

D4 - Multi-Document Summarization

...

Maria Sumner, Micaela Tolliver, Elizabeth Cary

CONTENT SELECTION - UPDATES

- Tried different TFIDF Measures
 - Log normalization, smoothing methods
 - Substantially decreased scores, even with different weights
- Weighting Measures
 - Needed to find best downweight strategy
 - Old strategy: Look at a couple downweight factors for devtest
 - Overfit to devtest
 - New strategy: Test a range of weights on devtest and devtrain, average ROUGE results between the two for the final selected weights

INFORMATION ORDERING - UPDATES

- Distance calculation
 - Preceding, current, succeeding sentences
- If > 6 sentences, generate 1000 possible orderings
 - Based on random and single-swap orderings

The study attributed the reason to the increasing incidence of obesity among UAE nationals, which is strongly linked to diabetes.

Doctors said common sense should be applied and they suggested nationals to eat more fruits, cereals and vegetables which were cited as the best sources to avoid diabetes.



Fifteen percent of female adults suffer from obesity, while the level among male adults 10.68 percent.



However, more boys suffer from obesity than girls.

FORAY INTO CONTENT REALIZATION

- Compressed sentences inserted into SumBasic Framework
- Stanford Parser
 - Abbreviations
 - Appositives
 - Adjectives
- Attributives

Arafat no longer has fever or nausea, but the stuffy nose remains, his aide said.
Arafat no longer has fever or nausea, but the stuffy nose remains.

STANFORD DEPENDENCY PARSER

The first batch of 100-ton cargoes was airlifted to Colombo, capital of Sri Lanka, Wednesday morning.

```
det(batch-3, The-1)
amod(batch-3, first-2)
nsubjpass(airlifted-8, batch-3)
amod(cargoes-6, 100-ton-5)
prep_of(batch-3, cargoes-6)
auxpass(airlifted-8, was-7)
prep_to(airlifted-8, Colombo-10)
appos(Colombo-10, capital-12)
nn(Lanka-15, Sri-14)
prep_of(capital-12, Lanka-15)
nn(morning-18, Wednesday-17)
tmod(airlifted-8, morning-18)
```

EXAMPLES

ABBREVIATIONS

Hong Kong's Department of Health (DH) on Thursday reported four more suspected cases of Influenza A H5N1, bringing the number to 10 in addition to nine confirmed cases.

APPOSITIVES

Phillips drove his boat 20 miles to Emory Hollow, a nearly two-hour trip, to participate.

Most days, the stainless steel sinks and tables inside their shuck house were shining, untouched.

Most days shining, untouched.

EXAMPLES

ADJECTIVES

Half of the assistance will be in the form of grants for **short term** relief and the remainder in loans for long term reconstruction, Howard said.

It was the biggest earthquake anywhere in the world...

It was the earthquake anywhere in the world ...

TYPE	ROUGE-1	ROUGE-2
baseline	0.3053	0.0888
appos+abbrev+amod	0.3050	0.0886
appos+abbrev	0.3063	0.0891
appos+amod	0.3044	0.0888
abbrev+amod	0.3052	0.0889
abbrev	0.3053	0.0889
amod	0.3059	0.0892
appos	0.3035	0.0886

RESULTS - devtest

D4 - Average recall

ROUGE-1	0.30482
ROUGE-2	0.08717
ROUGE-3	0.02879
ROUGE-4	0.01045

D3 - Average recall

ROUGE-1	0.29498
ROUGE-2	0.08520
ROUGE-3	0.03001
ROUGE-4	0.01209

RESULTS - devtest and evaltest

D4 - Average recall, DEVTEST

ROUGE-1	0.30482
ROUGE-2	0.08717
ROUGE-3	0.02879
ROUGE-4	0.01045

D4 - Average recall, EVALTEST

ROUGE-1	0.33469
ROUGE-2	0.09647
ROUGE-3	0.03300
ROUGE-4	0.01396

D1020-A.M.

D3

Bil Mar is a division of Sara Lee Corp. One recall was for milk. No illnesses have been reported. The outbreak began on Aug. 2. The investigators say they do not know how the listeria contaminated the cooked meats at Bil Mar Foods. The agency is studying whether some of the Bil Mar products might contain the Listeria bacteria. The Agriculture Department announced the recall Friday night. Since September, there have been nine known recalls due to listeria, mostly from deli meats and hot dogs. Any lunch meat or hot dogs should be thoroughly heated.

D4

More than 250,000 kilograms of ground beef in 33 U.S. states have been recalled because the meat might contain a deadly strain of E. coli. The Agriculture Department announced the recall Friday night. The investigators say they do not know how the listeria contaminated the cooked meats at Bil Mar Foods. The agency is studying whether some of the Bil Mar products might contain the Listeria bacteria. Club Pack hot dogs, and Grillmaster, Hygrade, Turkey, Sara Lee Deli Meat and Sara Lee Home Roast brands. One recall was for milk. Any lunch meat or hot dogs should be thoroughly heated.

D1015-A.M.

D3

Brazil and Malaysia are to jointly launch a series of remote sensing satellites (RSS) from 1998 to monitor tropical forests. China's largest tropical rain forest, in the Xishuangbanna nature reserve in Yunnan Province, will get further protection when the reserve is enlarged from 247,000 ha to 533,000 ha, according to Zhuang Yan, head of the Xishuangbanna Dai Autonomous Prefecture.

