LING 573 Project

Charlie Guo   qcg@uw.edu
Emma Bateman   ebateman@uw.edu
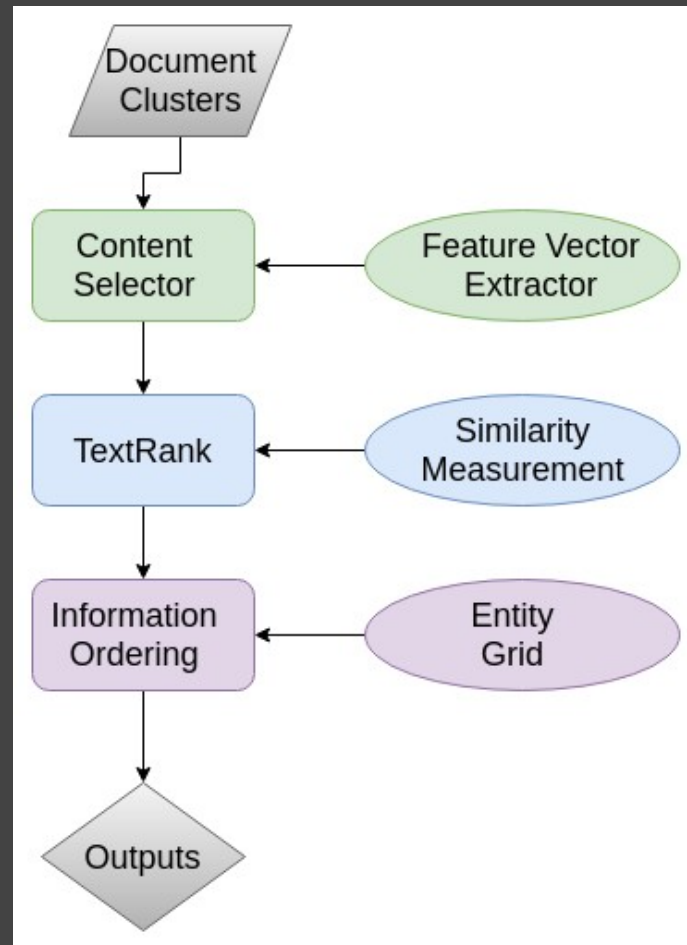John Dodson   jrdodson@uw.edu

Overview

- Graph-based sentence ranking with redundancy fix
- Entity grid information ordering

System Architecture

Approach

Content Selection

- Sentence salience estimated with graph-based algorithm
- Sentences converted to unigram feature vectors
- Ranked from highest to lowest relevance
- **New step: selected sentences subtracted from remaining feature vector**
- Sentences selected until word count is met

Information Ordering: Entity Grid

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	s	o	s	x	o	–	–	–	–	–	–	–	–	–	–	1
2	–	–	o	–	–	x	s	o	–	–	–	–	–	–	–	2
3	–	–	s	o	–	–	–	–	s	o	o	–	–	–	–	3
4	–	–	s	–	–	–	–	–	–	–	s	–	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	s	o	–	–	–	5
6	–	x	s	–	–	–	–	–	–	–	–	–	–	o	o	6

- Based on Barzilay and Lapata (2008)
- Find entities mentioned in summary
- For each entity, sentence pair determine grammatical role if any
- Assign weights: 1.0 for subject, 0.5 for object, 0.1 for none
- Calculate weight for each entity over all candidate sentences
- Pick highest weighted entity and select sentence with the highest weight for that entity
- Repeat until all sentences are sorted

ROUGE-2 scores

	Recall	Precision	F1
Baseline	0.05605	0.07022	0.06184
Entity Grid	0.05595	0.07012	0.06174

Issues and Future Work

- Ordering algorithm did not improve on baseline!

- Preprocessing needed: select based on linguistic data

- Postprocessing needed: resolve anaphoras and remove extraneous information

Thank you.