

COPLINK: Information and Knowledge Management for Law Enforcement

Hsinchun Chen^{a*}, Roslin V. Hauck^a, Homa Atabakhsh^a, Harsh Gupta^a, Chris Boarman^a,
Jennifer Schroeder^b, Linda Ridgeway^b

^aDept. of Management Information Systems, University of Arizona, Tucson

^bTucson Police Department, Tucson, Arizona

Photonics East Conference, SPIE, Technologies for Law Enforcement; Boston Nov. 5-8, 2000

ABSTRACT

The problem of information and knowledge management in the knowledge intensive and time critical environment of law enforcement has posed an interesting problem for information technology professionals in the field. Coupled with this challenging environment are issues relating to the integration of multiple systems, each having different functionalities resulting in difficulty for the end user. COPLINK offers a cost-efficient way of web enabling stovepipe law enforcement information sharing systems by employing a model for allowing different police departments to more easily share data amongst themselves through an easy-to-use interface that integrates different data sources. The COPLINK project has two major components: COPLINK Database (DB) Application and COPLINK Concept Space (CS) Application. The COPLINK DB design facilitates retrieval of case details based on known information. COPLINK CS is an investigative tool that captures the relationships between objects (e.g., people, locations, vehicles, organizations, crime types) in the entire database allowing investigators and detectives to perform investigative associations and case analysis. This paper describes how we have applied the design criteria of platform independence, stability, scalability, and an intuitive graphical user interface to develop the COPLINK systems. Results of user evaluations that have been conducted on both applications to study the impact of COPLINK on law enforcement personnel. The COPLINK DB Application is currently being deployed at the Tucson Police Department and the Concept Space is undergoing further modifications. Future development efforts for COPLINK project will also be discussed.

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

1. INTRODUCTION

1.1. Law Enforcement 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 lives¹. 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 information² 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 enterprising³, joint ventures⁴, aerospace engineering², and digital libraries⁵.

*Correspondence: Email: hchen@bpa.arizona.edu; WWW: <http://ai.bpa.arizona.edu>; Telephone: 520 621 2748;
Fax: 520 621 2433

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. The Artificial Intelligence Lab has gained wide recognition as a cutting-edge research unit and has been featured in *Science* and *The New York Times*, having received more than \$9M in research funding from various federal and industrial sponsors since 1989. The Lab sees COPLINK as 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.

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 Tucson Police Department (TPD) has encountered all the problems described in the previous section. Its information sources have included at least three distinct systems:

- The main incident-based system, Records Management System (RMS) captures the highlights of an incident in an Oracle 7.x database.
- A separate system by ImageWare Software Inc. captures mug shots (photos taken at the time of arrest) and limited related information in a Sybase database.
- A third information source, Criminal Information Computer (CIC) is a homegrown Microsoft Access-based application used to track gang activity. TPD officials attribute a disproportionate percentage of Tucson's criminal activity, especially homicides, to gang members and their known associates.

RMS contains approximately 1.8 million incident record sets and mug shot records (around 60000 mugs). CIC tracks the approximately 1200 individuals the department considers responsible for a majority of major crimes. Each of these systems has a different user interface, so accessing related information from any two or all three, has been difficult, cumbersome, and time-consuming:

- RMS has a cumbersome, difficult-to-navigate command-line driven system.
- CIC's gang database has been accessible only to certain detectives through a simple homegrown front-end interface.
- Mugshot database, a collection of arrest photographs, can only be integrated with information in RMS manually through a specific mug shot number.

As an NIJ-funded multi-year project, the major goals for the COPLINK project for TPD are:

- To develop an integrated system to allow TPD officers easy access to all the information contained in all three systems.
- To design a prototype system for use in developing similar systems at other police departments.
- To offer a model for allowing different police departments to share data easily.

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. While the scope of this project includes a multilevel development plan incorporating different information technologies, the focus of the research reported here is not only on the development of a multimedia database system to promote information sharing, but also the improvement of criminal intelligence analysis. The first step in this process was the evaluation of TPD's current Records Management System (RMS).

1.2.1 TPD's Records Management System (RMS)

The main database at TPD is the Records Management System (RMS), which stores a wide variety of data, including criminal case information and incident information from calls for service recorded from the Department's Computer-Aided Dispatch (CAD) system. RMS is a text-based system that is accessed using VAX terminals stationed in many central headquarters offices as well as substations located around the city.

