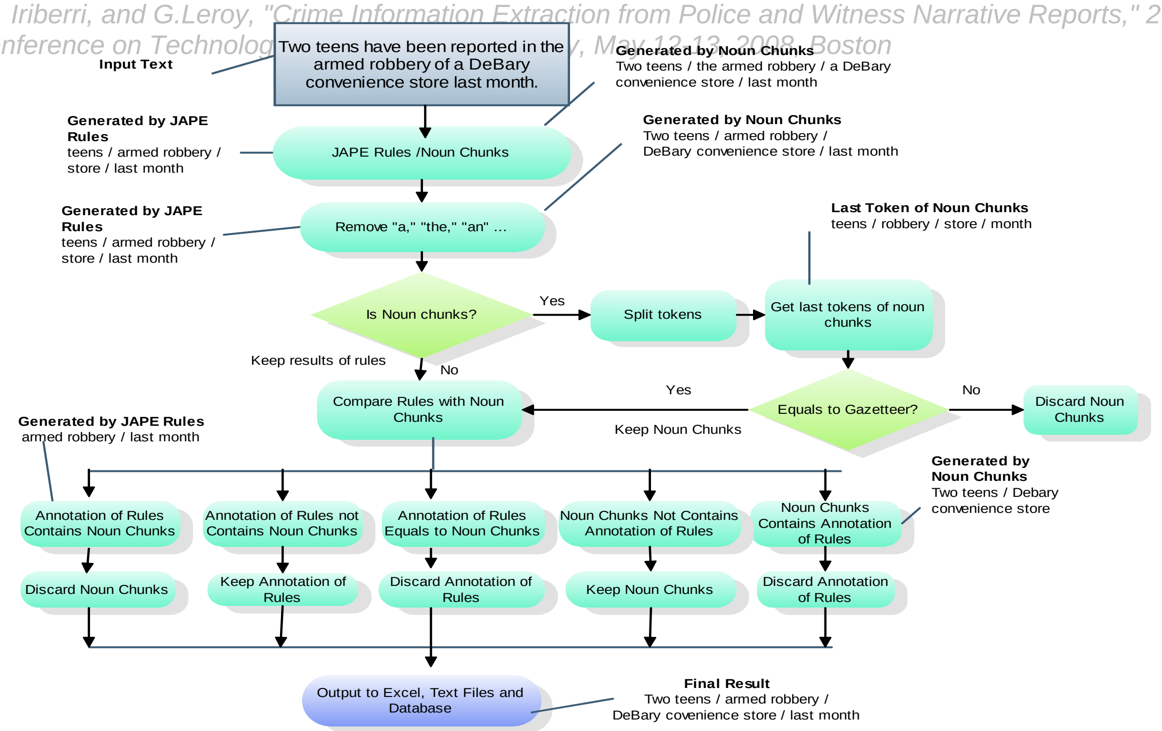
Input Text
Two teens have been reported in the armed robbery of a DeBary convenience store last month.
Generated by Noun Chunks
Two teens / the armed robbery / a DeBary convenience store / last month
Generated by JAPE Rules
teens / armed robbery / store / last month
JAPE Rules /Noun Chunks
Generated by Noun Chunks
Two teens / armed robbery / DeBary convenience store / last month
Generated by JAPE Rules
teens / armed robbery / store / last month
Remove "a," "the," "an" ...
Last Token of Noun Chunks
teens / robbery / store / month
Is Noun chunks?
Yes
Split tokens
Get last tokens of noun chunks
Keep results of rules
No
Compare Rules with Noun Chunks
Yes
Keep Noun Chunks
Equals to Gazetteer?
No
Discard Noun Chunks
Generated by JAPE Rules
armed robbery / last month
Generated by Noun Chunks
Two teens / Debary convenience store
Annotation of Rules Contains Noun Chunks
Annotation of Rules not Contains Noun Chunks
Annotation of Rules Equals to Noun Chunks
Noun Chunks Not Contains Annotation of Rules
Noun Chunks Contains Annotation of Rules
Discard Noun Chunks
Keep Annotation of Rules
Discard Annotation of Rules
Keep Noun Chunks
Discard Annotation of Rules
Output to Excel, Text Files and Database
Final Result
Two teens / armed robbery / DeBary covenience store / last month

**Figure 2 –** Phrase Filtering and Comparison Algorithm

The section *{Lookup.minorType == harm}* matches all words from crime-related 'harm' acts which were stored in our gazetteer lists. So this rules indicates that if a word earlier tagged as 'harm' is found, it needs to be annotated as an 'Act' of the subtype 'harm'.

