INFORMATION SCIENCE AS A LAW ENFORCEMENT UTILITY

Emil H. Levine
U.S. Department of Justice, Drug Enforcement Administration
Washington, D.C.

Abstract

Information science techniques have previously been used primarily to isolate documents in a collection to support other research. These techniques have an application to law enforcement, particularly in the study of organized crime, narcotics networks and major criminals. The transfer of this technology requires only that information on criminal elements be indexed, stored, manipulated and retrieved in the same manner as any other class of information. Techniques of ISR, Citation Indexing, bibliographic coupling, associative indexing, SDI, full-text processing, network analysis, modeling, simulation and predictive analysis have direct applications to the isolation and study of specific criminals or criminal networks and clusters. The concepts of the Invisible College, the Information Analysis Center and the half-life of information are highly useful. COM and a revolutionary firmware device, the associative processor, will open further information science technology to law enforcement.

INTRODUCTION

Information science techniques have classically been utilized for the retrieval of information to support research in some discipline. These techniques, as described by Borko (1), include determining information needs and uses, document creation and copying, language analysis, translation, abstracting, classification, coding and indexing, system design, analysis and evaluation, pattern recognition and adaptive systems. Related sciences, according to Debons and Otten (2), include mathematics, linguistics, engineering, management science, psychology and computer science. Garfield (5) has cited some uses of Citation Indexing for other than conventional retrieval. The author (9) has also cited one use of these techniques for management information retrieval. Undoubtedly, many other documented and undocumented cases exist where pure information science processes are being used for a variety of purposes other than document retrieval. Hence, information science is becoming, or has become, a generalized tool which can serve as a utility in both a conventional and unconventional mode.

The conventional potential of this utility in law enforcement is obvious. In fact, the Law Enforcement Assistance Administration operates the National Criminal Justice Reference Service using these techniques in the operation of a Federal Clearinghouse for criminal justice publications. The unconventional use of information science as a utility in law enforcement is less obvious, unless major criminals and criminal organizations are thought of as accessions in a collection. If this analogy is correct, and if it can be rationalized by the information scientist, then the technology can be transferred for the isolation and study of specific criminals or networks through associative processing and clustering techniques, and a variety of other sub-utilities. The application of a number of these procedures will be briefly discussed within the context of utilization against organized crime and narcotics, particularly as tools for criminal and narcotics intelligence analysts.

CONVENTIONAL

Information Storage and Retrieval

The most basic technique is conventional ISR for the retrieval of documents to support the classical analytical study of individuals or organizations. Operational systems provide for on-line, interactive queries based on name or other data elements. These queries result in the retrieval of criminal records and listings of manual case files related to the subjects of interest.

UNCONVENTIONAL

The following unconventional techniques are considered applicable.

Citation Indexing

The utilization of Citation Indexing to study science and to develop networks and clusters has been discussed by Garfield and Malin (6). If criminal records are properly structured to show the relationships between subject and associate, a potential exists to similarly show networks and clusters of criminal associations. This is particular applicable for the support of analysis of large narcotics organizations.

Bibliographic Coupling

The analogy to bibliographic coupling follows the applications of Citation Indexing. If

...bject A is associated with B, C and D, and subject E is associated with B, C, and D, there is a high probability that subjects A and E are associated. Because of the high degree of insulation surrounding major organized crime figures, such direct associations are often difficult to ascertain using conventional techniques.

Associative Indexing

The job of the criminal intelligence analyst, unlike that of his military counterpart, is often complicated by too much information rather than the fragments common to foreign intelligence analysis. The techniques of associative indexing introduced by Stiles (16,17,18) in 1961 can assist the analyst by preprocessing information and presenting the researcher with a priority for reading, based on statistical and linguistic analysis of all reports potentially relative to the research. In addition, Stiles' technique can, through a statistical and coupling process, point to reports that may be relevant even though they do not contain the keywords of the initial query. Mintz (10,11,12), in several technical papers, has discussed similar techniques and has isolated further appropriate algorithms for his "Associative Word Mapping." Thus, the analyst receives potentially relevant information that he did not directly request, and he is given an order for reading which will most likely satisfy his query based on statistical processing. William's (19) BROWSER will also perform the latter function.