Similarly to systems at many other law enforcement agencies, RMS has many problems pertaining to its interface, access to information and lack of knowledge management. Although users are able to search on name queries, location queries, vehicle queries, etc., they are not able to search multiple fields simultaneously. In addition, users of RMS complain that, depending on the type of query, RMS can take from a few minutes to a few hours to return its results.

1.2.2 Current TPD Knowledge Management Practice

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 1.8 million criminal case reports containing details from criminal events dating back to 1986. Although investigators can access RMS to tie together information needed to solve cases and crimes, they must manually search for connections or relationships in 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. The problem is not necessarily that the information has not been captured—any officer who fills out up to seven forms per incident can attest to that. The problem is one of access. Typically, law-enforcement agencies have captured data only on paper or have fed it into a database or crime information system. If the agency involved has more than one of these (that are possibly incompatible), information retrieval can be difficult or time-consuming.

Potent information retrieval tools can provide information sharing abilities as well as 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 two knowledge management systems that can provide the ability to access data from different systems as well as provide the functionality of intelligence analysis that currently does not exist in the RMS system. Real-life context evaluation of both systems, the COPLINK Database and COPLINK Concept Space, and future directions for the project are also discussed.

2. COPLINK DATABASE APPLICATION

After analyzing user requirements, we created the COPLINK Database (DB) application, employing a consistent and intuitive interface which integrates different data sources, such that the multiplicity of data sources remains completely transparent to the user, allowing law enforcement personnel to learn a single, easy-to-use interface. In addition to the interface design, we also developed a model that allows for information sharing both within and between law enforcement organizations.

2.1 Design Criteria

The main design criteria considered for the COPLINK project includes:

- **Platform independence:** Because not all police departments utilize the same hardware or software operating systems, platform independence was critical.
- **Stability and scalability:** The system also had to offer room for system growth and expansion.
- **Intuitive and ease of use:** The front-end user interface should be intuitive and easy to use, yet flexible enough to meet the equally demanding investigative needs of detectives and officers.

Typical law enforcement applications usually are legacy systems having out-dated performance and capability. For example, TPD's RMS took 30 seconds to answer simple requests and up to 30 minutes for more complex queries. Improved response time was critical to restoring departmental efficiency. To ensure application speed, issues of data and network communication, disk access and system I/O needed to be addressed. This also meant carefully distributing logic where it could be most quickly and efficiently executed, i.e., all user-input error checking should be done in the front end, and all database access logic achieved through pre-compiled stored PL/SQL procedures in the database.

Another critical issue, especially in designing a system that could be deployed across multiple law enforcement agencies, was acknowledging that no two agencies would store their incident data in exactly the same way. Therefore, it was important to come up with a data organization design that was flexible enough to be applied to any underlying data set. The database team designed a series of standardized “views” that fitted typical information search and presentation situations. For example, most of the data in the TPD systems were related to “Person,” “Location,” “Vehicle,” or “Incident” information. A set of views was developed for each of these areas of interest, with the underlying data sets mapped to those standard views, making the system more portable to other law enforcement agencies.

2.2 Database Design

Based on the criteria established and after much investigation, the COPLINK team decided upon a three-tier architecture (see Figure 1):

- A front-end interface: The front-end should be a thin client, consisting of a series of user-friendly query screens matching the four main areas previously discussed (Person, Location, Vehicle, and Incident). The front-end would generate query requests.
- A middle-ware application server: The middle-ware would handle secure requests from multiple clients, and execute the stored procedures in the database.
- A back-end database: Results from the database would be processed by the middle-ware, and be formatted into return data strings. These return strings would then be sent to the front-end where they would be parsed and displayed to the user.

