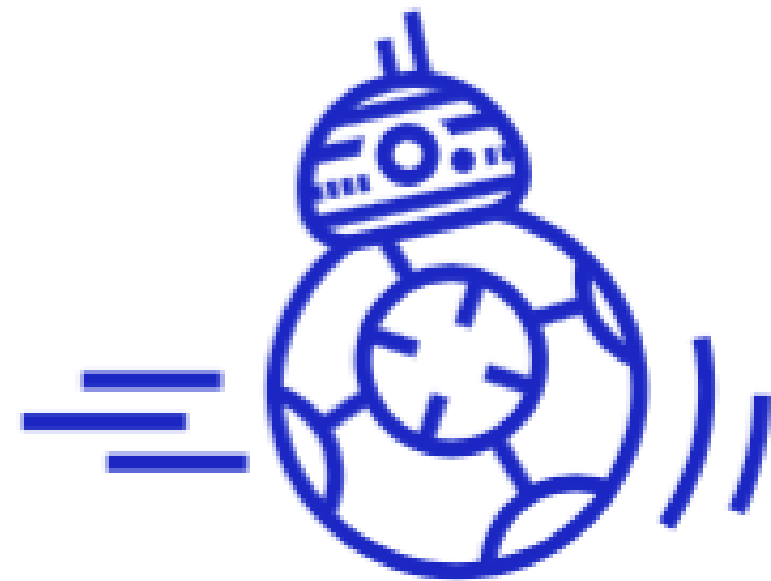


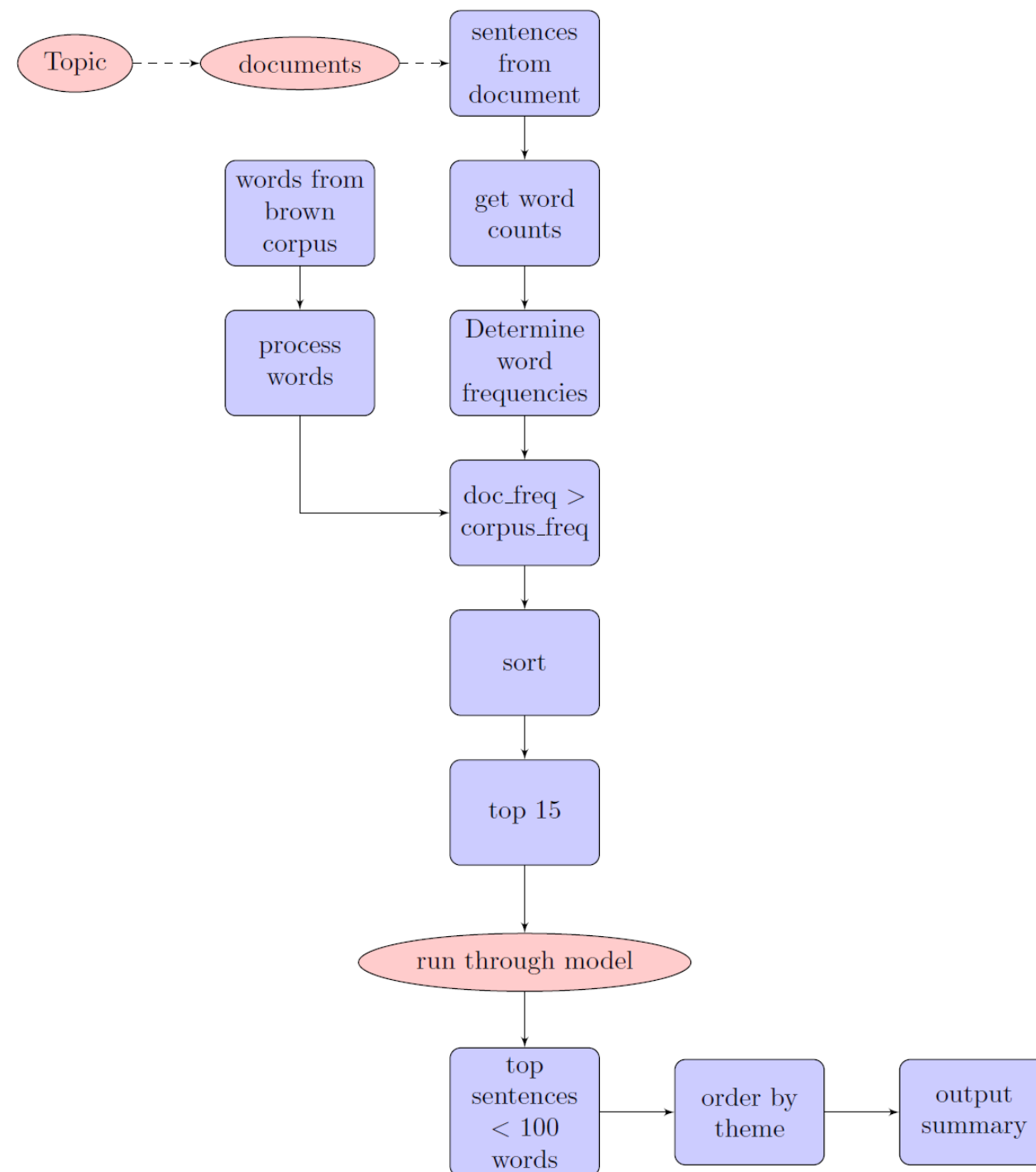
# Deliverables 4

Matt Calderwood

Kirk LaBuda

Nick Monaco





## Overall System Architecture (no changes from D3)



# System Changes

- Overall Refinements - several small bug fixes (*no empty summaries, regex fixes for preprocessing and content selection, etc.* )
- Content Selection - for summary vectors - normalized tf\*idf calculation, normalized sentence position in article. Settled on *RBF kernel*.