## CRIME IE ALGORITHM

During initial testing, we found that noun phrases sometimes provide irrelevant information such as 'need', 'favor', 'solution', 'efforts', 'terms', and 'experience'. In this case, we compared the last token of the noun phrase with the gazetteer lists to filter those irrelevant noun phrases. Furthermore, the JAPE rules and the noun phrase chunker may generate overlapping phrases because they are used in parallel. In this case, we would prefer to retain only the longest phrase. For example, the phrase 'DeBary convenience store' contains more information than the word 'store' alone. Richer information can help detectives solve crimes more efficiently.

To resolve these problems, we developed a filtering algorithm (see Figure 2). First, the algorithm removes determiners such as 'a', 'an', and 'the' from all phrases. Next, it captures those phrases generated by the noun phrase chunker but not by the JAPE rules and obtains the last token of each phrase. If this token is not found in any gazetteer list, the algorithm discards those noun phrases since they are most probably irrelevant. We can make this assumption because the gazetteer lists enumerate all relevant entities we are interested in. Next, phrases generated by both the JAPE rules and the noun phrase chunker are evaluated. The algorithm selects the longer phrases if the phrases have different lengths. If two phrases are equal, the algorithm will discard one of them. For our example, these rules result in:

*Two teens / armed robbery / DeBary convenience store / last month*

This output result is used to match questions that are pre-stored in the database. The matching is done based on the cognitive interview principles. The additional questions will lead to more information from users that will be processed in the same manner. Each time the extra information is stored in the database and additional questions are asked when necessary. We are currently testing the question generation component and report here on the information component.

## 5. EVALUATION

### METHODOLOGY

To develop and fine-tune our system, we collected a diverse corpus containing texts from Unsolved-Crimes[7], SUBA District Unit[8], True Crime Blog[10], Baltimore Crime[11], Chat LawInfo[9], and ExpertLaw[12].
For final testing of the information extraction modules, we collected two types of representative texts: police and witness reports. The police narratives were texts collected from alt.True-Crime[13], Secret Witness[14], SFGate Crime[15], Crime-Stopper[16], FreeAdvice[17], TheLAW.com[18], and LaborLawTalk[19].

[7] Unsolved-Crimes, http://groups.yahoo.com/group/Unsolved-Crimes/
[8] SUBA, http://groups.google.com/group/SPD_SDU/web/crime-bulletins
[9] Chat LawInfo, http://chat.lawinfo.com/
[10] True Crime Blog, http://laurajames.typepad.com/clews/
[11] Baltimore Crime, http://baltimorecrime.blogspot.com/
[12] ExpertLaw, http://www.expertlaw.com/forums/index.php
[13] alt.True-Crime, http://groups.google.com/group/alt.true-crime/topics
[14] Secret Witness, http://www.secretwitness.com/cases/
[15] SFGate Crime, http://www.sfgate.com/news/crime/
[16] Crime-Stopper, http://www.crime-stoppers.org/
[17] FreeAdvice, http://forum.freeadvice.com/

| | *cor* | *Err* | *mis* | *precision* | *recall* |
|---|---|---|---|---|---|
| People | 242 | 4 | 32 | 98% | 87% |
| Act | 75 | 1 | 9 | 99% | 88% |
| Scene | 101 | 20 | 5 | 83% | 80% |
| Time | 58 | 5 | 3 | 92% | 88% |
| Age | 7 | 1 | 12 | 88% | 35% |
| Face | 8 | 2 | 0 | 80% | 80% |
| Body Part | *5 | 1 | 0 | 83% | 83% |
| Personal Property | 19 | 0 | 2 | 100% | 90% |
| Physical Condition | 9 | 0 | 5 | 100% | 64% |
| Vehicle | 29 | 0 | 0 | 100% | 100% |
| Clothes | 18 | 0 | 0 | 100% | 100% |
| Weapon | 9 | 0 | 2 | 100% | 82% |
| Feature | 10 | 1 | 3 | 91% | 71% |
| Drug | *2 | 0 | 0 | 100% | 100% |
| Hair | *3 | 0 | 0 | 100% | 100% |
| Total | 595 | 35 | 73 | 94% | 85% |

**Table 1 –** Police Narrative Evaluation

Err-total error items
Mis-total missing items
Cor-total correct items
* - items extracted fewer than 7 times not discussed in this paper

These are mostly forums, blogs, news articles, or texts provided by police departments. For witness narratives, we collected texts from law-related forums such as Chat LawInfo, ExpertLaw, and FreeAdvice, where many people, victims or witnesses, ask for legal advice.

