LIVING IN A DIGITAL WORLD

“In order to be a competent, successful citizen, you need a new set of tools”

Lee Rainie
Pew Research Center

WHY INTERNET INTELLIGENCE?

“The Internet has breathed new life into the anarchist philosophy, permitting communication and coordination without the need for a central source of command, and facilitating coordinated actions with minimal resources and bureaucracy”

Canadian Security Intelligence Service
QUANTIFYING THE WEB
As of 2014, the Surface Web contains approximately 60 trillion pages according to Google.

Early estimates indicated the size of the Deep Web as being some 500 times larger than the Surface Web.

THE DEEP WEB
The “Deep Web” is the hidden part of the Web, containing a huge volume of content that is largely inaccessible to conventional search engines.

Approximately 95% of the Deep Web is publicly accessible information and not subject to fees or subscriptions.
THE DEEP WEB

- Databases
- Dynamic Content
- Non-textual Files
- Unlinked Content
- Contextual Web

WEB 2.0

Web 2.0 describes web applications that facilitate interactive information sharing and collaboration such as Social Networks, Wikis, Blogs and Social Bookmarks.

Significant to Web 2.0 is Collaboration Among Users and User Generated Content.

DNS ADVERTISED DOMAINS
**THE DIGITAL POPULATION**

2,802 Million Users Globally
(39% of the total population)

<table>
<thead>
<tr>
<th>Region</th>
<th>Users</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>1,265</td>
<td>22%</td>
</tr>
<tr>
<td>Europe</td>
<td>566</td>
<td>67%</td>
</tr>
<tr>
<td>Latin America/Carribean</td>
<td>302</td>
<td>49%</td>
</tr>
<tr>
<td>North America</td>
<td>300</td>
<td>85%</td>
</tr>
<tr>
<td>Middle East</td>
<td>104</td>
<td>45%</td>
</tr>
<tr>
<td>Africa</td>
<td>240</td>
<td>21%</td>
</tr>
<tr>
<td>Australia/Oceana</td>
<td>25</td>
<td>67%</td>
</tr>
</tbody>
</table>

Source: Internet World Stats - December 2013

**INTERNET USERS BY LANGUAGE**

- **English**: 29%
- **Others**: 17%
- **Chinese**: 10%
- **Spanish**: 9%
- **Japanese**: 8%
- **French**: 6%
- **Russian**: 3%
- **Arabic**: 3%
- **German**: 3%
- **Portuguese**: 2%
- **Italian**: 2%

**WEB CONTENT BY LANGUAGE**

- **English**: 67%
- **Others**: 15%
- **Portuguese**: 2%
- **Italian**: 2%
- **French**: 2%
- **Chinese**: 2%
- **Spanish**: 1%
- **Japanese**: 1%
- **Russian**: 1%
- **German**: 1%
The Middle East and Africa (MEA) region will continue to have the fastest growing internet protocol (IP) traffic until 2017
Source: Cisco Systems, September 2013

The Latin American online population grew faster than any other global region in the past year, rising 12 percent to 147 million visitors
Source: Comscore, May 2013

The value of a network grows as the square of the number of its users increase
Robert M. Metcalfe
Co-Inventor of Ethernet
WHAT IS A SEARCH ENGINE?

They search the Internet (or select pieces of the Internet) based on important words.

They keep an index of the words they find, and where they find them.

They allow users to look for words or combinations of words found in that index.

WHAT IS A SEARCH ENGINE?

Spider based technology is essentially “dumb”.

Because search engines cannot evaluate context, unrelated sites are often returned in search results.
KEY CONSIDERATIONS

Coverage varies dramatically
Indexing Patterns Vary
Only a Decreasing Fraction of the Web is Indexed
SEARCH ENGINE COVERAGE

Search Engine Rankings
March 2014

Percentage of Web users

<table>
<thead>
<tr>
<th>Search Engine</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>67.5%</td>
</tr>
<tr>
<td>Microsoft</td>
<td>18.1%</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>11.2%</td>
</tr>
<tr>
<td>Ask</td>
<td>2.6%</td>
</tr>
<tr>
<td>AOL</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

GOOGLE SEARCH VOLUME

Average Searches Per Day

<table>
<thead>
<tr>
<th>Year</th>
<th>Average Searches Per Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>1,000,000</td>
</tr>
<tr>
<td>2000</td>
<td>2,000,000</td>
</tr>
<tr>
<td>2007</td>
<td>4,000,000</td>
</tr>
<tr>
<td>2008</td>
<td>5,000,000</td>
</tr>
<tr>
<td>2009</td>
<td>6,000,000</td>
</tr>
<tr>
<td>2010</td>
<td>7,000,000</td>
</tr>
<tr>
<td>2011</td>
<td>8,000,000</td>
</tr>
<tr>
<td>2012</td>
<td>9,000,000</td>
</tr>
<tr>
<td>2013</td>
<td>10,000,000</td>
</tr>
</tbody>
</table>

USING MULTIPLE SEARCH ENGINES

“Meta Search” sites run your query through more than one search engine at the same time yielding a single list of combined search results.

The better Meta sites organize the hits according to relevance and eliminate duplicates.
“It's impossible to move, to live, to operate at any level without leaving traces, bits, seemingly meaningless fragments of personal information.”

William Gibson
“Johnny Mnemonic”

“A computer lets you make more mistakes faster than any other invention in human history, with the possible exception of handguns and tequila.”

Mitch Ratcliffe
“If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place... the reality is that search engines - including Google - do retain this information for some time and it's important, for example, that we are all subject in the United States to the Patriot Act and it is possible that all that information could be made available to the authorities.”

Eric Schmidt, Chairman, Google

**THE BROWSER CACHE**

Web browsers download web pages viewed into a temporary *cache* file.

When revisiting a Web site, the browser will compare the components of current Web document against the cache file.

When possible, the browser will display document components stored in the temporary file rather that re-download those components.

A browser’s cache file can be examined and analyzed to determine what sites a user has been visiting and what keywords were entered into a search engine.
ONLINE SECURITY
A Web site may permanently record your IP address and use multiple HTTP cookies, Web Beacons, Javascript and Flash LSOs to track your movements not just on that site but possibly elsewhere as well...

ONLINE SECURITY
...and combined with readily available information about your computer system and Traffic Pattern Analysis you or your organization could be easily positively identified

ONLINE SECURITY
Ensure that when using the Internet as an Investigative research tool you have a security plan in place!
In a recent survey, 18% of respondents said they were always able to find what they want...

5% of respondents said they were rarely or never able to find what they wanted

In this same survey, 29% of respondents stated they found Internet searching “very frustrating”

Only 9% found searching the Internet searching “not at all frustrating”

1.2 TRILLION SEARCHES IN 146 LANGUAGES

Google’s Fastest Rising Search Terms for 2014 (Global)

1. Robin Williams
2. World Cup
3. Ebola
4. Malaysia Airlines
5. Ice Bucket Challenge
6. Flappy Bird
7. Conchita Wurst
8. ISIS
9. Frozen
10. Sochi Olympics

Source: Google Zeitgeist
FREE OSINT NEWSLETTER

Welcome to the newsletter!

Sign up at www.toddington.com