# System Changes (cont.)

- Info Ordering - adopted hybridized theme-modeling and cosine readability approach.
- Shallow approach for theme modeling (D3), SciPy for cosine distance.
- Content Realization – Machine learning approach using compression corpus and classification



# Content Realization

- Machine learning
- Compression corpus (Clark & Lapata, 2008)
- Tree based compression
- Classification (keep, partial, omit)
- Trainer: MaxEnt
- Tools: Stanford CoreNLP, NLTK, MALLET



# Features

- Word (leaf node only)
- POS
- Parent/Grandparent POS
- Left/right sibling POS
- First/last two leaves
- Is left-most child of parent
- Is second left-most child of parent
- Contains negation



# Example

## Without:

A hurricane watch on the mainland was extended from the Miami area northward all the way to near Brunswick, Ga.

``We'll order heavy on those items tomorrow, because the next truck won't come until Tuesday and if it's coming it'll be in full swing by then.

As night fell on South Florida, shelters and hotel rooms inland, especially around Palm Beach, began to fill; cruise ships left for safer waters to the south; long flotillas of pleasure craft snaked along canals looking for safe harbor, as lines grew at hardware and grocery stores.

## With:

Many Floridians took advantage of the weekend's final day to take careful inventory of their hurricane supplies.

A hurricane watch on the mainland was extended from northward to near Brunswick Ga.

``We'll order heavy on those items tomorrow because the next truck if it's coming it'll be in full swing by.

More than 200,000 people on Florida's east-central coast were told to evacuate and another 200,000 were evacuated from coastal areas of Miami-Dade County.



# Examples

- Prosecutors meanwhile raided the house of a former head of the spy agency, **the National Intelligence Service (NIS)**, late Thursday and seized documents and computer discs believed to be related to the unlawful bugging.
- Sipadan is a world-renowned diving island **off the northeast coast of Sabah, the Malaysian side of Borneo Island, which is shared with Indonesia.**
- Aruban authorities have **defended their investigation, saying police work takes time.** Joran lived in an apartment attached to **the main house.**





# Possible Improvements

- Use language model to improve grammaticality and coherence
- Test combinations of node removals
- Better/additional features
- Context based features
- Rule based overrides



# Successes

- Machine Learning/SVR- helped us consistently improve ROUGE scores in D3 and D4. RBF kernel was best.
- Info Ordering - shallow approach seems reasonable, has yielded some good results.



## Successes (*cont.*)

- Content Realization – Output is usually coherent. Summaries include more sentences since most are shorter.
- Overall System - vastly improved from inchoate D2 system. Sometimes produces decent summaries.
- Substantially fewer “no idea” summaries.



# Issues



- Content Realization vs. Content Selection - slight conflict between these stages - didn't train on post-content realization summaries
- Info Ordering - more refinements could be made to fine-tune readability, (this stage is also dependent on content selection)
- Content Realization – Removal of important information, ungrammatical summaries
- Runtime- 30 min. runtime



# Issues



- Intermodule Interaction - Still improvements that could be made to make smoother cooperation between
- (a) content realization and content selection
- (b) content selection and info ordering



# Qualitative results (*dev set*)

Good: *could use reordering.*

D1046-A.M.100.H.1:

The ambassador said his government would conduct a survey this week in Aceh province in order to find out what is the most needed for survivors in the tsunami-hit areas. The killer tsunami was touched off by an 8.7-magnitude earthquake has climbed to over 145,000 with over 90,000 deaths in Indonesia. announced his country will host a January 6 summit of world leaders to discuss the aftermath of Asia's disaster. A Chinese quake rescue team arrived in tsunami-ravaged Indonesia to help emergency aid efforts.

Bad:

D1001-A.M.100.A.1:

I went three hours. To accommodate the school day some of Columbine's 158 teachers met for the first time since two student gunmen shot 12 fellow students and a teacher. In the crowd were people who had driven of to pay. ``Love is stronger than death.



# Qualitative results (*eval set*)

Good:

D1106-A.M.100.A.1:

Concern has been growing about overfishing of tuna and other fish species. The tuna talks will for the first time bring together the five regional fishery management organizations, including the oversight bodies for tuna fishing in the Atlantic, Pacific and Indian Oceans. The agency said participants are focusing on of. Japan eats a quarter of the world's tuna faces pressure from environmentalists, who have blamed the global fad for Japanese food for bringing tuna numbers to the brink of eventual extinction.

Bad:

D1107-A.M.100.B.1:

who were doing field work in the firm developed cramps vomiting diarrhea and were taken to afternoon and diagnosed with suspected food poisoning. The plan was worked out by the State Food and Drug Administration authorities for environmental protection technology commerce agriculture health. Two months after four people sued the city for delays in warning people not to eat snails infested with parasites, Beijing has come up with a series of alerts and actions it will take in the event of an outbreak of illness caused by contaminated food.



# ROUGE results (*dev set*)

## Devset Results:

	<u>D2</u>	<u>D3</u>	<u>D4</u>
<i>ROUGE-1 Average_R</i>	0.10987	0.12951	<b>0.15066</b>
<i>ROUGE-2 Average_R</i>	0.01891	0.02431	<b>0.02844</b>
<i>ROUGE-3 Average_R</i>	0.00502	0.00612	<b>0.00717</b>
<i>ROUGE-4 Average_R</i>	0.00129	0.00198	<b>0.00265</b>





# ROUGE results (*eval set*)

## Evalset Results:

	<u>D4 Eval Set</u>	<u>D4 Dev Set</u>
<i>ROUGE-1 Average_R</i>	<b>0.16779</b>	0.15066
<i>ROUGE-2 Average_R</i>	<b>0.02968</b>	0.02844
<i>ROUGE-3 Average_R</i>	<b>0.00705</b>	0.00717
<i>ROUGE-4 Average_R</i>	<b>0.00212</b>	0.00265



# Works Cited

- Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Li, Chen, Yang Liu, Fei Liu, Lin Zhao, and Fuliang Weng. "Improving Multi-documents Summarization by Sentence Compression Based on Expanded Constituent Parse Trees." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014).
- Smola, Alex J., and Bernhard Schölkopf. A Tutorial on Support Vector Regression \* (n.d.): n. pag. [Http://alex.smola.org/papers/2003/SmoSch03b.pdf](http://alex.smola.org/papers/2003/SmoSch03b.pdf). 30 Sept. 2003. Web. 16 May 2016.
- Yu, Pao-Shan, Shien-Tsung Chen, and I-Fan Chang. "Support Vector Regression for Real-time Flood Stage Forecasting." Journal of Hydrology, 328 (3–4), Pp. 704–716, Sept. 2006. Web. 16 May 2016.



