AN ANALYSIS OF THE VALIDITY OF RED-FLAG SEARCH TERMS AND IF THEY CAN BE USED TO RELIABLY IDENTIFY PERSONS OF INTEREST IN RELATION TO NATIONAL SECURITY

Crazy Eights

for IST 331, Spring 2014

Daniel Couillard, John Dori, John Greiner, Tyler Lipshutz

Pennsylvania State University

February 4, 2014/revised March 2014

Abstract

This paper explores the effectiveness of "red flag" search terms in a post-9/11 American environment. Today, it is easy as walking into a library and Googling what it takes to join Al Qaeda or how to produce weapons on a budget. This fear of *any* citizen being potentially radicalized on the Internet (the same Internet we all use), has led to the emergence of a system in which certain search queries are flagged due to the terms used therein. This study analyzed a pre-curated dataset of search queries from the now defunct AltaVista search engine that was gathered in the early 21st century. For the study, the researchers identified five search terms from the Department of Homeland Security's Desktop Binder (Stone, 2011) on media monitoring, specifically related to terrorism, and attempted to simulate the task of identifying users to flag for terrorism. This paper examines just how difficult it is to flag for anything based on search terms.

-John Dori

Introduction

Terrorism is a very serious subject today and a major concern for many departments of the government. With the various attacks like the ones on September 11, 2001 and the Boston Marathon bombings, it is fairly easy to see why. With all the threats against human lives, and the public's demand for safety, the government is exploring all avenues for terrorism prevention. One of them, which is the topic for this paper, is analyzing all Internet search terms and flagging any that may be suspicious. To qualify as suspicious, it should have some remote link to harming another individual, or a terroristic act in general. Oftentimes, searches are flagged but the person searching has no intent to commit any terrorist act. This was the case in August, 2013, immediately after the Boston Marathon Bombings, when Michele Catalano and her husband simultaneously (coincidentally) searched Google for pressure cookers and Google backpacks, respectively (Bump, 2013).

The misinterpretation of this potential flag is obvious and raised the question to the American public - "Why can the government see what I search for on Google?". If surveillance is to be done by our government, such at this, to prevent radicalization of citizens and terroristic threats, we still need the human element when it comes to seeing if a person is actually a threat to national security or the flag is just a false alarm.

-John Greiner and John Dori

Hypothesis

We expect to find that the volume of searches that will be flagged will make it too difficult to do an adequate analysis without using advanced tools. With a database of over 3.5 million queries, there will be too large a number of users searching each term to do an in-depth analysis of each user. We expect that our original intuition will be correct; in that flagging people for relations to terrorism based on search queries is a very difficult task to automate and that humans should be involved in this process as a final say as to whether a person should be flagged.

-Daniel Couillard and John Dori

Method

To begin the analysis, the Alta Vista database from 2002 was downloaded and parsed into a SQL database. The researchers identified search terms that are considered to be red flag words by the Department of Homeland Security (Stone, 2011). The researchers identified five terms as being particularly of interest to an organization tasked with flagging individuals for terrorism based on their search queries. These terms included Car bomb, Dirty bomb, Bomb, Ammonium Nitrate, and Weapon.

Each word was chosen individually and ran through the database to determine how many users appeared to have searched each term. After looking

through the list of users obtained via this method, their search terms were evaluated for their relevance to the topic of terrorism. Each search term was labeled as either (1) not related, (2) unclear, or (3) related to terrorism. For those users that searched for terms potentially related to terrorism, the researchers inspected other past queries to determine if a pattern could be determined. If a pattern of searches potentially related to terrorism was present, this user should have been flagged for further investigation. For search terms that returned far too many users, the top 50 tuples were analyzed as a subset and extrapolated to give a rough estimate of the total query.

Ultimately, the term that we are dealing with in these queries is “Locate Tagged”. It refers to the instance where the user isn’t quite sure what they are looking for but is searching with terms that aren’t exactly accurate. (Byrne, 1991). The researchers applied this by inspecting potentially flagged users’ other recorded queries. This is obviously a non scalable solution, but a necessary one in that a computer cannot reliably flag queries as being related to terrorism - the related search terms are simply too broad. For example, ammonium nitrate is used in legal explosions, ice packs, and fertilizers - all seemingly innocent contexts.

-Daniel Couillard and John Dori

Terms and Data

The following terms were gleaned from the Desktop Binder issued to analysts at the Department of Homeland Security - these are the official terms analysts are told to watch for as red flags of terrorist activity (Stone, 2011).

Terror - terrorism: violence or the threat of violence carried out for political purposes.

Dirty bomb - bomb spreading radioactive waste: a bomb containing radioactive nuclear waste dispersed by means of conventional explosives.

Ammonium nitrate - ingredient of fertilizers and explosives: a white crystalline solid.

Car bomb - bomb concealed in car: an explosive device concealed inside or under a vehicle and detonated by remote control or when the engine is started.

Bomb - explosive projectile: a missile containing explosive or other destructive material.

-Daniel Coulliard

Results

Our findings show that people used this search engine in a naive manner. Some users assumed nobody was recording what queries are searched for - some users did. Queries are present in the dataset that proved quite alarming at first glance but turned innocuous after peering into other queries ran by the same user. Some queries were genuinely alarming.

The researchers never flagged any query as being definitely related to terrorism out of the principal that further investigation and information is almost always needed. In the case of a government agency conducting this surveillance, and having access to all exhaustive data related to a query, a conclusion could be drawn -

but as students, the researchers could not claim that any particular query was related to terrorism with confidence.

