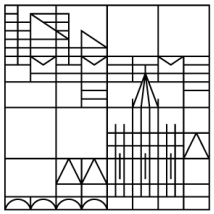


Visualizing Segment Matches between News Articles

Philipp Meschenmoser, MA

ISG Seminar, 01/02/18

Motivation

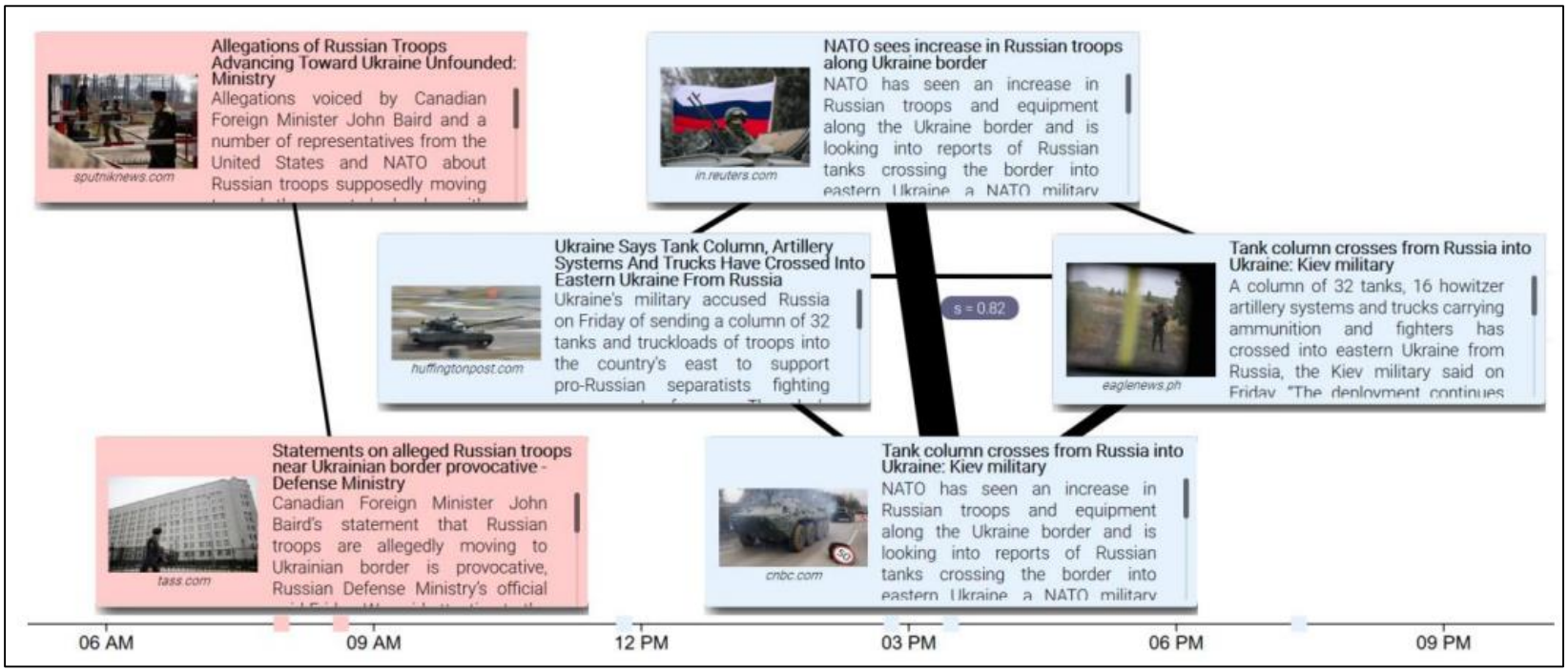
- Channel Theory and Gatekeeping Bias [1, 2, 3]: **omitting/selecting** information

  -> sociopolitical effects? [4, 5, 6]

- Many more kinds of biases [7]

- Biases at large news aggregators? novel requirements [8,9,10]
- Understand information flow in news cycle

Motivation (cont‘d)



- Temporal flow and overall semantic similarity
- **Detailed, segment-based similarity?**

