Geovisualization of News Stories

Julie Dillemuth
PhD Candidate
Geography Dept.
UC Santa Barbara

February 27, 2007
OSU Geovis Series

Collaborator:
Alex Villacorto,
Statistics and Applied Probability
Interactive Digital Multimedia
IGERT, UCSB

11 departments

30 students, 20 faculty

3 collaborative projects

http://media.igert.ucsb.edu/
Projects

Spheres of Influence II
Geospatial and news data integration & visualization with a camera-tracking interface

MAT, ECE, Geography, Statistics, Psych, CS, Art
Projects - 2

Out of the Ether II
Audio, video, and electric field sensors for cueing control during live musical performance

MAT, ECE, Music
Minotour: a mobile, spatially aware Wiki
A PDA-based tour guide that weaves a narrative for a user’s journey based on Wikipedia article snippets

Geography, Psych, Film Studies, CS
Overview

Driving Questions

Big Picture

Our approach: single story case study
national front-page stories

‘Spheres of Influence’ project
Driving Questions

How is news reported around the country; who has access to what news stories and when?

• How can semantic meaning and geography be related through visualization?

• Does news spread like a disease; can it be modeled?
Big Picture

3 ‘spaces’ of news events
  • geographic
  • temporal
  • semantic

Geovisualization for:
  • data exploration & pattern i.d.
  • presentation to broader audience

Spatial -- Temporal -- Semantic analysis of news stories
Buzztracker
www.buzztracker.org

Newsmap
www.marumushi.com/apps/newsmap

Spacecast
geog.ucsb.edu/~sara

ThemeRiver
www.pnl.gov/infoviz/technologies.html
**GeoViz**

Graphic display of complex spatial data

- Audience
- Data relations
- Interaction

Our Approach

- case study – single story, compare 2 states
- what can geovisualization do
- feasibility of automation
The Story

Semantic: Terri Schiavo (right-to-die case in FL)
Temporal: Feb. 05 - June 05
Geographic: reporting in Texas & California

Data source: America’s Newsbank digital archive
Keyword search

Newspapers: 32 CA papers
13 TX papers
Constraints

• newspaper archive incomplete
• archives have inconsistent structure – impedes parsing
• newspaper circulation areas hard to determine
Process

1) America’s Newsbank keyword search

1367 articles CA
716 articles TX
2) Exported text parsed to excel file

- Paper
- City
- Date
- Headline
- Word count

<table>
<thead>
<tr>
<th>Paper</th>
<th>City</th>
<th>Date</th>
<th>Headline</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventura County Star (California)</td>
<td>Ventura</td>
<td>23-Oct-2003</td>
<td>Bush's stay of execution; Florida case highlights value of the Schiavo</td>
<td>682</td>
</tr>
<tr>
<td>Sacramento Bee</td>
<td>Sacramento</td>
<td>25-Oct-2003</td>
<td>The Schiavo intrusion Governor, lawmakers tread on court</td>
<td>500</td>
</tr>
<tr>
<td>The Daily Review (Hayward, CA)</td>
<td>Hayward</td>
<td>3-Nov-2003</td>
<td>Shocked by statements</td>
<td>572</td>
</tr>
<tr>
<td>The San Diego Union-Tribune</td>
<td>San Diego</td>
<td>6-Nov-2003</td>
<td>Guardian at issue in right-to-die fight</td>
<td>105</td>
</tr>
<tr>
<td>San Mateo County Times (San Mateo, CA)</td>
<td>San Mateo</td>
<td>19-Nov-2003</td>
<td>Right to live</td>
<td>368</td>
</tr>
<tr>
<td>Ventura County Star (California)</td>
<td>Ventura</td>
<td>25-Jan-2005</td>
<td>Court rules on Schiavo, asks for reviews of cases</td>
<td>575</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>24-Feb-2005</td>
<td>The Nation; Schiavo Feeding Tube Stays for Now; Judge</td>
<td>641</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>26-Feb-2005</td>
<td>Judge Orders Feeding Tube Removed; A three-week delay</td>
<td>722</td>
</tr>
<tr>
<td>The San Diego Union-Tribune</td>
<td>San Diego</td>
<td>11-Mar-2005</td>
<td>Schiavo's husband offered $1 million; Local businesses</td>
<td>380</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>13-Mar-2005</td>
<td>Passionate Debate Over the Right to Die</td>
<td>426</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>19-Mar-2005</td>
<td>Schiavo Taken Off Food Supply; Congress tries to stop it</td>
<td>1987</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>20-Mar-2005</td>
<td>Weary Mother Bags for Her Daughter's Life; As demonstrated</td>
<td>621</td>
</tr>
<tr>
<td>Ventura County Star (California)</td>
<td>Ventura</td>
<td>21-Mar-2005</td>
<td>Bush signs Schiavo legislation</td>
<td>893</td>
</tr>
<tr>
<td>Inland Valley Daily Bulletin (Ontario, CA)</td>
<td>Ontario</td>
<td>22-Mar-2005</td>
<td>Few California legislators on hand for Schiavo vote</td>
<td>297</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>22-Mar-2005</td>
<td>THE TERRI SCHIAVO CASE; COST OF CARE; Financ</td>
<td>237</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>23-Mar-2005</td>
<td>THE TERRI SCHIAVO CASE; Judge Raises Doubts About</td>
<td>1013</td>
</tr>
<tr>
<td>San Gabriel Valley Tribune (San Gabriel V)</td>
<td>San Gabriel V</td>
<td>23-Mar-2005</td>
<td>It's murder...</td>
<td>945</td>
</tr>
<tr>
<td>The Daily News of Los Angeles</td>
<td>Los Angeles</td>
<td>22-Mar-2005</td>
<td>FEW CALIFORNIA LAWMAKERS CAST SCHIAVO VOTE</td>
<td>309</td>
</tr>
<tr>
<td>The San Diego Union-Tribune</td>
<td>San Diego</td>
<td>23-Mar-2005</td>
<td>The Schiavo case, Congress was wrong to intervene</td>
<td>494</td>
</tr>
<tr>
<td>Ventura County Star (California)</td>
<td>Ventura</td>
<td>23-Mar-2005</td>
<td>Judge defers decision in Schiavo case</td>
<td>969</td>
</tr>
<tr>
<td>The San Francisco Chronicle</td>
<td>San Francisco</td>
<td>23-Mar-2005</td>
<td>Life, death and politics</td>
<td>514</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>23-Mar-2005</td>
<td>THE TERRI SCHIAVO CASE; MEDICAL PERSPECTIVE</td>
<td>1084</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>23-Mar-2005</td>
<td>THE TERRI SCHIAVO CASE; POLITICAL AFTERMATH</td>
<td>681</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>Los Angeles</td>
<td>23-Mar-2005</td>
<td>THE TERRI SCHIAVO CASE; Parents Lose Appeal in S</td>
<td>1896</td>
</tr>
<tr>
<td>San Gabriel Valley Tribune (San Gabriel V)</td>
<td>San Gabriel V</td>
<td>23-Mar-2005</td>
<td>Newspapers weigh in on Schiavo case</td>
<td>591</td>
</tr>
<tr>
<td>San Gabriel Valley Tribune (San Gabriel V)</td>
<td>San Gabriel V</td>
<td>23-Mar-2005</td>
<td>Disgusting Intervention</td>
<td>1240</td>
</tr>
</tbody>
</table>
3) Geo-coded papers, mapped in ArcGIS
4) Annotated and animated in Flash
Viz 1: presence of ‘Terri/y Schiavo’ keyword

