DataQ7
Forensic Name Matching Process
October 24, 2010

Society's Problem:

In today's technology world, an individual has access to a vast array of information about themselves and others. In seconds, one can search the internet and find authors, writers, high school friends, etc. Many companies have profited from helping individuals search hundreds of millions of records to find a piece of information to help track down a specific person.

Even on a much narrower scale, organizations have been collecting millions of records on individuals. Financial, medical, and marketing information are just a few examples of the type of information organizations keep on individuals. As society becomes technology focused, more and more companies are realizing that they need to have the capability to narrow down mailing lists, to merge many databases, and to allow for quick searching of individuals in their database, as well databases owned by others.

Finally, within the criminal justice system the primary focus of the system is to ensure that the criminals have been positively identified. With so many justice related systems (criminal history, courts, law enforcement, public defenders, etc), the difficulty is ensuring the same name and date of birth is used and entered into the appropriate systems. It is also critical that the information is easily accessed and ensures future positive identification.

One of the key components of searching millions of records is ensuring that you match individuals correctly. The two primary pieces of information that one uses to match individuals is the first, middle, and last name, and date of birth (DOB). In some cases, individuals have only a partial name and no DOB, while other times they have a full name with a DOB. Other times one may use many other factors that can be used to ensure that the matching process is accurate (i.e. address, phone number, driver license number, medical numbers, etc).

The key piece of information is the name (first, middle, and last names). There are many problems associated with matching names in or between any data bases. Data entry clerks do not always enter data correctly. In any data base, one can expect that two to five percent of the records will have some type of data entry error. This is especially true for full names and long names.

In addition to the misspellings, many times there are non-letter characters have been added to the names by the data entry person or sometimes a faulty computer programs. It is common to have funky characters in large data sets, such as numbers (0,1,2…9), duplicate letters (allan vs alan), or other characters (#;%*). This can also be the nature of some names (i.e. McDonald vs Mc Donald), so it is difficult to know if there is a space, hyphen or other characters that should actually be included in the name or if it is a data entry error.

To complicate matters, it is common for individuals to change names. Females in particular usually change their last names when they get married and/or divorced. This becomes especially difficult when individuals take on both last names (Mary Jo Anderson-Nelson). This person can be known by three names (Anderson-Nelson, Nelson, or Anderson).

Individuals can formally change their names through the courts, such as Bob Dylan (born Robert Allen Zimmerman), Angellian Jolie (born Angelina Jolie Voight), and Prince (born Prince Rogers Nelson). Individuals may even change their names to avoid financial obligations or even to avoid their past criminal records. Immigrants have changed their names or had their names changed when entering the country over the last century, but could use either or both names (Tom Mackenowski vs Tom Mack). One individual could have thousands of aliases that they have officially or unofficially changed, and even stolen through identify theft.

Criminals use elaborate methods to change their name, legally and illegally to avoid being on the grid. They have learned through their encounters with the justice system that sometimes small variations in their name or date of birth can cause them to "get lost" in the justice system. If there is no official record of the criminal's activity, they will be treated as a first time offenders, even though they are repeat offenders.

There are also regional differences in the county that have unique challenges in name matching. In the Midwest, a large number of settlers were named Anderson, Nelson, and Olson. This makes it more difficult given the vast number of individuals that would have to be filtered through to find a specific person. To accommodate these regional differences, a system will need to have the capability to have broad name search capabilities for unique names (Avalance Gordan Kerkhoff) and very narrow name search capability for common names (Bob Anderson born January 1, 1960). This is compounded by today's world environment, which requires matching of names from all counties and all nationalities.

The gold standard in database name matching process would require a biometrics be paired with the name. Biometrics could include the taking of fingerprint, obtaining a DNA sample, eye retina scanning, hand geometrics, and even brain scanning. Grocery stores are allowing individuals to pay by the touch of a fingerprint. Photographs are a silver standard in matching individuals in a database, as individuals can drastically change their appearance over time. Facial recognition is being used more and more by industries today.

The problem with biometrics and pictures is how little they are used today. Even if our society moved to gold or silver standard for name matching for every piece of data being collected today and hence forward, there would still be an incredible amount of data in existence that would still need to be matched by name and DOB.

Many companies and individuals have been developing computer algorithms to make these matches, with varying success. Soundexing is one of the primary methods organizations have used in developing a name matching search program. The "American" Soundexing was patented in April 2, 1918 by Robert C. Russell of Pittsburgh, Pennsylvania. This process consists of creating a code that takes the first letter of the name, followed by three digits (i.e. William = W452, Lee =L000, etc).

The process has been improved on over the last century still relies on a numeric code (4 to 7 digits long). While this process does narrow down the list of potential matches, it is far from foolproof. Given the list of problems of name matching mentioned above, through the Soundexing design, many individuals may be missed and/or too many individuals may be found. This requires a vast amount of human intervention (costs to companies) to review the matches, but Soundexing does narrow the list down.

