Google: Tomorrow’s Silicon (not Crystal) Ball

July 15th, 2009 — Scott Hartley

The Silicon Valley has yet to create true forecasting technology, but certain online tools are providing voyeurs with the ability to interpret political events in terms of observable changes over time. In some cases, comparing relative change over time with against an expected baseline of activity can indicate predictive deviations. Explosive growth in use of the term “SBY” across Internet platforms corroborated what polling said off-line: Incumbent President Yudhoyono was re-election bound. And Google Trends data stood in contrast to polling expectations that Jusuf Kalla would lead Megawati in second place. Google, and not polling data, corroborated actual electoral ordering.

Last Wednesday, on July 8th, Indonesia swiftly completed the second direct democratic presidential election in the country’s history. According to a national polling group, Lembaga Survei Indonesia (LSI), incumbent Susilo Bambang Yudhoyono (popularly referred to as “SBY”) won 60.82 percent of the vote, with the opposing Democratic Party of Struggle (PDI-P) candidate Megawati winning 26.57 percent, and the Golkar party candidate Jusuf Kalla taking 12.61 percent. Despite tepid claims that 5.9 million fictitious names had been
included among the eligible voters (made by Megawati’s billionaire financial contributor, Hashim Djojohadikusumo), the election took place without incident.

<table>
<thead>
<tr>
<th>NO</th>
<th>ZONA</th>
<th>MASUK</th>
<th>MEGA-PRABOWO</th>
<th>SBY-BOEDIONO</th>
<th>JK-WIN</th>
<th>PARTISIPASI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NAD</td>
<td>100.00%</td>
<td>2.63%</td>
<td>93.37%</td>
<td>3.95%</td>
<td>77.44%</td>
</tr>
<tr>
<td>2</td>
<td>SUMUT</td>
<td>100.00%</td>
<td>22.28%</td>
<td>72.12%</td>
<td>5.61%</td>
<td>64.44%</td>
</tr>
<tr>
<td>3</td>
<td>SUMBAR</td>
<td>100.00%</td>
<td>6.76%</td>
<td>80.43%</td>
<td>12.81%</td>
<td>68.81%</td>
</tr>
<tr>
<td>4</td>
<td>RAU-KPEPI</td>
<td>100.00%</td>
<td>21.33%</td>
<td>68.24%</td>
<td>12.42%</td>
<td>66.71%</td>
</tr>
<tr>
<td>5</td>
<td>JAMBI</td>
<td>98.00%</td>
<td>25.65%</td>
<td>64.61%</td>
<td>9.74%</td>
<td>72.51%</td>
</tr>
<tr>
<td>6</td>
<td>SUMSEL</td>
<td>98.39%</td>
<td>37.46%</td>
<td>56.74%</td>
<td>5.81%</td>
<td>75.66%</td>
</tr>
<tr>
<td>7</td>
<td>BENGKULU</td>
<td>100.00%</td>
<td>22.48%</td>
<td>62.02%</td>
<td>14.80%</td>
<td>67.87%</td>
</tr>
<tr>
<td>8</td>
<td>LAMPUHO</td>
<td>100.00%</td>
<td>24.62%</td>
<td>68.87%</td>
<td>5.50%</td>
<td>72.87%</td>
</tr>
<tr>
<td>9</td>
<td>BABEL</td>
<td>100.00%</td>
<td>37.80%</td>
<td>50.24%</td>
<td>11.96%</td>
<td>67.73%</td>
</tr>
<tr>
<td>10</td>
<td>DKI JAKARTA</td>
<td>100.00%</td>
<td>19.79%</td>
<td>71.25%</td>
<td>8.96%</td>
<td>68.46%</td>
</tr>
<tr>
<td>11</td>
<td>JABAR</td>
<td>99.68%</td>
<td>27.09%</td>
<td>69.94%</td>
<td>8.97%</td>