Xishuangbanna, one of China's largest tropical rain forest reserves, will almost double its area to bring more wild plants and animals under protection. A tropical rain forest project is to start soon in south China's Hainan province.

D4

A tropical rain forest project is to start soon in south China's Hainan province. It first set up a tropical primary forest protection zone in the area in 1951. **Angola, Liberia, Nigeria, Gabon, Congo, Cameroon, Sao Tome and Principe, Tanzania, Ghana, Cote d'Ivoire, Zaire, Equatorial Guinea and the Central African Republic are members of the African Lumber Organization.** China's largest tropical rain forest, in the Xishuangbanna nature reserve in Yunnan Province, will get further protection when the reserve is enlarged from 247,000 ha to 533,000 ha, according to Zhuang Yan, head of the Xishuangbanna Dai Autonomous Prefecture.

FUTURE ASPIRATIONS

- More sophisticated parsing heuristics
 - Current parsing heuristics only improve ROUGE scores very marginally and decrease readability
 - Improving parsing methods could help with the drop in readability
- Other potential improvements
 - Starting sentence in D1118: “She then called Hawkins’ mother.”
 - Coreference problems, especially with pronouns
 - Either utilize a coreferent tool, or penalize pronouns at the beginning of the summary

RELATED READING

- Vanderwende, Lucy, et al. "Beyond SumBasic: Taskfocused summarization with sentence simplification and lexical expansion." *Information Processing & Management* 43.6 (2007): 1606-1618.
- Zajic, D., B. Dorr, J. Lin, C. Monz, and R. Schwartz. 2005. A Sentence-Trimming Approach to MultiDocument Summarization. In *Proceedings of DUC 2005*.