A close-up shot of a man with curly brown hair and a distressed expression. He has a bloody wound on his left cheek and a bloody mark on his chin. The background is a bright, fiery orange and yellow, suggesting a battle scene or a fire. The lighting is dramatic, highlighting his features and the blood.

# D4: It's Done

Laurie Dermer – Stephanie Peterson – Katherine Topping

# System Changes



# Preprocessing changes

- Stripped any remaining metadata and formatting
- Had to account for the different structure of the evaltest data set
- Thankfully didn't have to think about xml :):):)

# Selection Changes

- $tf \cdot idf$  was refined.
  - Used the Reuters IDF. The Reuters corpus is a news corpus provided by NLTK.
  - Accounted for rare words – treated anything not seen in the idf corpus as a singleton.
  - These changes brought ROUGE-2 up to .042 (from .018!)
- Added LexRank for sentence selection
  - ROUGE-2 went up to .058 when LexRank is used, compared to  $tf \cdot idf$
  - LexRank uses the idf improvements when calculating idf modified cosine similarity.
- Used LexRank's similarity measure to check for similar sentences and skip them (.9 or more idf-modified cosine similarity) - since the function was already there and gave scores from 0-1

## Selection Changes cont.

- Implemented LLR with backing Reuters corpus
  - ROUGE-2 Average\_R: 0.03395 (95%-conf.int. 0.02535 - 0.04323)
  - ROUGE-2 Average\_P: 0.03564 (95%-conf.int. 0.02666 - 0.04541)
  - ROUGE-2 Average\_F: 0.03469 (95%-conf.int. 0.02591 - 0.04405)
- Better than initial scores using this selection method, but lexrank outranked
- LexRank was based on idf-based cosine similarity, so we did not pursue this avenue

# Ordering Changes

- Changed from frequency-based theme selection to cosine similarity theme selection
- Sentences are grouped together with the sentence(s) they have the highest cosine similarity with
- Redundancy is already controlled in selection, not a big worry here
- The old frequency-based ordering performs better with tf\*idf selection, BUT this new method performs better with LexRank selection
  - As far as ROUGE is concerned
- Otherwise, ordering structure works the same (themes ordered by "popularity", sentences ordered chronologically)



# Ordering Changes cont.

- Our previous experiment tried incorporating headlines into the ordering process, which proved faulty
- As a final experiment, we expanded common headline terms into their synonym sets using WordNet, and tried query based ordering techniques with these headline synsets as the query
- ROUGE scores were not improved so this technique was abandoned

# Realization Changes

- Focused on shallow realization
  - (attempted some deeper realization, to be discussed shortly)
- Loosely based on CLASSY realization
- Catches and removes surface-level entities
  - Ages
  - Sentence-initial prepositional/adverbial/conjunctive phrases
  - Parentheticals
  - Clauses separated by hyphens
  - Quotations

# NP-complete: The NP Saga



# The NP Saga

- We tried really hard to implement NP handling
  - Really, really hard
- But ran into some roadblocks
  - Wanted to use parser with NLTK, but struggled with internal NLTK parsers/finding an appropriate grammar to use
  - Then attempted slightly-shallower approaches using POS tags (and even, in a moment of desperation, RegExes on the hunt for capital letters)
  - Ultimately, a failure – approaches interfered with non-NPs
  - Followed by weeping and gnashing of teeth
- Probably tried adding this in a bit too late, probably could have made some of the online resources work with a bit more time dedicated to the problem

# DevTest ROUGE Scores

- D3

- ROUGE 1 - 0.11429
- ROUGE 2 - 0.01891
- ROUGE 3 - 0.00410
- ROUGE 4 - 0.00077

- D4

- ROUGE 1 - 0.24808
- ROUGE 2 - 0.06112
- ROUGE 3 - 0.01531
- ROUGE 4 - 0.00472



# DevTest vs EvalTest ROUGE Scores

- DevTest

- ROUGE 1 - 0.24808
- ROUGE 2 - 0.06112
- ROUGE 3 - 0.01531
- ROUGE 4 - 0.00472

- EvalTest

- ROUGE 1 - 0.27375
- ROUGE 2 - 0.07550
- ROUGE 3 - 0.02503
- ROUGE 4 - 0.01208

# Error Analysis



# Columbine Summary (D3)

- In an age when so many Americans regularly lament the breakdown of community, the many communities that the Columbine massacre has produced are proving that the notion, at least in time of crisis, still thrives.

Jefferson County has 500,000 residents, but today our community is much larger," county commissioner Patricia Holloway said Sunday at a shopping-center parking lot service attended by 70,000 people -- a hastily stitched-together community unto itself.

There are myriad mini-communities created by the bloodshed: Denver-area students, their rivalries suddenly rendered irrelevant; emergency personnel, united in their harrowing experiences; towns like Jonesboro and Paducah and Springfield and Edinboro, who understand Columbine's anguish but never asked to be members of this kind of community.



# Columbine Summary (Final Version)

- To accommodate the 1,965 Columbine students, the school day is being split with Chatfield students beginning early in the day and Columbine students showing up shortly before 1 p.m.

Jefferson County school officials said Columbine's 1,800 students would return to classes Thursday a few miles south at Chatfield High School, a school originally built to accommodate Columbine's overflow.

Columbine students returned to classes at Chatfield High School on Monday. But not all Columbine students were as welcome as others.

Students at Chatfield, which has a sports rivalry with Columbine, went out of their way to welcome the 1,900 Columbine students.

Chatfield students will attend classes in the morning and Columbine students in the afternoon.

# Panda Summary (D3)



- China will soon finish building its first blood bank for pandas, which will assist researchers in studying the endangered animals' blood types and chances of accepting blood transfusions, state media said Friday.