<td>77.72%</td>
</tr>
<tr>
<td>12</td>
<td>JATENG</td>
<td>100.00%</td>
<td>37.72%</td>
<td>52.44%</td>
<td>9.84%</td>
<td>70.86%</td>
</tr>
<tr>
<td>13</td>
<td>DIY</td>
<td>100.00%</td>
<td>28.75%</td>
<td>61.29%</td>
<td>10.60%</td>
<td>75.01%</td>
</tr>
<tr>
<td>14</td>
<td>JATIM</td>
<td>100.00%</td>
<td>30.48%</td>
<td>60.58%</td>
<td>8.96%</td>
<td>71.17%</td>
</tr>
<tr>
<td>15</td>
<td>BANTEN</td>
<td>100.00%</td>
<td>24.70%</td>
<td>67.19%</td>
<td>8.11%</td>
<td>74.40%</td>
</tr>
<tr>
<td>16</td>
<td>BALI</td>
<td>100.00%</td>
<td>51.12%</td>
<td>44.69%</td>
<td>4.30%</td>
<td>74.94%</td>
</tr>
<tr>
<td>17</td>
<td>NTB+TT</td>
<td>100.00%</td>
<td>24.40%</td>
<td>62.64%</td>
<td>12.68%</td>
<td>77.72%</td>
</tr>
<tr>
<td>18</td>
<td>KALBAR+KALTEN+KALTIM</td>
<td>96.77%</td>
<td>34.64%</td>
<td>53.09%</td>
<td>12.27%</td>
<td>69.48%</td>
</tr>
<tr>
<td>19</td>
<td>KALSEL</td>
<td>100.00%</td>
<td>22.39%</td>
<td>64.41%</td>
<td>13.17%</td>
<td>73.96%</td>
</tr>
<tr>
<td>20</td>
<td>SULU+SULTENG+SULTRA+SULBAR</td>
<td>98.53%</td>
<td>14.11%</td>
<td>51.94%</td>
<td>33.85%</td>
<td>73.86%</td>
</tr>
<tr>
<td>21</td>
<td>SUSEL</td>
<td>100.00%</td>
<td>3.59%</td>
<td>82.06%</td>
<td>14.35%</td>
<td>74.80%</td>
</tr>
<tr>
<td>22</td>
<td>GORONTALO</td>
<td>98.00%</td>
<td>6.05%</td>
<td>44.86%</td>
<td>49.28%</td>
<td>82.11%</td>
</tr>
<tr>
<td>23</td>
<td>MALUKU+MALUT+PAPUA+RIJABAR</td>
<td>94.34%</td>
<td>13.58%</td>
<td>66.72%</td>
<td>20.70%</td>
<td>72.89%</td>
</tr>
<tr>
<td></td>
<td>INDONESIA</td>
<td>95.49%</td>
<td>26.57%</td>
<td>66.82%</td>
<td>12.61%</td>
<td>72.59%</td>
</tr>
</tbody>
</table>

LSI Regional Presidential Election Data.

Observation of relative trends over time has been used in many contexts. For example, Raymond Fisman, co-author of Economic Gangsters, observed corruption by monitoring stock prices and news. Under the Indonesian Suharto regime, insider-information as to Suharto’s health moved the then-Jakarta Stock Exchange before news became public. In this case, insider information, driven by concealed concern over political change, facilitated opportunistic buying and selling of stock that, in moving the market price, helped reveal corrupt practices.

Today, Internet users reveal themselves publicly in a variety of ways that collaboratively paint a picture of preferences and concerns that, if not generally applicable to the populous, are immediately reflective of the online demographic in the region. In Internet ecosystems such as the Netherlands, where 90 percent of the population is online, or in the United States, where 72 percent of Americans have access, Internet trends can be more widely extrapolated to indicate the public ethos. In Indonesia, despite its low Internet penetration of roughly 5 percent, the Internet is still a useful tool to observe opinion on important issues as well as regional strongholds of support.