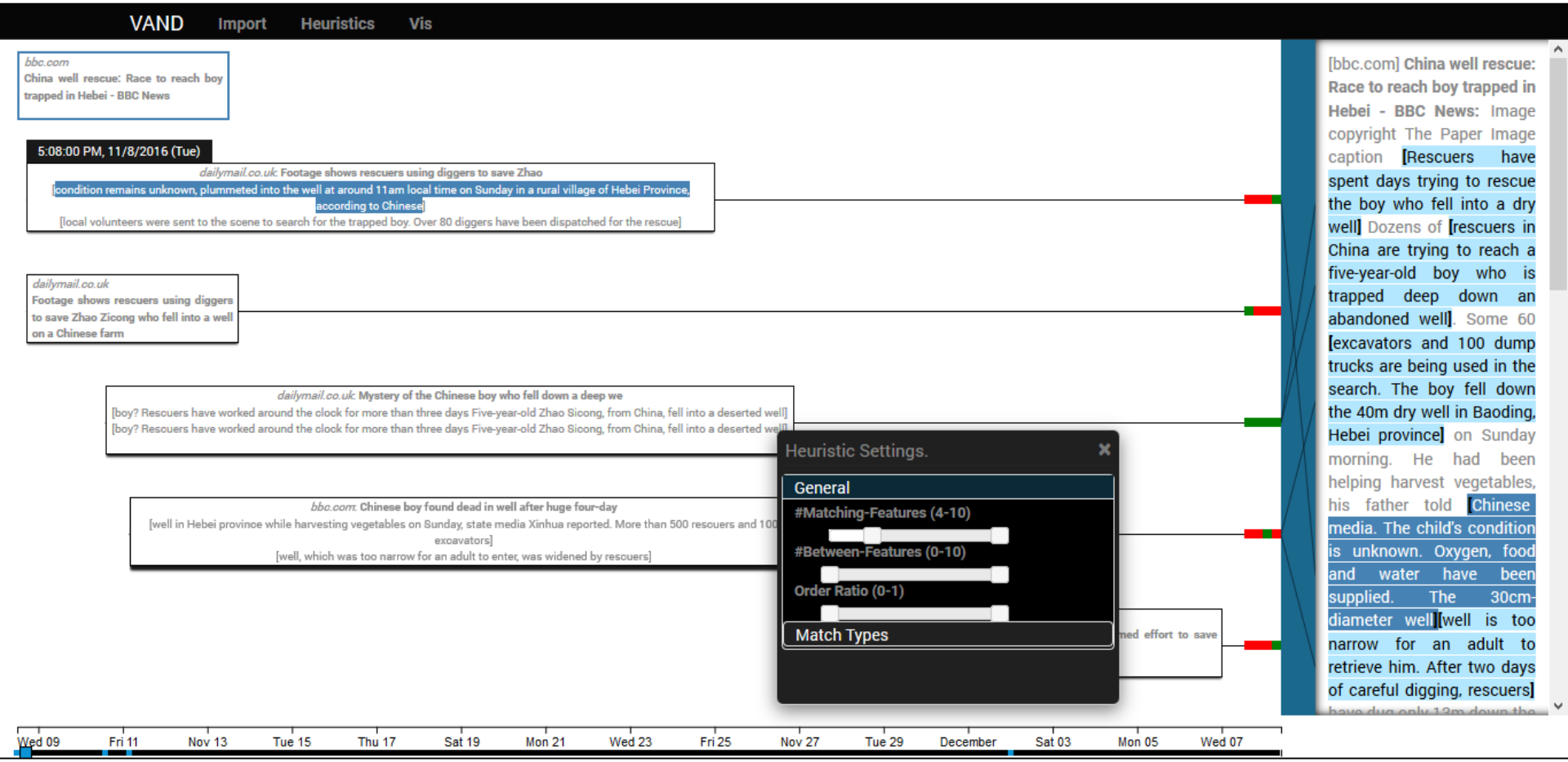
Research Objective

Develop a system to visualize **information re-use** within a **segment-based 1:n article comparison**, having a **research-oriented focus** for parameter evaluation.