Located in the giant panda breeding lab, the bank will help researchers answer questions such as how many blood types pandas have and whether they reject blood transfusions, centre sources said.

b'<TEXT> Taipei City Government will form a task force soon to facilitate its bid to host the two giant pandas that China has offered as gifts to the Taiwan people, Mayor Ma Ying-jeou said Wednesday.

On the decision of Shoushan Zoo in the southern port city of Kaohsiung to compete for the right to house the Chinese pandas, Ma said that Kaohsiung is welcome to enter the competition, although he added that he believes Taipei City is more capable of taking care of the pandas.

# Panda Summary (Final version)

- The two giant pandas at the city's zoo retired to their favorite spot under a few bushes and mated over the past two days — the only successful natural insemination of a panda this year in the United States, officials said.

The Qinling panda has been identified as a sub-species of the giant panda that mainly resides in southwestern Sichuan province.

A total of 273 wild giant pandas have been spotted in an area of 347,864 hectares, which officials say means there are 7.8 pandas on per 100 sq km, the highest density among all pandas' habitats in China.

The Qinling pandas are believed to have separated from the giant panda about 50,000 years ago, Chinese researchers said.

# Error Analysis Goes Here

- Our ROUGE scores went way up! Overall, more information that is more relevant gets put into the summaries with LexRank.
- However, the summaries read less well.
- This may be due to the lesser importance of keywords, and the (based on looking at the summaries) relatively lower likelihood of picking a long sentence with LexRank compared to with the tf\*idf system we were using.
- The shorter sentences really might have an impact on score, because more topics get covered in the summary.



some lower-scoring  
summaries  
for comparison....

# D4 tf\*idf & best scoring order

Jefferson County school officials said Columbine's 1,800 students would return to classes Thursday a few miles south at Chatfield High School, a school originally built to accommodate Columbine's overflow.

Welcoming signs and banners decorated Chatfield High School today as students arrived from former rival Columbine, who hadn't been to class since the devastating shooting attack nearly two weeks ago.

Chatfield, about 3 miles from Columbine, was decorated to welcome the Columbine students, with unity signs incorporating the Chatfield Chargers' signature burgundy with the Columbine Rebels' navy blue.

Monday, Columbine and Chatfield senior high schools came together yet again, neither as rivals nor as mourners, but as partners in helping Columbine students to reclaim their lives as normal teenagers for whom school is a place to learn, not to flee.

## D4 tf\*idf & best scoring order

Ma said that Kaohsiung is welcome to enter the competition, although he added that he believes Taipei City is more capable of taking care of the pandas.

Stressing that the panda is an animal protected by the Convention on International Trade in Endangered Species and that there are only about 1,000 left in the world, Ma said that having pandas at Taipei City Zoo would not be simply for entertaining visitors, but also to show the zoo's capabilities in conserving, nurturing and studying special wild animals.

The two pandas, very likely to be provided by the Wolong Panda Conservation District of China's western province of Sichuan, are expected to attract more than 1 million extra visitors to Taipei City Zoo each year if indeed they are allowed to be brought to Taipei, Ma said.

# Further Error Analysis

- Not sure whether our impressions would actually be borne out on a larger scale.
- There's an element of coincidence here, where you might get a relevant and important background tidbit or name that makes something else clear.
- There's also an element of luck involved when  $n=2$ .
- The types of terms  $tf*idf$  favors (names, places) also might lead to more relevant-seeming summaries and sentences.



# Abstractive Work (future work... grumble)

- Option 1) Create java module using Stanford Dependency Parser
- Option 2) Utilize <https://github.com/dasmith/stanford-corenlp-python>
- Generate MRS representations from modified DELPH-IN English resource grammar
- Follow [Liu et al, 2015](#) semantic graph reduction strategy (taking into consideration is was following MRS not AMR)
- Compare this technique to our selection and realization strategies
- Definitely tried adding this way too late. So now it goes in our future work section. Grumble. Gnashing teeth. Grumble.