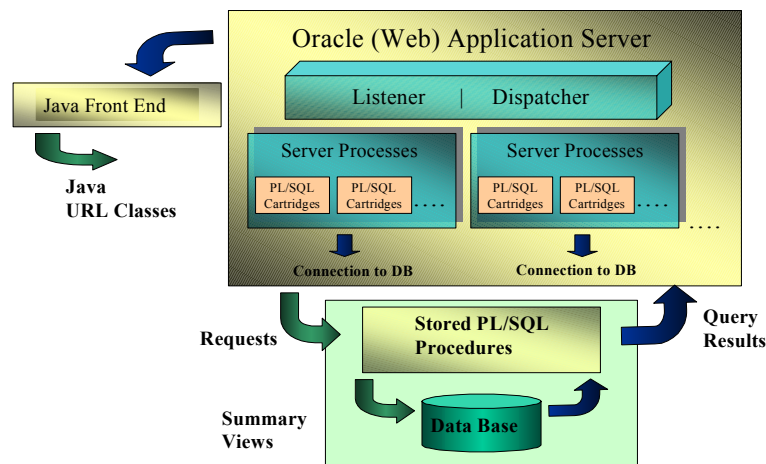


Figure 1: COPLINK Database Three-tier Architecture

As mentioned, the front-end had to be a platform-independent thin, stable client, based on a popular programming language. Our current prototype, created using Java 1.1, allows for client-side analysis, avoiding the overhead incurred by database operations. Oracle’s Application Server (OAS) met our middle-ware needs. It has versions available for both Windows NT and UNIX-based systems and utilizes a CORBA-based “cartridge server” system. A cartridge server is a shared library that either implements program logic or provides access to program logic stored elsewhere, such as in a database. In implementing the COPLINK application, we utilized the PL/SQL cartridge system of the OAS, which gives access to the logic stored as pre-compiled PL/SQL procedures in the database. The procedures actually execute the queries in the database, and return the results to the front-end application as HTTP-based strings. Although this system appears to be Oracle-centered, it has flexibility that allows us to access non-Oracle databases whereas such a cartridge as ODBC could only be used to access an ODBC-compliant database. The database system was designed to be compatible with either Oracle 7.3 or 8.0, and different versions of the data sets have been run on Windows NT and Dec Alpha UNIX platforms. The major portions of the database consist of tables and indices that contain incident-based information, the set of views discussed previously, a series of procedures used by the middle-ware to query the database, and the packages necessary to execute queries from the OAS.

There are four main query screens, each resulting in a summary listing of information related to an initial query. Figure 2 illustrates relationships among queries. For example, if a user initiates a search on a particular first-name/last-name

combination, a summary table is presented as a result of a dynamic SQL query, listing all possible matches, as well as the number of incidents associated with each individual match. From there, the user can select either a secondary listing of incidents related to a particular individual or can access a more detailed summary of the personal information on the individual. For an incident summary, all the pertinent case detail information on a particular incident is presented. For a detailed person summary, the user can select the incident summary for that individual, and from there obtain case details for any incident listed.

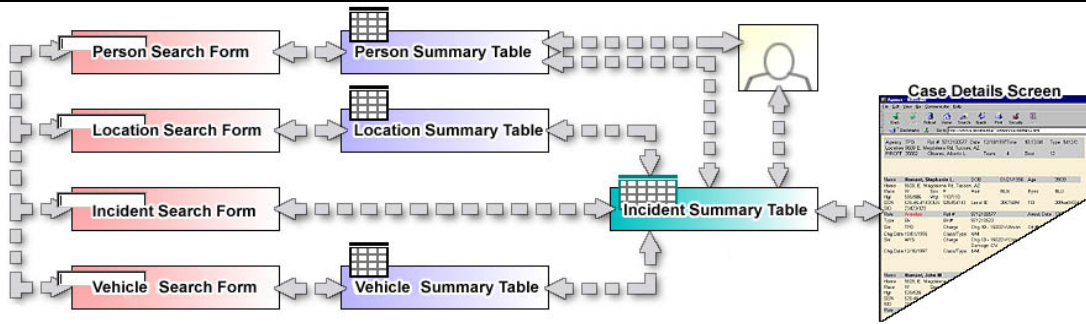


Figure 2 - Screen Flowchart

An officer wanting to know more about a particular incident or person can enter a query in the search form, query further through the summary table to see details about a person, or select an incident from the incident summary table to view on the case details summary screen. In previous screens, information could be displayed in formatted rows, but a more dynamic display was needed. For example, mug shots needed to be displayed both as person details and on the case-details screen. To accommodate this feature, screens have been laid out in clusters, grouping information for easier understanding. This in turn required manipulating the data retrieved and capturing pictures from the database, a problem solved by constructing a cyclical procedure that would loop through the data and build a hierarchical tree. We could then apply display patterns to the nodes of the tree, navigate the tree and place the information on the screen.

