COPLINK Concept Space: An Application for Criminal Intelligence Analysis

Roslin V. Hauck, Homa Atabakhsh, Pichai Ongvasith, Harsh Gupta, Hsinchun Chen

University of Arizona, Management Information Systems Dept., Artificial Intelligence Lab.
http://ai.bpa.arizona.edu/

Revision submitted to IEEE COMPUTER on September 24, 2001.

Abstract
Across the research area of knowledge management, there are a number of common problems dealing with barriers to access and usage. This paper illustrates an approach to these problems in the specific domain of law enforcement. Government agencies across the United States have begun to focus on innovative digital technologies to aid in knowledge management and intelligence analysis. In the domain of law enforcement agencies, the analysis of criminal information is often hampered by its knowledge-intensive and time-critical environment. This atmosphere fosters the need for intelligence tools to combat criminal activity by aiding in case investigation. Funded by the National Institute of Justice and the National Science Foundation, the University of Arizona's Artificial Intelligence Lab has teamed with the Tucson Police Department (TPD) to develop the COPLINK Concept Space application, which uncovers relationships between different types of information that exist in TPD's records management system. In this paper, we present the technology behind the COPLINK Concept Space as well as its usage in real life criminal investigation activities. Future directions of the COPLINK project and development of other advanced technologies for law enforcement are also discussed.

Keywords: Information Systems, Knowledge Management, Information Retrieval, Intelligence Analysis, Information Sharing

1. Introduction

1.1 Intelligence Analysis and Knowledge Management

In this era of the Internet and distributed multimedia computing, new and emerging classes of information technologies have swept into all areas of business, industry and government. As information technologies and applications become more overwhelming, pressing, and diverse, persistent information technology problems have become even more urgent. *Information overload*, a result of the ease of information creation and rendering via Internet, the WWW, and organizational data sources, has become more evident in people's lives[1]. This phenomenon is nowhere more evident than in government, specifically in criminal justice information systems. Federal, state, and local criminal justice entities possess vast repositories of information, but the explosive growth in digital information and the need for access within government agencies have made information overload increasingly significant.

Agencies' knowledge management problems frequently stem from barriers to access and utilization resulting from incompatible content and format of information[2] that make creation and utilization of knowledge management a complex and daunting process. Nevertheless, a number of different applications and approaches to knowledge management technologies are emerging, among them: virtual enterprising[3], joint ventures[4], aerospace engineering[2], and digital libraries[5].

Several government initiatives have been established to address some of the problems of the law enforcement sector of the digital government. The Office of Justice Programs (OJP) Integrated Justice Information Technology Initiative is involving five bureaus including the National Institute for Justice (NIJ) in an effort to use wired-information technologies to improve the effectiveness and fairness of the justice system through better information sharing. An NIJ wireless initiative,

the AGILE program of the NIJ Office of Science and Technology primarily addresses interoperability issues (other government initiatives are described on http://www.ojp.usdoj.gov).

These government initiatives motivated a proposal to unite the technical expertise of the University of Arizona's Artificial Intelligence Lab with the law enforcement domain knowledge of Tucson Police Department to develop cutting edge technologies for the law enforcement community in the COPLINK project. This paper describes the COPLINK Concept Space application, one of those technologies which, although originally funded by NIJ, has received additional funding from both NIJ and the National Science Foundation (NSF) under its Digital Government Initiative. COPLINK is an opportunity to serve the community by bridging the gap between research in developing technologies and solving such real-world problems as helping police officers fight crime. This paper highlights the development and application of a technology that began as a research project, but has become a real-time system being used in everyday police work.

1.2 A Case Study: The Tucson Police Department

The Tucson Police Department (TPD) recently evaluated its information technology and identified problems of lack of information sharing, integration, and knowledge management. The department agreed to participate in research to investigate the potential of current state-of-the-art, near-term, and cost-effective database, Intranet, and multimedia technologies to make computer justice information database integration, management, and access more effective.

The COPLINK project attacks several problems existing in many law enforcement agencies, including TPD, by developing a model integrated system that allows law officers both within and between different agencies to access and share information. An additional goal of COPLINK is to develop consistent, intuitive and easy-to-use interfaces and applications that support specific and often complex law enforcement functions and tasks. Although the scope of this project includes a multilevel development plan incorporating different information technologies, the focus of the research reported here is on the improvement of criminal intelligence analysis.

A function of the daily routine of many crime analysts and detectives at TPD is to create knowledge from information by analyzing and generalizing current criminal records that consist of approximately 1.5 million criminal case reports containing details from criminal events dating back to 1986. Although investigators can access the Records Management System (RMS) to tie together information needed to solve cases and crimes, they must manually search for connections or relationships existing in the data. Combining information to create knowledge is often hampered by voluminous information examination of which requires exorbitant time and effort on the part of the investigator. Compounding this problem is the variability of individual investigator's ability to locate relevant information.