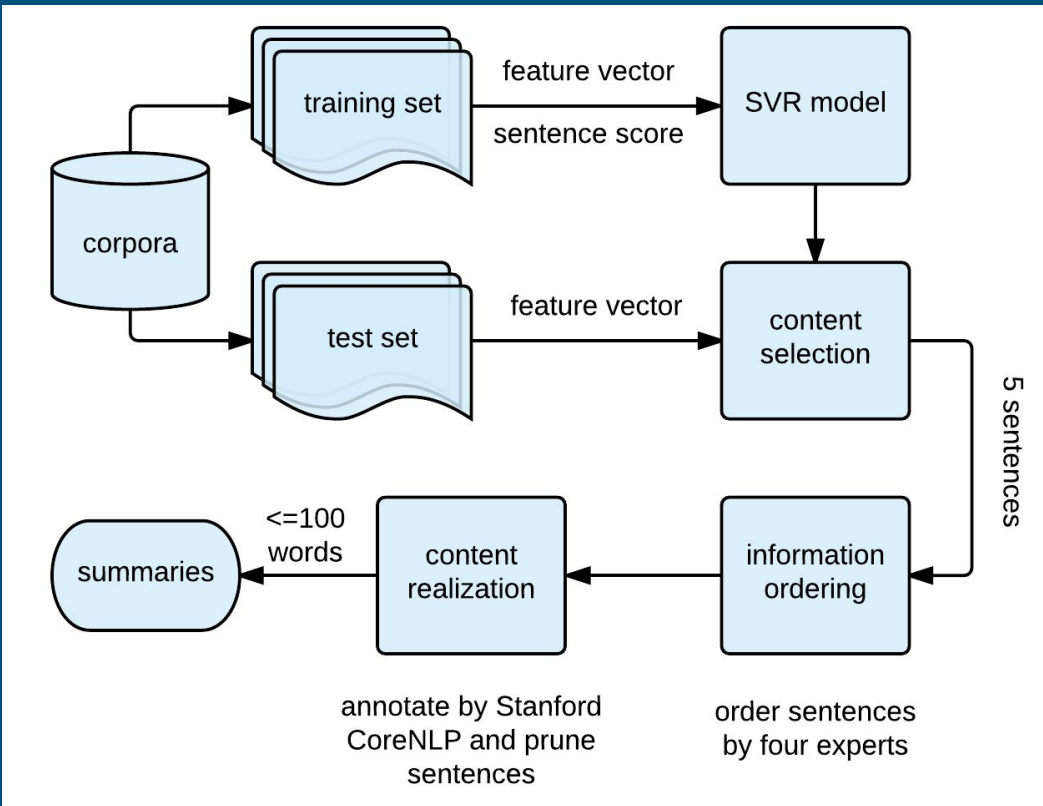
# Ling573 Project D4 System



Xiaosu Xue   Yveline Van Anh   Alex Cabral



# System Architecture



# Preprocessing (experiment)

---

- **Regular expressions to prune down sentences (based on CLASSY):**
  - Extra words (e.g. (AP) --, urls, numbers in parentheses)
  - Relative temporal phrases (e.g. before, after)
  - Words such as 'however', 'also' in the middle of a sentence
  - Ages (e.g. age 50)
  - Gerund phrases
  - Relative clause attributives (e.g. whom, which)
  - Attributions without quotes (e.g. police said)

# Preprocessing (experiment)

---

- Did not improve ROUGE scores

	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4
D3 + fill 100 + pruning	0.25209	0.07122	0.02601	0.01089
D3 + fill 100	<b>0.26483</b>	<b>0.07532</b>	<b>0.02860</b>	<b>0.01336</b>

# Content Selection

---

- **Based on the SIEL algorithm:** iiit hyderabad at tac 2009
- **Approach:** extract sentences with the highest predicted scores given by the SVR model (RBF kernel)
- **Features:**
  - sentence position:  $1 - n/1000$  if  $n \leq 3$ ;  $n/1000$  otherwise
  - query score
  - document frequency score
  - Kullback–Leibler divergence

# Content Selection - smoothed LMs for KLD

- **Bayesian smoothing using Dirichlet priors:**

$$D_{\text{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

$$p_\mu(w | d) = \frac{c(w; d) + \mu p(w | \mathcal{C})}{\sum_w c(w; d) + \mu}$$

- **Collection unigram model:**
  - used AQUAINT corpora, computed by maximum likelihood estimation
- **Parameter  $\mu$ :**
  - set to 2000.0 (Zhai & Lafferty, 2001)

# Content Selection - centroid score as a feature

---

- **Centroid Score:**
  - ten centroid words in a cluster -- each has a centroid value
  - centroid score scaled by the sentence length
- **Comparison with Document Frequency Score:**
  - computing time
  - words covered



# Content Selection - fill the 100-word limit

## Diallo Trial

D3  
+fill

1/ Diallo's father, Saikou Amad Diallo, arrived here Wednesday from the West African nation of Guinea and said he was anxious to see the officers not only charged, but brought to trial.

2/ Through their lawyers, the officers have said they thought Diallo had a gun.

3/ ...

4/ The four New York City police officers charged with murdering Amadou Diallo returned to work with pay Friday after attending a morning court session in the Bronx in which a Jan. 3 trial date was set.

5/ ``We grieve for Amadou Diallo and the four officers involved and pray they get a fair trial," Safir said.

D3

1/ Diallo's father, Saikou Amad Diallo, arrived here Wednesday from the West African nation of Guinea and said he was anxious to see the officers not only charged, but brought to trial.

2/ The four New York City police officers charged with murdering Amadou Diallo returned to work with pay Friday after attending a morning court session in the Bronx in which a Jan. 3 trial date was set.

# Information Ordering

---

- **Same as D3:**
  - Sentences ordered by experts in Bollegala et al. minus probabilistic expert
- **Experimenting with different weights:**
  - Chronology:  $0.33 \rightarrow 0.2$
  - Topicality:  $0.03 \rightarrow 0.2$
  - Precedence:  $0.2 \rightarrow 0.3$
  - Succession:  $0.44 \rightarrow 0.3$
- Most orderings did not change
- Ultimately kept weights the same for D3

- **Original weights:**

The bank at southwest China's Giant Panda Protection and Research Centre in the Wolong Nature Reserve in Sichuan province will be completed this year, the China Daily said.

Located in the giant panda breeding lab, the bank will help researchers answer questions such as how many blood types pandas have and whether they reject blood transfusions, centre sources said.

The giant panda is one of the world's most endangered species, with an estimated 1,000 living in the mountainous regions of Sichuan, Shaanxi and Gansu provinces.

On Dec. 14 last year, Feng Shiliang, a farmer from Youfangzui Village, told the Fengxian County Wildlife Management Station that he had spotted an animal that looked very much like a giant panda and had seen giant panda dung while collecting bamboo leaves on a local mountain.

Twenty-two giant pandas living in parts of Baishuijiang State Nature Reserve in the northwestern province of Gansu will be moved to other locations with better food, the China Daily said, quoting Zhang Kerong, director of the reserve.

- **New weights:**

The bank at southwest China's Giant Panda Protection and Research Centre in the Wolong Nature Reserve in Sichuan province will be completed this year, the China Daily said.

The giant panda is one of the world's most endangered species, with an estimated 1,000 living in the mountainous regions of Sichuan, Shaanxi and Gansu provinces.

On Dec. 14 last year, Feng Shiliang, a farmer from Youfangzui Village, told the Fengxian County Wildlife Management Station that he had spotted an animal that looked very much like a giant panda and had seen giant panda dung while collecting bamboo leaves on a local mountain.

Located in the giant panda breeding lab, the bank will help researchers answer questions such as how many blood types pandas have and whether they reject blood transfusions, centre sources said.

Twenty-two giant pandas living in parts of Baishuijiang State Nature Reserve in the northwestern province of Gansu will be moved to other locations with better food, the China Daily said, quoting Zhang Kerong, director of the reserve.

# Content Realization

---

Text based pruning

Deep processing based on algorithm described in Zajic et. al

- Select Root S node
- Remove preposed adjuncts
- Remove complementizer that
- Apply the XP over XP rule for NPs
- Remove temporal expressions

# Removing Conjunctions

---

Two years ago, in response to reports of adverse reactions, the Food and Drug Administration proposed restricting...

