Mapping Cyberspace: Tracking the Spread of Ideas on the Internet

• Ming-Hsiang Tsou  mtsou@mail.sdsu.edu,
• Associate Professor, San Diego State University,

The Department of Geography, San Diego State University, USA
Funded by NSF Cyber-Enabled Discovery and Innovation (CDI) program, Award # 1028177. (four years 2010-2014)

http://mappingideas.sdsu.edu/

**PI:** Dr. Ming-Hsiang (Ming) Tsou (Geography)
mtsou@mail.sdsu.edu

Co-PI: Dr. Dipak K Gupta (Political Science),
dgupta@mail.sdsu.edu

Co-PI: Dr. Jean Marc Gawron (Linguistic),
gawron@mail.sdsu.edu

Co-PI: Dr. Brian Spitzberg (Communication)
spitz@mail.sdsu.edu

Senior Personnel: Dr. Li An (Geography)
lan@mail.sdsu.edu

**Graduate Research Assistants:**
Ick Hoi (Rick) Kim, Sarah Wandersee,
Kellen Stephens, Jennifer Smith,
Amit Nagesh, Vickie Mellos, & Ting-Hwan Lee,
(Undergraduates: Alejandra Coronado, Luke Kemper)
Spatial Web Automatic Reasoning and Mapping System (SWARMS) prototype
WEB SEARCH ENGINE & SEMANTIC DATABASES

- Microsoft SQL server with Web-based GeoLocating services.
- Access Bing and Yahoo search engines (search for 1000 results)
- WSDL to access IPPAGE.COM for geolocation conversion.
CONVERTING URLs TO GEOLOCATIONS

WHOIS’ databases → host registrant → street address → latitude/longitude
Visualization maps (Testing Keyword: "Jerry Sanders" in Yahoo.com on MARCH 9, 2011, 978 records, 81 failed in geocoding, 91.7% successful geocoding rate.)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Search</th>
<th>Keyword</th>
<th>URL</th>
<th>Title</th>
<th>Latitude</th>
<th>Longitude</th>
<th>IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://en.wikipedia.org/wiki/Jerry_Sanders">http://en.wikipedia.org/wiki/Jerry_Sanders</a></td>
<td>Jerry Sanders (politician) - Wikipedia, the free encyclopedia</td>
<td>27.7788</td>
<td>-82.6823</td>
<td>208.80.152.2</td>
</tr>
<tr>
<td>3</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://en.wikipedia.org/wiki/Jerry_Sanders">http://en.wikipedia.org/wiki/Jerry_Sanders</a></td>
<td>Jerry Sanders (businessman) - Wikipedia, the free encyclopedia</td>
<td>27.7788</td>
<td>-82.6823</td>
<td>208.80.152.2</td>
</tr>
<tr>
<td>5</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://www.daylife.com/topic/Jerry_Sanders">http://www.daylife.com/topic/Jerry_Sanders</a></td>
<td>Jerry Sanders - News, photos, topics, and quotes</td>
<td>40.7519</td>
<td>-73.9763</td>
<td>64.147.115.80</td>
</tr>
<tr>
<td>6</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://www.huffingtonpost.com/jerry-sanders">http://www.huffingtonpost.com/jerry-sanders</a></td>
<td>Jerry Sanders</td>
<td>40.6888</td>
<td>-74.0203</td>
<td>69.60.14.6</td>
</tr>
<tr>
<td>7</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://features.rr.com/topic/Jerry_Sanders">http://features.rr.com/topic/Jerry_Sanders</a></td>
<td>Road Runner: News, photos, topics, and quotes</td>
<td>40.7519</td>
<td>-73.9763</td>
<td>64.147.115.89</td>
</tr>
<tr>
<td>8</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://topics.treehugger.com/topic/jerry-sanders">http://topics.treehugger.com/topic/jerry-sanders</a></td>
<td>Jerry Sanders - From Treehugger</td>
<td>40.7519</td>
<td>-73.9763</td>
<td>64.147.115.89</td>
</tr>
<tr>
<td>9</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://www.latimes.com/topic/economics">http://www.latimes.com/topic/economics</a></td>
<td>Jerry Sanders: Jerry Sanders News and Phelps</td>
<td>41.9288</td>
<td>-87.6315</td>
<td>144.142.224.43</td>
</tr>
<tr>
<td>10</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://www.amdboard.com/sanders">http://www.amdboard.com/sanders</a></td>
<td>?? AMDboard.com - Jerry Sanders Special</td>
<td>37.4249</td>
<td>-122.0074</td>
<td>68.180.151.76</td>
</tr>
<tr>
<td>11</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://www.tonic.com/people/jerry-sanders">http://www.tonic.com/people/jerry-sanders</a></td>
<td>Jerry Sanders - Tonic Profiles</td>
<td>47.6914</td>
<td>-122.33</td>
<td>204.236.228.15</td>
</tr>
<tr>
<td>13</td>
<td>Yahoo</td>
<td>Jerry Sanders</td>
<td><a href="http://twitter.com/MayorSanders">http://twitter.com/MayorSanders</a></td>
<td>Jerry Sanders (Mayor Sanders) on Twitter</td>
<td>30.5072</td>
<td>-98.5748</td>
<td>67.192.110.104</td>
</tr>
</tbody>
</table>

Table Of Contents

- SD-Sanders
- Top-30-cities-1990
- burn-Koran-Jan30
- US-cities
- Cntry94
- states48
Kernel point density function was performed in the ArcGIS using 3 map unit threshold (radius) and 0.5 map unit output scale. 1 map unit =~ 50 miles.