Table 1 shows the breakout of the data collected during analysis. Some of the search terms yielded a high volume of searches so a sample size of 50 was used for the terms that returned over 500 hits. For those search terms yielding a small number (10 or less) every hit was analyzed for malicious intent.

The more general terms, weapon and bomb, returned the most values. Among these searches, 88% of the "bomb" hits were determined to be innocuous where the remainder required further analysis. For the "weapon" search there was a slightly higher return of suspicious tangential searches leading to a 34% of users requiring further review up from 22% for "bomb".

Search Term	Unlikely (sample)	Undetermined (Sample)	Sample Size	Total Searches
Car Bomb	3	2	5	5
Dirty Bomb	2	1	3	3
Bomb	44	6	50	1628
Ammonium Nitrate	1	3	4	4
Weapon	33	17	50	649

Table 1

Figure-1 shows a breakout of the analysis where the height of the column is the total sample size. The red shows the portion requiring further analysis while the blue area is the portion of users showing non-threatening behavior in other searches.

Figure 1

-Daniel Couillard

Discussion

The main limiting factor to our research was the sheer size of the data we analyzed. For our queries that had more than 50 results, we used a restricted sample size to subset the data. This could have a slight affect on the results of our research, but we assumed a normal distribution to be present for all queries. It is worth noting that Alta Vista's usage in the years this data was gathered was most likely declining, so the sample may not be as representative of the American public as possible.

Another area of concern would be that you have to rely on the researchers' opinion on if it had to do with terrorism. This is a necessary evil because at this moment, there are currently no algorithms that could completely take out the human element. Computer flags are still unreliable and human logic is still needed to determine of the flags are valid.

Given more time and resources to complete this analysis, we would have invested much more time into attempting to train algorithms to more accurately detect terroristic flags in search queries. Our method worked as a quick-glance method, but our findings are not conclusive. We believe that semi-accurate prediction of these flags is possible from a machine-learning approach. The algorithm would rely heavily on human input and verification before it would be self-sustaining, but a pattern seems to be evident that often reveals queries particularly suspicious of being related to terrorism. With the help of a human saying "yes, that's probably terrorism" or "no, this user is fine - we can't investigate further with confidence," the algorithm would learn over time which queries to flag.

-John Greiner and John Dori

Analysis from the Perspective of a User

Everyday users should be able to use this data to adjust their search habits. In today's world, where nearly every bit of information traveling across the Internet is documented and neatly stored away for future analysis, an intelligent user can adapt browsing methods that could potentially keep him off the big data radar.

These browsing methods could include anything from using VPNs to mask the users IP to visiting related sites that might hide the true meaning of a search. An example might be someone who actually has malicious intent but does not wish to be caught. He or she could simply look up safety related websites along with bomb making materials as we saw in the ammonium nitrate search above.

From this perspective, it is clear that this system of flagging queries based on search terms and past, related searches is flawed. The researchers' hypothesis could be interpreted as being correct here, in that a human is needed as the final say to determine if a user should be flagged - gaming the system is simply too easy from a user's perspective.

-Daniel Couillard

Analysis From Big-Data Perspective

Data sets of this magnitude are very difficult to sift through with any degree of accuracy. So how do large companies use these data sets to their advantage? In the

Barnes(1996) paper reviewed in class, the importance of self-managed teams is referenced. In this discussion it claims that a group should strive to be autonomous and reduce cost wherever possible (Smith, 1999).

Even though this paper is nearly two decades old, the idea still holds true today. Datasets this large are nearly impossible to fully analyze using the methods we discussed above. In 2014, however, scripts can be easily written to quickly parse plain text to illuminate patterns, which are much easier to digest from a big data point of view.

The future of surveillance of this sort, from a big-data perspective, is either in machine-learning algorithms far more complex than any heuristic method or retaining the human final-say element. Without sufficiently complex search term analysis, a human is needed to verify potential flags before action is taken.

-Daniel Coulliard

References

Alastair Smith, (1999) "Information Seeking in Context: *Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts*, 14-16 August 1996, Tampere, Finland", Asian Libraries, Vol. 8 Iss: 5, p. 163

Behr, I. V., Reding, A., Edwards, C., & Gribbon, L. (2013). The use of the Internet in 15 cases of terrorism and extremism. *Radicalisation in the digital era*, *1*.

Bump, P. (n.d.). Now We Know Why Googling 'Pressure Cookers' Gets a Visit from Cops. *The Wire*. Retrieved January 30, 2014, from http://www.thewire.com/national/2013/08/government-knocking-doors-because-google-searches/67864/

Byrne, M., John, B., Wehrle, N., & Crow, D. (1999). The Tangled Web We Wove: A Taskonomy of WWW Use. *Human Factors in Computing Systems: Proceedings of CHI 99*, *1*, 544-551.

Department of Homeland Security, National Operations Center, Media Monitoring Capability. (2011). *Analyst's Desktop Binder*. Andrea Stone.

Date: February 4, 2014

Group Name: Crazy Eights

--

Member 1 Name: John Dori

What they did for this write-up: Researched APA guidelines, formulated plan for data extraction, came up with hypothesis, edited all sections, managed references, and wrote the abstract, results, and part of the introduction, method, discussion.

Member 2 Name: John Greiner

What they did for this write-up: Contributed to the hypothesis, Created the graphs, and wrote the introduction, the discussion, and part of the results.

Member 3 Name: Tyler Lipshutz

What they did for this write-up: Imported AltaVista Data into searchable database (Microsoft Access), reported data findings.

Member 4 Name: Daniel Couillard

What they did for this write-up: Wrote the hypothesis, methods as well as the analysis sections from both perspectives.