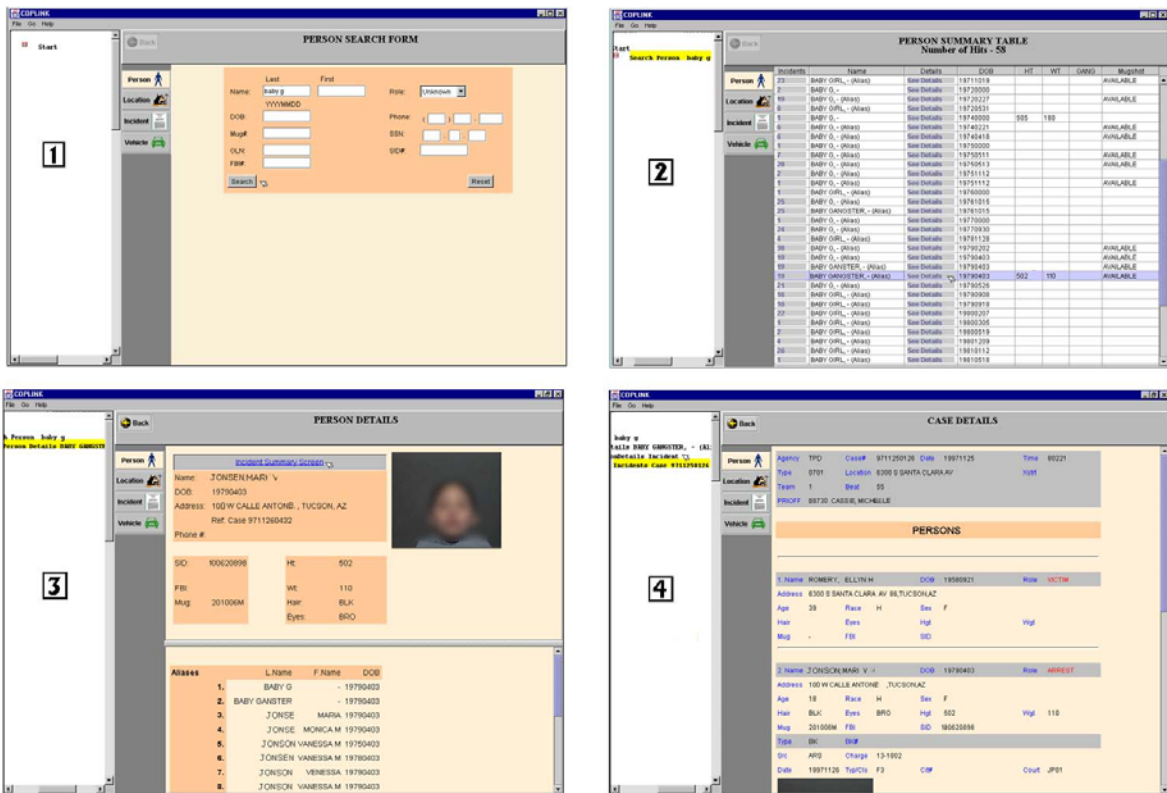


Figure 3 – COPLINK Database Interface Screen Samples-- Sample Scenario: An officer is trying to identify a suspect involved in an automobile theft. A confidential informant has reported that the suspect goes by the street name of “Baby Gangster,” is about 20 (probably born in 1979), and is around 5’3” tall.

Screen 1- COPLINK DB Search. The officer can choose one of the four types of information upon which to search: Person, Location, Incident, or Vehicle. The officer selects the Person search screen and enters “baby g” in the COPLINK DB system. Note the left panel history screen, which keeps track of the user’s searches. Screen 2 - Person Summary. The system returns 58 listings referring to “baby g.” (all of the returns include the name “baby g.”) The system permits sorting by any of the column headings in the table. The officer chooses to sort by date of birth and finds an entry for “baby gangster,” born in 1979, whose height is 5’2”. The officer then clicks on the “See Details” button to find out more about this particular “Baby Gangster”. Screen 3- Person Details. This screen contains personal information about the selected person, including real name, latest description information, latest home address, other identifiers that the person may use, and a mug shot, if available. The officer now has a real name of a person who matches the description of the possible suspect he was given. The officer then decides to go to the incident summary screen to get an idea of the cases in which this person has been involved. Screen 4- Incident Summary. This screen displays all the incidents in which the selected person has been involved. The officer sorts by crime type, looking for cases of stolen vehicles (0701) and finds the suspect has been involved in four such incidents, either as a suspect or as an arrestee. The officer selects Case #9711250126 to look at the actual case information.