In addition, Identify Access Management (IAM) has been and continues to be one the top five problems addressed by Chief Information Officers. Gartner, a well known technology think tank, has hosted several IAM summits for national and international Chief Information Officers. While the focus of this group is to ensure that the individual accessing their systems is a person who has the right to access the system, it has lead to some advancement in verification that the person is the correct person. It is vital to validate who has the rights to get into the system and ensuring that only that person does access that system. While this is slightly different than name matching, it does require a similar process.

Other companies have started to market name matching processes. IBM has obtained a multi-million dollar grant from the federal government to match names, DOB and many other factors (address, phone numbers, social security numbers, etc). Their product focuses on names through root name analysis (i.e Anders vs Anderson). Other companies, such as SAS, SPSS, and MicroSoft have also developed modules for name and DOB matching, as well as other factors.

Today, there is not a single product on the market that has an ideal name matching process that addresses all of the problems associated with names. Each product addresses a piece of the problem with some success. A more comprehensive forensic approach is needed that will focus specifically on first, middle, and last names. This product also needs to be designed to be customer user-friendly.

Solution:

The uniqueness of the DataQ7 forensic name matching methodology is the seven layered name matching algorithm that is linked to a sequential process. Each of these seven layers provides an expansion on the name matching possibilities. This provides users with the ability to narrow or broaden their search criteria, depending on their need.

While the above seven layers provide a narrow and broad view of the name and DOB matching, the key to making this happen is the newly created sequential processes. This process incorporated many different aspects to match names, either between multiple databases, or a name search on one or many databases. It also provides a complex sequencing of the seven layers described above. Because of the name rules used in each layer, the name matching sequential process is critical to ensure that proper matching occurs.

The final component of the DataQ7 forensic name matching process is to incorporate connections between databases and user-friendly web-sites. This interactive website is easy to use, provides responses in seconds, and allows the users the ability to narrow or expand the search.

Business Expertise:

Over the last decade, the designer of this new forensic name matching alga rhythm and process has designed a new process to provide users with a more comprehensive method of electronic name matching, with a new design for a user-friendly website.

Mr. Storkamp has nearly 20 years experience working in the area of criminal justice information systems. He was the Director of the Criminal Justice Statistical Analysis Center at the Minnesota Planning Agency. In the first 7 years at Minnesota Planning, he obtained large databases to track offenders at they passed through the stages of the criminal justice system (Cops, Courts, and Corrections). This analytical work required the establishment of an Offender Based Tracking System (OBTS). Through six year development process, Mr. Storkamp reviewed and analyzed all criminal justice databases. He has reviewed tens of thousands of names and DOB to match them between differing databases.

Part of his work with the Criminal Justice Statistical Center required him to apply for and obtain grant funding to support the analytical work. These grants were obtained from federal and state granting entities.

Mr. Storkamp has spent the last 12 years on the Minnesota Department of Corrections (DOC) management team in a variety of positions. He has worked as the agency Director of Planning, Director of Research, Assistant Commissioner, and Director of the Information and Technology Division. These experiences have allowed him opportunities to understand the intricate problems in dealing with multi-agency and multi-discipline data.

He has held the position of legislative liaison for two years and was the Department's representative with legislature during that time. He has served as the Assistance Commissioner of Management Services for nearly a year and was responsible for the agency's policy, legal, human resource, fiscal, information technology, and office support. Today, Mr. Storkamp is the director information and technology for the DOC.

One of the systems Mr. Storkamp lead for the DOC was the establishment of the Statewide Supervision System. This system is a compilation of all individuals under probation, individuals booked into jails, and individuals admitted to the prison systems from hundreds of different operational databases and from different jurisdictions throughout the state. While the system displays individuals, there is no component that provides for enhanced matching by name and DOB. Today, over 7,000 criminal justice professional utilize this system regularly to support their work instituting justice.

As a consult, Mr. Storkamp worked with a professor from the University of Minnesota, Steve Simon. Professor Simon had obtained access to the Minnesota Driver's License (DL) file to complete work on analyzing DWI recidivism rates. During the seven year consulting effort, Mr. Storkamp worked with Professor Simon on analyzing millions of driver and vehicle records. In addition, special projects were completed to match the large, unmanageable driver's license file with smaller program databases.

One of these projects was to match the DL file with a staggered sentencing program7 database. Through this process, Mr. Storkamp identified and documented a new method and process to match names. While crude at the beginning, this method and process helped match name and DOB from a relatively small number of individuals, approximately 100.

In the last 7 year, Mr. Storkamp has greatly enhanced this name matching method and process to support his consulting efforts. This process has been expanded into a seven layered name matching effort. In addition, the sequencing of how the seven stage name matching effort will be used is critical in making the name matching method work.

Please visit www.dataQ7.net for additional information and to try an on-line live demonstration of the products versatility.