UNIVERSITY *of* WASHINGTON

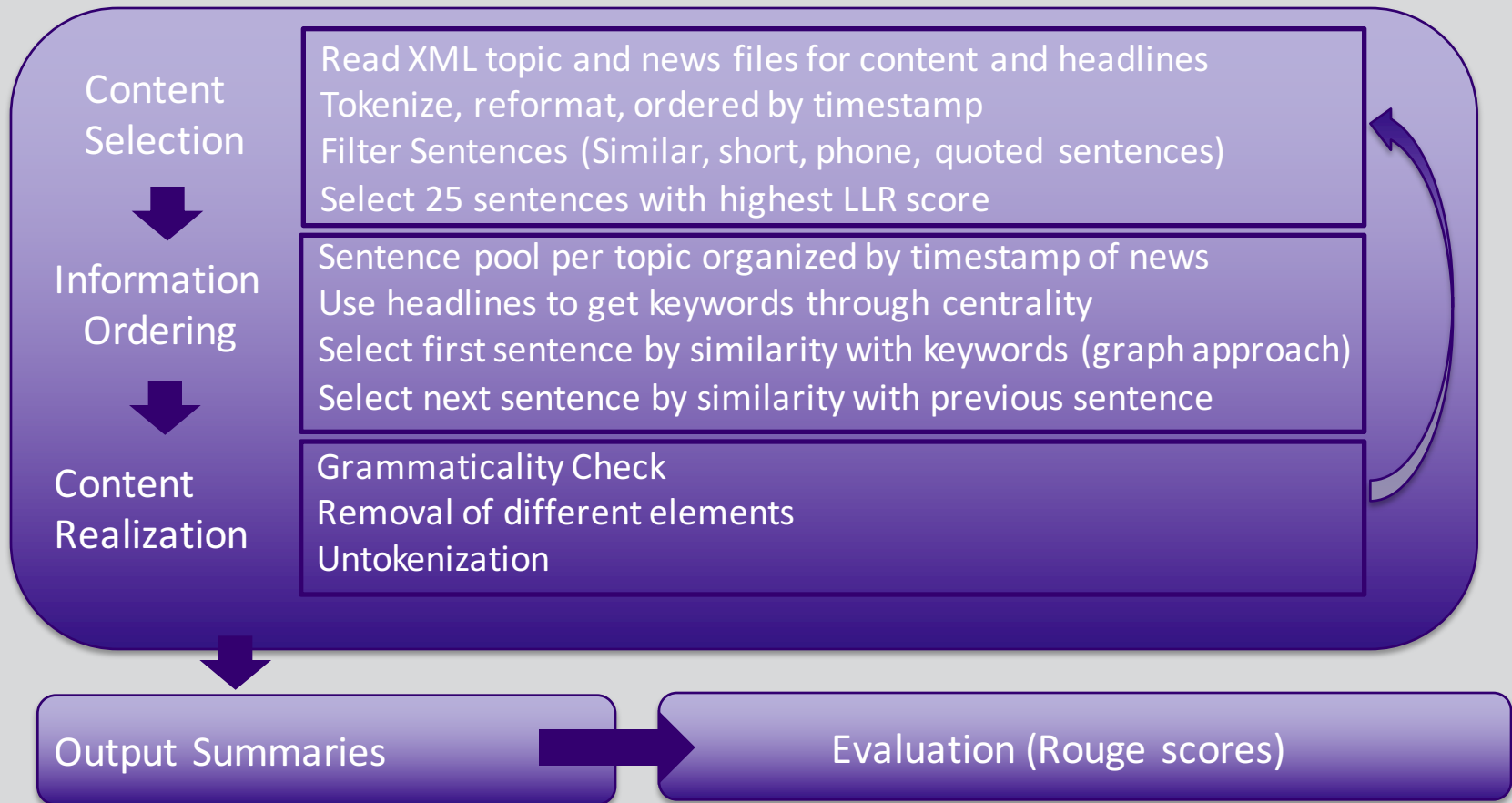
Summarization Task

Deliverable 4

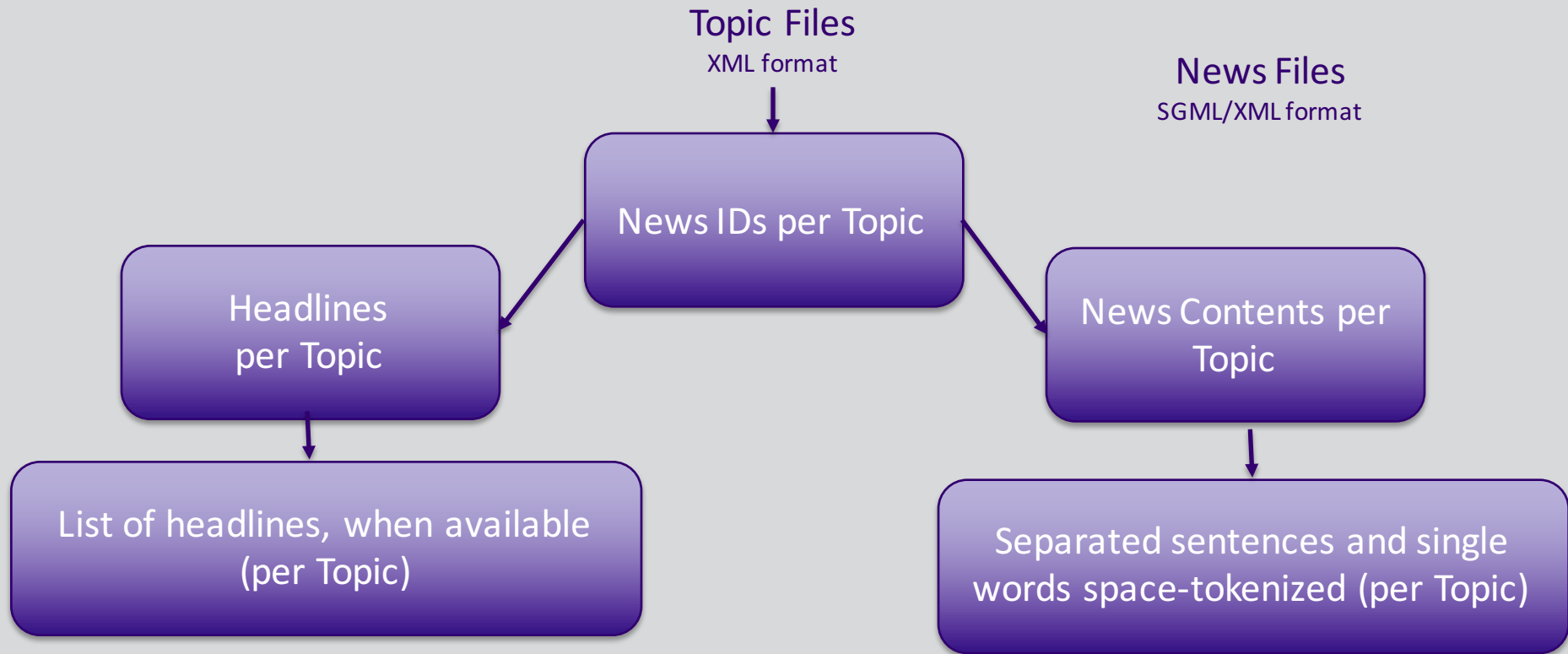
LING 573 – Spring 2016



System Architecture – Diagram



Content Selection – Information Extraction



Content Selection - Log Likelihood Ratio

Obtain words tokens per sentence (excluding stopwords)

Used English Gigaword as background corpus

$$L(n, k, p) = p^k (1 - p)^{n-k} \rightarrow \text{weight}(w) = \begin{cases} 1, & \text{if } -2 \log \lambda > 10 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{weight}(s_i) = \frac{\sum_{w \in s_i} \text{weight}(w)}{|\{w \mid w \in s_i\}|}$$



Content Selection – Grammaticality Check

Many of our intermediary outputs resulted in sentence fragments

→ Use of a syntactic parser to determine syntactic category of whole “sentence/fragment”

ACE/ERG parser
(pydelphin)