Agenda

- The Application and its Pipeline
- Live Demo
- Architecture
- Challenges & Next Steps

The Application

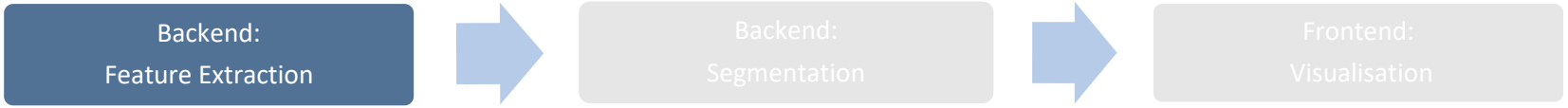


The Application

- *d3.js*-based web app with *node.js express* backend

- effective multi-comparison and **match overview**
- in-depth analysis at **1:1 article view**
- **temporal alignment** + overall article similarity

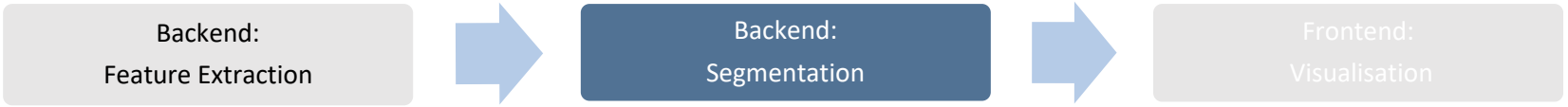
- Realisation by VA paradigms:



1) Feature Extraction

- Tokenization (Penntree Bank 3)
- POS Tagging
- NER
- Stemming or Lemmatization

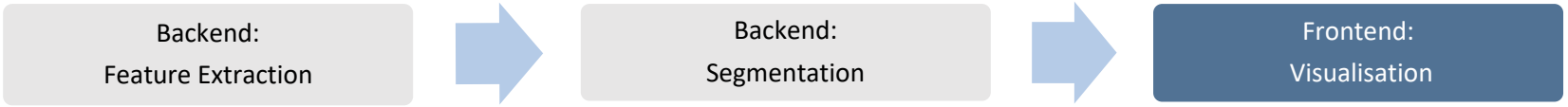
nouns, verbs, adjectives, adverbs,
people, places, organisations ,….



2) Segmentation

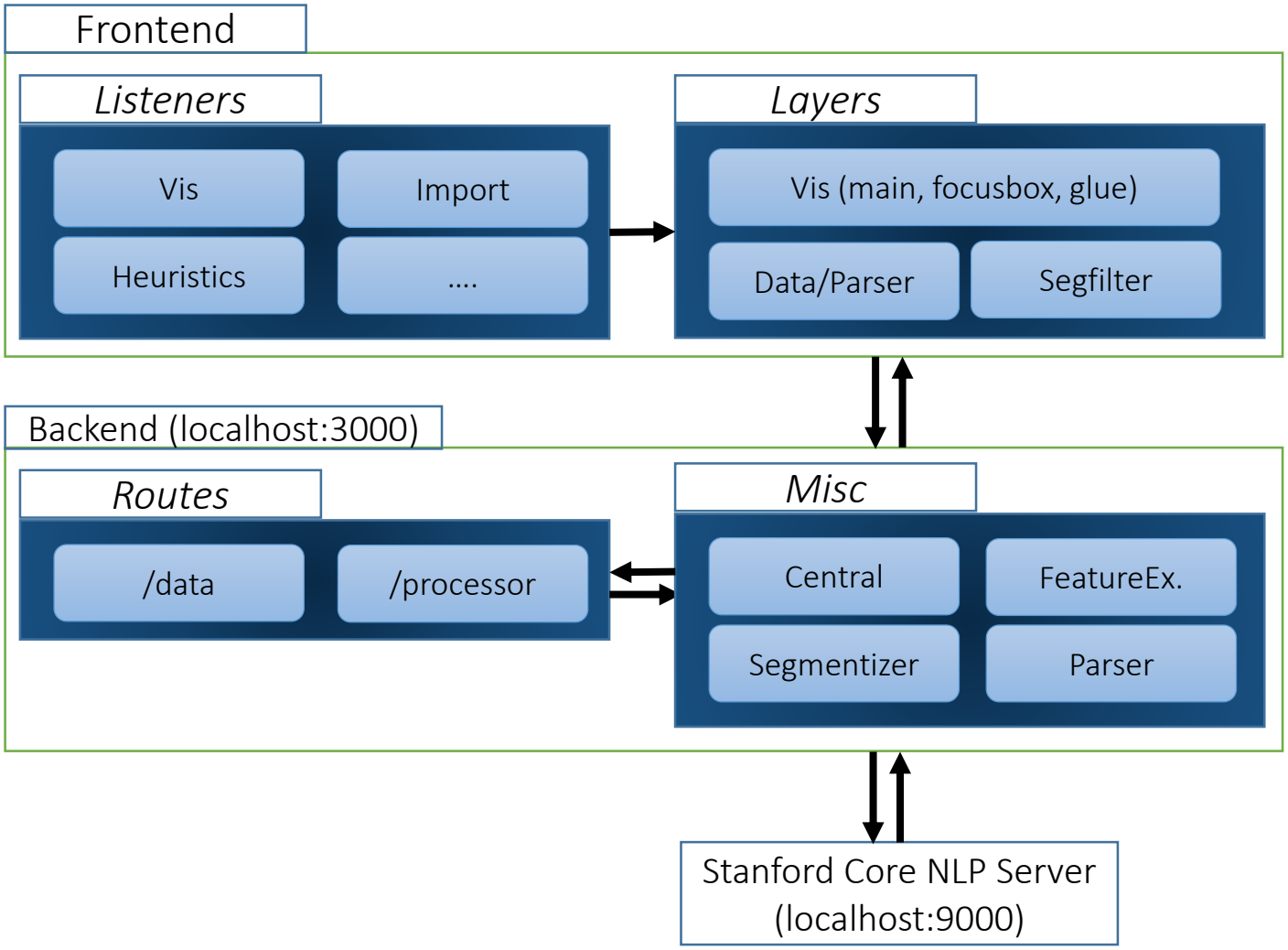
- pack features of **main** and **reference documents** into **n-grams**
- **match candidate**: if number of **intersecting features** above threshold
- match is **visualized**: if match candidate pass through frontend filter

