

Summarization System D3

Katherine Topping – Stephanie Peterson – Laurie Dermer

Changes to our system

Preprocessing

Selection Additions

Ordering & Theme Order

Command Line & Trial Logging



Error Analysis

The good, the bad, the irresponsibly incorrect.

Updated Rouge Scores

- Then

- ROUGE-1 : 0.12271
- ROUGE-2 : 0.02196
- ROUGE-3 : 0.00522
- ROUGE-4 : 0.00183

- Now

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



Columbine: the D2 version

b"<DOC> APW19990503.0128 1999-05-03 15:55:11 washington Congress Looking at Youth Violence \tWASHINGTON (AP) -- Pressured to help stop kids from killing, Congress is opening hearings on the causes of a ``crisis among our young" amid a thorny political question of what government should do to prevent massacres like the one in Littleton, Colo.

\t``The tragedy at Columbine High and the ongoing carnage on our inner city streets presents us with a complicated cultural moment and an important opportunity to thoroughly examine the root causes of a crisis among our young," House Judiciary Committee Chairman Henry Hyde told reporters on Monday.

b"<DOC> NYT19990424.0231 NEWS STORY 1999-04-24 21:37 A2024 tad-z u a BC-SCHOOL-WRAP25-COX 04-24 0784 BC-SCHOOL-WRAP25-COX
`Please comfort this town' By Rachel Sauer c.1999 Cox News Service
LITTLETON, Colo. _ Lynda Pasma and Kerry Herurlin stopped halfway down Mt.

Our first D3 system's results

...left out some important details

and highlighted some irrelevant ones.

about shoes, for example.




If it bleeds..... meh.



\t`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.



\tThe baseball team has received an estimated \$5,000 worth of clothing and gear from Reebok, Mizuno, Denver Athletic Supply and other sports companies.



Changing theme ordering helped

A bit, at least.

But we're still pretty far off from the target summaries.

Columbine: The official version

\tIn 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.

\t`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: the target

In the worst school killing in U.S. history, two students at Columbine High School in Littleton, Colorado, a Denver suburb, entered their school on Tuesday, April 20, 1999, to shoot and bomb.

At the end 15 were dead and dozens injured.

The dead included the two students, Eric Harris and Dylan Klebold, who killed themselves.

Harris and Klebold were enraged by what they considered taunts and insults from classmates and had planned the massacre for more than a year.

The school is a sealed crime scene and Columbine students will complete the school year at a nearby high school.

Here's the error analysis

- Wow! That's... not really the most pertinent facts of what happened at Columbine.
- But it's *coherently* irrelevant.
- The second one was also better than the first one.
- We'll see whether that continues to hold true once we start shortening sentences – which will also allow more content into the summaries and give our ordering system more opportunities to fail.
- It looks like our ROUGE scores may have been artificially boosted by tf*idf picking first sentences... due to high-scoring metadata

Preprocessing

- Remaining metadata: removed!
- Now process each headline in the same vain as we do sentences
- Processed headline associated with doc_id (and is passed onto ordering)

Selection Additions

- Added LLR as an option for word/sentence weighting scheme
 - Probability of observing w in cluster taking into account probability of observing w in background corpus
 - In our model cluster is just a document
- Added downweighting strategy in an effort to control redundancy
 - Multiplies sentence scores by a specified float if the sentences contain non-stop-words already present in selected sentences
 - Helps with redundancy, but tanks ROUGE scores and coherence of themes in output summaries

Ordering & Theme Ordering

- Lots of experimentation, loosely based off of Barzilay et al, '02 (discussed in class)
- Themes are chosen using word frequency in selected sentences
 - Also experimented with extra weighting for words that appear in headlines, though found this generally lowered ROUGE scores
 - Want to better tune the similarity measure/headline weighting moving forward
- Themes are ordered based upon "popularity" -- how many sentences fall under that theme, in descending order
 - Also experimented with ordering themes by chronology using their first appearances, but this yielded some wacky summaries
- Sentences within themes are ordered chronologically