2.3 Graphical User Interface for COPLINK Database

The graphical user interface (GUI) for the COPLINK Database Application is shown in Figure 3, although actual information has been altered to maintain data confidentiality. The Java front-end consists of two major parts, the input and display of data and the processing of information. Working closely with TPD officers, the COPLINK team first made low-fidelity, paper prototypes of the screens used to obtain feedback on the display and organization of the information, which was used to modify the design and functionality of the interface. Display of results was important to the front-end. We learned that a user’s idea of what constitutes a manageable and intuitive display varied with the query type and sometimes required formatting in a different way. We responded by creating a dynamic text table, using the Java API to make the interface more flexible.

2.4 User Evaluations for the COPLINK Database Application

A usability evaluation, involving 52 law enforcement personnel, was conducted to assess the achievement of a number of the goals that guided the design and development of the COPLINK Database. Items on the questionnaire used to assess and compare the COPLINK and RMS systems were based upon user perceptions of such widely used measures of usability as: *effectiveness* (impact of system on job performance, productivity, effectiveness of information, and information accuracy), *ease of use* (measures of effort required to complete a task, ease of learning how to use the application, ability to navigate easily through the different screens, and satisfaction with the interaction), and *efficiency* (speed of completing tasks, organization of the information on the screens, ability to find information and the interface design itself)⁶.

Benchmark levels from TPD’s current RMS system for all three usability factors were established and compared with COPLINK DB ratings. In addition to written questionnaires, observation of the data collection methods and structured interviews were used both to supplement findings and to provide feedback for further development efforts.

Data analysis of the usability questionnaire support a conclusion that use of COPLINK DB provided improved performance over use of the current RMS system. On all usability measures (effectiveness, ease of use, and efficiency), participants rated COPLINK DB higher than RMS, with the average rating for COPLINK being 4.1 and RMS being 3.3 (1=strongly disagree to 5=strongly agree). Statistical analyses revealed that this ratings difference was significant for all measures.

From both questionnaires and interviews, participants indicated that the quality and quantity of information from COPLINK DB surpassed those of RMS. In a review of current RMS practices, a number of detectives and officers were actually unable to use RMS but were able to use COPLINK DB to conduct searches. It is evident from this evaluation that COPLINK DB allowed a population of TPD personnel to access information that would have been quite difficult for them to acquire using the RMS system. From both the questionnaire and the interview data collected from this evaluation, it is evident that many participants rated the information found in COPLINK as more useful than the information in RMS. This finding is very interesting, because most of the information contained in COPLINK has been taken from RMS.

3. COPLINK CONCEPT SPACE

To complement the functionality of the COPLINK Database application, our next phase of the COPLINK project was to develop a knowledge management tool specifically designed to aid investigators and detectives in criminal intelligence analysis. Drawing upon artificial intelligence techniques and algorithms, the COPLINK Concept Space was created.

Concept space, or automatic thesaurus, is a statistic-based, algorithmic technique used to identify relationships between objects (terms or concepts) of interest⁷. 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^{5,8}. 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 functions⁹ 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 CS, detailed case reports are the underlying documents 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. Figure 4 provides samples of the COPLINK Concept Space application.

3.1. 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 ⁹). 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.

3.1.1 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 TPD personnel, we created a set of term types for the COPLINK Concept Space. Table 1 shows the term types supported by Concept Space and the size of each.

