You Are What They Post
Identity Construction on the Internet

Naomi S. Baron
American University
Washington, DC USA

Internet Research 5.0 (Association of Internet Researchers)
Sussex, England
September 19-22, 2004
Searching for Your Doppelgänger

“But I’m hot on the trail of my Google doppelgänger.”

The New Yorker, February 10, 2003  p. 44
Searching for Yourself

O wad some Power the giftie gie us
To see oursels as ithers see us

Robert Burns
“To a Louse” (1786)
The Web as a Source of Identity Construction

- Creating Our Own Identities
  - Erving Goffman: presentation of self in everyday life
  - Cybertalk at play: MUDs, MOOs, Chat, screen names, IM profiles

- Having Our Identities Created by Others
  - Profile emerging from Web searches

- Uses of Web Identity Constructions
  - Tracking down former acquaintances
  - Scoping out would-be teachers, dates, employees
Properties of Web Identity Construction

- End-users work alone (no external input or correction)
- End-users typically only check initial hits (out of thousands)
- Different search engines produce different results
- Results from individual search engines shift over time
Being Found on the Web

- **Problem Cases**
  - Common names: Sally Smith
  - Names meaning something else: Bill Banning
  - Person not involved in/subject of posted events/organizations

- **Telling You Apart from Your Doppelgänger**
  - Easy: thyroid surgeon vs. porn queen
  - Hard: author A vs. author B

- **Vagaries of Search Engines**
  - What is most important to know about person (often buried far down on list)
  - Getting pushed down in results (or out entirely) by better known people
Research Questions

1. Doppelgänger (are there any?)
2. Revelations (about yourself)
3. Stalkers (who knows what about you?)
4. Space Junk (what is still out there?)
5. Transience (what earlier hits disappeared?)
6. Lost in Translation (posts you can’t read)
7. Guilt (un-kept promises made public)

8. In sum, what do others know (or think they know) about you?
In Search of Naomi Baron: Methodology

- Used Four Search Engines
  - Google
  - Yahoo!
  - Alta Vista
  - Lycos

- Analysis
  - Time 1: May 2004
    - first 300 hits for Google, Yahoo!, and Alta Vista
    - first 50 hits for Lycos
  - Time 2: August 2004
    - first 50 hits for Google, Yahoo!, Alta Vista, Lycos
    (NOTE: new similarities of Yahoo!, Alta Vista, Lycos)
In Search of Naomi Baron: Analysis Issues

- **Profile of Each Data Set**
  - number of hits
  - location of Naomi Susan Baron (NSB) homepage
  - number of matches in first 50 hits
    - incorrect matches to “Naomi” + “Baron”
    - number of Doppelgänger hits
    - number of NSB hits
  - content of first 10 hits

- **Comparison of Data Sets**
  - across time
  - across search engines

- **8 Research Questions**
### Time 1: May 2004
Search Engine Comparison: First 50 Hits

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Yahoo!</th>
<th>Alta Vista</th>
<th>Lycos</th>
</tr>
</thead>
<tbody>
<tr>
<td># of hits</td>
<td>c. 58,200</td>
<td>c. 93,300</td>
<td>16,780</td>
<td>16,793</td>
</tr>
<tr>
<td>NSB homepage</td>
<td>1\textsuperscript{st} hit</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Not “Naomi” + “Baron”</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Doppelgänger</td>
<td>1 (1 entity)</td>
<td>6 (2 entities)</td>
<td>6 (1 entity)</td>
<td>1 (1 entity)</td>
</tr>
<tr>
<td></td>
<td>(entry includes both NSB and Doppelganger)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSB hits</td>
<td>56/50</td>
<td>41/50</td>
<td>38/50</td>
<td>40/50</td>
</tr>
</tbody>
</table>
## Time 2: August 2004

### Search Engine Comparison: First 50 Hits

<table>
<thead>
<tr>
<th></th>
<th>Google</th>
<th>Yahoo!</th>
<th>Alta Vista</th>
<th>Lycos</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong># of hits</strong></td>
<td>c. 86,700</td>
<td>c. 165,000</td>
<td>165,000</td>
<td>18,180</td>
</tr>
<tr>
<td><strong>NSB homepage</strong></td>
<td>1&lt;sup&gt;st&lt;/sup&gt; hit</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; hit</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; hit</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; hit</td>
</tr>
<tr>
<td><strong>Not “Naomi” + “Baron”</strong></td>
<td>2</td>
<td>11</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td><strong>Doppelgänger</strong></td>
<td>2 (1 entity)</td>
<td>3 (2 entities)</td>
<td>6 (2 entities)</td>
<td>3 (2 entities)</td>
</tr>
<tr>
<td><strong>NSB hits</strong></td>
<td>46/50</td>
<td>36/50</td>
<td>36/50</td>
<td>22/50</td>
</tr>
</tbody>
</table>
### Time 1: May 2004
#### Search Engine Comparison: First 10 Hits

