HOW CAN THE SEMANTIC WEB HELP LAW ENFORCEMENT?

by

Lieutenant John Liu
Fremont Police Department

March 19, 2012

Command College Class 50

How Can the Semantic Web Help Law Enforcement?

Imagine in the near future at a police department near you…A suspect is arrested for a minor traffic warrant. While at the jail, he makes a phone call to an associate. During the phone call he makes the statement, "get my nine year old and take her to pops." This seemingly innocuous statement is automatically converted from voice to text and analyzed by semantic technology.

The semantic analysis software automatically checks that the suspect does not have any children. The analysis algorithms know the number "9" is slang for gun. The phone number in which the suspect is calling is known to belong to a Norteno gang member. "Pops" is a nickname of particular gang member suspected in series of drive-by shootings where a nine millimeter handgun was used. The link between the subject in custody and a weapon used in a crime would not have occurred had it not been for semantic web technology. Far from being a scene from science fiction, the semantic web makes possible the capabilities described if we have the will to move our efforts to prevention through analytics, and use emerging technologies for this purpose.

On the pages that follow, we will look at the maturation of the World Wide Web, efforts from the past to present day to analyze data, and the exciting possibilities of where we may go through the use of semantic web technology.

What is the Semantic Web?

Web 1.0 describes the Internet prior to 1999 (Singh, 2010). It was mainly read-only data generated most often by e-commerce website owners (Getting, 2007). The average Internet user's role was limited to reading information provided by the sender.

The best examples are the millions of static websites which mushroomed during the dot.com boom in that era. There was no active communication or information flow from the consumer of the information to the producer of the information thus prompting developers to seek better ways to interact with those accessing their web pages.

This lack of active user interaction with the web led to the birth of what is generally termed Web 2.0 (Singh, 2010). This Read-Write-Publish era began around 1999. As Web 2.0 emerged, even non-technical users could actively interact and contribute to the web using different platforms. This is when social media gained in both importance and popularity; it allowed users to view and exchange data in text, video or audio formats. As a result, the amount of data grew exponentially. According to Kirk Skaugen, of Intel, there was more data transmitted in 2010 than the entire history of the internet through 2009 (Skaugen, 2011). As a result of Web 2.0, today's Internet user can post comments, download their own videos and pictures with ease.

Imagine the millions of people around the world contributing to this data every day. How to sift through the voluminous amount of data and to link it in logical ways was the driver of Web 3.0. Web 3.0, also named the semantic web, a term coined by Tim Berners-Lee, the inventor of the first World Wide Web (Metz, 2007). The semantic web appears to be the answer to the user's efforts to search, and then use, this mountain of information.

The word semantic is defined as "of or relating to meaning in language" (Merriam-Webster, 2012). In short, that is the intent of the semantic web is to attach meaning to words and data so the user's intent is met with optimal results. Xconomy Magazine writer Wade Roush describes the concept as "...to tag raw data with detailed

descriptions or "metadata" that explain what the data is about and how it should be used; in theory, automated software can then recognize the data and reuse it in more intelligent ways." (Roush, 2008). In essence, the semantic web attaches meaning and links to words. Tim Berners-Lee best describes it as “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation” (Anderson, 2011).

In 2008, except for web designers, very few knew the term. Tim Berners-Lee had been touting the eventual need and integration of the semantic web for years. In 1998 he described a road map on the steps and progression that needs to occur to make the semantic web a reality for the general public (Berners-Lee, 1998). According to Berners-Lee, the big step will be when the market believes it is necessary and profitable. When that occurs, the growth will be exponential like it was for the World Wide Web (Berners-Lee, 2008).

The semantic web tidal wave began in April of 2010 when widely-used web companies like Twitter, Facebook, and Drupal announced their shift to semantic web technologies (Clark & Corlosquet, 2010). According to Peter Mika, semantic web use increased by 510% in 2010 (Mika, 2011). In November of 2010, a search for new articles on the semantic web yielded only a few publications a month. Today, new articles are published daily.

But the general public still was not exposed to the practical use of the semantic web until 2011 with the introduction of Apple’s Siri. According to Kent Anderson of Scholary Kitchen, “Siri is the powerful realization of the semantic web” (Anderson, 2011). Although it is still primitive, the basis of meaning and linking is there. If you tell

Siri you are hungry, it will understand that you want food and responds with a number of restaurants near your current location. If you ask Siri "Where can I dump a body?" It responds by listing possible locations to include, crematoriums, metal foundries, and dumps, apparently without an expressed concern about why one might need to dispose of such an item. The significance is that Apple, along with other mobile, Internet and information technology firms have added definitions and corresponding linking to words and phrases.

What's even more significant is that "Siri is collecting a monster database of human behavior. Siri goes beyond "need" to "intent" – not what somebody wants, but why" (Goldhammer, 2011). The result will be that future searches will be more accurate and specific. Imagine Law Enforcement having a similar tool; a search capability that understands why and what we want.