Potent intelligence tools can alleviate crime analysts' information overload, reduce information search time required for analysis of available criminal records, and advance the investigation of current cases. This paper introduces a knowledge management system that can provide the functionality of intelligence analysis that currently does not exist in the RMS system. This system is designed to serve as a knowledge tool that serves the same purpose as current knowledge practices, but systematically and robustly gives crime analysts and investigators power to explore the entire data set for relationships that may exist. Real-life context evaluation of our system, the COPLINK Concept Space, and future directions for the project are also discussed.

2. The COPLINK Concept Space

Concept space, or automatic thesaurus, is a statistic-based, algorithmic technique used to identify relationships between objects (terms or concepts) of interest[6]. The technique is frequently used to develop domain-specific knowledge structures for digital library applications.

In the University of Arizona Artificial Intelligence Lab, the idea of concept space was generated to facilitate semantic retrieval of information. Several user studies showed concept space to improve searching and browsing in the engineering and biomedicine domains. In the biosciences, the concept space approach was applied to the Worm Community System (WCS) and the FlyBase system. There also have been successful results in the Digital Library Initiative studies conducted on the INSPEC collection for computer science and engineering and on Internet searching[5,7]. Current on-going concept space research is being conducted in geographical information systems, law enforcement, and medicine.

A concept space is a network of terms and weighted associations that represent the concepts and their associations within an underlying information space that can assist in concept-based information retrieval. In addition, co-occurrence analysis uses similarity and clustering functions[6] to weight relationships between all possible pairs of concepts. The resulting network-like concept space holds all possible associations between objects, which means that every existing link between every pair of concepts is retained and ranked.

In COPLINK, detailed case reports are the underlying space and concepts are meaningful terms occurring in each case. Concept Space provides the ability to easily identify relevant terms and their degree of relationship to the search term. The relevant terms can be ranked in the order of their degree of association so that the most relevant terms are distinguished from inconsequential terms. From a crime investigation standpoint, Concept Space can help investigators link known objects to other related objects that might contain useful information for further investigation. For instance, like people and vehicles related to a given suspect.

Information related to a suspect can direct an investigation to expand to the right direction, but a case report that reveals relationships among data in one particular case fails to capture those relationships from the entire database. In effect, investigators need to review all case reports related to a suspect, which may be a tedious task. In the COPLINK project, we introduce Concept Space as an alternative investigation tool that captures the relationships between objects in the entire database.

To date, we have successfully adopted our techniques to create a COPLINK Concept Space based on a collection of 1.5 million case reports from the current Tucson Police Department Records Management System. These cases span a time frame from 1986 to 1999 (the entire case record collection for the City of Tucson). Based on careful user requirement analysis, five entity fields from the database were deemed relevant for Concept Space analysis: Person, Organization, Location, Vehicle, and Incident type. The purpose of this tool is to discover relationships between and among different crime-related entities. It is important not only to know that there is a relationship, but also to know what each relationship is.

3. Applying the Concept Space Technique to Criminal Data

In general, there are three main steps in building a domain-specific Concept Space. The first task is to identify collections of documents in a specific subject domain; these are the sources of terms or concepts. For Tucson Police Department, we are using the case reports in the existing database. The next step is to filter and index the terms. The final step is to perform a co-occurrence analysis to capture the relationships among indexed terms. The resulting Concept Space is then inserted into a database for easy manipulation (for a more in-depth analysis of the Concept Space algorithm, see Chen and Lynch[8]). The last two steps have been customized for COPLINK. After optimizing the code and tuning the database, we found that the total time required for building a COPLINK Concept Space is approximately five hours, which is acceptable in the given situation.

3.1 Term filtering and indexing

Due to the nature of the data residing in TPD's database, each piece of information is categorized in case reports and stored in well-organized structures. Theoretically, concept space can contain

any number of term types (e.g., person names, organizations, locations, crime types, etc.). In practice, however, the size of the database, the time required to build a Concept Space, and the response time of queries are major constraints that limit the number of term types. To balance performance and comprehensiveness, a Concept Space should contain only meaningful types frequently searched by users. With the collaboration of personnel from the Tucson Police Department, we identified and created a set of term types for the COPLINK Concept Space.

Term types in Concept Space can be divided into five main categories. For a Person, Organization, Location, and Incident type, only one piece of information, such as a person's full name, street address, or crime type, is descriptive enough to be a search term. On the other hand, for a Vehicle, one piece of information, such as color, make or type, typically is comparatively common and when used as a search term would generate a large number of relevant terms. This problem can be avoided by combining two or more non-specific terms into composite terms.

The index maintains the relationship between a term and the document in which it occurs. Both index and reverse index are required for co-occurrence analysis. The index contains the links from term to document; the reverse index contains the links from document to term.

3.2 Co-occurrence Analysis

After identifying terms, we first computed the term frequency and the document frequency for each term in a document, based on the methodology developed by Chen[8]. Term frequency, *tf*, represents the number of occurrences of term j in document i. Document frequency, *df*, represents the number of documents in a collection of *n* documents in which term *j* occurs.

