

# Suspect Description Model Based On Natural Language Processing

Prof. M.U. Karande<sup>1</sup>, Jasvantsing S. Patil<sup>2</sup>, Atul L. Misal<sup>3</sup>, Ashish Sarkate<sup>4</sup>, Kunal More<sup>5</sup>  
<sup>1,2,3,4,5</sup> Department of Computer Science & Engineering, Dr. V. B. kolte College of Engineering Malkapur, Maharashtra, India

## ABSTRACT

*Crime reporting needs to be possible 24/7. Although 911 and tip-lines are the most publicized reporting mechanisms, several other options exist, ranging from in-person reporting to online submissions. Internet-based crime reporting systems allow victims and witnesses of crime to report incidents to police 24/7 from any location. However, these existing e-mail and text-based systems provide little support for witnesses' memory recall leading to reports with less information and lower accuracy. These systems also do not facilitate reuse and integration of the reported information with other information systems. We are developing an anonymous Online Crime Reporting System that is designed to extract relevant crime information from witness' narratives and to ask additional questions based on that information. We leverage natural language processing and investigative interviewing techniques to support memory recall and map the information directly to a database to support information reuse. We report on the evaluation of the Suspect Description Module (SDM) of the system. Our interface captures 70% (recall) of information from witness narratives with 100% precision. Additional modules will follow the design and development methods used with this module.*

**Keyword:** - Tokenizer, Crime report system, Gazetteer, Natural language processing, JAPE rules.

## 1. INTRODUCTION

Our mission statement directs us to promote the public interest while balancing the Law. Obviously we'll not rush to judgment, but will look at the facts and render a thoughtful decision based on these facts. Organization is an independent statutory body which was created to maintain an effective and efficient police service for the public. Its primary task includes securing continuous needs of the public in an efficient manner. As part of the responsibility, 'our committee' is responsible for performance monitoring decided to look at the contribution and effectiveness of the organization within the public. This includes several pieces of key Literature in the area of android applications for crime area detection. Reference [1] proposes a design for mobile communication infrastructure to help the police and general public fight the crime within a metropolis. The infrastructure serves as a common platform for both the police and general public to interact and exchange information about criminal activities and to track criminals.

Every year millions of crimes are committed in the United States. In 2003, the Federal Bureau of Investigation reported that 10.3 million property crimes and 1.38 million violent crimes were committed [1]. However, the Department of Justice reported that in the same year only half of all violent crimes and a third of all property crimes were reported [2]. Reporting crime to police is important for authorities and citizens because more accurate information allows policy makers, law enforcement officials, and police departments to control violence and allocate resources (legislation, and program evaluation) more effectively. Accurate information benefits citizens as well because with it they can identify locations with high and low crime rates, take preventive measures, and make informed decisions on where to live.

## 2. RELATED WORK

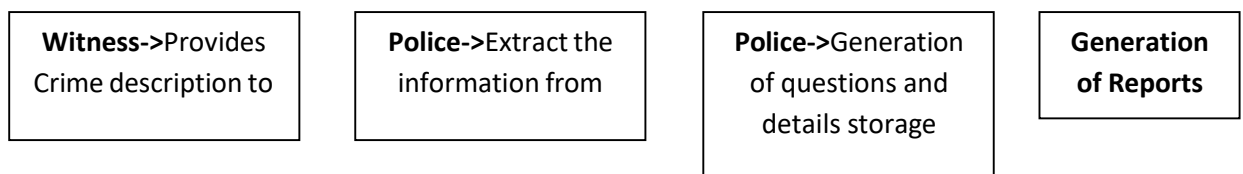
This chapter reviews established strategies, self-administered questionnaires and indirect questioning techniques, for increasing the willingness of respondents to report stigmatizing behaviors. While these methods improve reporting, each has shortcomings and burdens which limit their effectiveness. A new computer-based self-interviewing approach which incorporates recorded audio playback of questions offers improved self-administered interviewing. The chapter discusses this technology, audio computer-assisted self-interviewing (audio-CASI), describing its features and positive results from the early research tests of the method. A group of