For our evaluation described here, we randomly selected twenty police narratives and twenty witness narratives. For each we established a gold standard. One author created the gold standard for each document and marked the phrases according to pre-defined categories such as ‘Weapon’, ‘Vehicle’, and ‘Time’. The results generated by our system were compared with this gold standard. We use precision and recall to evaluate our approach. Precision is the ratio of the correctly extracted features to the total extracted features. Recall is the ratio of the correctly extracted features to all of the correct features in the documents.

$$\text{Recall} = \frac{\text{Correctly Extracted Features}}{\text{All Correct Features in Document}}$$

$$\text{Precision} = \frac{\text{Correctly Extracted Features}}{\text{Total Extracted Features}}$$

## 6. Results and Discussion

**Table 2 –** Witness Narrative Evaluation

| | *cor* | *Err* | *mis* | *precision* | *recall* |
|---|---|---|---|---|---|
| People | 639 | 11 | 26 | 98% | 95% |
| Act | 71 | 1 | 13 | 99% | 84% |
| Scene | 99 | 12 | 7 | 89% | 84% |
| Time | 62 | 8 | 2 | 89% | 86% |
| Age | 7 | 5 | 1 | 58% | 54% |
| Face | *1 | 1 | 0 | 50% | 50% |
| Body Part | 10 | 0 | 1 | 100% | 91% |
| Personal Property | 21 | 2 | 7 | 91% | 70% |
| Physical Condition | 10 | 0 | 3 | 100% | 77% |
| Vehicle | 24 | 0 | 1 | 100% | 96% |
| Clothes | *1 | 0 | 0 | 100% | 100% |
| Weapon | 7 | 0 | 0 | 100% | 100% |
| Feature | *4 | 1 | 1 | 80% | 67% |
| Drug | 9 | 0 | 4 | 100% | 69% |
| Hair | *0 | 0 | 0 | NA | NA |
| Total | 965 | 41 | 66 | 96% | 90% |

Err-total error items
Mis-total missing items
Cor-total correct items
* - items extracted fewer than 7 times not discussed in this paper

The average length of the 20 police narratives is 130 words while that of the 20 witness narratives is 240 words. The open-source spell checker Ekit was used to correct typos in the witness narratives. The first author selected the best alternative word for each typo.

Precision was very high for both types of narratives: 94% for police narratives and 96% for witness narratives. Recall was also very high: 85% for police narratives and 90% for witness narratives. Table 1 and Table 2 provide an overview of precision and recall for each entity we extracted. For police narratives, we achieved 100% precision for ‘Personal Property’, ‘Physical Condition’, ‘Vehicle’, and ‘Clothes’, but lower precision (80%) for ‘Face’. For witness narratives, we achieved 100% precision for ‘Body Part’, ‘Physical Condition’, ‘Vehicle’, ‘Weapon’, and ‘Drug’ and encountered the lowest precision (58%) for ‘Age’. Recall for ‘Age’ was low for both police narratives and witness narratives. This is due to text such as ‘Steven Warrichiet, 40,’ and ‘14yr-Girl/15yrs-Boy’ from which the system did not extract age correctly.

The witness narratives often contain slang or street language, such as ‘weed roach’ or ‘daddy’ and unorganized syntax such as sentence fragments. Therefore we expected lower recall and precision for the witness narratives. Surprisingly, precision and recall were higher for the witness narratives than for the police narratives. A partial explanation is that, the spell checker removed most of the typos. Only one typo was found in the 20 police narratives while 51 typos were

[18] TheLAW.com, http://www.thelaw.com/forums/
[19] LaborLawTalk, http://www.laborlawtalk.com

found in the 20 witness narratives without the spell checker. Additionally, many more correct items were available and extracted for 'People' from the witness narratives (639 items) than from the police narratives (242 items). The system extracted most pronouns that appeared in the witness narratives but could not extract some personal names such as 'Keisharra Abercrombie' in the police narratives.

## 7. Conclusion

We achieved high precision and recall when testing our modules with police and witness narratives. We plan to collect additional witness narratives using crime video system to further test our system and test the question-interaction components. Our final goal is to provide a reliable online crime reporting system people can use to report crime anonymously, that will encourage people to recall more crime information, and will provide a meaningful summary and a graphical result for police investigators to solve crimes more quickly and efficiently.

## 8. References

[1] M. Reiter and P. Rohatgi, "Homeland Security," *Internet Computing, IEEE,* vol. 8, pp. 16-17, 2004.