Two years ago, in response to reports of adverse reactions, the Food proposed restricting....

# Removing Conjunctions Cont'd

---

Tensions in the region escalated Thursday when tribesmen ambushed a convoy of troops and fired rockets at a military base on the outskirts of Dera Bugti, 300 kilometers southeast of Baluchistan province's capital, Quetta.

Tensions in the region escalated Thursday when tribesmen ambushed a convoy of troops.

# Removing Modal Verbs

---

The health department said that residents should avoid contacting birds in case they might be infected with the deadly virus of bird flu known as H5N1.

The health department said that residents should avoid contacting birds in case they infected with the deadly virus of bird flu known as H5N1.

Maryland may be rushing crucial research on introducing Asian oysters into the Chesapeake Bay, the states of Delaware and New Jersey said this week...

Maryland rushing crucial research on introducing Asian oysters into the Chesapeake Bay, the states of Delaware and New Jersey said this week...

# Removing PPs without Named Entities

---

The Papua New Guinea (PNG) Defense Force, the police and health services are on standby to help the victims of a tsunami that wiped out several villages, killing scores of people, on PNG's remote north-west coast Friday night.

The Papua New Guinea (PNG) Defense Force, the police and health services are to help the victims that wiped out several villages, killing scores, on PNG's remote north-west coast Friday night.



# Removing PPs from SBARs

---

In May, Metabolife International Inc., sued WCVB-TV Channel 5, its anchor Susan Wornick, and Dr. George L. Blackburn of Harvard Medical School after the Boston station reported on health risks related to the use of ephedrine-containing weight-loss dietary supplements...

In May, Metabolife International Inc., sued WCVB-TV Channel 5 after the Boston station reported.

# Good Example

---

On Sunday, about 3,000 people, mostly women and children of local Bugti tribesmen, left Dera Bugti, a day after 3,000 government employees and their families, fearing more fighting in the town, located some 300 kilometers (185 miles) southeast of Quetta, Baluchistan's provincial capital.

On Sunday, about 3,000 people left Dera Bugti, fearing more fighting in the town, located some 300 kilometers southeast of Quetta, Baluchistan's provincial capital.

# Another Good Example

---

The shooting at Jokela High School in Tuusula, some 50 kilometers (30 miles) north of the capital, Helsinki, shocked the Nordic nation because gun violence is rare.

The shooting shocked the Nordic nation because gun violence is rare.

Highest overall ROUGE-1 score, 0.41850

# Not So Good Example

---

“Personal information on IRS computers is at risk to unauthorized disclosure, destruction or modification, and most alarmingly, to identity theft,” Thompson said Tuesday.

Personal information on IRS computers is at risk to unauthorized disclosure to identity theft,” Thompson said Tuesday.

# Not So Good Example

---

Of the 5,743 known species of toads, frogs, salamanders, newts and worm like amphibians, 1,856 (32.5 percent) are under threat, according to the work by 500 researchers in 60 countries.

Of the 5,743 known species, 1,856 are under threat, according to the work by 500 researchers in 60 countries.

Lowest overall ROUGE-1 score, 0.05761

# Results

---

	Devtest		Evaltest	
	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2
MEAD (baseline)	0.22437	0.06144	0.24932	0.07134
D3	0.24145	0.07059	0.27584	0.07918
D4	0.24168	0.06870	0.27757	0.07707
D3+fill 100 (best)	<b>0.26483</b>	<b>0.07532</b>	<b>0.31096</b>	<b>0.08955</b>

# Discussion

---

- D3+fill\_100 system performs better than D4, but D4+fill\_100 could potentially perform better
- Experimented with different weights (higher emphasis on topicality, less on chronology) for information ordering, but this seemed to hurt readability
- Content realization via parse based pruning improved ROUGE-1 over D3, but didn't prove to be as beneficial for ROUGE-2

# (Hypothetical) Future Work

---

- Perform deep pruning in the preprocessing step, combine it with the fill-100-word extraction strategy
- Use KLD ratio as a feature for content selection
- Resolve coreferences
- Incorporate more regex based pruning
- Prune sentences based on dependency parses



# Reference

---

Bollegala, Danushka, Naoaki Okazaki, and Mitsuru Ishizuka. "A preference learning approach to sentence ordering for multi-document summarization." *Information Sciences* 217 (2012): 78-95.

Varma, V., Bysani, P., Kranthi Reddy, V. B., Santosh GSK, K. K., Kovelamudi, S., Kiran Kumar, N., & Maganti, N. (2009, November). iit hyderabad at tac 2009. In Proceedings of Text Analysis Conference 2009 (TAC 09).

Radev, D. R., Blair-Goldensohn, S., & Zhang, Z. (2001). Experiments in single and multi-document summarization using MEAD. Ann Arbor, 1001, 48109.

Radev, D. R., Jing, H., Styś, M., & Tam, D. (2004). Centroid-based summarization of multiple documents. *Information Processing & Management*, 40(6), 919-938.

Zhai, C., & Lafferty, J. (2001, September). A study of smoothing methods for language models applied to ad hoc information retrieval. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 334-342). ACM.

Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out: Proceedings of the ACL-04 workshop (Vol. 8).



Thank you!

---

---

# Ling 573 - Multi- document Summarization System : D4

Martin Horn, William Lane, Ryan Lish, Spencer Morris

---

---

---

# Content Realization

- Coref
  - Entity resolution
  - Sentence pruning
  - 100-word cutoff
-

---

# Coreference Resolution

- cort: **coreference resolution toolkit** (Martschat et al. 2015)
  - Uses model developed on CoNLL-2012 shared task data
  - We use end-to-end option to process coreference chains on raw text
  - Outputs sentence-split, tokenized text with XML-like markup
  - Ex: `<mention id="13" span_start="43" span_end="43" entity="12" antecedent="11">its</mention>`
-

---

# Coreference Resolution

- Coref performed in Content Selection module
    - Ran on original documents (one doc at a time)
    - “Main” (first) mention of entity used as comparison key between documents
      - Exact match of main entity → combine referent sets
    - Sentence-splitting from coref used downstream
  - If coref breaks on document: default to original sentence splitting; no coref resolution later
  - “Main” mention of entity and set of referents sent to Entity Resolution component of Content Realization
-

---

# Visualization example of co-referred mentions of entities

- 
1. Government scientists raised the alert level for Mount St. Helens after its second steam eruption in two days was followed by a powerful tremor .
2. They said the next eruption was imminent or in progress , and could threaten life and property in the remote area near the volcano .