150 psychiatric patients were administered the Diagnostic Interview Schedule (DIS) on two occasions, once by a trained layperson and once using a computerized interview format in which the patient interacted directly with the computer. Agreement between the two methods on 15 diagnoses was relatively modest, as indicated by a mean kappa score of .51, but was similar to agreement rates reported in other studies of the DIS. The discrepancies may have been due to the high number of acutely ill inpatients studied, patients' reporting more symptoms in one of the interviews, and difficulty translating some of the DIS questions to the computer. Patients had positive feelings about both methods, but a significant majority liked the computer interview better and found it less embarrassing. The authors conclude that computerized administration of the DIS is as reliable as other methods but that exclusive reliance on the DIS for clinical diagnosis is inappropriate.

We are developing a Crime Reporting System that will address not only the cited reasons for not reporting, but also the need of police departments for more accurate, complete, and reusable information that may free up their time and resources to allocate them to policing the streets. Our system will incorporate the convenience of the Internet, the support of techniques for memory recall, the information extraction capabilities of natural language processing (NLP) technologies and the utilities for storage, integration and reuse of information of database and electronic technologies. With our approach, witnesses will provide more accurate and complete crime information that can be stored in a database so police can create up-to-date, ad-hoc reports and overviews, and can combine this information with other sources, such as geographic information, to provide more in-depth new insights on crime. With more information gathered, police departments will be one step closer to achieving a more effective allocation of resources to better prevent and solve crimes. With natural language processing and the memory enhancing techniques used in investigative interviews we aim to design a Crime Reporting System that is a convenient and safe way for victims and witnesses to provide more information correctly in a format that is immediately reusable. We report here on one module of our system: the Suspect Description Module (SDM). Additional modules including location, vehicle and weapon descriptions will follow the same design and development methods as this one.

### 3. SOFTWARE REQUIREMENT SPECIFICATION

#### 3.1 Operating En3. Proposed Work



#### 3.2 Design and Implementation Constraint

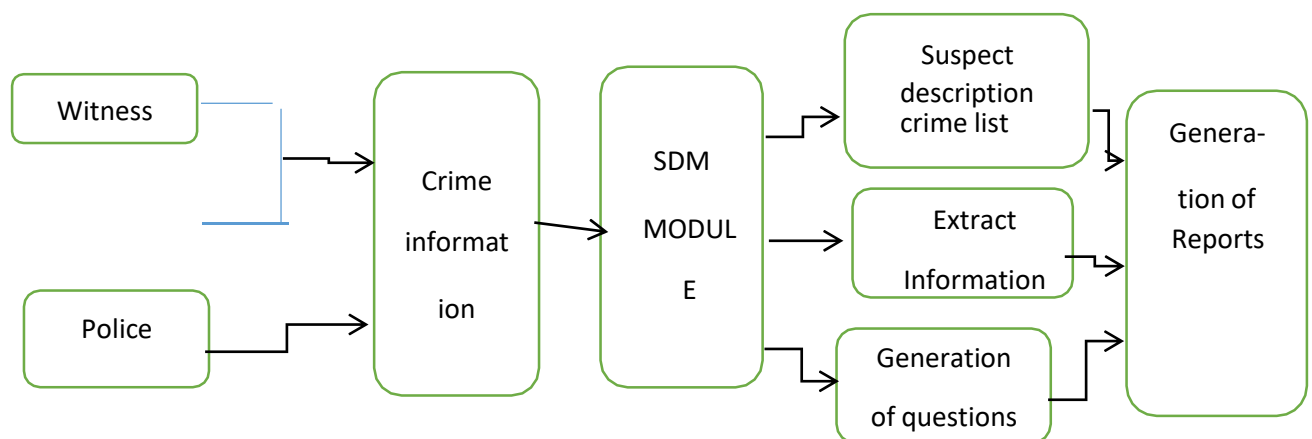
This document intended to bring the architecture goals and the various ways in which the Implementation is divided to meet those goals. Following is descriptions of how these goals are met using the Object Oriented System Design pattern and modular design. The architecture goals considered are:

- Scalability: The system shall have ability to simulate any kind of the system.
  - Modularity: Breaking a design into modules that interact through well-defined interfaces allows developers to work independently, enhances maintainability and testability
    - Common look-and-feel: Applications are easier to use if the user can always guess where to look for desired information.
- Portability: The System shall have ability to portable to other hardware and software platforms with extra estimation.
- Usability: The system shall be user friendly.
  - Maintainability / Extensibility: The system shall accommodate changes and enhancements but will require tweaking in the application with extra efforts. The components should be more cohesive and less coupled.
  - Reusability: Component reuse decreases the cost for new development, provides incremental quality improvements (as software flaws are repaired in long-lived components), and establishes design best practices that everyone in the organization understands.
- Visibility: Provide a standalone system for enabling users to have complete visibility of simulation of selected system.

### 3.3 Data Flow Diagram

A Data Flow Diagram (DFD) is a graphical representation of the “flow” of data through an information system. It can also be used for the visualization of data processing.

## 4. SYSTEM ARCHITECTURE



## 5. CONCLUSIONS

The changes in governance structure introduced during the evaluation period, including the establishment of the War Crimes Steering Committee, the Program Coordination and Operations Committee and its File Review Subcommittee, have contributed to a more cohesive and coordinated program. These changes have contributed to clearer guidance on resource allocation, clearer definitions of departmental roles and responsibilities, and more rigorous criteria for assigning cases to different remedies. We reported on the design and evaluation of one module of our Crime Reporting System: the Suspect Description Module. This module is based on Cognitive Interview techniques and Natural Language Processing. The module leverages GATE and uses 39 Gazetteers and 85 pattern rules to extract information from witnesses’ narratives.

## 6. REFERENCES

1. Waterton, J.J. and J.C. Duffy, A Comparison of Computer Interviewing Techniques and Traditional Methods in the Collection of Self-Report Alcohol Consumption Data in a Field Survey. *International Statistical Review*, 1984. 52(2):p. 173-182.
2. Greist, J.H., et al., Comparison of Computer and Interviewer-Administer Versions of the Diagnostic Interview Schedule. *Hosp Community Psychiatry*, 1987. 38: p. 1304- 1311.
3. Fisher, R.P., et al., Enhancing Enhanced Eyewitness Memory: Refining the Cognitive Interview. *Journal of Police Science and Administration*, 1987. 15(4): p.291- 297.
4. Fisher, R.P., R.E. Geiselman, and M. Amador, Field test of the Cognitive Interview: Enhancing the Recollection of Actual Victims and Witnesses of Crime. *Journal of Applied Psychology*, 1989. 74: p.722-727.
5. Lessler, J.T. and J.M. O'Reilly, Mode of Interview and Reporting of Sensitive Issues: Design and Implementation of Audio Computer-Assisted Self-Interview. *National Institute on Drug Abuse: Research Monograph*, 1997. 167: p. 366-382.
6. Fellbaum, C., ed. *WordNet: An Electronic Lexical Database*. 1998, The MIT Press.
7. Peiris, R., P. Gregor, and N. Alm, The Effects of Simulating Human Conversational Style in a Computer-based Interview. *Interacting with Computers*, 2000. 12: p. 635-650.
8. Cunningham, H., et al. 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)*. 2002. Philadelphia.
9. FBI, *Crime in the USA*. 2003, Federal Bureau of Investigation
10. Thomson, R. and J. Langley, Who Do Young Adult Victims of Physical Assault Talk to about their Experiences? *Journal of Community Psychology*, 2004. 32(4): p. 479-488.