Concept of Half-Life

Several studies have been conducted on criminal recidivism to justify the retention or purging of criminal records in criminal justice systems. The analogy of the concept of the half-life of information is obvious. Information on major organized crime and narcotics figures will generally have a long half-life, while that on petty and minor criminals an extremely short half-life.

Information Analysis Center

The concept of the Information Analysis Center as related to law enforcement was inherent in Executive Order 11676, which established the Office of National Narcotics Intelligence and called for the creation of a National Narcotics Intelligence System for the pooling and sharing of Federal, State and local narcotics information. Reorganization Plan Number Two, which established the Department of Justice Drug Enforcement Administration, merged and incorporated the requirements of the Executive Order into DEA. Included in the Drug Enforcement Administration is the Office of Intelligence, which will eventually serve as an Information Analysis Center. DEA is the first Federal law enforcement agency to utilize such a philosophy.

SDI

The volume of current criminal reports which are of interest to major law enforcement agencies is massive. A primary method of assisting in the intelligence analysis of this information is through timely and efficient dissemination of these reports to appropriate authorities and analysts. Conventional mail routing techniques are inadequate. If the majority of the reports can be captured in a digital form, either during creation or transmission, an automated SDI can be implemented.

Full-Text Processing

Most law enforcement investigative reports are in a narrative form, with limited spaces for formatted information. In 1965, Farell (4) and Isaacs (7) conducted research with the Los Angeles Police Department on methods of processing this full-text as an investigative aid. Of particular interest were ways of correlating criminal methods of operation by comparing full-text phrases which described the criminal act (e.g., "gun in brown paper sack"). Advances in methods of full-text processing, the development of extensive ADP telecommunications systems and the increased cost of manual indexing has renewed interest in this application to the classical method of preparing law enforcement investigative reports. Sophistication of this technique would assist in the development of SDI and ISR systems.

The Invisible College

Because of the mobility, and often world-wide nature of organized crime, particularly narcotics, several law enforcement agencies may be independently working against the same target without knowledge of each other's efforts. Occasionally, law enforcement officers in the same organization find themselves in this situation. This problem is more complicated than in the scientific world. The law enforcement officer is occasionally forced to work in an environment which has a high degree of corruption, in an undercover role. The very act of seeking information through conventional means may jeopardize the investigation or the life of the undercover agent. An additional factor is the hesitancy of criminal investigators to place certain categories of information into an automated system. Thus, if a counterpart to the scientist's Invisible College could be developed, a number of conventional law enforcement problems could be solved.

A concept, perhaps called the "Stored Query Routine", can illustrate this counterpart. In consideration of this concept, a short discussion on automated criminal justice systems is necessary. These systems are generally broken down as criminal justice records systems and criminal intelligence systems. The former, such as the

FBI NCIC, are based on public records of arrests and convictions. The latter, which are investigative tools, serve as indexes to files which are often based on informant information and other unverified reports. As such, they are extremely sensitive and release of information in these systems is highly restrictive. It is in the context of the latter in which the "Stored Query Routine" is discussed. At the user's option, a query against a name or other retrievable data element in the criminal intelligence system would be stored as part of the data base, regardless if a hit was obtained or not. These stored elements would be periodically purged. A second user, working against the same target, would be informed by the system of the previous query and provided a pointer to the first user. Thus, the system would emulate certain functions of the Invisible College. Users could maintain security through a variety of techniques, such as being identified by codewords known only to a central authority. Names placed in the system as a result of a query would be periodically and automatically purged. Notwithstanding the validity of the several excellent arguments which could be presented by those opposing automated criminal justice systems, the implementation of a "Stored Query Routine" could significantly enhance law enforcement efforts against major organized crime.

Zipf's Law

There would appear to be little if any relationship between linguistic analysis and forensic analysis. Recent developments in forensic science include the computerization of the characteristics of seized and purchased illicit drugs. These characteristics include the names and percentage of purity of each drug and diluent combination found in the analyzed sample. (For instance, brown heroin hydrochloride 9%, procaine 15%, lactose 62%, quinine hydrochloride 14%.)

