Processing Capstone Email
Using Predictive Coding
Brent West
Outline

- Capstone
- Barriers
- Predictive Coding
- Projects
What is Capstone?

• Simplified email management
  ➢ Reduced burden
  ➢ Reduced risk

• Retention based on position
  ➢ Preserve permanent email
  ➢ Address legacy accounts

Barriers to Access

- Mixed content
- Organization
- Search
- Sensitivity
- Trust
Predictive Coding

• Technology-assisted review
  ➢ Classification
  ➢ Redaction

• Continuous active learning
  ➢ Training
  ➢ Recall
Predictive Coding

Target population:
- 77985 (14.4%)
- 316578 (45.4%)
- 296120 (43.2%)

Documents (excluding attachments):
- 53393 (18.2%)
- 141251 (47.7%)
- 100676 (34.1%)

Emails:
- 118399 (38.9%)
- 193179 (62.0%)
- 0 (0.0%)

Attachments:
- 5567 (7.7%)
- 78156 (92.3%)
Predictive Coding

Review-recall ratio

Parameters

<table>
<thead>
<tr>
<th></th>
<th>Review</th>
<th>Recall</th>
<th>Next relevant</th>
<th>Total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>84.6%</td>
<td>$13.60</td>
<td>$135.7K</td>
</tr>
</tbody>
</table>

Distribution by relevance score

Details

Collection

- Files in collection: 605,501
- Richness: 5.20%
- Error-margin: 1.17%

Review set parameters

- Read %: 20%
- Files in review set: 135,659
- Cutoff score: 7.42

Review set statistics

- Recall: 84.59% ±7.91%
- Precision: 22.23% ±6.77% -4.44%
- F Measure: 35.21% ±8.21% -7.91%

Review cost estimates

- Review cost per file: $1.00
- Marginal cost to capture an additional relevant file in review: $13.60
- Average cost to capture relevant file in review: $4.50
- Estimate total review cost: $135.7K

Family

- Files in review set with family text files below cut-off: 20,348
- Files in review set with family text an non-text files below cut-off: 108,527
- Estimated total review cost with family text files below cut-off: $20,348.00
Predictive Coding

<table>
<thead>
<tr>
<th>Index status</th>
<th>Basic search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Search phrase</td>
</tr>
</tbody>
</table>

Showing 1-30 out of 200 results. New search

<table>
<thead>
<tr>
<th>Search phrase</th>
<th>View: Themes Actions Sort by number of relevant files</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe, concerned, fact, reason, cases</td>
<td></td>
</tr>
<tr>
<td>electricity, california, utility</td>
<td></td>
</tr>
<tr>
<td>developments, manager, business, technologies</td>
<td></td>
</tr>
<tr>
<td>fare, prices</td>
<td></td>
</tr>
<tr>
<td>party, contracts, terminal</td>
<td></td>
</tr>
<tr>
<td>direction, cpuc, decisions, sce, access</td>
<td></td>
</tr>
<tr>
<td>rule, commission, required, authority, state</td>
<td></td>
</tr>
<tr>
<td>year, saying</td>
<td></td>
</tr>
<tr>
<td>draft, comments, revised, review</td>
<td></td>
</tr>
<tr>
<td>confidentiality, intends, message, privileges, recipients</td>
<td></td>
</tr>
<tr>
<td>bills</td>
<td></td>
</tr>
<tr>
<td>point, talked, problem</td>
<td></td>
</tr>
<tr>
<td>don, going</td>
<td></td>
</tr>
</tbody>
</table>

Results per page: 10 - 50
Preserving Email Messages of Enduring Value

- Analysis phase
  - Observations
  - Workflows
  - Management

- Pilot phase
  - Ingest
  - Repository
  - Filtering tools
Mining Large Email Collections Using Text Analytics

- Collect email
- Test tools
- Improve tools
  - Active use
  - FOIA
  - Archives
Processing Capstone Email Using Predictive Coding

• State of Illinois officials
• Test tools
  ▶ Identify sensitive content
  ▶ Generate descriptive metadata
  ▶ Efficiency
• Public access
Next Steps

• Equivio preview
• Email
• Experts
• Testers

go.uillinois.edu/Capstone