<table>
<thead>
<tr>
<th>Google</th>
<th>Yahoo!</th>
<th>Alta Vista</th>
<th>Lycos</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Alphabet to Email</td>
<td>3. Visible Lg. article</td>
<td>3. Fall 2002 course</td>
<td>3. Fall 2002 course</td>
</tr>
<tr>
<td>5. ESU Award</td>
<td>5. Fall 2002 course</td>
<td>5. Work cited</td>
<td>5. Work cited</td>
</tr>
</tbody>
</table>
## Time 2: August 2004

### Search Engine Comparison: First 10 Hits

<table>
<thead>
<tr>
<th>Google</th>
<th>Yahoo!</th>
<th>Alta Vista</th>
<th>Lycos</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Work cited</td>
<td>2. <a href="#">Homepage</a></td>
<td>2. <a href="#">Homepage</a></td>
<td>2. <a href="#">Homepage</a></td>
</tr>
<tr>
<td>5. ESU Award</td>
<td>5. <a href="#">Alphabet to Email</a></td>
<td>5. “Art and Science of Handwriting”</td>
<td>5. Work cited</td>
</tr>
<tr>
<td>7. <a href="#">Other Naomi Baron</a></td>
<td>7. <a href="#">Alphabet to Email</a></td>
<td>7. Growing Up with Language</td>
<td>7. <a href="#">Bewitched Baron</a></td>
</tr>
<tr>
<td>8. <a href="#">Alphabet to Email</a></td>
<td>8. <a href="#">Alphabet to Email</a></td>
<td>8. <a href="#">Alphabet to Email</a></td>
<td>8. <a href="#">Naomi Katz</a></td>
</tr>
<tr>
<td>10. <a href="#">Alphabet to Email</a></td>
<td>10. <a href="#">Alphabet to Email</a></td>
<td>10. <a href="#">Alphabet to Email</a></td>
<td>10. <a href="#">Alphabet to Email</a></td>
</tr>
</tbody>
</table>
Comparisons across Time

- More engines included Homepage in Time 2 (August) than in Time 1 (May)
- Persistence of entry “Art and Science of Handwriting”
- Significant divergence in first 50 hits across time
- Variance in number of hits for each engine
Comparisons across Engines

- Lycos most imprecise (Time 1 and Time 2)
  - most not “Naomi” + “Baron”

- Yahoo!, Alta Vista, and Lycos highly similar by T2, due to business mergers

- Variance in number of hits
  - however: after 200-300, does anyone care?
### Research Questions:

1. **Doppelgänger** (are there any?)

1. **Baltimore**: widow; author of book on managing alone
2. **Key West**: restaurateur
3. **Toledo**: officer in synagogue
4. **1851**: died in England
5. **Tudor Grange** (UK): high school alumna
6. **St Peters Lutheran** (Australia): high school senior
7. **Egypt Trip**: Q: same as Baltimore? Toledo?
8. **Georgia Family**: is “Naomi” also a “Baron”?
9. **Seeing Eye Dog**: Naomi, Baron and Grandma Shirley
# Doppelgänger Hits in Data

<table>
<thead>
<tr>
<th></th>
<th>May (Red = hits 1-50 Black = hits 51-300)</th>
<th>August (Red = hits 1-50)</th>
</tr>
</thead>
</table>
| 1. Baltimore | G: 1, 0  
Y: 1, 0  
AV: 0, 1  
L: 0 | G: 0  
Y: 0  
AV: 0  
L: 0 |
| 2. Key West | G: 0, 7  
Y: 5, 5  
AV: 6, 3  
L: 0 | G: 0  
Y: 2  
AV: 5  
L: 0 |
| 3. Toledo | G: 0, 0  
Y: 0, 2  
AV: 0, 1  
L: 0 | G: 0  
Y: 0  
AV: 0  
L: 0 |
| 4. 1851 | G: 0, 1  
Y: 0, 1  
AV: 0, 1  
L: 0 | G: 0  
Y: 0  
AV: 0  
L: 0 |
| 5. Tudor Grange (UK) | G: 0, 1  
Y: 0, 1  
AV: 0, 0  
L: 0 | G: 0  
Y: 0  
AV: 0  
L: 0 |
| 6. St Peters Lutheran (Australia) | G: 0, 0  
Y: 0, 1  
AV: 0, 0  
L: 0 | G: 0  
Y: 0  
AV: 0  
L: 0 |
| 7. Egypt Trip | G: 0, 1  
Y: 0, 1  
AV: 0, 0  
L: 0 | G: 0  
Y: 0  
AV: 0  
L: 0 |
| 8. Georgia Family | G: 0, 0  
Y: 0, 1  
AV: 0, 1  
L: 0 | G: 0  
Y: 0  
AV: 0  
L: 1 |
| 9. Seeing Eye Dog | G: 0, 3  
Y: 0, 1  
AV: 0, 1  
L: 0 | G: 2  
Y: 1  
AV: 1  
L: 2 |
Naomi, Baron and Grandma Shirley