Matches of drug samples based on such a data base point to a possible single source of supply, a "connection" which may not have been obvious through conventional investigative techniques. Originally, only the more common diluents were being characterized; unusual diluents were coded only as "other". This could result in a large number of low correlations. A discussion of Zipf's (20) law has led to consideration for a greater effort to analyze previously undetermined diluents, since the lesser used diluents will have a greater information and correlation factor than the more common diluents. Thus, those drug samples obtained in widely separated areas of the country which contain extremely unusual diluents will have a high degree of correlation and could assist in determination of a single source of supply.

Microform

The increased reliance in law enforcement information utilities will result in a proper increased requirement for storage of docu This will create a larger demand for micr storage and retrieval techniques. User ance of microform is predicated on several the most important being:

1. Indexing
2. Quality of microform
3. Quality of image blowback
4. Rapid retrieval

Given a large digital data base, a COM capability, and small portable readers which operate on self-contained or automobile generated power, a situation exists which allows the law enforcement officer to carry with him enormous amounts of information to support any on-going investigation. Such data bases would be particularly applicable for use during large scale surveillance which could occur during investigations of a major crime organization. The ability to immediately identify and obtain background information on subjects observed during such a surveillance could increase the efficiency and resources management aspects of such operations.

Network Analysis

The study of organized crime is a study of networks. The criminal intelligence analyst attempts to develop and define as many links in the network as possible. He is, therefore, interested in other applications of networks analysis and similar associative techniques, such as clustering studies. The networks displayed in Korfhage's (8) "Informal Communications of Scientific Information" resemble those which might be used by a narcotics intelligence analyst who is attempting to determine a source of supply by graphically portraying all associative information developed in an investigation. Mintz (15) has discussed such techniques in this very context. Thus, the formulas used by the information scientist to describe network and clustering techniques would appear to have great applications to the analysis of criminal organizations. If such equations could be made to operate on a computerized data base, if output were graphically, and if the system were interactive, thousands of hours of manual research time could be eliminated in the annual studies of major criminal organizations. Such methodologies will undoubtedly be developed and sophisticated in the near future.

Modeling, Simulation and Predictive Analysis

The study of organized crime, unfortunately is primarily related to past, and sometimes, present acts. If the techniques for network analy

previously discussed, could be implemented, then modeling of criminal organizations and simulation of criminal acts could be accomplished. One form of such simulation would be demographic and epidemiological studies of patterns of crime and narcotics flow; for instance, the effect of a stronger (or weaker) state drug law on the incidence of drug abuse in surrounding states. An ultimate information science sub-utility would be an effective technique of predictive analysis, one finite enough to support not only enforcement action, but also the general allocation and management of resources. While some larger city police forces are using rudimentary prediction techniques, the applications have yet to be fully accepted even in the intelligence community, which has been conducting research in this field for over a decade. Of particular interest are methods of utilizing Bayes' Theorem to support, reinforce and perhaps enhance the value of the human processor. These procedures, as developed by Edwards (3), Zlotnick (21) Mintz (15) and others, have been shown to have substantial potential value in the foreign intelligence field. The transfer of this sub-utility to the study of major organized crime is logical.

Associative Processing Applications

A computer science technique, associative processing, has recently reached a marketable state. This "firmware" will allow specific and powerful applications previously only theoretically discussed, especially by Mintz (13, 14). Advanced publicity, and a limited number of operational equipments, show promise of revolutionizing conventional ADP approaches, especially as related to law enforcement.

Natural Language Processing

A major application for associative processing is searching through large volumes of natural language. While not as rapid as most inverted file searches, the technique is usually fast enough to be considered "real-time", and is interactive. The low cost of these systems, relative to competitive conventional systems, and the ease of data maintenance, show promise that associative techniques can provide an economical and viable, if not a preferable method of full-text processing. Such a capability makes more attractive several of the previously discussed sub-utilities related to full-text and associative indexing.

Voice Recognition

An associative processor, according to one vendor, can also be used for waveforms correlation, including waveforms from voice prints. Such a capability could tremendously enhance one of the major tools of conventional law enforcement, the Title III wiretap under the Omnibus Crime Bill. The ability to match voices and to compare full-text transcripts, simultaneously, and in real or near real-time, could significantly assist the law enforcement officer assigned to monitor or process material received under such methods.