Command Line & Trial Logging

- We added the ability to toggle our various options for selection/ordering from the command line
- We used this to run numerous tests
- We got some unexpected and/or heartbreaking results
- Just our hearts broken right there in text with numbers
- This probably means our selection strategy is to blame for our low ROUGE scores – back to the drawing board there
- Our ROUGE numbers were all over the place until we realized that we were overwriting our summary output while it was being read by the ROUGE evaluation script – we needed separate run IDs
- Other ROUGE variation seemed to be due to tiebreaking in theme ordering

Various System Scores

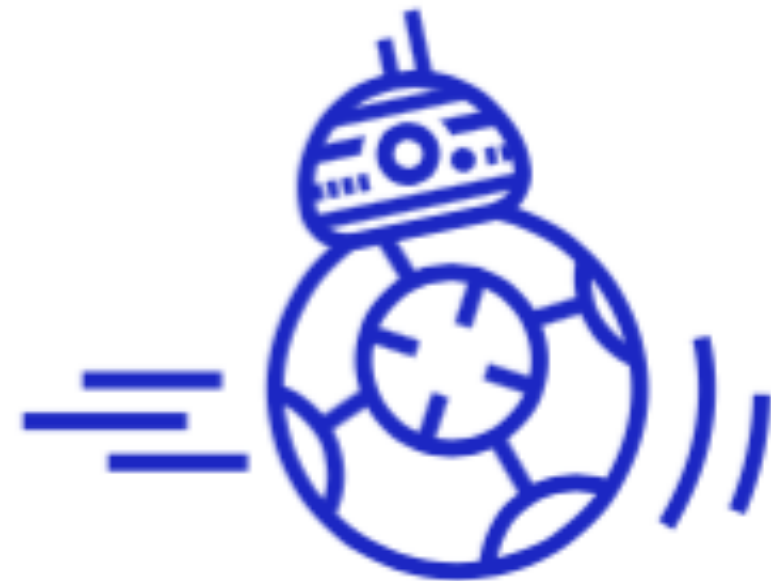
- The first term is the selection algorithm
 - The numbers indicate the redundancy multiplier to suppress sentences with words that have been chosen already (1.0 means redundancy handling was turned off)
 - “On” and “off” refer to boosting "themes" with headline words when ordering sentences by theme
- | | |
|---------------|----------------------------|
| tfidf+1.0 on | ROUGE-2 Average_R: 0.01897 |
| tfidf+1.0 off | ROUGE-2 Average_R: 0.01862 |
| llr+.9 on | ROUGE-2 Average_R: 0.00976 |
| tfidf+.9 on | ROUGE-2 Average_R: 0.00987 |
| llr+.9 off | ROUGE-2 Average_R: 0.01626 |
| tfidf+.9 off | ROUGE-2 Average_R: 0.00982 |