“Government scientists”: {“Government scientists”, “They”, ...}

“Mount St. Helens”: {“Mount St. Helens”, “its”, ...}

“the remote area near the volcano”: {“the remote area near the volcano”, ...}

“the volcano”: {“the volcano”, ...}

---

---

# Entity Resolution

for each mention in each sentence:

if previous\_mention == entity\_id or entity already mentioned in sentence:

    If mention is a pronoun:

        leave as pronoun

    else :

        Use shortest form of name

else if entity already mentioned in summary:

    use shortest form of name

else:

    use first mention form of name

if mention contains another mention and current mention name not changed:

    recursively resolve nested mentions

---



---

# Pruning and cutoff

- Prune with set of regexes
  - Until under 100 words:
    - Take out lowest scoring sentence
  - Fill the gap with the highest scoring sentence that will fit
    - Unless sentence is too small
      - Only take >4 word sentences
-

---

---

# Content Selection

- Incorporate headline info into Focus LexRank
    - Topic words higher weight
    - Headline words lower weight
    - Basic implementation didn't help scores
-

---

# Results

ROUGE average recall scores for D3 and D4

	<b>D3</b>	<b>D4 Devtest</b>	<b>D4 Evaltest</b>
<b>R-1</b>	0.2211	0.2273	0.2745
<b>R-2</b>	0.0552	0.0572	0.0749
<b>R-3</b>	0.0184	0.0186	0.0267
<b>R-4</b>	0.0068	0.0066	0.0116

---

---

---

# Successes

- We have coref!
  - Entity resolution makes sentences easier to understand
  - Does decent job of maintaining possessive form
  - Summaries very close to 100 words
-

---

## D1038-A.M.100.G.1: ~~D3~~->D4

Just before noon on Friday, seismometers at the Cascades Volcano Observatory in Vancouver, Wash., wiggled in a familiar pattern.

~~They~~Government scientists said the next eruption was imminent or in progress, and could threaten life and property in ~~the remote area near the volcano~~ this area.

Blah blah blah... (the rest)

---

---

# Issues

- Because coref done to each document instead of each topic, entities often not linked across documents; long form of name repeated
  - Many mentions linked to wrong entities
  - Mention sometimes overly greedy with modifiers: “the Bronx district attorney on Wednesday”
  - Some summaries over 100 words
-

---

## D1102-A.M.100.A.1: ~~D3~~->D4

Australia said Tuesday it was engaged in an unprecedented diplomatic campaign to dissuade Japan from trying to escalate its killing of whales.

Tokyo will not yield to foreign pressure seeking to stop it from whaling a campaign against Japan's annual hunt in the name of scientific research.

~~Green Party lawmakers in Australia and New Zealand~~ ~~Japan~~ are considering urging consumers to boycott Japanese products to protest Tokyo's plan to expand its annual whale hunt, Green Party's co-leader said Monday.

An animal rights group on Friday lost a bid to sue a Japanese whaling company for allegedly killing hundreds of whales inside an Australian whale sanctuary.

---

---

# D1028-A.M.100.E.1: ~~D3~~->D4

~~Unabomber suspect Theodore J. Kaczynski~~ pleaded innocent today via video to charges he sent the mail bomb killing an advertising executive exactly two years ago.

The judge in the trial of ~~Unabomber suspect Theodore Kaczynski~~ ~~UNABOMber suspect Theodore Kaczynski~~ turned down a series of defense requests for revisions in jury selection.

A federal judge has rejected a motion to exclude key evidence from the Sacramento, California trial in November of ~~UNABOMber suspect Theodore J. Kaczynski~~.

Authorities have moved ~~UNABOMber suspect Theodore Kaczynski~~ from Sacramento County Jail to a federal prison 20 miles southeast of Oakland in California.

Lawyers for ~~UNABOMber suspect Theodore Kaczynski~~ ~~UNABOMber Kaczynski~~ are asking for special measures to find northern California jurors who aren't biased against him by news coverage.

---



---

---

# Related reading which influenced our approach:

Sebastian Martschat and Michael Strube. 2015. Latent structures for coreference resolution.

*Transactions of the Association for Computational Linguistics*, 3, 405-418.