Stanford CoreNLP
Parser

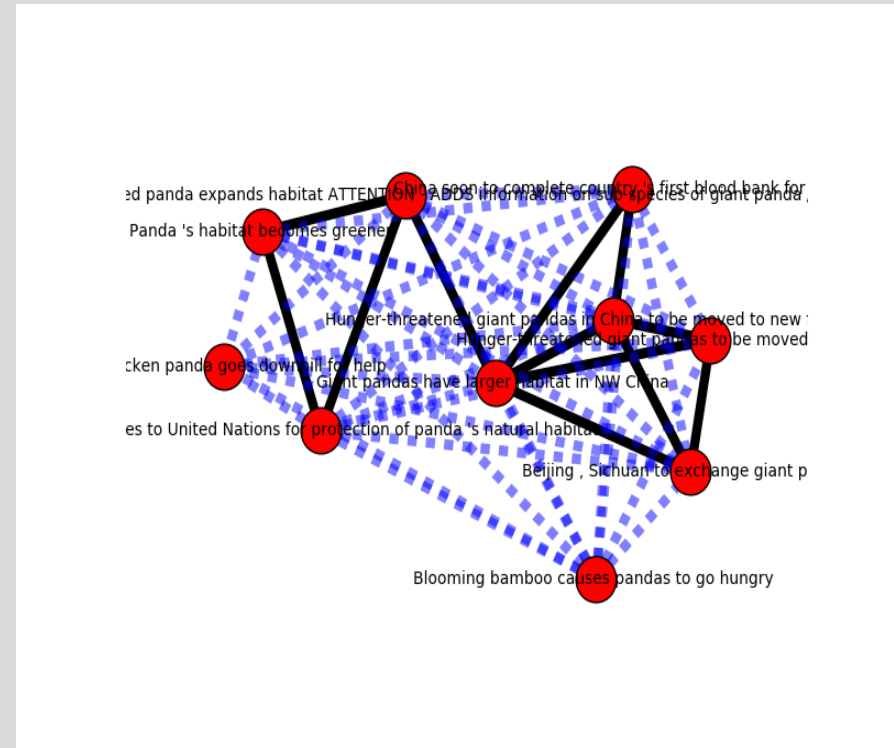
Berkeley
Parser

→ Only full sentences were further allowed down the process

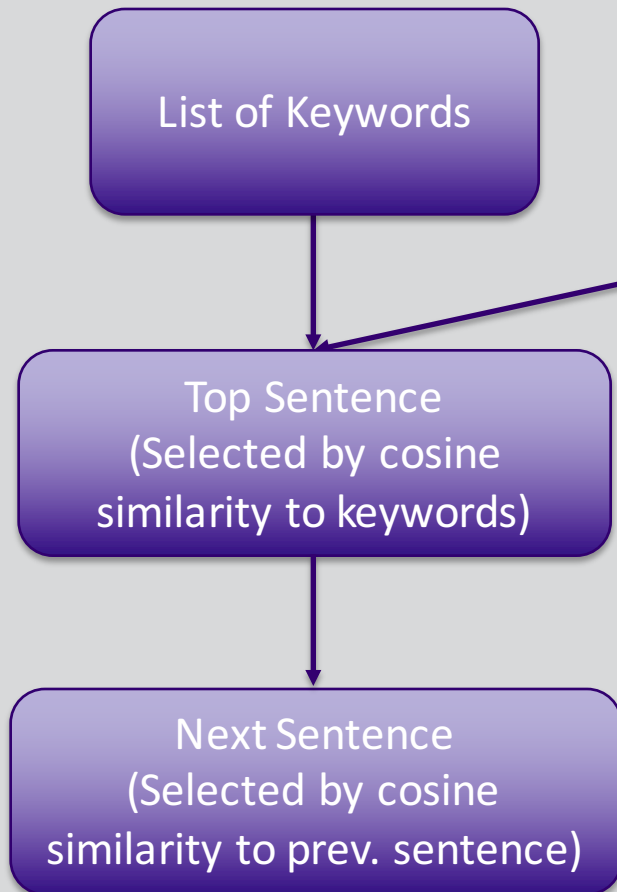


Headline Content Selection

- > Focus on words instead of sentences
- > Each word is scored based on how many headlines it appears in
- > Pass on top n keywords
- > A keyword is only passed on if it appears in more than one headline



Content Selection – Information Ordering



Updates

- Added minimum cosine similarity when checking chronologically later sentences
- Fixed minor bug involving selecting the most similar first sentence based on keywords
- Parameter adjustments



Content Realization

- > Regex based removal of some fragments
- > Initial adverbials
- > Bylines
- > He/she said phrases at start and end
 - Only if the phrase does not contain a keyword



Evaluation of Results

Recall ROUGE scores for 1,2,3,4 N-grams.
Here average scores over all summary topics:

DevTest

	R-1	R-2	R-3	R-4
D3	0.17798	0.04681	0.01615	0.00658
D4	0.19131	0.05375	0.02087	0.01039

EvalTest

	R-1	R-2	R-3	R-4
D4	0.23346	0.06468	0.02486	0.01273



Output Analysis - The good...

D3 – Baseline

'That's what I want to look like,'" says Tracie Strucker, director of La Que Sabe, a Central Texas eating disorders outpatient program. The Tulane University-affiliated River Oaks Hospital Eating Disorders Treatment Center in New Orleans uses several approaches simultaneously in treating anorexics. For instance, the diagnosis of Asperger's syndrome, a variant of autism, was pulled from a collection of disorders previously labeled Pervasive Personality Disorder Not Otherwise Specified.

D4 – Final System

According to the National Eating Disorder Association, about 7 million U.S. females and 1 million males suffer from an eating disorder. Heidi Strickler, an eating disorder specialist at Providence Healthcare Network's DePaul Center who arrived after Montaner started Lifelines, helped write the materials.

Men, too, are being diagnosed as bulimic or anorexic at an increasing rate, says Dr. Sharon Alger-Mayer, who works in the division of clinical nutrition at Albany Medical Center Hospital. Television shows such as The Swan or Extreme Makeover equate thinness and beauty with happiness.



New pre-selection Content Realization features, along with improvements/fixes to the Content Selection, resulted in a much better choice of first sentence, which in turn led to much improved summaries.



Output Analysis - The good...