We then computed the combined weight of term *j* in document *i*, d_{ij}, based on the product of "term frequency" and "inverse document frequency" as follows:

$$d_{ij} = tf_{ij} \times \log(\frac{N}{df_j} \times w_j)$$

where *N* represents the total number of documents in a collection and w_j represents the weight of words in descriptor *j*. In general, some term types are more descriptive and more important than others and deserve to be assigned higher weights so as to ensure that relationships associated with these types are always ranked reasonably. In COPLINK Concept Space, crime types are assigned comparatively higher weights.

We then performed term co-occurrence analysis based on the asymmetric "Cluster Function" developed by Chen and Lynch[8].

$$W_{jk} = \frac{\sum_{i=1}^{n} d_{ijk}}{\sum_{i=1}^{n} d_{ij}} \times WeightingFactor(k)$$

$$W_{kj} = \frac{\sum_{i=1}^{n} d_{ikj}}{\sum_{i=1}^{n} d_{ik}} \times WeightingFactor(j)$$

W_{jk} indicates the similarity weights from term j to term *k* and W_{kj} indicates the similarity weights from term *k* to term *j*. d_{ij} and d_{ik} were calculated based on the equation in the previous step. d_{ijk} and d_{ikj} represent the combined weight of both descriptors *j* and *k* in document *i*. However, they were computed slightly differently due to their different starting terms. They are defined as follows:

$$d_{ijk} = tf_{ijk} \times \log(\frac{N}{df_{jk}} \times w_j)$$

$$d_{ikj} = tf_{ijk} \times \log(\frac{N}{df_{jk}} \times w_k)$$

where tf_{ijk} represents the number of occurrences of both term *j* and term *k* in document *i* (the smaller number of occurrences between the terms was chosen); df_{jk} represents the number of documents (in a collection of *N* documents) in which terms *j* and *k* occur together.

In order to penalize general terms (terms which appeared in many places) in the co-occurrence analysis, we developed the following weighting scheme, which is similar to the inverse document frequency function.

$$WeightingFactor(k) = \frac{\log \frac{N}{df_k}}{\log N}$$

$$WeightingFactor(j) = \frac{\log \frac{N}{df_j}}{\log N}$$

Terms with a higher df_k or df_j value (more general terms) had a smaller weighting factor value, which caused the co-occurrence probability to become smaller. In effect, general terms were pushed down in the co-occurrence table (terms in the co-occurrence table were presented in reverse probabilistic order, with more relevant terms appearing first.)

Significant research needs to be conducted to investigate using Concept Space with our proposed noun phrasing and entity extraction techniques. In the above example, entity types from database fields were identified manually by human analysts. In addition, the Tucson Police Department does not yet capture free text narratives. Many law enforcement agencies have begun to incorporate content-rich narratives in their record management systems (e.g., Phoenix Police Department has complete narratives about each case). These narratives will provide a fertile test bed for combining noun phrasing and Concept Space analysis for intelligence identification.

4. Graphical User Interface for COPLINK Concept Space

The graphical user interface for the COPLINK Concept Space Application is shown in Figures 1-3 (actual information has been altered to maintain data confidentiality). Search terms can be entered from any of the four search forms namely *Person, Organization, Location, and Vehicle*. For two or more search terms, each search term can be typed in the relevant search forms and can be added to the list through the *'Add'* button. The list of search terms is displayed in *'display relationships between'* box.

Relationships displayed between the entered search terms are organized by the five categories namely: *Person, Organization, Location, Vehicle, and Crime Type.* The Concept Space Application is also capable of displaying case reports with detailed information including case number, team beat, crime type etc. To illustrate the interface and usage of the Coplink CS system, the following is a possible scenario for an officer's or analyst's investigation of a crime.

Scenario: A detective is investigating a robbery at a local convenience store. The only witness, the night store clerk, only remembers that the suspect drove away in a white pickup truck.

