

Towards Automatic Linkage of Analyst's Claims with Associated Evidence from Screenshots

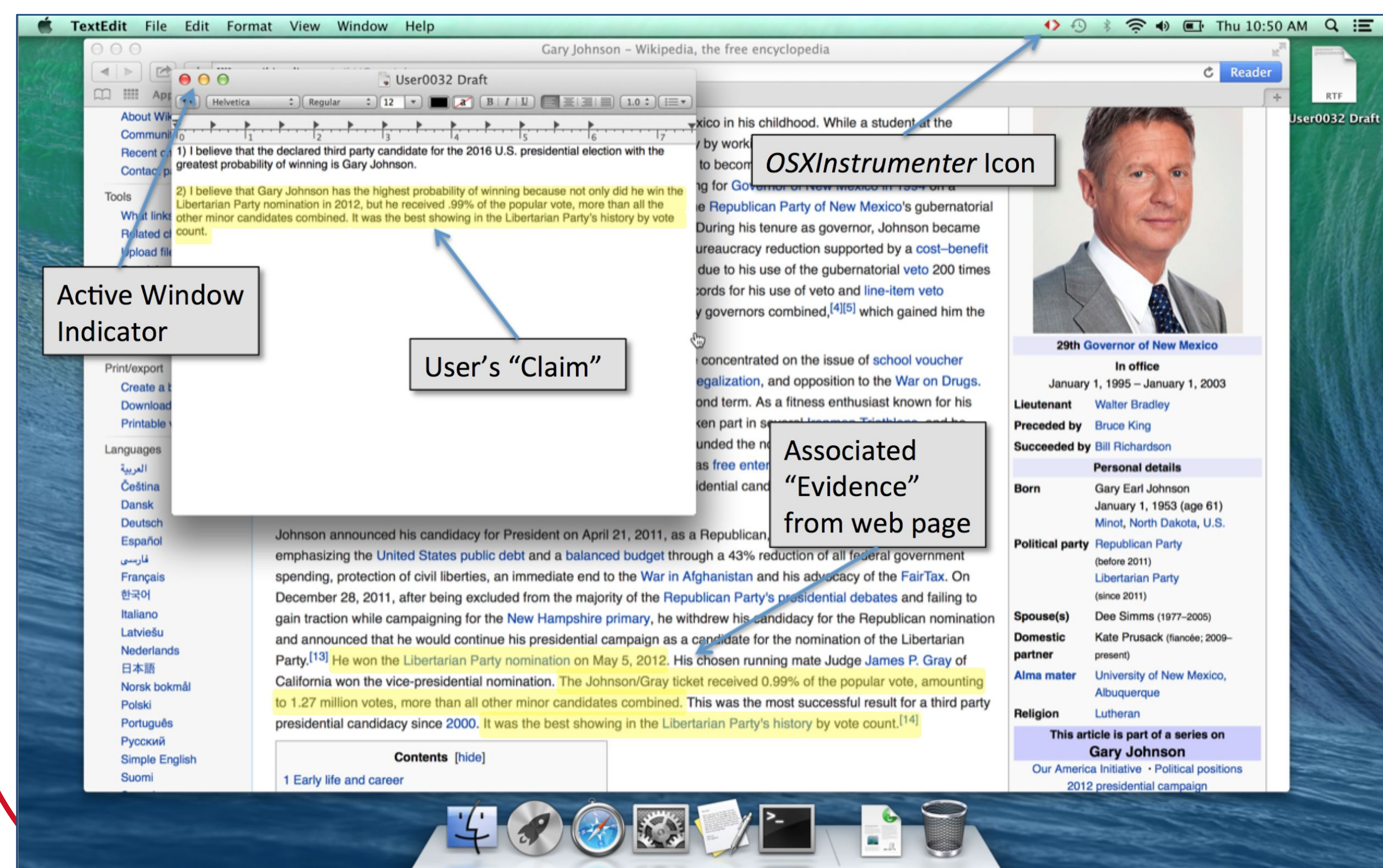
Paul Jones, Dakota Medd, Sreekanth Ramakrishnan, Rajat Shah, Joann Keyton, Nagiza F. Samatova, CSC & CHASS, NC State University