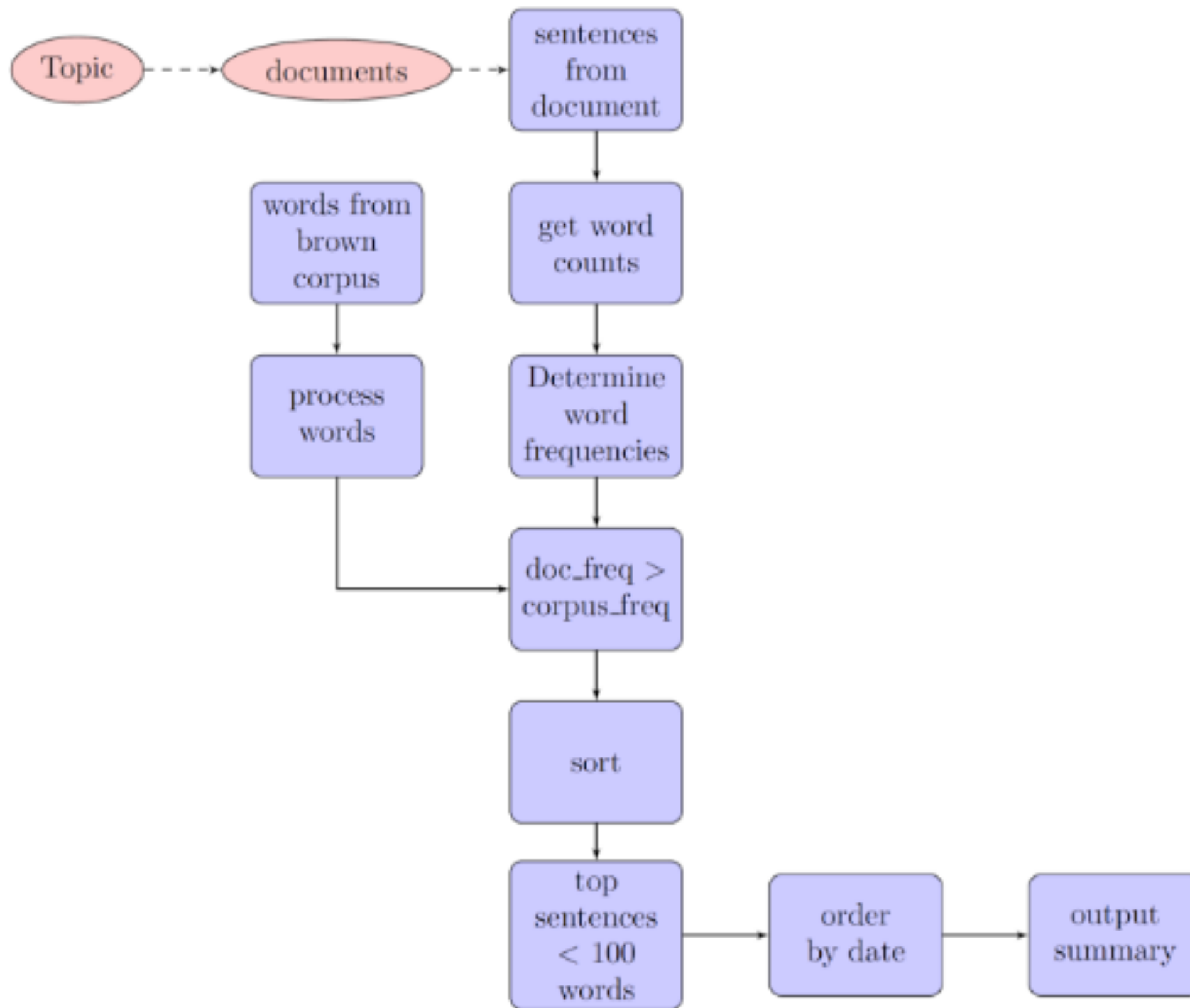
Deliverables 3

Matt Calderwood

Kirk LaBuda

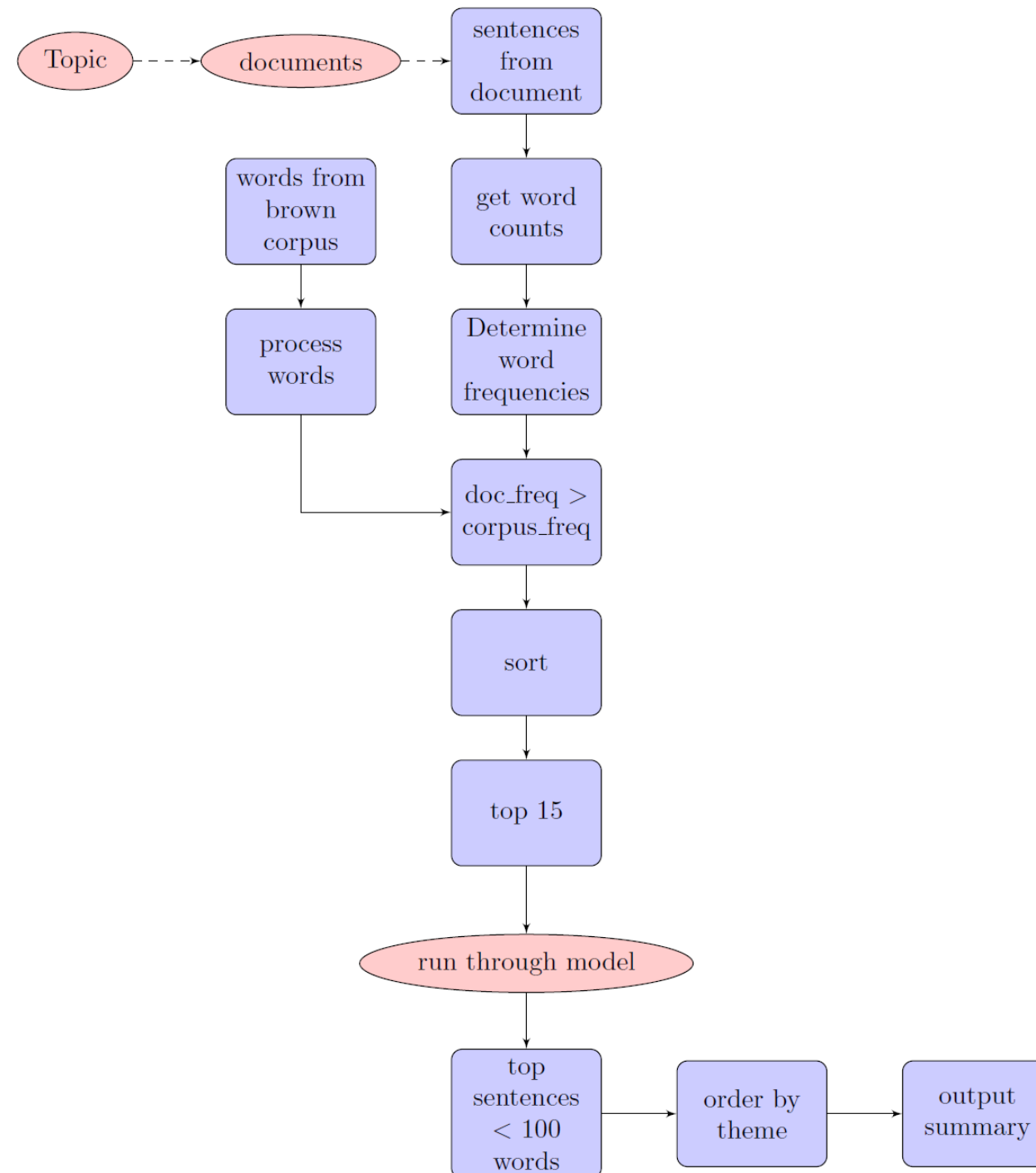
Nick Monaco





D2 System Architecture Diagram