*{**segid**:1, **type**: ‚pos‘,*
***maindoc**:0, **refdoc**:5,*
***mainLeftOffset**: 10, **mainRightOffset**: 112,*
***refRightOffset**:40, **refRightOffset**:86*
***mainIntersections**:[…], **refIntersections**:[…]}*



It‘s demo time!

Architecture



Challenges

- Performance

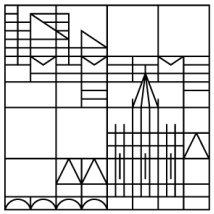
  -> efficient feature extraction

  -> matching algorithm runtime

- Node arrangement
- Browser-based bugs: e.g. scrolling in *SVG:ForeignObject*

Next Steps

- Work on Small Multiples
- Fine-Tuning of visualization
- Further feature types and matching parameters



Thank you!

References

1. Lewin, Kurt. "Constructs in psychology and psychological ecology." *University of Iowa Studies in Child Welfare* 20 (1944): 1-29.
2. White, David Manning. "The “gate keeper”: A case study in the selection of news." *Journalism Bulletin* 27.4 (1950): 383-390.
3. Shoemaker, Pamela J., and Timothy Vos. *Gatekeeping theory*. Routledge, 2009.
4. D'Alessio, Dave, and Mike Allen. "Media bias in presidential elections: a meta-analysis." *Journal of communication* 50.4 (2000): 133-156.
5. DellaVigna, Stefano, and Ethan Kaplan. "The Fox News effect: Media bias and voting." *The Quarterly Journal of Economics* 122.3 (2007): 1187-1234.
6. Eberl, Jakob-Moritz; Wagner, Markus; Boomgaarden, Hajo G. (2016-10-23). "Are Perceptions of Candidate Traits Shaped by the Media? The Effects of Three Types of Media Bias". The International Journal of Press/Politics. 22 (1): 111–132.
7. Eberl, Jakob-Moritz, Hajo G. Boomgaarden, and Markus Wagner. "One bias fits all? Three types of media bias and their effects on party preferences." *Communication Research* (2015): 0093650215614364.
8. Ulken, A. "Question of Balance: Are Google News search results politically biased." *unpublished thesis, USC Annenberg School for Communication* (2005).
9. Diaz, Alejandro. "Through the Google goggles: Sociopolitical bias in search engine design." *Web search* (2008): 11-34.
10. Bui, CamLy. "How Online Gatekeepers Guard Our View--News Portals' Inclusion and Ranking of Media and Events." *Global Media Journal: American Edition* 9.16 (2010).