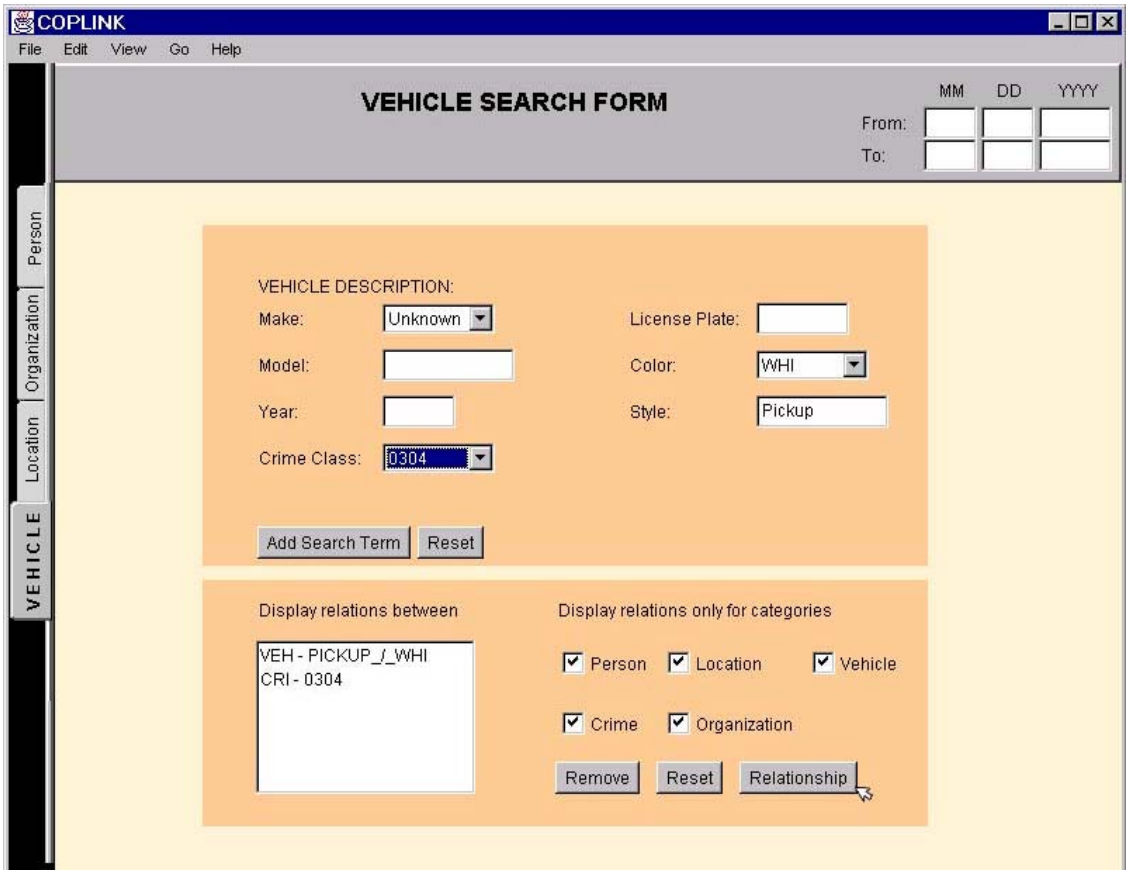

Figure 1: COPLINK Concept Space Search Screen

Using COPLINK Concept Space, users are able to enter one of four information types as a search term. In this example, the detective needs to generate a lead, given the type of crime and the use of a white pick-up truck. The detective selects the Vehicle search screen and enters "White" for color, "Pickup" for style, and "0304", the universal crime report classification code for robbery of a convenience store. After adding the search terms to the relations box, the detective selects the Relationship button to enter the Concept Space. Note that the user can choose to select or deselect the types of relations returned by the system. This allows the user to choose only relevant categories and control for information overload.