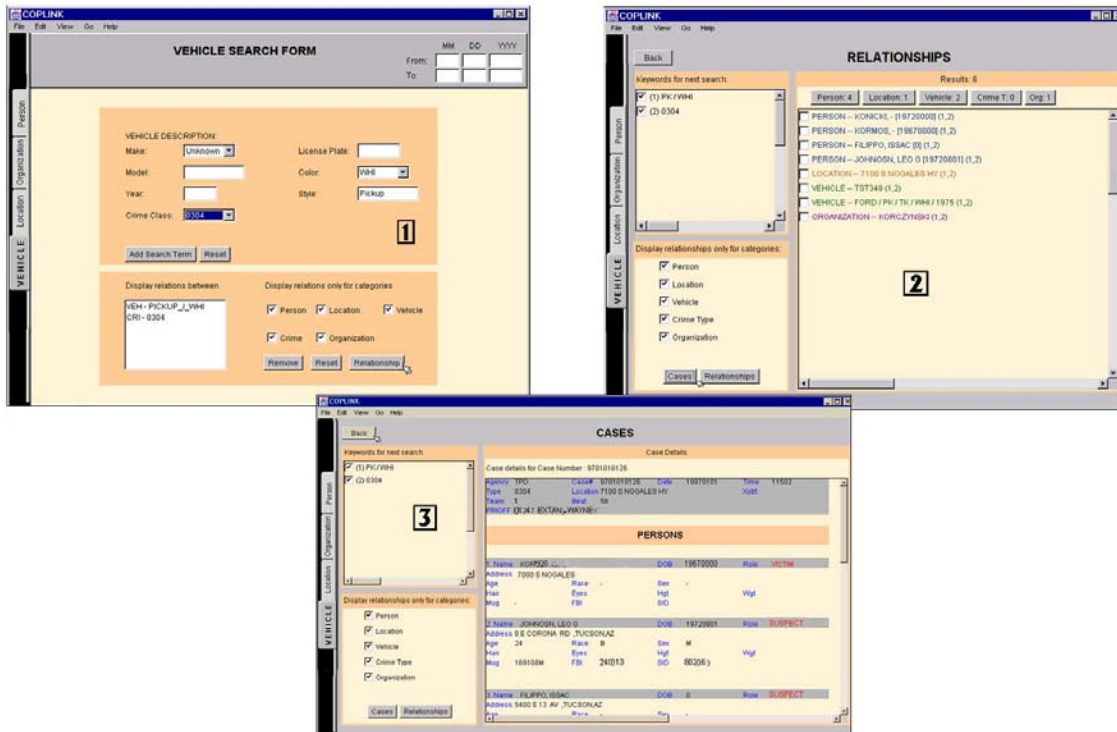


Figure 4 – COPLINK Concept Space Interface Sample Screens-- Sample Scenario: A detective is investigating a robbery at a local convenience store. The only witness, the night store clerk, only remembers that the suspect drove away in a white pickup truck.

Screen 1 - COPLINK CS Search. 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. Screen 2 – COPLINK CS Results. 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 there are 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. Screen 3 – CS Case Details. 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.

Term types in Concept Space can be divided into the 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.

General	Category	Type	Size
Total Case Reports			1,504,838 cases
Total Unique Terms			1,267,776 terms

	<i>Person</i>	Full Name	744,250 terms
	<i>Organization</i>	Organization Name	26,517 terms
	<i>Location</i>	Street Address	141,875 terms
		Street Address with Apartment	58,664 terms
	<i>Vehicle</i>	License Number	202,996 terms
		Vehicle Identification Number	57,547 terms
		Make / Color	1,924 terms
		Make / Model	908 terms
		Make / Model / Color / Year	30,142 terms
		Make / Model / Style	2,749 terms
		Make / Model / Year	6,500 terms
		Make / Style / Color	7,334 terms
		Make / Style / Color / Year	39,708 terms
		Make / Style / Model / Color / Year	41,757 terms
		Make / Year	2,773 terms
		Style / Color	938 terms
		Style / Year	1,609 terms
	<i>Incident Type</i>	Crime Type	374 terms
		Team / Beat (Geographical area)	147 terms
Average Number of Terms per Case			5.43 terms
Number of Associated Terms			27,707,675 pairs

Table 1: Statistical Information on the COPLINK Concept Space

3.1.2 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 Chen⁹. 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. A few adjustments were made to the standard term frequency and inverse document frequency measures.

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:

$$d_{ij} = tf_{ij} \times \log\left(\frac{N}{df_j} \times w_j\right) \quad (1)$$

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 Lynch⁹.

$$W_{jk} = \frac{\sum_{i=1}^n d_{ijk}}{\sum_{i=1}^n d_{ij}} \times \text{WeightingFactor}(k) \quad (2)$$

$$W_{kj} = \frac{\sum_{i=1}^n d_{ikj}}{\sum_{i=1}^n d_{ik}} \times \text{WeightingFactor}(j) \quad (3)$$

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:

$$d_{ijk} = tf_{ijk} \times \log\left(\frac{N}{df_{jk}} \times w_j\right) \quad (4)$$

$$d_{ikj} = tf_{ijk} \times \log\left(\frac{N}{df_{jk}} \times w_k\right) \quad (5)$$

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.

$$WeightingFactor(k) = \frac{\log \frac{N}{df_k}}{\log N} \quad (6)$$

$$WeightingFactor(j) = \frac{\log \frac{N}{df_j}}{\log N} \quad (7)$$

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 for each case). These narratives will provide a fertile test bed for combining noun phrasing and Concept Space analysis for intelligence identification.