Motivation

- Volume and variety of data is continuing to outpace automated methods that help people to analyze it.
 - As a result, knowledge workers falling victim to cognitive biases, particularly **confirmation bias**.
- Need to build computer systems to understand the **sensemaking** process of analysts and pro-actively help them achieve their goals. This is the purpose of **instrumentation** but app-specific code is needed to extract content from applications, which becomes complex and brittle; screenshot analysis provides an alternative approach.
- Initial research question: **can we automatically associate claims with evidence used to support them?**
- Core assumption: workers are analyzing new information under time pressure; they are not making claims based on prior knowledge of a particular topic, and they have minimal time to contemplate and refactor new knowledge that they learn. Hence we assume that *claims and associated evidence will be found close together in time and in space on the computer screen*.

Screenshot Analysis Approach

- Screenshot corpus collected during 2014 LAS/CHASS instrumentation experiment:
 - 150 participants (in 54 groups) performed a controlled analysis task...
 - 'Which 3rd party candidate stands the highest chance of winning in the 2016 US Presidential Election?'**
 - Users wrote their report in *TextEdit* app, browsing for evidence in *Safari*.
 - 121,000 screenshots captured by **OSXInstrumenter**
 - Periodically every 10 seconds, and on mouse-click and Enter key-pressed events.
- Key steps in algorithms, and approaches used:..
 - Active window detection and boundary detection - template matching from *OpenCV*.
 - Text extraction/OCR and segmentation - *Tesseract* and new sentence-length filtering algorithm.
 - Sentence similarity methods - Char sequence (DiffLib), BLEU, Word matching (Jaccard), TF/IDF, WordNet.
- Two algorithm variants evaluated:
 - Stateless** (claim and evidence are both in same screenshot e.g. example below)
 - Stateful** (evidence can occur in any screenshot prior to the claim being written in TextEdit).



Algorithm 1 Stateless Algorithm: extract claim-evidence pairs when they both occur in the *same* screenshot

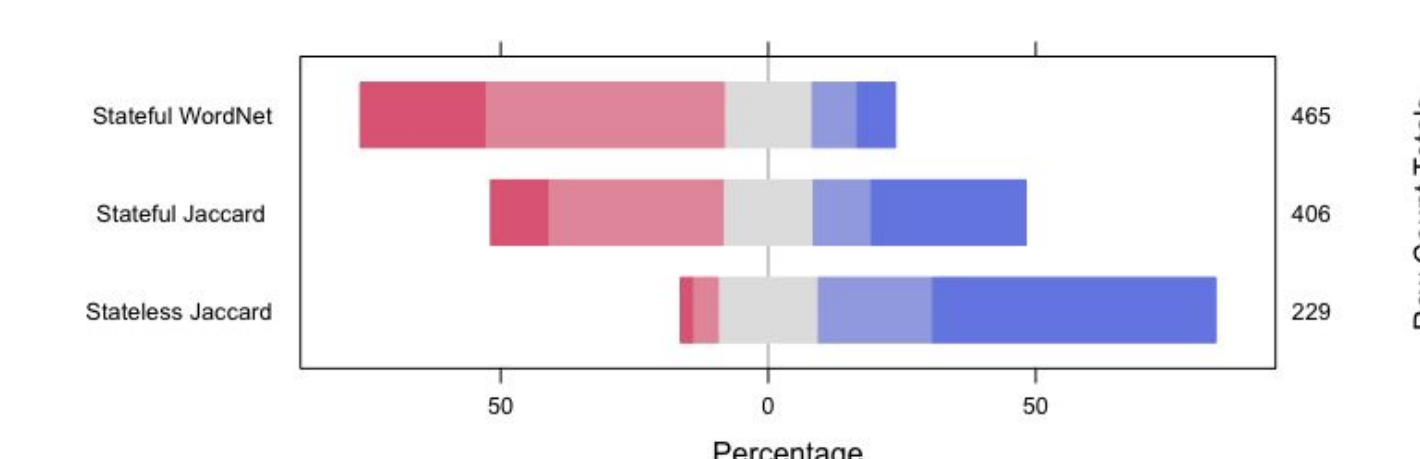
```
1: function FINDSTATELESSPAIRS(ScreenShot s)
2:   pairs ← empty set
3:   if Editor is active in s then
4:     EditorBoundary ← extract boundary from s
5:     BrowserBoundary ← extract boundary from s
6:     claims ← OCR on EditorBoundary
7:     evidences ← OCR on BrowserBoundary
8:     for each claim ∈ claims do
9:       for each evidence ∈ evidences do
10:        sim ← similarity(claim, evidence)
11:        if sim ∈ [MIN, MAX] then
12:          pairs.append(claim, evidence, sim)
13:        end if
14:      end for
15:    end if
16:  end if
17: end function
```