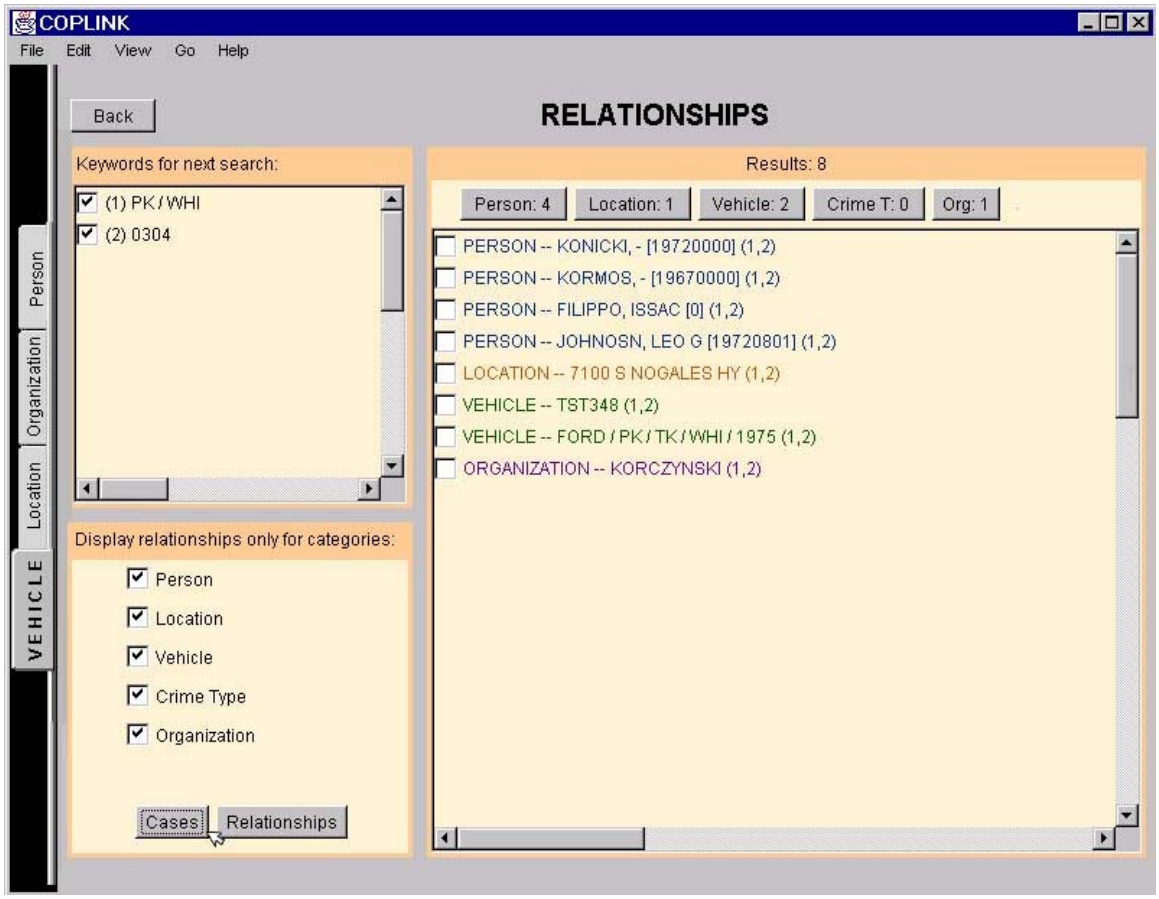

Figure 2: COPLINK Concept Space Result Screen

The system returns eight terms related to both a white pick-up and the 0304 crime type. Note that the Concept Space can return elements for each of the five information object types. The detective now knows not only that are there four people somehow related to both this type of crime and vehicle, she also has a license plate number for a vehicle. The detective can always add any of the Concept Space terms to the search or remove one of the two keywords from the search. As on the initial search screen, the panel in the lower left-hand corner allows users to control the amount of information returned by the Concept Space. The detective decides to view the cases that underlie the relationships uncovered by the Concept Space.

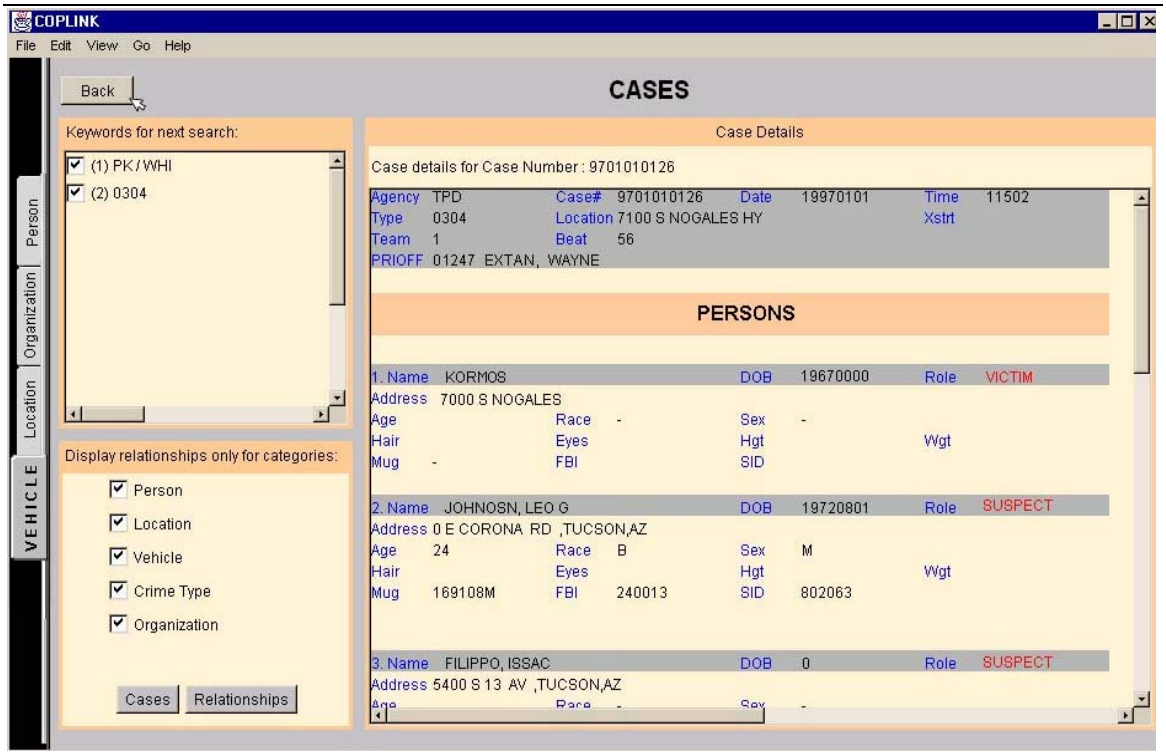

Figure 3: COPLINK Concept Space Case Details Screen.

The Cases view displays actual case reports; in this example, there is only one case in the database. The detective can view the details of the prior incident, including the role of each person involved, and their home addresses. At any time, the detective can choose "Back" to review previous screens or modify the search keywords by selecting another type of search term or deselecting the current search terms.