Pattern Recognition

Given such a processing capability, a number of other law enforcement related possibilities appear, all of which could be classed under the category of pattern recognition.

Rapid processing of fingerprints is an obvious application. Another procedure related to identification includes the potential to match photographs of mug shots. In regard to photography in general, it would appear that associative processing techniques have definite applications for analysis and comparison of overhead photography of narcotic related crops such as poppy and coca.

Within the forensic lab, other uses become possible. The on-line comparison of photographs of ballistics markings seems potential. However, the term "ballistics" now encompasses applications other than just the comparison of bullets. The detailed ballistics comparison of illicit pills can lead to identification of a single source of supply based on unique markings left by pill punches. A file of all photographs of illicit pills could be rapidly searched to determine if drugs from the same source had been previously obtained.

Other potential applications include the rapid searching of various types of analog to digital files, such as those which might be obtained through spectographic or chemical analysis.

REFERENCES

(1) Borko, H. Information Science: What Is It? American Documentation. January 1968, 3-5.

(2) Debons, Anthony; Otten, Klaus Towards a Metascience of Information: Informatology. Journal of the American Society for Information Science. January-February, 1970, 93.

(3) Edwards, Ward. Decision Making Is Too Simple for Executives to Waste Time on It. Innovation. Number 5, 1969, 38.

(4) Farell, Jules. Textir: A Natural Language Information Retrieval System. System Development Corporation. TM-2392. May 5, 1965.

(5) Garfield, Eugene. Citation Indexing for Studying Science. Nature. August 15, 1970, 227: 669-671.

(6) Garfield, Eugene; Malin, Morton V. Diagonal Display-A New Technique for Graphic Presentation of Network Diagrams. Inst. for Scientific Information. Presented at American Association of Cost Engineers 13th National Meeting, Pittsburgh, Pennsylvania, June 30 - July 2, 1969.

(7) Isaacs, Herbert H. Crime Pattern Recognition in Natural Language. System Development Corporation. SP-2077. May, 1965.

(8) Korfhage, Robert R. Informal Communication of Scientific Information. Journal of the American Society for Information Science. January-February 1974, 25:1, 25-27.

(9) Levine, Emil H. MIS Spinoff of a Conventional STINFO Retrieval System. Journal of the American Society for Information Science. July-August 1973, 24: 4, 305.

(10) Mintz, Stanton E. Towards a Basic Understanding of Problems Associated With the Production of Intelligence. Litton AT-TM-4-64, 1964.

(11) Mintz, Stanton E. Associate Information Retrieval. Litton Technical Note TRN 535, 1965.

(12) Mintz, Stanton E. Collected Notes on Association Maps. Litton Technical Note, 1966.

(13) Mintz, Stanton E. Associative Processors and Association Maps. Volume I, Litton, 1968.

(14) Mintz, Stanton E. Associative Processors and Association Maps. Volume II, Litton, 1969.

(15) Mintz, Stanton E. An Introduction to EWAC, Network Analysis and Personal Probabilities. Operations Research Society of America National Meeting, Nov. 8, 1972.

(16) Stiles, Edmund H. The Association Factor in Information Retrieval. Journal of the Association of Computing Machinery. Volume 8, 1961, 271-279.

(17) Stiles, Edmund H. Machine Retrieval Using the Association Factor. Machine Indexing. American University, 1962, 192-206.

(18) Stiles, Edmund H. Progress in the Use of the Association Factor in Information Retrieval. Unpublished. Washington, D.C. November 15, [illegible]

(19) Williams, J. H. Jr. Browser: An Automatic Indexing On-Line Text Retrieval System. AD 693 143. September, 1969.

(20) Zipf, G.H. Human Behavior and the Principle of Least Effort, Addison-Wesley. Cambridge, Massachusetts. 1949.

(21) Zlotnick, Jack. Bayes' Theorem for Intelligence Analysis. Conference on the Diagnostic Process. Ann Arbor, Mich. June 18, 1970.