Algorithm 2 Stateful Algorithm: extract claim-evidence pairs from *different* screenshots

```
1: function FINDSTATEFULPAIRS(Screenshots S)
2:   pairs ← empty set
3:   sE ← most recent editor screenshot
4:   sB ← subset of S preceding sE
5:   EditorBoundary ← extract boundary from sE
6:   claims ← OCR on EditorBoundary
7:   for each s ∈ sB do
8:     BrowserBoundary ← extract boundary from s
9:     evidences ← OCR on BrowserBoundary
10:    for each claim ∈ claims do
11:      for each evidence ∈ evidences do
12:        sim ← similarity(claim, evidence)
13:        if sim ∈ [MIN, MAX] then
14:          pairs.append(claim, evidence, sim)
15:        end if
16:      end for
17:    end for
18:  end for
19: end function
```

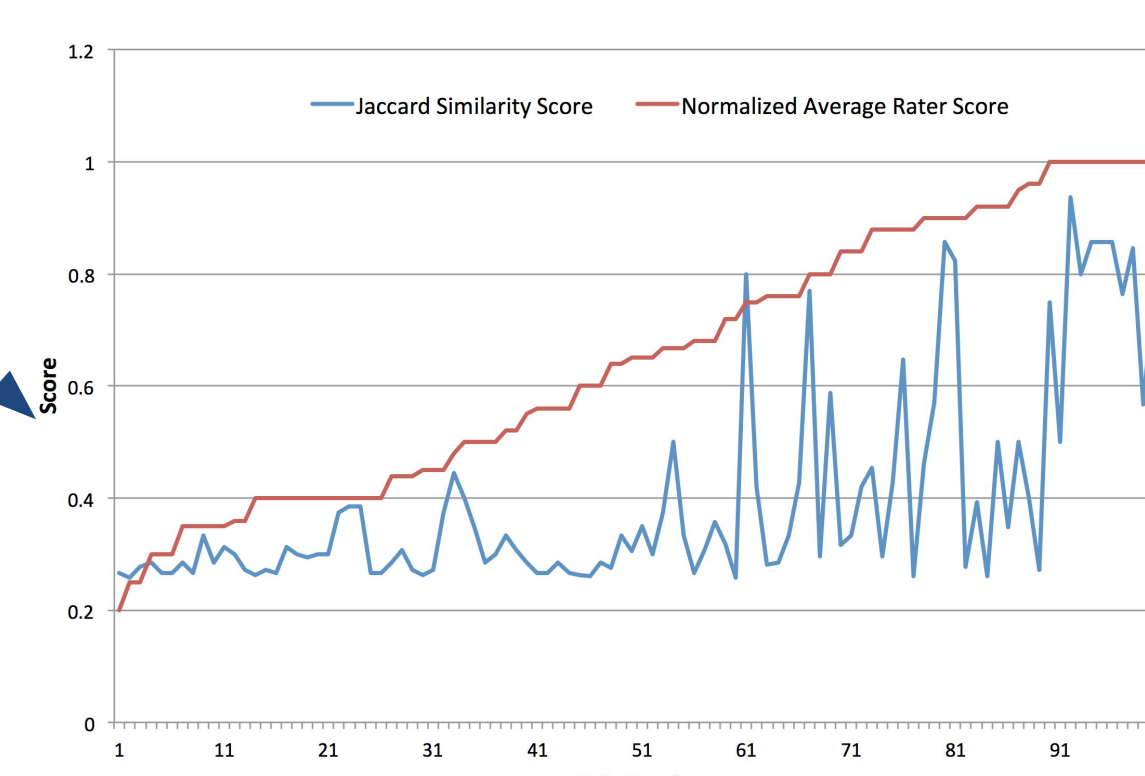
Evaluation Results

- Extracted putative claim and evidence sentences and matched using 5 different sentence similarity approaches.
- For the most promising ones, extracted top-100 for human evaluation:
 - 5 raters, each providing a score for the quality of claim/evidence associations using 5-point Likert scale.
 - Compared human to computer scores - pearson correlation of 0.66.
 - Kappa analysis of inter-rater variation showed good agreement for scores 1, 2 and 5.
 - Rater scores:
 - 1/2 - Red
 - 3- Gray
 - 4/5 - Blue