5. User Evaluations for the COPLINK Concept Space

We conducted user evaluations to examine the effects of COPLINK CS on law enforcement investigation and work practices[9]. Twelve crime analysts and detectives, participated in the four-week longitudinal evaluation, during which they were asked to complete journal entries on searches they had conducted using COPLINK CS. By utilizing data collection methods of documentation, structured interviews, and direct observation, we were able to evaluate the function and design of the COPLINK CS system.

The journals and interviews revealed three major areas in which COPLINK Concept Space provided support for intelligence analysis and knowledge management.

5.1 Link analysis and Summarization

Participants indicated that Concept Space served as a powerful tool for acquiring information and cited its ability to determine the presence or absence of links between people, places, vehicles and other object types as invaluable in investigating a case[10]. The impact of link analysis on investigative tasks is crucial to the building of cases. An officer assigned to investigate a crime has to have enough information to provide a lead before he/she can begin working. Too many cases have to be closed because of lack of information or inability to utilize information existing elsewhere in the records management system. Concept Space manages all the data in the records system in such a way that it can be used as knowledge about the suspect. Link analysis can represent one of three types: directly linking known information, indirectly linking known

information, and linking unknown information. Participants also reported they could use a concept space as a summary of the different information types related to a search term.

5.2 Interface Design

In general, users reported that the web-based interface of the COPLINK Concept Space was engaging and quite easy to use. Officers said the use of color to distinguish different object types and a graphic user interface provided a more intuitive tool than the text-based RMS system. Additionally, the ability to have results returned as either the concept space or case details allowed them to specify the type of information they needed. Participants reported that the data fields chosen for the Concept Space embody the basic necessary information for an investigation. They also reported that the separation between different fields in the output effectively encouraged easy comprehension of the information. A criminal investigation usually requires officers to make specific connections between people, places, vehicles, etc. in order to build a complete picture. The ability to aggregate information fields for searching provides a potent tool for problem solving and crime investigation.

5.3 Efficiency

Perhaps one of the most crucial benefits of the use of COPLINK Concept Space in law enforcement is its speed. As one of our participants explained, identifying a suspect between 48 to 72 hours after a crime is difficult. Beyond this time frame, a suspect is able to destroy evidence that may tie him/her to the crime or change his/her appearance to avoid identification. Witness/victim memory of the suspect's appearance also fades within this period. Identification of the suspect ideally should occur within 48 hours of the crime, so establishing useful links for identifying and locating the suspect is a crucial step. A number of interview and journal comments indicated that use of COPLINK Concept Space increased productivity by reducing time spent per information search.

In journals and interview sessions, each participant was asked to report the time it took to complete at least one particular search task using both RMS and COPLINK CS. The data indicated that in direct comparison of 15 searches, use of COPLINK Concept Space required an average of about 30 minutes less per search than with the use of RMS. However, review of other qualitative data from the journals and interviews indicated that subjects perceived much quicker response to a query from COPLINK CS, especially when multiple search entries and query expansion were involved.

5.3.1 Multiple Search Entries

The COPLINK Concept Space allows entry of multiple search terms, whereas this search capability is not possible in the current RMS system that forces an officer to conduct a number of single searches, then manually compare them. This can take a few hours of work to accomplish what Concept Space can do within seconds, as was the case for searches demonstrating large differences between Concept Space and RMS reported times.

5.3.2. Query Expansion

Users are able to add to the search any terms returned from the concept space. Point-and-click action to add any number of search terms allows users to expand searches quickly and easily and allows officers to explore more searches in a shorter amount to time. In addition, users can view concept spaces or documents on terms returned from previous searches without having to actually type in the query.

For more information on user evaluations for the COPLINK project, refer to [9]

.

During user evaluation, we also looked at application of COPLINK Concept Space to real-life crimes. An example is the real-life case of a shooting reported to us by a Tucson Police Department Crime Analyst:

"The Tucson Police Department had responded to a shooting with the victim in critical condition. Although there were no witnesses at the crime scene, an anonymous caller contacted the police with some information. The caller did not know the shooter's name, however he did know that the shooter had a sister (name unknown) who was a victim in a domestic violence case. The caller was able to provide the identity of the arrestee in the domestic violence (DV) case.

Without the COPLINK Concept Space, success for this type of search would be difficult. I would have to query the DV arrestee in the system, pull up each case to see if it is a DV case; if so enter the case to view the people involved. If I found possible females, I would then have to repeat the process for each of their names. Depending on the number of cases each person is involved in, this manual process could easily take a few days to complete.

Using the COPLINK Concept Space, I entered the DV arrestee's name and the crime code for a domestic violence incident and searched the Concept Space. In a few seconds, I was looking at a list of people associated with the DV arrestee. I located a woman who was a victim in a DV case and ran the Concept Space on her. Sure enough, I found another male who was associated with her in a prior case. Checking his background and previous cases, I found out that he was indeed her brother and was a likely suspect in the shooting. The entire process using COPLINK Concept Space took about five minutes. It is very rare that with such limited information, we are able to generate a probable lead. COPLINK Concept Space will definitely help us to develop leads in even the most difficult of cases."