3.2 User Evaluations for the COPLINK Concept Space

We conducted user evaluations to examine the effects of COPLINK CS on law enforcement investigation and work practices¹⁰. 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.

3.2.1 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 case¹¹. 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.

3.2.2 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.

3.2.3 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.

4. FUTURE DIRECTIONS FOR COPLINK

Criminals do not bound themselves by county borders or jurisdictions. Furthermore, criminals are creatures of habit and being able to understand their habits and close associations is important¹². The COPLINK Database and Concept Space applications take advantage of these characteristic by not only promoting information sharing between stovepipe information sources and different agencies, but also by capturing connections between people, places, events, and vehicles, based on past crimes. Our evaluation of these knowledge management and intelligence analysis applications support its potential for transforming law-enforcement practices in this age of digital governments. In addition to these two projects, we are currently developing a number of other technologies for the law enforcement community.

Large collections of unstructured text as well as structured case-report information exist in 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.

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.

In relation to the COPLINK project, the concept of a distributed database system has important implications. The most important of these is accessibility to and dissemination of law enforcement records and information. Currently, the vast majority of criminal data collection and compilation is done on a community level but may not be in a format that is readily available and accessible to local law enforcement officers. A distributed COPLINK prototype is under development using three COPLINK database servers to simulate the independent nodes in a distributed environment. Work is under way to include functionality that will provide interoperability among the different DBMS platforms that may support future COPLINK nodes. In the immediate future, we plan to begin deployment and testing of a Distributed COPLINK prototype

with the Tucson and Phoenix police departments. 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 a grant from the National Institute of Justice, Office of Science and Technology (Tucson Police Department, "COPLINK: Database Integration and Access for a Law Enforcement Intranet", \$1.2M, 1997-1999) and the National Science Foundation ("COPLINK Center: Information and Knowledge for Law Enforcement", \$530K, 2000-2001). We would also like to thank the Digital Equipment Corporation External Technology Grants Program, agreement #US-1998004, for its award of a \$198,451 equipment grant allowance toward the purchase of a DecAlpha Server for the COPLINK Project. Appreciation also goes to all the 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* **28**, pp. 289-299, 1985.
2. P. Jones and J. Jordan, "Knowledge Orientations and Team Effectiveness," *Int'l J. of Technology Management* **16**, pp. 152-161, 1998.
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* **6**, pp. 225-244, 1998.
4. C. Inkpen, and A. Dinur, "Knowledge Management Processes and Joint International Ventures," *Organization Science* **9**, pp.454-468, 1998.
5. H. Chen, "Artificial Intelligence Techniques for Emerging Information Systems Applications: Trailblazing Path to Semantic Interoperability," *J. of the American Society for Information Science* **49**, pp. 579-581, 1998.
6. R. V. Hauck, "COPLINK: Exploring usability of a multimedia database application for law enforcement," Report. Available: <http://ai.bpa.arizona.edu/go/datawarehousing/publications/nij.pdf>. 1999.
7. M. Lesk, *Practical Digital Libraries*, Morgan Kaufmann, Los Altos, 1997.
8. 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," *J. of the American Society of Information Science* **46**, pp. 348-369, 1995.
9. H. Chen and K. J. Lynch, "Automatic Construction of Networks of Concepts Characterizing Document Database. *IEEE Transaction on Systems, Man and Cybernetics* **22**, pp. 885-902, 1992.
10. R. V. Hauck and H. Chen, "COPLINK: A Case of Intelligent Analysis and Knowledge Management," *Proc. of the 20th Annual International Conference on Information Systems*, P. De and J. I. DeGross, (Eds.), pp. 15-28, International Conference of Information Systems, Atlanta, 1999.
11. W. R. Harper and D. H. Harris, "The Application of Link Analysis to Police Intelligence," *Human Factors* **17**, pp. 157-164, 1975.
12. N. Joyce, "ICAM: Chicago's Newest Crime-Fighting Tool," *Proc. of the Conference in Technology, Community Policing*, National Law Enforcement and Corrections Technology Center. Available: <http://www.nlectc.org/textfiles/confprt.html>. 1996.