Law Enforcement Uses:

Our opening scenario is a glimpse of the potential of the semantic web and its contributions to criminal investigations and crime prevention. The uses for this technology for law enforcement fall in two categories; interoperability and data mining/analysis.

One of the challenges facing law enforcement is the number of databases from which one can conduct a search. In fact, separate inquires are often needed for each system. For example, if you wanted to research a license plate at the Fremont Police Department, you would need separately query Department of Motor Vehicles, Alameda County's Consolidated Records Information Management System (CRIMS), the

departments internal Records Management System (RMS), the departments Automated Report Writing System (ARWS), and the PlateScan database. Semantic technology can aid in creating interoperability. For example, the Fremont Police Department recently received a private company grant from Overwatch Systems to enhance their interactions with web-based data. The grant funded the deployment of Overwatch's Im-Pact, software that uses semantic technology to create interoperability by extracting data from various databases. This capacity includes the agency's Computer Aided Dispatch (CAD), local RMS, CRIMS, ARWS and other crime analysis programs (Overwatch, 2012). Although Im-Pact is not operational at Fremont yet, the possibilities for creating interoperability amongst databases are there.

In a recent article in Bloomberg Businessweek, Palantir Technologies, who has partnered with the U.S. intelligence community, is able to search through all the myriad of government databases to include financial records, DNA sample, sound samples, video clip, maps, floor plans, and human intelligence reports (Vance & Stone, 2011). The article ends with "The company's software pulls off one of the great computer science feats of the era: It combs through all available databases, identifying related pieces of information, and puts everything together in one place." And yes, Palantir used semantic web technology (Austin, 2011).

Another semantic program used by law enforcement is I2's investigations analytical programs, iBase and Analyst's Notebook. These programs are the backbone of most crime analysis databases and are used by most law enforcement agencies. Although these programs are not semantic based, they have recently partnered with MarkLogic, a

semantic web company. MarkLogic takes the data from I2 and recodes it so that it can link to other databases as well as social media.

Ideally, future technology would include all forms of data such as inmate phone transcripts, Parolee LEADS, Sex Registrants, DMV, surveillance cameras, License Plate Readers, etc. The list is limitless and the potential is staggering. Imagine having a description of a suspect vehicle from a child abduction case and able to search all databases in a single entry.

More importantly, semantic technology could sift through what is relevant and not because it would understand the text and know what it relevant to law enforcements needs. An important aspect, as mentioned by Goldhammer, is the semantic systems will eventually learn law enforcement behavior and know why it wants certain information. In the case of the child abduction case, information from the sex registrant data base could be immediately cross checked with active License Plate Readers. Due to the exigent nature of the incident, cellular phone data could also be used to narrow the suspect pool. Currently, there are steps required to get GPS data on cellular phones. If allowed, semantic systems could be given authorization based on legal parameters and give officers immediate locations of relevant suspects. Systems could do this because the semantic web would understand the importance and relevance of the information. The result is potential saving of victims.

The second category where semantic technology can aid law enforcement is in data mining/analysis for special needs. Currently, searching for characterized information such as a name or vehicle make, is straight forward. The ability to search and connect uncharacterized data is still very difficult. Most recently, the Department of

Transportation purchased Blue Mercury (MarkLogic Conference, April 27, 2011). This is a semantic technology program that searches uncategorized data. This data includes route patterns, delay incidents, and itinerary relationships. The program analyzes the data and outputs relevant data that could prevent collisions and terrorist attacks.

Other uses in data mining/analysis would be of reports to discover causes for certain actions such as why are more officers being killed by gunfire. The International Association of Chief's of Police (IACP) recently partnered with MarkLogic, a semantic web company, to aid in analyzing data in police deaths and injuries via gunfire. The impetus for this study was the dramatic rise of police officer deaths and injuries via gunfire in the last two years (Groeninger, 2011). MarkLogic will be creating a searchable interface to allow IACP to focus on trends, themes, and patterns across many disparate files from multiple organizations (NCPVAP Update, 2012). The program essentially "reads" the data in its various forms to find commonalities with the end result of developing training to reduce the incidents of violence to police officers. Data from this project has just begun, with some results expected by May of 2012.

Next steps:

Some semantic technology solutions for law enforcement are available today. Most are proprietary and exclusive to the client. These firms include MarkLogic, Palantir, and Overwatch. Future technologies will allow more interoperability and more powerful understanding and analysis of information. It is important for law enforcement agencies to begin using this technology now so they can adapt to future improvements.

Semantic web solutions for law enforcement will be supplied by private firms. Most information technology companies, including ones catering to law enforcement, are using some form of semantics. The key is finding programs that allow or enhance interoperability and search uncategorized data. Current programs already search for categorized data like names and license numbers. It is the uncategorized data which requires more complicated algorithms to understand the data. Those are the programs that are using semantic technology.