D3 – Baseline

But U.S. Rep. Tom Davis of Virginia, Republican chairman of the House Government Reform Committee, sent a letter to FDA's acting commissioner, Dr. Lester Crawford, giving him two weeks to answer questions about how FDA tracks problems with drug safety, how it dealt with issues surrounding Vioxx and whether the agency will study potential risks in similar drugs. Dr. David J. Graham, associate director for science in the FDA Drug Center's Office of Drug Safety, told Senate investigators he faced stiff resistance within the regulatory agency to his findings.

D4 – Final System

On Friday, Pfizer Inc issued a warning that its Cox-2 drug Bextra may increase cardiovascular risk for some patients. In addition to Bextra, Pfizer makes Celebrex, the best-selling COX-2 drug on the market. Less than a week after Merck & Co. yanked its popular Vioxx arthritis drug off the market, the New England Journal of Medicine voiced strong concerns Wednesday about the safety of Pfizer's bestselling Celebrex. Researchers writing in the New England Journal of Medicine voiced their concerns as well with such drugs as Pfizer's popular Celebrex.

W

Output Analysis - The bad...

D3 – Baseline

Lyle Vail, director of the Lizard Island Research Station on Australia's Great Barrier Reef, said damage to coral reefs from a tsunami would likely be similar to that from a cyclone.

Tamelander said many coral reefs in the Indian Ocean were just beginning to recover from damage caused by the changing water temperatures generated by the El Nino weather pattern.

The agency reported one case in Thailand where the mangroves also damaged shallow coral reefs. After the tsunami, authorities in Tamil Nadu state launched a massive forestation program along the coast.

D4 – Final System

After the tsunami, authorities in Tamil Nadu state launched a massive forestation program along the coast.

Swaminathan Research Foundation, which helped the government persuade the Irulas to manage the Pichavaram forest.

In the Maldives, dive operator Norbert Schmidt said the eastern part of the island was hit the worst, with dead coral and sand covering the runway at Hulule International Airport.

The Shompen live along riverbanks in the dense mangroves surrounding the settlement of Campbell Bay, 800 kilometres south of the Andamanese capital Port Blair.



Because we removed the said phrases, the initial sentences - which actually had some good, informative content- were removed, and the summary became less coherent and less informative overall



Output Analysis - The ... meh...

D3 – Baseline

BOULDER, Colo. _ Sheets of white paper covered windows in the courtroom where a grand jury met Thursday to pursue its two-week-old investigation into the murder of JonBenet Ramsey.

BOULDER, Colo. (AP) -- Prosecutors received nearly \$ 80,000 Thursday to help finance the grand jury investigation into the JonBenet Ramsey slaying through March.

BOULDER, Colo. (AP) -- Prosecutors posted a photograph of a white, furry bear dressed as Santa Claus on an Internet site Thursday in a plea for public help with the JonBenet Ramsey slaying investigation.

D4 – Final System

Police say her parents, John and Patsy Ramsey, remain under suspicion.

Boulder Police Chief Mark Beckner has said they are under an umbrella of suspicion.

In two separate letters, they accuse Boulder District Attorney Alex Hunter of unwarranted delays in prosecuting the case and of being overly accommodating to the Ramseys.

Boulder District Attorney Alex Hunter appointed Bruce Levin and Mitch Morrissey, in part, because of criticism he has mishandled the Ramsey case. Susan Gordy of Atlanta said the Ramseys' stated reason for waiting makes sense.



The bylines in the D3 summary meant that the sentences chosen were very informative, but the summary had no coherency. The removed bylines in the D4 summary meant that a much more coherent summary was formed, but it was less informative overall



Conclusion – looking to the future....

Summarization system should be able to distinguish between type of information supplied in base documents, in order to select strategy

➔ Different performance of system on different doc sets because documents are different

MRS-representations together with lexical semantic tools for a more abstractive approach, using the ERG grammar for sentence generation

➔ How can we generate different sentence styles with a grammar out of an abstract meaning representation?