5.4 Current Status

Currently, COPLINK Concept Space has been deployed at Tucson Police Department, where crime analysts, officers, detectives and sergeants from at least 16 different departmental units are voluntarily using the technology as part of their daily investigative routine. This cross-section population comprises approximately 30-40% of the investigative units at TPD, and includes 90% of the crime analysts. Investigative units included sexual assault (adult and child), aggravated assault, auto theft, elder abuse, community office, fraud, gangs, homicide, undercover, neighborhood crimes, night detectives, patrol, robbery, and warrants. While information and training sessions were offered, installation and use of COPLINK at TPD is voluntary. A number of users learned of the availability of the application from their colleagues, and figured out how to install and use the system on their own.

During a seven-week period, we closely monitored the usage of COPLINK Concept Space, documenting statistics on pattern and frequency of use, as well as following up with real case activities. During the seven-week period, 965 queries were conducted using COPLINK Concept Space. Eighty-four percent of the searches were completed in less than three seconds. The most common type of search conducted consisted of person queries (79%). These measures of efficiency are supported by a number of comments from the users (e.g., "It took less than one minute to find all the associations and incidents associated with the suspect," and COPLINK Concept Space is "100% quicker").

Evidence of COPLINK Concept Space's ease of use was also found. Training that was provided took little time (in minutes) and a number of users installed and learned to use the system on their own (e.g., "Very useful; very easy; very user friendly," and "I could use it without training"). Perhaps most important is the effectiveness of COPLINK Concept Space in the investigation of real crimes, with an improved case closure and solvability (e.g., "You get more information utilizing less effort"). One detective investigating a gang shooting, informed us of the impact that COPLINK Concept Space had on her investigation. The information that the detective had was a moniker or alias of the suspected shooter. Unfortunately, the moniker was one commonly used by a number of gang members. Using COPLINK Concept Space, the detective cross-referenced the moniker with the name of the victim, searching for associations. The detective then generated a "photo line-up" using COPLINK criminal mug shots. The witness identified the suspect from the line-up, leading to the apprehension and arrest of the suspected shooter. The

investigation using COPLINK and the generation of the photo line-up took the detective about 5 to 10 minutes. Without the Concept Space, the search could take days with the assistance of crime analysts.
The Tucson Police Department is in the process of deploying COPLINK Concept Space for use by its 100+ detectives and crime analysts in the entire department. In addition, several large law enforcement agencies in the states of California, Texas, and Michigan have contacted our research team for use of COPLINK in their investigative work.

5.5 Security and Privacy

An important consideration for the law enforcement domain is security and privacy of the transference of information. With COPLINK, we have ensured security and privacy by a number of different measures. First, users access the system by hitting only the COPLINK node; this ensures that the users' queries never penetrate directly to the existing underlying database systems. This allows agencies to enhance security by being able to select the portion of information they wish to share, while protecting other information, such as local personnel data and pending investigations. Second, utilizing intranet/extranet solutions, such as firewalls, IP address check, user password authorization, and database audit trails, allows for the tracking of system access and usage. Information is encrypted and compressed using HTTPS/IP over a dedicated line or virtual private network (VPN). For example, in Tucson Police Department, COPLINK resides behind a firewall. Only workstations with certain IP addresses are allowed access, with individual user password protection. In addition, audit trails of all queries run on COPLINK are captured, allowing system administrators to monitor all user activities. Finally, privacy of the information is controlled locally, in that the host agency determines the sharing via a policy-guided solution and no intelligence information or sensitive information is ported into the node. For example, although the gang unit captures much information on potential gang members, some of this information may be classified as intelligence (i.e., sensitive tactical and strategic information on potential but not verified criminal activity), and therefore cannot be released to the entire department. Following Arizona statutes, gang information selected to be included in COPLINK must meet at least two of the seven Gang Membership Identification Criteria (GMIC) codes.

6. Future Directions for COPLINK

Criminals are creatures of habit and being able to understand their habits and close associations is important[11]. The COPLINK Concept Space takes advantage of this characteristic by capturing connections between people, places, events, and vehicles, based on past crimes. Our initial evaluation of this intelligence analysis application supports its potential for transforming law-enforcement knowledge management practices in this age of digital governments. We have also designed the COPLINK Concept Space to be a scalable and powerful tool for other federal and local law enforcement agencies. In addition to the Concept Space, we are currently developing a number of other technologies for the law enforcement community.

Although this particular project does not utilize entity extraction techniques since the data was drawn from a structured database system, large collections of unstructured text as well as structured case-report information exist in many police records systems. These textual sources contain volumes of information for investigators that are often not captured in the structured fields. One future research direction is to explore the development of textual mining approaches that support knowledge retrieval from such sources for law enforcement case reports. In order to perform a fine-grained content analysis, we will investigate the development of linguistic analysis and textual mining techniques that make intelligent use of large textual collections in police databases.