NOTE: Search engines do not recognize punctuation.
Research Questions:

2. Revelations (about yourself)

- Appearance in Anthology
  - “The Art and Science of Handwriting” in *Writing Material*

- Paper Mill, Digital Library
  - AcaDemon ($71.95 for 8.1 pages, 10 references)
  - Questia.com (who gets the royalties from the $19.95/month?)

- Public Reaction to Away Messages Study
  - Blogs bashing “away messages” as topic of serious study
    (“Away messages? I’m going to do an independent study on e-mail signatures. I want a B.A. in answering machine anthropology.”)

- Research Citations, Syllabi, Blogs
  - Who knew they knew?
  - cf. Amazon “Search inside this book”
Research Questions:

3. Stalkers (who knows what about you?)

- Issue: Individuals have little control over
  - what gets posted about them
  - who reads these posts

- Example of Internet stalking
  - Army Specialist in Virginia requesting advice regarding studying computer languages
    (“OK, ma’am. I see….I send my best regards to you and the future 2003 Linguist Conference in Spain”)

- Cf. Stalking from other “public” appearances
  - email response to NPR show on English as a global language
    (“I do not understand what motivates your xenophobia and your hatred of the United States….May you burn in hell….”)
Research Questions:

4. Space Junk (what is still out there?)

- Outdated Material
  - old lectures, syllabi
  - MA, PhD theses advised 20-30 years ago
  - online book vendor ad for 1977 book (within first 10 hits; wrong title)

- Personal Material
  - 5k race from 1997 (including time)
  - conference photo
Research Questions:

5. Transience (what earlier hits disappeared?)

- Heraclites of Ephesus Updated:
  - “You can’t step into the same Web search twice.”

- Examples
  - NPR “Car Talk”
  - Ohio natural foods market raffle winner
    - Q: the synagogue officer?
    - Q: the Egypt trip-taker?
Research Questions:

6. Lost in Translation (posts you cannot read)

- Web sites in other languages
  - e.g., Spanish, French, Latvian, Serbian

- Web sites in other scripts
  - e.g., Japanese, Thai

- “Translate this page”
  - e.g., Italian: “little pessimist of the Baron”
    (original: “Meno pessimista della Baron, il professor David Crystal … nega che…” = “Less pessimistic than Baron, Professor David Crystal says that…”)


Research Questions:

7. Guilt (un-kept promises made public)

- Pending book review
  - listed on author’s web site

- Reference to forthcoming book
  - “qui paraîtra bientôt” (= “which will appear soon”)
Research Questions:

8. In sum, what do others know (or think they know) about you?

- Which Naomi Baron am I?
  - Do I run a restaurant?
  - Did my husband die?
  - Have I been to Egypt?

- Even if I’m the right one, what is important to know about me?
  - What information comes up in the early hits?
  - Is that the information I personally want people to know?

- Ignoring the public or defending your honor?
  - Should inaccurate blogs, citations be ignored (like unfavorable and/or inaccurate book reviews)?
  - When do you pursue copyright protection? (e.g., Questia)
“T” is for Terrorist?

(“T. Kennedy” on Airport Security Watchlist, March 2004)
Concluding Remarks

- **Dangers of Mistaken Identity**
  - potential teachers, dates, employees

- **Learning the Scope of Your Reach**
  - ego-trip or heightened sense of responsibility?

- **The End of Privacy?**
  - Web stalking
  - the Web as the new panopticon
    - average Londoner is captured on surveillance cameras more than 300 times a day (Electronic Privacy Information Center, 2002)
  - can we opt out of Web identities?
  - if not, what kind of people do we become?
I assure you, the author really is …

Naomi S. Baron
American University
Washington, DC   USA

nbaron@american.edu