Viz 2: event-related vs. tangential articles
Viz 1: presence of ‘Terri/y Schiavo’ keyword

- First analysis: salient events
- Big cities first to carry story, smaller ones follow on
- Plot of articles-per-day similar to disease pattern
- Short attention span, or ‘immunity’ for news story
Viz 2

Viz 2: event-related vs. tangential articles

- Article classes:
  - events (176 CA, 184 TX) = fact-based
  - tangential (356 CA, 306 TX) = broader impact
  - unrelated (59 TX, 533 CA) = affected

- differences in two classes of article, 2 states
  - event-related until Congress & Bush involved (3/21)
  - western TX picks up story only then
  - tangential in TX before CA (3/21)
  - at TS death, more tangential than event
V i z 2

Evaluative questions, further analysis:

• what does the spatial component provide that graphs don’t?

• how similar are the stories across papers and states?
Limitations

• potential for misinterpretation, overinterpretation
  incomplete paper coverage
  publication locations, not readership coverage
  content type of article

• automation challenges: classifying articles
Observations

• Major cities ‘early adopters’ of story

• Uniformity in reporting only for major or unusual events

• Story growth and decay similar to disease

• Short attention span for story

• Event-related versus tangential stories have different patterns
Story: All front page news
Duration: Aug. ’99,’01,’03
Location: Nationwide
Newspapers: 15 papers
Data: America’s Newsbank online database

Filter Criteria
- Top 15 most circulated (excluding multiple cities)
- > 3 papers carry story
- > 5 publications total
Avg. Circulation Size 2001 & 2003:

1,140,000 – 850,000: NYT, LA Times, Chicago Trib, WA Post
770,000 – 500,001: Denver, Detroit, Dallas, Houston, San Fran.
500,000 – 370,000: Boston, Atlanta, Ariz., Newark, Minneapolis, Philly

Bar height represents the percentage of all national-level stories carried by each paper. Color indicates the average 'response time' for each paper, or number of days after a story breaks before it first appears on the front page.
National News 2001

- 40 national-scale stories, 640 articles
National News 2003

- 27 national-scale stories, 781 articles
• On average, domestic stories are carried about the same.
• International stories are carried more frequently in east papers.
Implications

• contributes to understanding of communication channels

• Integrates semantic, spatial and temporal domains of dataset

• Towards an automated, interactive tool to visualize any text-based database query

• Inspired ‘Spheres of Influence’ concept & content
Spheres of Influence II
Geospatial and news data integration & visualization with a camera-tracking interface

MAT, ECE, Geography, Statistics, Psych, CS, Art
Global News Reporting

Alex Villacorta statistics
Anne-Marie Hansen art
Emily Moxley electrical & computer engineering
Jonathan Ventura computer science
Joriz De Guzman computer science
Karl Grossner geography
Matt Peterson psychology
Video demo
Spheres of Influence 2007

**Story Statistics:**

- Total number of reports: **211**
- Avg number of reports per day: **75**
- Total time in database: **2 weeks**
- Rank in database: **22nd**
- Highest rank: **3rd**
- Lowest rank: **131st**
- Max/Min reports: **249/6**
- Location w/ most reports: **North America**
- % of all known locations: **2%**
- Number of keywords: **25**
Thanks to:

Keith C. Clarke, Advisor
Shane Kendrick, CNSI Apprentice Researcher Program
Sara Fabrikant
UCSB IGERT in Interactive Digital Multimedia
NSF Graduate Research Fellowship
Dr. Dawn Wright
OSU Geosciences, Eco-IGERT
Questions, comments?