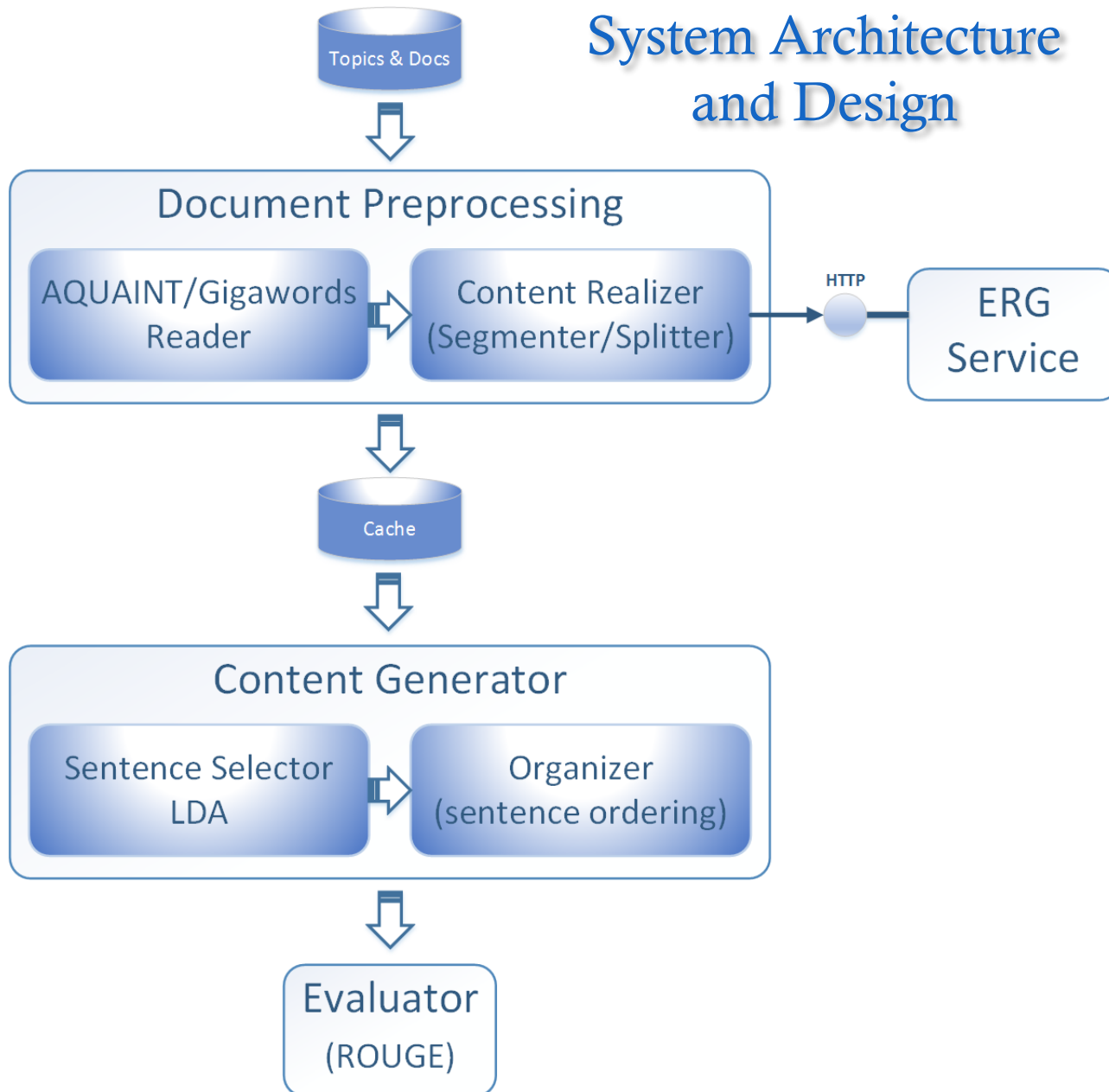
MultiDocSummarizer

LING 573 – D3½

Natalia Rodnova, Cade Bryant, Kevin Wonus
(The Calamity Crew)



System Architecture and Design



Components

- ◆ Mostly Python 3
- ◆ Some C#
- ◆ NLTK for tokenization
- ◆ English Resource Grammar (ERG) to assist with CR
- ◆ Gensim LDA for sentence selection
- ◆ Theme-based ordering (*per Barzilay et al, 2001*)

Preprocessing Content Realization

- ◆ English Resource Grammar (<http://www.delph-in.net/erg/>):
ERG is used to realize complete sub-phrases of an otherwise complete sentence. The assumption is that splitting complete sentences into shorter complete sentences increases the density of information. By performing this step prior to sentence selection, extraneous information is eliminated by the subsequent LDA step since such extraneous data will not be selected.
- ◆ Entity-based content realization is planned, but not yet fully implemented.

Example – Preprocessing Content Realization

	Example Sentence
BEFORE	And in 1964, when a tsunami hit Alaska, news reports noted that baby salmon were killed, although it's unclear how many.
AFTER	And in 1964, when a tsunami hit Alaska, news reports noted that baby salmon were killed. It's unclear how many.

Sentence Selection using Latent Dirichlet Allocation

- ◆ First introduced by David Blei, Andrew Ng and Michael Jordan in 2003. Paper is called “Latent Dirichlet Allocation”
- ◆ Gensim algorithm utilized (created by Hoffman et , 2010)
 - ◆ LDA represents documents as a mixture of topics that share words with certain probabilities
 - ◆ For N most important topics, pick K most probable sentences
 - ◆ Assuming this generative model for a collection of documents, LDA then tries to backtrack from the documents to find a set of topics that are likely to have generated the collection.

Sentence Ordering using Chronology and Themes

- ◆ Chronology plays a role
- ◆ But ordering is augmented by themes (Barzilay et al, 2001)
 - ◆ Sentences are normalized by removing stopwords & punctuation
 - ◆ Sentences compose a *theme* if their content is similar
(*determined by cosine similarity*)

D3 $\frac{1}{2}$ compared with Baseline (devtest)

ROUGE	Baseline (D2)	Current (D3 $\frac{1}{2}$)
1	0.15280	0.15742
2	0.03258	0.03713
3	0.00860	0.01252
4	0.00212	0.00382

Possible Future Work

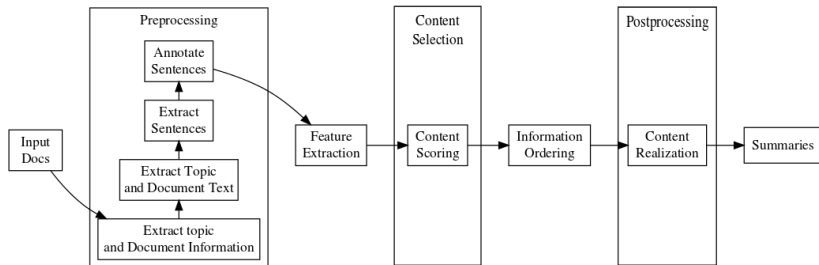
- ◆ Use machine learning for sentence ordering (based on “A preference learning approach to sentence ordering for multi-document summarization” by D. Bollegala, N. Okazaki and M. Ishizuka). Or cluster adjacency method proposed by J. Donghong and N. Yu in “Sentence Ordering based on Cluster Adjacency in Multi-Document Summarization” since we already have sentences clustered around topics by LDA.
- ◆ Use sentence compression and/or fusion
- ◆ Use anaphora resolution for better coherence