Over the course of 90 days leading up to the election, top Google queries across Indonesia almost exclusively included references to popular networking sites such as Facebook and Friendster. As such, among connected Indonesians, use of social networking platforms is important. “Facebook Lexicon,” a tool that allows one to observe trends of terms or topics used in “wall posts” between friends, becomes relevant in indicating shifting ethos. Within the Indonesian Facebook demographic — admittedly a small and likely young group—Facebook Lexicon reveals that over the last year there has been significant change in topics of on-site political discussion.

While Facebook is a networking and discussion platform and its use typically involves conversation, active use of Google search engine indicates explicit interest. And over the same period, data from Google Insights for Search queries confirms a swelling Internet interest in incumbent candidate SBY. Since January 1, 2009, Google queries across Indonesia on “SBY” grew by 625 percent, compared with 40 percent on “Mega.” While relative search on “Jusuf Kalla” increased by 1100 percent, his absolute search volume was roughly 90 percent lower than SBY.
Telephone polling without regard for demographics.

Retrospective analysis is always problematic. One must be wary of ex-ante conditions observed ex-post, and the ease of false attribution. And as was seen in the telephone polls of American voters prior to the 1948 presidential election, the misinterpretation of niche trends for alterations in public opinion can yield headlines such as “DEWEY DEFEATS TRUMAN,” perhaps the greatest Chicago Daily Tribune gaffe. But as Google.org and the CDC have shown, aggregated search engine query data can, by observing online health-seeking activity, “accurately estimate the current level of weekly influenza activity in each region of the United States.” In politics, if the frequency of Internet user queries about candidates correlates with the percentage of votes cast, then perhaps this information could also be applied to pre-election statistics, which could augment offline polling data on connected demographics.

Pre-election polling by LSI and other Indonesian polling groups showed SBY leading followed by Jusuf Kalla. Putative belief was that Megawati was entirely out of the running. Pre-election Google search query data pointed to an online reality that Megawati was still garnering much attention. In fact, nationally, the results for actual votes cast more closely followed the numbers of Internet search query data—not domestic polling numbers. Perhaps such observation is mere coincidence. Perhaps Internet Search Analytics is an increasingly important data point to cohere with offline demographic polling. Today the online information-seeking behavior of a geographically diverse sample of connected Indonesians is perhaps illustrative of broader pre-electoral interests. It’s not perfect, and its scope is limited to Internet penetration, but I’m betting that tomorrow’s crystal ball could be made of silicon.
• Subscribe to the I&D Blog!

Click Here

• I&D Delicious Feed

- Cashing in on Internet censorship - CNN.com
- Блоги (сетевые дневники) российских правозащитников | Права человека в России
- 20 самых популярных чиновников-блогеров. О чем они пишут? Какие делают ошибки? На что жалуются? Кто из них создает посты сам? | Forbes Russia
- Findings - People Share News Online That Inspires Awe, Researchers Find - NYTimes.com
- James Fallows

• Blogroll

- Andrew Sullivan
- Citizen Media Law Project
- DigiActive
- Digital Natives
- Ethan Zuckerman
- FP Passport
- George Packer
- Global Voices
- GV Iran
- GV Russia
- Harvard’s Berkman Center
- John Palfrey
- Marc Lynch
- Net.Effect
- Open Net Initiative
- Progressive Realist
- Rebecca MacKinnon
- Steve Coll
- TechPresident
- The Cable

• Tech Tools Feed

- 'New York' Mag Hooks Up With Foursquare
- Orchestra seeks audience by SMS
- Great White Sharks Text Their Whereabouts, For Science and Swimmer Safety
- Links for 2010-07-21 [del.icio.us]
- MySpace and Facebook: How Racist Language Frames Social Media (and Why You Should Care)

• Archives

- April 2010
- March 2010
- February 2010
- January 2010
- December 2009
- November 2009
- October 2009
- September 2009
- August 2009
- July 2009
- June 2009
- May 2009
- April 2009
- March 2009