[2] A. Iriberri, G. Leroy, and N. Garrett, "Reporting On-Campus Crime Online: User Intention to Use," in *Proceedings of the 39th Hawaii International Conference on System Sciences*, 2006, p. 82a.

[3] E. D. d. Leeuw, "Reducing Missing Data in Surveys: An Overview of Methods," *Quality and Quantity,* vol. 35, pp. 147-160, 2001.

[4] A. M. Lyons, J. Michael W. Packer, M. B. Thomason, J. C. Wesley, P. J. Hansen, J. H. Conklin, and D. E. Brown, "Uniform Crime Report "SuperClean" Data Cleaning Tool," *Systems and Information Engineering Design Symposium, 2006 IEEE,* pp. 14-18, 2006.

[5] M. Ginet and J. Py, "A Technique for Enhancing Memory in Eye Witness Testimonies for Use by Police Officers and Judicial Officials : the Cognitive Interview " *Le Travail Humain,* vol. 64, pp. 173-191, 2001.

[6] A. Iriberri and G. Leroy, "Natural Language Processing and e-Government: Extracting Reusable Crime Report Information," in *Information Reuse and Integration, 2007. IRI 2007. IEEE International Conference on*, Las Vegas, NV, USA, 2007, pp. 221-226.

[7] M. Konchady, "Information Extraction," in *Text Mining Application Programming*: Charles River Media, 2006, pp. 151-182.

[8] D. G. Gregg and S. Walczak, "Exploiting the Information Web," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 37, pp. 109-125, 2007.

[9] F. Yuan, B. Liu, and G. Yu, "A Study on Information Extraction from PDF Files," in *ICMLC*, 2005, pp. 258-267.

[10] S. Pyysalo, F. Ginter, J. Heimonen, J. Björne, J. Boberg, J. Järvinen, and T. Salakoski, "BioInfer: A Corpus for Information Extraction in the Biomedical Domain," *BMC Bioinformatics,* vol. 8, pp. 1-24, Feb. 2007.

[11] S. V. Nath, "Crime Pattern Detection Using Data Mining," in *2006 IEEE/WIC/ACM International Conference on*, 2006, pp. 41-44.

[12] M. Chau, J. J. Xu, and H. Chen, "Extracting Meaningful Entities from Police Narrative Reports," in *ACM International Conference Proceeding Series*. vol. 129 Los Angeles, California: Digital Government Research Center, 2002, pp. 1-5.

[13] P.-J. Cheng, H.-C. Chiao, Y.-C. Pan, and L.-F. Chien, "Annotating Text Segments in Documents for Search," in *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence* Compiègne University of Technology, France, 2005.

[14] Y. Liu, Y. Lin, and Z. Chen, "Using Hidden Markov Model for Information Extraction Based on Multiple Templates," in *Natural Language Processing and Knowledge Engineering, 2003. Proceedings*, 2003, pp. 394- 399.

[15] H. Chen, W. Chung, Y. Qin, M. Chau, J. J. Xu, G. Wang, R. Zheng, and H. Atabakhsh, "Crime Data Mining: An Overview and Case Studies," in *Proceedings of the 2003 Annual National Conference on Digital Government Research*, 2003.

[16] R. Feldman, B. Rosenfeld, and M. Fresko, "TEG - A Hybrid Approach to Information Extraction," *Knowledge and Information Systems,* vol. 9, pp. 1-18, Jan. 2006.

[17] D. Maynard, V. Tablan, K. Bontcheva, and H. Cunningham, "Rapid Customization of an Information Extraction System for a Surprise Language," *ACM Transactions on Asian Language Information Processing (TALIP),* vol. 2, pp. 295-300, Sep. 2003.

[18] R. Srihari and W. Li, "Information Extraction Supported Question Answering," in *Proceedings of the Eighth Text Retrieval Conference (TREC-8)*, 1999.

[19] A. Memon and R. Bull, "The Cognitive Interview - Its Origins, Empirical Support, Evaluation and Practical Implications," *Journal Of Community & Applied Social Psychology,* vol. 1, pp. 291-307, 1991.

[20] M. Mantwill, G. Kohnken, and E. Aschermann, "Effects of the Cognitive Interview on the Recall of Familiar and Unfamiliar Events," *Journal of Applied Psychology,* vol. 80, pp. 68-78, 1995.

[21] H. Cunningham, "GATE, a General Architecture for Text Engineering," *Computers and the Humanities,* vol. 36, pp. 223-254, May 2002.