D4: Summarization with Neural Networks and Learning2Rank

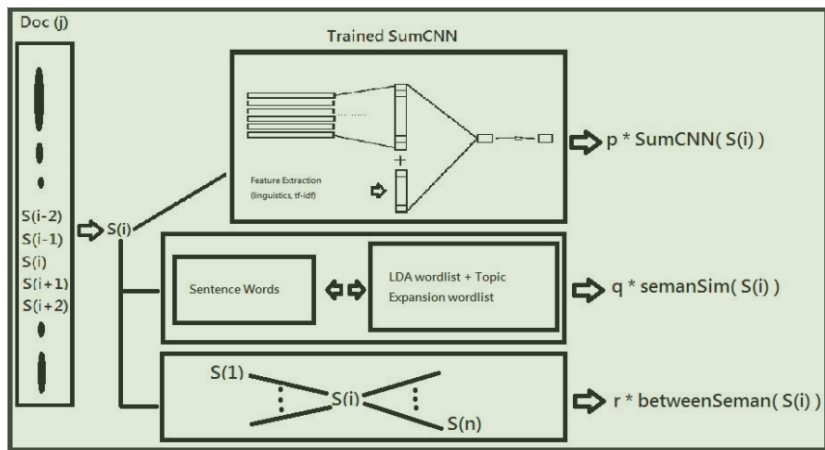
Tony P. and Ernie C. and Jason B.

June 2, 2016

System Architecture



Model Architecture



Information Ordering (Part 1)

- Conceptually applies principles of single document summarization to multi-document summarization.
 - Order by salience and then by position
 - Two ordering passes
- All topic sentences sorted first by saliency score.
- Salience summary (1st pass) built from saliency sorted sentences limited by compression value (max sentences parameter) and redundancy threshold parameter.

Information Ordering (Part 2)

- Position ordering (2nd pass) uses position information from other sentences in the input documents.
- Inspired by Barzilay et. al., 2002 Majority Ordering.
- Each sentence in salience summary is considered a theme.
- Sentences left out of salience summary are clustered to these theme sentences.
- Cluster members then use their document positions to vote on summary precedence between pairs of themes (i.e. salience summary sentences).

Information Ordering (Part 3)

- Overall votes determine path score between theme pairs.
- Best (max) path through salience summary is then determined producing ordered summary.
- If length of salience summary prevents exhaustive path calculation (TSP problem) then a sliding lookahead window is used.
- Exhaustive search within window.
- Parameter setting for window size to keep computationally tractable.
- Fixed starting point for window and only top new sentence is kept for each sliding window ordering.

Content Realization (Part 1)

- Creates final summary from position ordered summary.
- With the goal of a more content-rich summary, first attempts some simple, rule-based sentence compression for all sentences in the position ordered summary.
- Headers : "ARVADA, Colo. (AP) – "
- Leading adverbials : "Fortunately, "
- Within parentheses : "(U.S. dlr 83 million)"
- Relative date/time : "on Sunday"
- Trailing attribution : ", police said."

Improved ROUGE scores 3% (ROUGE-1) to 10% (ROUGE-4).

Content Realization (Part 2)

- Starting with top-ranked, possibly compressed, sentences adds sentences to final summary if the addition will not cause the final summary to exceed the summary word limit.
- Attempts to add all position summary sentences to final summary. Potential to have a lower scoring, but shorter sentence added to final summary – because it fits. Excludes sentences with a word length of five or less.
- The final summary is re-ordered using cosine similarity on 3 by 4 skipgrams (tri-grams, 4 word skips) to improve coherence.
- Again, a sliding lookahead window is used if exhaustive best path calculation is not computationally tractable.

Trial and Error

- Training Set

Model	Info	Rouge-1	Rouge-2	Rouge-3	Rouge-4
SumCNN	(0.0001, 1000, 200, 282)	0.23118	0.05905	0.01898	0.00797
LDA	LDA + ngram	0.31014	0.08566	0.02967	0.01295
Learning2Rank	semanSim + betweenSeman	0.31065	0.08586	0.02956	0.01305
New SumCNN	SumCNN+Learning2Rank	?	?	?	?

Results (ROUGE results table for devtest)

D	Name	Average R	CI Lower	CI Upper
2.0	ROUGE-1	0.07118	0.05750	0.08601
2.0	ROUGE-2	0.01484	0.01011	0.01998
2.0	ROUGE-3	0.00359	0.00146	0.00600
2.0	ROUGE-4	0.00046	0.00000	0.00103
3.0	ROUGE-1	0.19325	0.17105	0.21344
3.0	ROUGE-2	0.04657	0.03734	0.05547
3.0	ROUGE-3	0.01423	0.00989	0.01895
3.0	ROUGE-4	0.00436	0.00214	0.00684
3.1	ROUGE-1	0.26900	0.24814	0.28852
3.1	ROUGE-2	0.06284	0.05342	0.07218
3.1	ROUGE-3	0.01992	0.01493	0.02567
3.1	ROUGE-4	0.00676	0.00361	0.01136
4.0	ROUGE-1	0.27308	0.25360	0.29258
4.0	ROUGE-2	0.06404	0.05458	0.07321
4.0	ROUGE-3	0.02074	0.01545	0.02749
4.0	ROUGE-4	0.00724	0.00405	0.01202

