

Processing Capstone Email Using Predictive Coding

Brent West



UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

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



Source: www.archives.gov/records-mgmt/bulletins/2013/2013-02.html

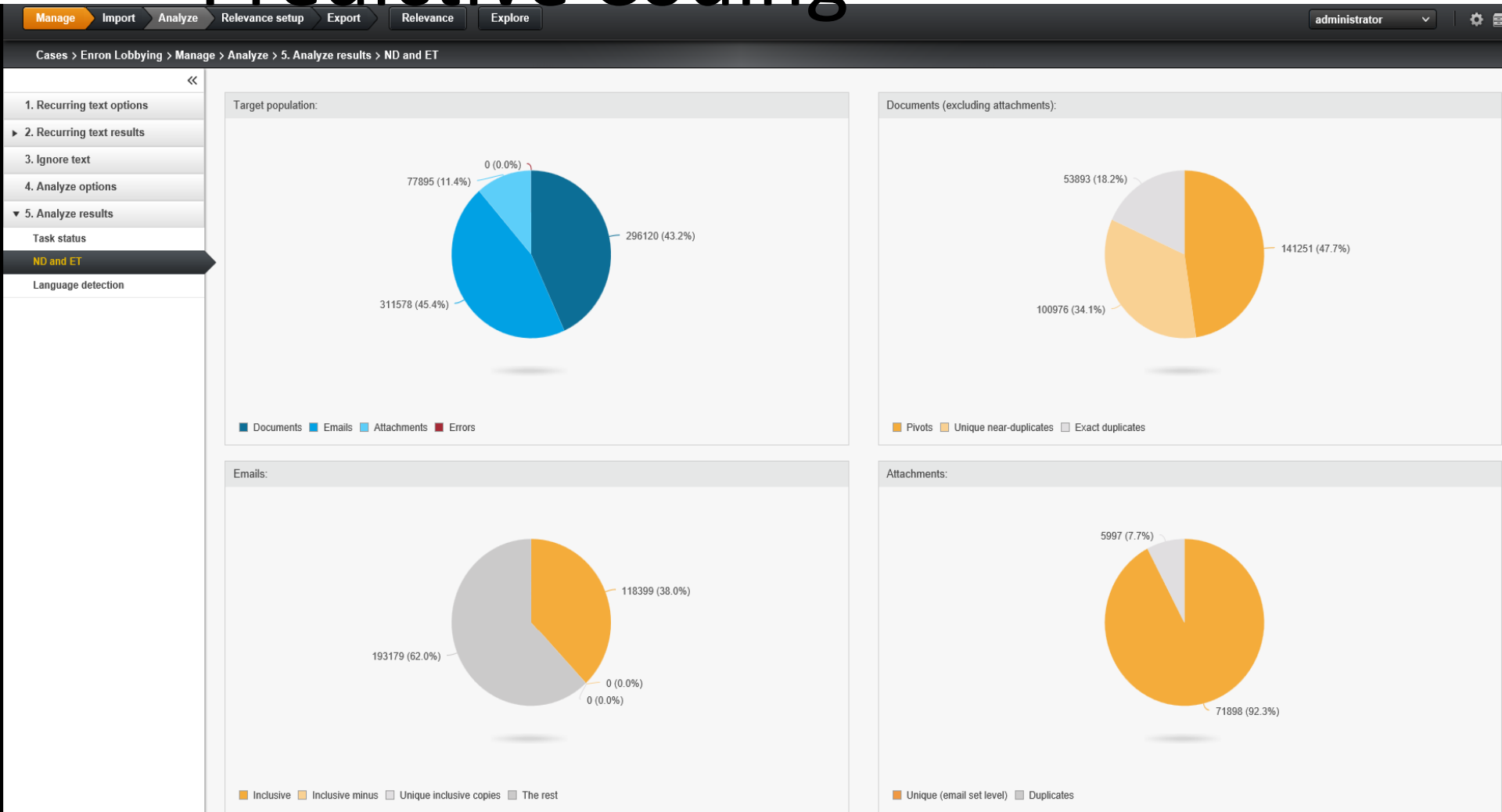
Barriers to Access

- Mixed content
- Organization
- Search
- Sensitivity
- Trust

Predictive Coding

- Technology-assisted review
 - Classification
 - Redaction
- Continuous active learning
 - Training
 - Recall