The ranking of search results were considered as the "popularity" and the "population" in the kernel density algorithm. 

\[
\text{population} = (1001 - \text{rank#}).
\]

A website ranked #1 will be assigned to "1000" (1001 - 1) for its population parameter.

Compare two keywords:

Jerry Sanders (San Diego Mayor)  
Antonio Villaraigosa (L.A. Mayor)
Map Algebra (Raster-based).

Differential Value = (Keyword-A / Maximum-Kernel-Value-of-Keyword-A) - (Keyword-B / Maximum-Kernel-Value-of-Keyword-B)

The red hotspots in the new map indicate that "Jerry Sanders" is more popular than "Antonio Villaraigosa" in these areas and the blue color areas indicate that "Antonio Villaraigosa" is more popular than "Jerry Sanders".

The differential information landscape map illustrated important geospatial fingerprints hidden in the text-based web search results depending on the context of selected keywords.
Many web pages located in Denver are created by very conservative republicans or anti-illegal immigrant groups. These web pages created a “negative popularity” hotspot in the information landscape. These anti-immigrant groups dislike Villaraigosa very much because he is one of few Hispanic mayors in big cities in the United States.
The following settings of kernel density thresholds for detecting spatial fingerprints at different map scales were used.

- **6 - 8 map units** for detecting the State level spatial fingerprints.
- **2-3 map units** for detecting the County level spatial fingerprints.
- **0.5 - 1 map units** for detecting the City level spatial fingerprints.
- **0.1 - 0.2 map units** for detecting the Zipcode level spatial fingerprints.

*Figure 9.* Comparing six different settings of radius distances (threshold) and output grids in the differential maps between “Jerry Sanders” and “Antonio Villaraigosa.”
Radical Ideas Analysis: “Burn Koran”

The kernel density of “burn Koran” keyword search results and 1000 associated websites (red dots) with weighted ranks (radius: 3.0 map units, output grid: 0.5 map units) using Yahoo search engine on January 30, 2011.

How to “standardize” these information landscapes?

1. Compare two similar keyword maps.
2. Standardized by the population density (U.S. maps).
The U.S population density map was used to standardize the popularity density map of “burn Koran”. After the standardization, the red color hot spots indicate that San Jose, Houston, and the middle of Kansas State are the popular areas of "burn Koran" keywords. The blue color hot spots indicate the negative value (less popular) of "burn Koran" standardized by city population density.

WHY the hotspot in the middle of Kansas? Near the City of Topeka? (after the original event happen in the church located in Gainesville, Florida (green symbol), another church in the city of Topeka, Kansas claimed that they will continue the action of “burn Koran”. )
Comparing “burn Koran” between Jan 30 and April 03, 2011 (immediately actual incident date).

[April 03] – minus [ Jan 30]

Hot spots: Saint Louis, Pittsburgh, Philadelphia – NEW trends?

RED: Increased density of web pages on April 03, 2011 (compared to 1.30.11)
BLUE: Decreased density of web pages on April 03, 2011 (compared 1.30.11)
Different language search top 1000 hits for "Osama bin Laden"

English
"Osama bin Laden"

Chinese (S)
奥萨马本拉登

Arabic
" أسامة بن لادن "

LANGUAGE DEPENDENT!
"Osama bin Laden" (Geronimo) – (minus) -- Background Constant

Note 1: Hotspots in San Francisco and New York.

RED: high density of web pages related to “Osama bin Laden” (comparing to the average web page density in U.S.)

BLUE: low density of web pages related to “Osama bin Laden” (comparing to the average web page density in U.S.)
300 random chosen keywords to create 50,000 records (hits).

Punctuation is ignored
Words that are commonly used, like 'the,' 'a,' and 'for,' are usually ignored (these are called stop words).
Suggestion: Cyberspace Cartography
New Research Direction for Cartographers

- New Visualization Methods for Cyberspace?
- New networking symbols for social networks? 2D? 3D?

Other domain applications:
- Marketing (iPhone, Android, iPad, GIS, GPS with keyword search).
- National Security: militia groups, terrorism, radical movements.
- Infectious Diseases: SARS, Flu, Whooping Coughs.
- Natural Disaster Responses and Recovery: Earthquake, Wildfires, Hurricanes…

- Dynamic Cartographic Representation for SPACE-TIME relationships among these concepts, ideas and events.
Space – Time Animation (Whooping Cough)

Whooping Cough
Radius: 3 Cell Size: 0.5

January 17 2011
The following three research topics are essential to the future development of cyberspace mapping.

1. How to analyze the spatial relationships among points (websites and individual web pages), lines (hyperlinks within web pages), and polygons (community groups or social networks) on cyberspace?

2. How to develop effective cartographic representation methods and map symbols to illustrate the dynamic flows of ideas and concepts on the Internet?

3. Which spatial scale is the best scale to represent specific ideas or concepts on maps?
“iPhone (RED)” popularity [Minus] “Android (Blue)” popularity = differences between iPhone and Android?
(Use Map Algebra method)
“Cartography” popularity (red) MINUS “GIS” popularity (blue)
Thank You

Q & A

Ming-Hsiang Tsou
• mtsou@mail.sdsu.edu

Funded by NSF Cyber-Enabled Discovery and Innovation (CDI) program. Award # 1028177. (four years 2010-2014) http://mappingideas.sdsu.edu/