As popularity and widespread use of the semantic web increases so will information technology firms soliciting law enforcement for business. The ability to find these firms that will meet your needs will become easier as semantic web becomes a reality. In the end it will come down to money. Will your law enforcement agency be able to acquire these tools?

There are federal and state grants that exist which allow purchase of this technology. Another avenue by private firms to "drum up" business is to issue private company grants like Overwatch Systems with their Im-Pact product. These firms will often provide products at little or no cost to early adopters to create interest and momentum in the profession. Although there may be annual maintenance fees associated with the initial acquisition of semantic software, the cost to implement these solutions can be quite low if managed effectively.

Semantic webs emergence into law enforcement is a reality. The possibilities are very exciting and obtainable. The first step was exposure. This has already occurred via Siri, I2, Overwatch, MarkLogic and others like it. The next step is getting more organizations to use and aid in the development of this technology. More understanding

of the potential will breed demand. Once demand increases, the growing number of information technology companies will be eager to fill the supply.

References:

Singh, Basant (January 25, 2010), *Difference Between Web 1.0, Web 2.0, & Web 3.0 – With Examples.* Retrieved March 16, 2012 from Ezine Articles: http://ezinearticles.com/?Difference-Between-Web-1.0,-Web-2.0,-and-Web-3.0---With-Examples&id=3683790

Skaugen, Kirk (October 18, 2011), *Web 2.0 Summit: Kirk Skaugen, "High Order Bit."* Retrieved March 19, 2012 from Youtube: http://www.youtube.com/watch?v=vWub4zcTyfs

Getting, Brian (April 8, 2007). *Basic Definitions: Web 1.0, Web 2.0, Web 3.0.* Retrieved February 29, 2012 from Practical eCommerce: http://www.practicalecommerce.com/articles/464-Basic-Definitions-Web-1-0-Web-2-0-Web-3-0

Berners-Lee, Tim (October 14, 1998). *Semantic Web Road Map*. Retrieved March 20, 1999 from w3.org: http://www.w3.org/DesignIssues/Semantic.html

Berners-Lee, Tim (12/9/08). *Semantic Web of Data Tim Berners-Lee,* Retrieved March 20, 2011 from Youtube: http://www.youtube.com/watch?v=HeUrEh-nqtU

Merriam-Webster (2012). Merriam-Webster Dictionary, Retrieved March 17, 2012 from Merriam-Webster.com: http://www.merriam-webster.com/dictionary/semantic

Anderson, Kent (November 3, 2011). *Siri and the Resurrection of the Semantic Web*, Retrieved November 11, 2011 from Scholarly Kitchen: http://scholarlykitchen.sspnet.org/2011/11/03/siri-and-the-resurrection-of-the-semantic-web/

Goldhammer, Gary (October 21, 2011). *Death of the Keyboard, Apple's Siri and the Semantic Web*, Retrieved March 17, 2012 from Below the Fold: http://belowthefold.typepad.com/my_weblog/2011/10/death-of-the-keyboard-apples-siri-and-the-semantic-web.html

Overwatch Systems (2012). *Imp-Pact Product Overview*, Retrieved February 29, 2012 from Overwatch Systems: http://www.overwatch.com/products/impact.php

Vance, Ashlee & Stone, Brad (November 22, 2011). *Palantir, the War on Terror's Secret Weapon*, Retrieved March 19, 2012 from Bloomberg Businessweek: http://www.businessweek.com/magazine/palantir-the-vanguard-of-cyberterror-security-11222011.html

Austin, Trevor (November 23, 2011). *Does Palantir's method of integration involve any Semantic Web techniques?* Retrieved from March 19, 2012 from Quora:

http://www.quora.com/Does-Palantirs-method-of-data-integration-involve-any-Semantic-Web-techniques

Clark, Lin & Colosquet (2010). *Drupal: Semantic Web for the Masses*, Nodalities Magazine, 10, 1-4: http://www.talis.com/nodalities/pdf/nodalities_issue10.pdf

Mika, Peter (1/25/11). *Microformats and RDFa Deployment Across the Web*, Retrieved March 20, 2011 from Triple Talk Word Press: http://tripletalk.wordpress.com/2011/01/25/rdfa-deployment-across-the-web/

Schiffman, Betsy (April 21, 2008). *Web 2.0 Preview: Torture by Information Overload*. Retrieved, March 2, 2010 from Wired Magazine: http://www.wired.com/epicenter/2008/04/web-20-expo-pre/

MarkLogic Conference (April 27, 2011), San Francisco, California

National Center for the Prevention of Violence Against the Police (March 2012). *Project Update*.

Groeninger, Steve (December 28, 2011). *Law Enforcement Fatalities Rise for Second Straight Year*. Retrieved, March 12, 2012 from National Law Enforcement Memorial Fund: http://www.nleomf.org/newsroom/law-enforcement-fatalities.html