Predictive Coding



Predictive Coding

Manage

Relevance

Track

Tag

Decide

Test

Explore

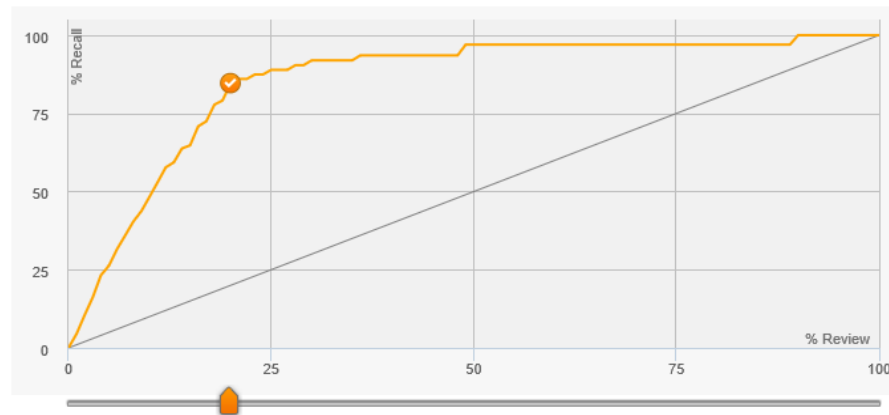
administrator



Cases > Enron Lobbying > Relevance > Decide

Issue Super-issue – Lobbying

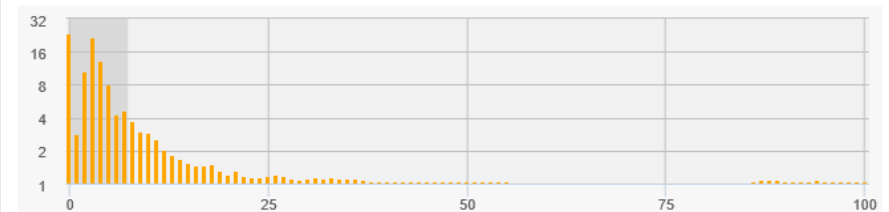
Review-recall ratio



Parameters

Review	Recall	Next relevant	Total cost
20%	84.6%	\$13.60	\$135.7K

Distribution by relevance score



Details

Collection	Review set parameters	Review set statistics	Review cost estimates	Family
Files in collection: 685,591	Read %: 20%	Recall: 84.59% Error-margin: ±7.91%	Review cost per file: \$1.00	Files in review set with family text files below cut-off: 20,348
Richness: 5.20% Error-margin: 1.17%	Files in review set: 135,659	Precision: 22.23% Error-margin: +6.77%, -6.44%	Marginal cost to capture an additional relevant file in review: \$13.60	Files in review set with family text in non-text files below cut-off: 108,527
	Cutoff score: 7.42	F Measure: 35.21% Error-margin: +8.21%, -7.91%	Average cost to capture relevant file in review: \$4.50	Estimated total review cost with family text files below cut-off: \$20,348.00
			Estimate total review cost: \$135.7K	

Predictive Coding

Basic search

Showing 1-30 out of 200 results, *New search** View: Themes Actions Sort by number of relevant files

believe, concerned, fact, reason, cases	
electricity, california, utility	
developments, manager, business, technologies	
ferc, prices	
party, contracts, terminal	
direction, cpuc, decisions, sce, access	
rule, commission, required, authority, state	
year, saying	
draft, comments, revised, review	
confidentiality, intends, message, privileges, recipients	
bills	
point, talked, problem	
don, going	

Search phrase

Index status

Add Filter... Search

Page 1 of 7 Results per page: 10 30 50

PROJECTS

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