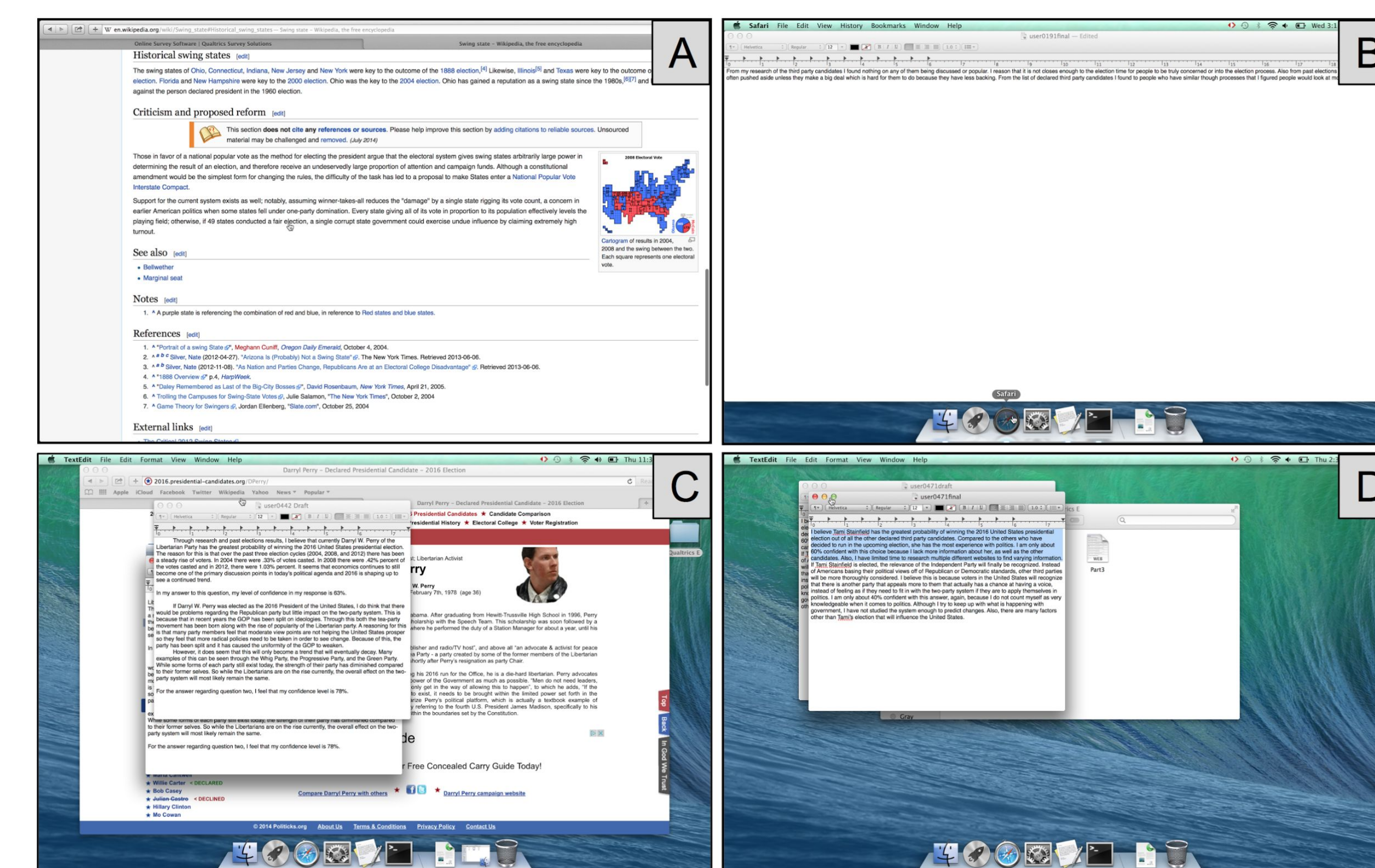
Metric	Stateless	Stateful
TextEdit sentences extracted ('claims')	109,465	1,450
Safari sentences extracted ('evidence')	89,669	100,124
Claim-evidence matches (0.3 threshold)		
...from DiffLib char sequence similarity	19,558	110,457
...from BLEU similarity	45,001	297,931
...from Jaccard (word) similarity	212	275
...from TF/IDF and cosine similarity	-	887
...from WordNet similarity	-	37,251
Claim-evidence matches (0.15 threshold)		
...from DiffLib char sequence matching	74,013	485,648
...from BLEU similarity	58,198	371,560
...from Jaccard (word) similarity	1,340	3,711
...from TF/IDF and cosine similarity	-	9,701
...from WordNet similarity	-	478,738

Score	Description
1	Poor - nonsense, no association between sentences.
2	Fair - some association but evidence does not match claim.
3	Good - associated sentences and weak claim/evidence.
4	Very good - looks likely that evidence matches claim.
5	Excellent - evidence clearly matches claim.



- Some high scoring pairs:

Claim	Evidence
In the 2012 election, Johnson received almost 1% of the popular vote, which was more than the other third party candidates combined.	The Johnson/Gray ticket received 0.99% of the popular to 1.27 million votes, more than all other minor candidates combined.
He supports energy exploration, gun rights, gay rights, and a free market.	He is fiscally conservative and socially libertarian, believing in energy exploration, gun rights, gay rights and a free market.
When I last voted, I felt that I was voting for the lesser of two evils.	How many people felt they were voting for the lesser of two evils?



- Some problematic cases:
 - A: maximized window
 - B: wrong active window
 - C: multiple TextEdit
 - D: highlights and spell check make OCR harder

Future Work and Potential Mission Impact

- Investigate **argumentation structure approaches** for better claim/evidence detection.
- Investigate more **sentence similarity approaches** - possibly *doc2vec* and *fasttext*.
- Extract more context for evidence (maybe surrounding sentences).
- Attempt multi-source aggregation of evidence -> closer to understanding interpretations.
- Investigate **Convolutional Neural Nets** for object detection and semantic analysis.
- Key long-term research question: can we build prototypes for analysts that perform automated **provenance** gathering and maybe automated **fact-checking / cross-reference / recommendation** capabilities?