

What's the Difference, Really?

A closer look at 'in-person' research versus 'screen scraping'

By: Vince Brodt, SJV & Associates

For members of NAPBS that were able to attend the Annual Conference this past March in San Antonio, TX, a Panel Discussion was held which provided a closer look at two distinctly different methods of performing research for the purposes of criminal background checks. On one side of the coin, a spotlight was placed on 'live' or 'in-person' research, traditionally performed by local, regional, and nationwide research providers. On the other side of the coin, the panel discussed 'screen scraping', or more scientifically, 'web data extraction/web-harvesting'. This type of research is performed not by a direct human presence in a courthouse, but rather a highly sophisticated technology-based solution. The purpose of this article is to briefly summarize the differences between these two research methods, while outlining their respective strengths and perceived weaknesses.

Historically speaking, in-person research has been a prominent fixture in the background screening process. The gist or upshot of this particular research method rests upon the physical presence of a skilled and experienced human being, possessing an intimate knowledge of the subtle nuances specific to case management practices in a given jurisdiction, conducting research directly in a courthouse. This is seen as a simple concept, until the many variations between courthouses are focused upon. The way data is stored or maintained physically and electronically, as well as the assistance of employees of the court all constitute wildcards that a trained investigator uses his or her skills and experience to handle.

In terms of web harvesting, the actual technology utilized has long existed in numerous capacities, but is now being applied to web-based courthouse interfaces. In other words, there are websites hosted by courthouses that provide access to criminal data, in many ways similar to a public access terminal in the courthouse. Just as there are many variations within a courthouse as to how data is stored and maintained, there exist many variations amongst these websites, as well. It is the representation of specialists behind this technology that can safely help someone navigate the court records available via the web, while minimizing the chance for human error and reducing the labor costs that would be consumed using in-house employees to perform research by going to these websites directly.

When a trained investigator in a courthouse is performing research, the advantages mirror that of any experienced employee simply going about their job's daily responsibilities. An investigator is able to make decisions based on empirical, logical, and research-generated data, in order to produce a highly detailed account of what is represented in the courthouse concerning an individual's criminal background. They're able to inquire as to the nature of partial or missing case information, abbreviations

within a charge or sentence, and produce copies of criminal records with the assistance of the court's employees.

When web harvesting technology is used to generate research based on a courthouse's web-based interface, results can return in minutes, or even seconds, depending upon the responsiveness of the website being accessed. The technology behind the scraping process mimics the steps that a human user would go through, from logging onto the website, to keying in the name to be searched, and retrieving the results. So long as the technology used is of sound programming, a common or 'human' typographical is not likely to be made. The results, perhaps including certain details of the cases found, are copied or even 'scraped' directly from the court's available online resource, minimizing the opportunity for a human researcher to make a typographical error while transcribing the criminal record itself.

These are but a few brief examples of the numerous strengths that exist amongst both in-person and web harvesting based research. The question is begged, then, what of the perceived weaknesses that each of these research methods possess?

For in-person research, the obvious answers would be the overall average turnaround time, the capacity for human error, and quite frankly, the cost. These are the three most common sources of concern when utilizing an investigator. Courthouses are only open during certain hours of the day, and not on weekends. There exists a limited time span in which an investigator can physically be on-site to perform research at a courthouse. As all humans are aware, people make mistakes from time to time. Regardless of one's level of training or years of experience, mistakes do happen. Whether it's a small typo, or a larger error, no investigator is without error. Finally, the industry-wide tightening of the belt appurtenant to the nationwide loss of millions of jobs during the 2009 recessionary period led background screeners to take a hard look at their bottom line, and the premium associated with utilizing a trained professional investigator led screeners to seek lower-cost alternatives.

In terms of web harvesting, the prominent potential weakness and risk centers around the absence of any human involvement in the harvesting of information that may potentially be used to deny employment, tenancy, or some other benefit to a consumer. Often times, court-specific abbreviations, cryptic codes, and other jurisdictional shorthand are used to condense the length of criminal records. If the technology used is simply copying this information, and doesn't have the ability to discern what the full scope of the information being reported is, then this can be a perceived weakness. Other issues of expertise exist in terms of the length of scope ordered for a criminal background search. Is the website listing all the exact, same information, historical and present, that the courthouse does? The answer is by default, no, because many records are not available on public access that are filed beyond a specific date, and thus, not available online.

Both of these particular research methods have their respective strengths and weaknesses, and advocates within NAPBS of both approaches are working towards

continually improving their particular methodologies. Given the significance of the important decisions affecting consumers that will ultimately be made using the information that we as an industry furnish, and the sensitive political climate which we as an industry now find ourselves in as we face the specter of increased government scrutiny under the watchful eye of a potential new regulatory authority, it's important that we fully understand the features and limitations of the research methods we ultimately bet our businesses on.

Vince Brodt currently serves as Business Development Manager for SJV & Associates, an Atlanta-based Nationwide Wholesale Research Firm which maintains an ISO 9001:2008 Certified Quality Management System. Vince is the current Chairman of the NAPBS Provider Committee. Their mission is to establish guidelines and best practices for Providers, while promoting overall education within association space, and is composed of Research Providers, Database Providers, and Software/Technology Solution Providers.