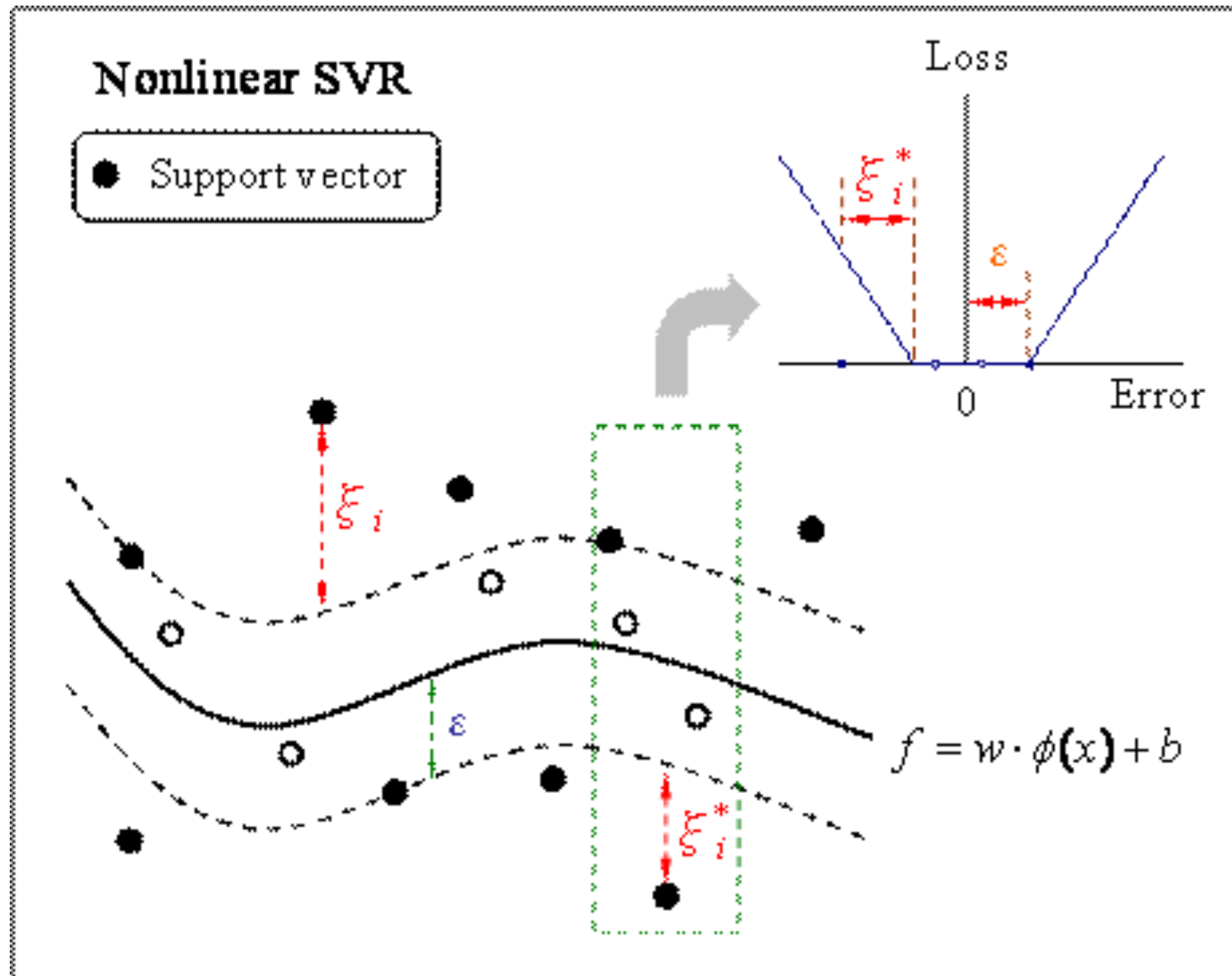
Updated diagram for D3



System Changes

- Content Selection - incorporated new machine learning approach with support vector regression. Currently using *linear kernel*.
- Plot summaries in 3D space - use hyperplane to predict ROUGE score and choose summary with best probable ROUGE score