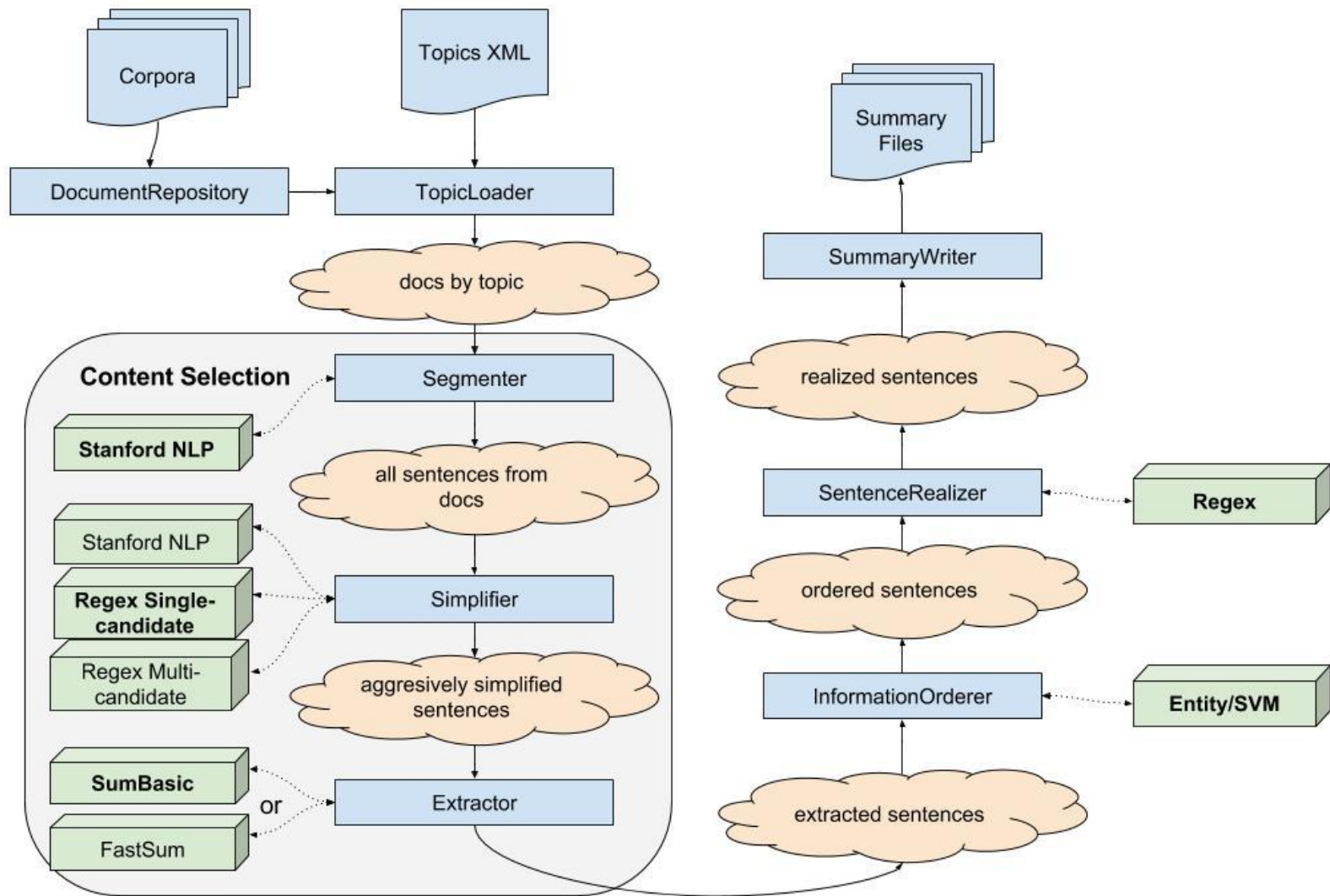
---

# Summarization System Improvements

Alex Burrell, Robert Gale, and Chris LaTerza

# Improvements for Deliverable #4

- Making ordering less crude (though still fairly crude)
- FastSum implementation
- Stopwords numbers experiment
- Improved readability through sentence filtering and realization
- Multiple levels and types of simplification



# FastSum

- **Learning algorithm:** Support Vector Regression with a linear kernel
- **Goal:** predict a score in range  $[0,1]$  showing how summary-worthy a sentence is, the higher the better.
- **Training:** each sentence in a cluster was scored by its normalized content word overlap with gold standard summaries of that cluster
- **Features:** Schilder and Kondadadi (2008) use LARS to show the best features, and optimal number of features, for FastSum. Our best run used 4 of their top features: document frequency, content word frequency, topic title frequency, and headline frequency

# FastSum results on dev set

	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4
Recall	0.19938	0.05580	0.02007	0.00737
Precision	0.22067	0.06143	0.02202	0.00808
F-score	0.20901	0.05834	0.02095	0.00769

# Stopwords Numbers Experiment

- Observation: Number showing up as significant tokens, e.g. 1985, 20,000
- Idea: Consider any token without alphabetical content a stopwords (regex)
- Outcome: Helped with dev set, hurt with evaluation set.

	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4
Dev Keep Numbers	0.26144	0.06701	0.02175	0.00734
<b>DEV Filter Numbers</b>	<b>0.26627</b>	<b>0.06812</b>	<b>0.02304</b>	<b>0.00735</b>
<b>EVAL Keep Numbers</b>	<b>0.31007</b>	<b>0.08679</b>	0.02947	<b>0.01339</b>
EVAL Filter Numbers	0.30894	0.08588	<b>0.02952</b>	0.01325

# Making Ordering Less Crude

Method: Entity Ordering (Barzilay and Lapata, 2005)

Concept: Devise “entity transitions” as features. Train on raw articles.

Tools: Libsvm linear regression

Previously:

- “Corefs” match on noun word
- Salience: present vs not present
- Average of 608-way tie for best ordering

Now:

- Match whole noun phrases
- Ranking by left-to-right position
- Average of 2-way tie for best ordering



# Improved readability

- After FastSum extraction, filter out sentences that:
  - Are too short
  - Are quotes
  - Are questions
  - Contains a subject that is not a pronoun, somewhere in the sentence
- Make sure the 1st reference to a name is the full name, after that it can be just the last name

# Simplification levels

- MINIMAL
  - Remove sentence altogether if it doesn't contain a verb
  - Remove newspaper-style junk
- CONSERVATIVE
  - Along with MINIMAL...
  - Remove all content inside parentheses and dashes
- AGGRESSIVE
  - Along with MINIMAL and CONSERVATIVE...
  - Remove comma-separated clauses that start with certain POSs that are deemed less useful in the final summary

# Single vs. Multi Candidate Simplification

- Single-candidate
  - Just one sentence passed forward from simplification step to extraction
  - All sentences are simplified at the same level, set by a flag
- Multi-candidate
  - Sentences simplified with the MINIMAL, CONSERVATIVE, and AGGRESSIVE flags are all computed and sent forward
  - Extraction step ranks all the sentence versions
  - Filtering step ensures that only the single highest-scoring version of each sentence can make it through to the final summary

# Final system: components

- Aggressive single-candidate regex sentence simplification
- SumBasic content selection
- Additional cosine redundancy filtering
- Entity-based content ordering

# Final system: results

	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4
Dev	0.26627	0.06812	0.02304	0.00735
Eval	0.30894	0.08588	0.02952	0.01325

Thanks!