ROUGE Results (D4 devtest vs evaltest)

test	Name	Average R	CI Lower	CI Upper
dev	ROUGE-1	0.27308	0.25360	0.29258
dev	ROUGE-2	0.06404	0.05458	0.07321
dev	ROUGE-3	0.02074	0.01545	0.02749
dev	ROUGE-4	0.00724	0.00405	0.01202
eval	ROUGE-1	0.35700	0.33624	0.37671
eval	ROUGE-2	0.09690	0.08291	0.11091
eval	ROUGE-3	0.03134	0.02273	0.04012
eval	ROUGE-4	0.01338	0.00824	0.01924

Results (devtest improvement for D1001)

- D2

LITTLETON, Colo. - Lynda Pasma and Kerry Herurlin stop halfway down Mt. Columbine on Saturday morning to pray.

LITTLETON, Colo. - AP - they stream in from all over -- hand-lettered sign and card and bouquet by the hundred, aim at comforting a town tear asunder.

LITTLETON, Colo. - AP - comforting each other with ceremony and song, Columbine High School's mourner turn a strip-mall parking lot into a arena of grief and promise each other they would reach past they pain.

- D4

Comforting ceremony song Columbine High School mourners turned parking lot arena grief promised would reach past pain. The day Columbine High School students return class delayed many attending funerals students killed April massacre administrator said. They streamed signs cards bouquets hundreds aimed comforting town torn asunder. The New York Times plans two pages stories photos graphics aftermath school shooting Denver suburb left dead. Republican presidential candidate Pat Buchanan says stricter gun laws could prevented deadly school shootings Littleton Colo. The sheriff initial estimate many dead Columbine High massacre mark apparently six SWAT teams swept building counted victims.

Results (devtest improvement for D1021)

• D2

tampa - with just a week to go before trial, the glare of celebrity have not dim.
tampa - investigator say it begin with a note, pass in a math class at Wharton High School.
tampa - the spotlight will fade, but she neighbor will always know.
the St. Petersburg Times plan to move the follow story for client of the New York Times News Service for edition of Wednesday, November 22, and thereafter.

• D4

Tampa Investigators say began note passed math class Wharton High School. The St. Petersburg Times plans move following stories clients New York Times News Service editions November thereafter. Tampa The spotlight fade neighbors always know. A female teacher pleaded guilty sex student avoiding prison part plea agreement. A hitch developed plea deal agreed former teacher pleaded guilty sex student spokesman prosecutor said. Prosecutors released photographs secretly recorded tapes support accusation Debra LaFave formerly schoolteacher sex student. A female teacher whose sexual liaisons middle school student made tabloid headlines avoided prison time pleading guilty sex boy classroom home

Results (Lowest scoring evaltest)

D1141 Table

ROUGE-1	R:0.19636	P:0.16667	F:0.18030
ROUGE-2	R:0.02583	P:0.02187	F:0.02369
ROUGE-3	R:0.00000	P:0.00000	F:0.00000
ROUGE-4	R:0.00000	P:0.00000	F:0.00000

D1141 Summary

Half century Mao Zedong Great Leap Forward brought irrigation arid grasslands remote corner northwest China government giving attempt make breadbasket increasingly become stretch scrub sand dunes. China unveiled policies improve sustainable agriculture raise incomes hundreds millions farmers. This first sandstorm hit Minqin County northwest China Gansu Province China Meteorological Administration CMA said. Eleven sand storms expected hit China spring last country experienced sand storms far less annual average. China hillside farm areas lose billion tons soil annually third nation total soil loss said Vice Water Resources Minister E Jingping.

Results (Highest scoring evaltest)

D1126 Table

ROUGE-1	R:0.51376	P:0.34146	F:0.41025
ROUGE-2	R:0.22897	P:0.15123	F:0.18215
ROUGE-3	R:0.14762	P:0.09688	F:0.11699
ROUGE-4	R:0.10194	P:0.06646	F:0.08046

D1126 Summary

President Bush made valedictory visit Iraq country largely define legacy trip likely remembered unscripted moment Iraqi journalist hurled shoes Bush head denounced live television dog delivered death sorrow nearly six years war. Missed shoe landing first one loud thud wall behind two leaders held ground journalists security officials wrestled ground. A man threw shoes President George W. Bush dragged away security officials president farewell trip Iraq. Iraqi reporter throws shoe Bush Baghdad BAGHDAD Dec. Xinhua An Iraqi reporter threw shoes visiting U.S. President George W. Bush called dog Arabic news conference Iraqi Prime Minister Nuri Baghdad.

An obvious improvement would be to use the original sentences, possibly compressed, in the summaries to make them more readable.

- Should have:
 - output original sentence to content realization.
 - requested for torch/keras to also be installed at the beginning

- Regina Barzilay and Noemie Elhadad. 2002. Inferring strategies for sentence ordering in multidocument news summarization.
- Peng Wang, Jiaming Xu, Bo Xu, Chenglin Liu, Heng Zhang, Fangyuan Wang, and Hongwei Hao. 2015. Semantic clustering and convolutional neural network for short text categorization.