In the entity extraction field, there are a number of researchers exploring different techniques. For example, the MENE system[12], developed at NYU, utilizes a rule-based approach in combination with a feature importance component supported by maximum entropy probability

calculation. Another approach is that of NetOwl™, which uses name recognition rules with a name database lookup[13]. Although both these systems have performed well on specific types of text, because it is rule-based, generalizability may be problematic when applied to different domains. With police report narratives, generalizability will be a key in system performance. We plan to use a unique combination of noun phrasing, finite state automata (FSA) system, and artificial neural networks (ANN) to accomplish the task of entity extraction. We are currently working with Phoenix Police Department with their collection of criminal narratives to develop and test our entity extraction technique.

Several Internet research projects have shown the power of a new "agent" based search paradigm. In addition to supporting conventional searches performed by users, search agents allow users automatically to establish search profiles (or create profiles for users) and extract, summarize, and present timely information content. We believe such a proactive search agent is well suited to use by investigative personnel in law enforcement agencies. Search agents for law enforcement can support conventional searching techniques and be profiled for specific investigations. We plan to develop a personalized law enforcement search agent that will support wide expansion in connectivity and information sharing between police agencies.

As distributed solutions and analysis tools are developed for law enforcement officers, a specific focus must be on providing tools within the constraints of a wireless environment. One of our future goals is to develop and refine applications to support the expansion of distributed and mobile law enforcement networks and inter-jurisdictional information retrieval as well as to investigate and study network security issues.

Acknowledgments

This project has been funded by grants from the National Institute of Justice, Office of Science and Technology (#97-LB-VX-K023) and the National Science Foundation (#9983304) with support from the Digital Equipment Corporation External Technology Grants Program, agreement #US-1998004, for its award of an equipment grant. Appreciation also goes to Joanne Martinez and Kristen Tolle for their feedback as well as Sgt. Jennifer Schroeder, Officer Linda Ridgeway and all the other personnel from the Tucson Police Department who were involved in this project.

References

1. D. C. Blair and M. E. Maron, "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System," *Communications of the ACM*, Vol. 28, No. 3, 1985, pp. 289-299.
2. P. Jones and J. Jordan, "Knowledge Orientations and Team Effectiveness," *International Journal of Technology Management*, Vol. 16, No. (1-3), 1998, pp. 152-161.
3. Y. M. Chen, C. C. Liao, and B. Prasad, "A Systematic Approach of Virtual Enterprising Through Knowledge Management Techniques," *Concurrent Engineering-Research and Applications*, Vol. 6, No. 3, 1998, pp. 225-244.
4. A. C. Inkpen, and A. Dinur, "Knowledge Management Processes and Joint International Ventures," *Organization Science* Vol. 9, No. 4, 1998 , pp.454-468.
5. H. Chen, "Artificial Intelligence Techniques for Emerging Information Systems Applications: Trailblazing Path to Semantic Interoperability," *Journal of the American Society for Information Science,* 1998, Vol. 49, No. 7, pp. 579-581.
6. M. Lesk, *Practical Digital Libraries*, Morgan Kauffmann, Los Altos, CA, 1997.
7. H. Chen, and D. T. Ng, "An Algorithmic Approach to Concept Exploration in Large Knowledge Network (Automatic Thesaurus Consultation): Symbolic Branch-and-Bound vs. Connectionist Hopfield Net Activation," *Journal of the American Society of Information Science*, Vol. 46, No. 5, 1995, pp. 348-369.
8. H. Chen and K. J. Lynch, "Automatic Construction of Networks of Concepts Characterizing Document Database. IEEE Transaction on Systems," *Man and Cybernetics*, Vol. 22, No. 5, Sept./Oct. 1992, pp. 885-902.

9. R. V. Hauck and H. Chen, “COPLINK: A Case of Intelligent Analysis and Knowledge Management,” *Proceedings of the 20th Annual International Conference on Information Systems,* 1999, pp. 15-28.
10. W. R. Harper and D. H. Harris, “The Application of Link Analysis to Police Intelligence,” *Human Factors,* 1975, Vol. 17, No. 2, pp. 157-164.
11. N. Joyce, “ICAM: Chicago's Newest Crime-Fighting Tool,” *Proc. of the Conference in Technology, Community Policing*. National Law Enforcement and Corrections Technology Center, 1996, available at http://www.nlectc.org/txtfiles/confrpt.html
12. A. Borthwick, J. Sterling, E. Agichtein, and R. Grishman, “NYU: Description of the MENE Named Entity System as used in MUC-7,” *Proceedings of the Seventh Message Understanding Conference (MUC-7)*, 1998.
13. G. R. Krupka and K. Hausman, “Description of NetOwl™ Extractor System as Used in MUC-7,” *Proceedings of the Seventh Message Understanding Conference (MUC-7)*,1998.