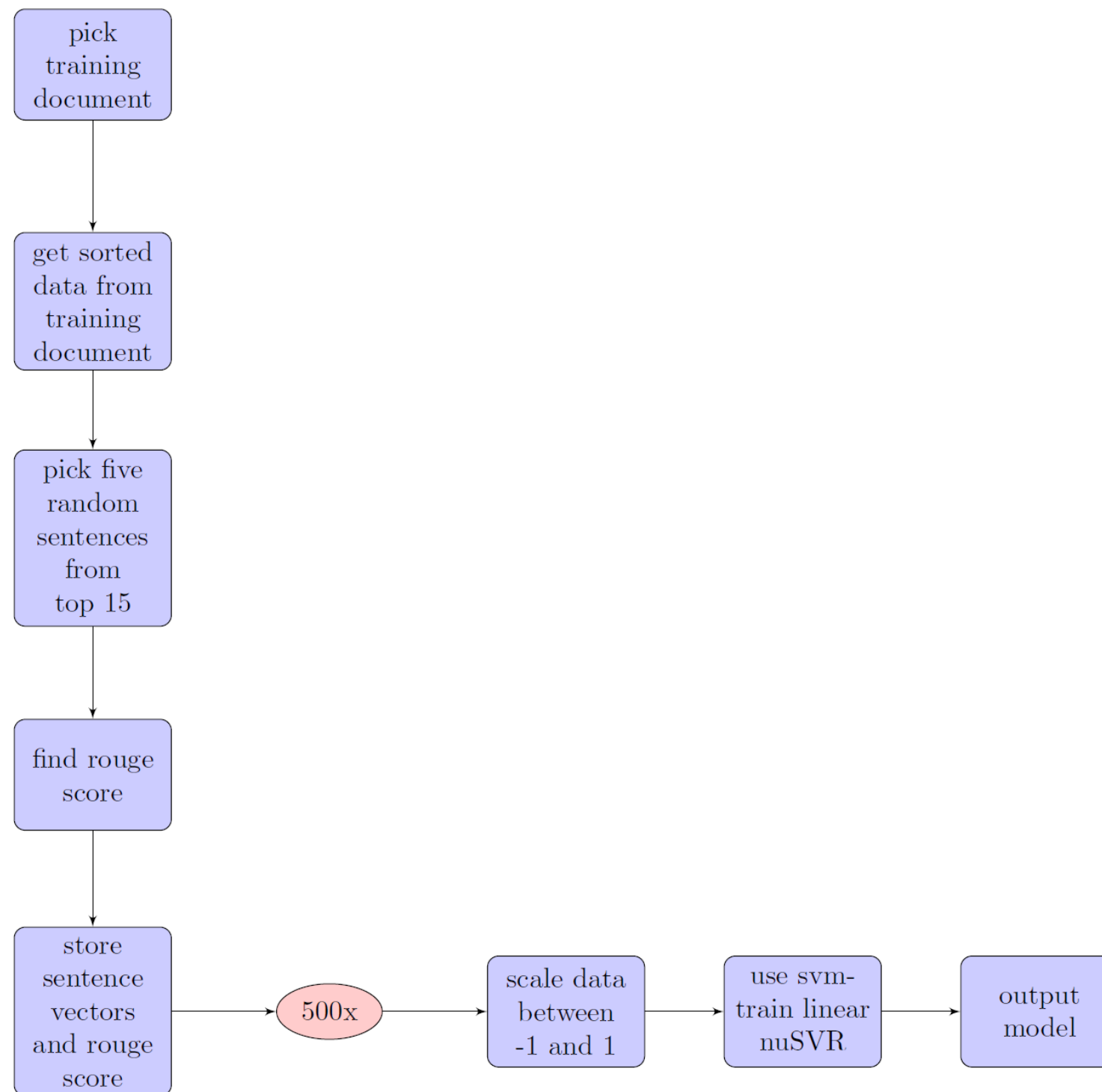




“

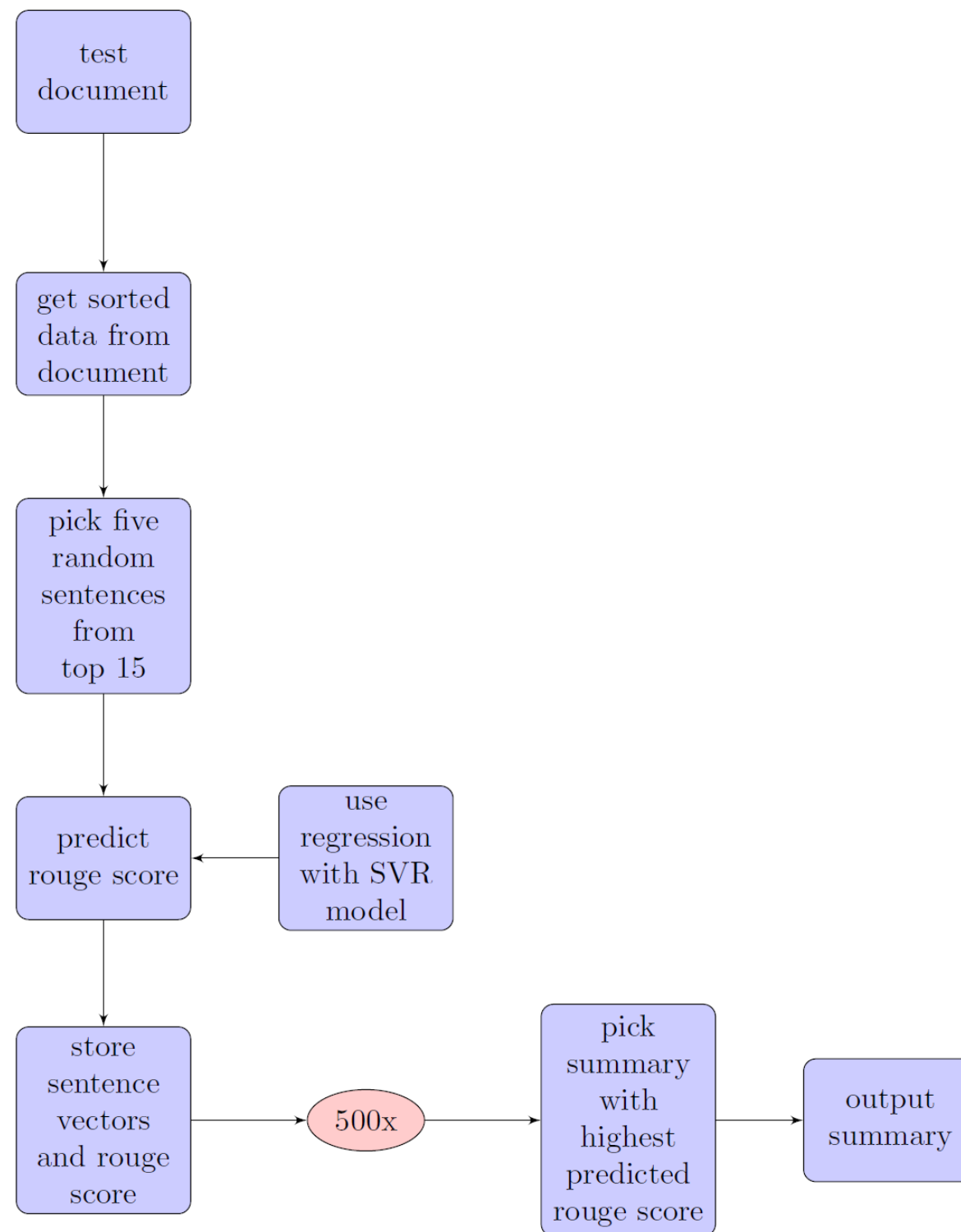
Suppose we are given training data $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset X \times \mathbb{R}$, where X denotes the space of the input patterns (e.g. $X = \mathbb{R}^d$). In ε -SV regression, our goal is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. ...we do not care about errors as long as they are less than ε ...





Support Vector Regression - Training Diagram





Support Vector Regression - Testing Diagram



System Changes (cont.)

- Information Ordering - Order summary sentences by theme. Shallow approach.
- Content Realization - no changes



Successes

- Machine Learning/SVR- approach seems promising.
- Info Ordering - shallow approach seems reasonable, has yielded some good results.



Issues



- Machine learning approach - still experimenting with different kernel fns for SVR. Planning to use more features.
- Info Ordering - multiple sentences occasionally registering as one sentence -skews results.
- Content Realization - could use sentence compression - some summaries contain long sentences.
- Runtime- room for optimization with runtime.



Good:

D1019-A.M.100.D.1

Six masked assailants have taken more than 20 people, including Americans and Europeans, hostage on a remote resort island in eastern Malaysia, police said Monday. Prior to the rebel announcement, Col. Ernesto de Guzman, chief of staff of the Philippine military's Southern Command, said officials were trying to determine whether there was a connection with the hostage situation on the southern Philippine island of Basilan, about 275 miles northeast of Semporna, the port from which tourists take boats for Sipadan.

(#meh):

D1030-A.M.100.F.1

And, more importantly, are they safe? The company introduced Cholestin at the end of 1996, after nearly three years of study. Current co-host Robin Hartl joined ``Hometime'' in July 1992. Obviously we are not breaking any laws. He pleaded guilty to using a telephone in the trafficking of drugs and was sentenced to five years' probation.



ROUGE results

	<u>D2</u>	<u>D3</u>
<i>ROUGE-1 Average_R</i>	0.10987	0.12951
<i>ROUGE-2 Average_R</i>	0.01891	0.02431
<i>ROUGE-3 Average_R</i>	0.00502	0.00612
<i>ROUGE-4 Average_R</i>	0.00129	0.00198



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>
- 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.



UNIVERSITY *of* WASHINGTON

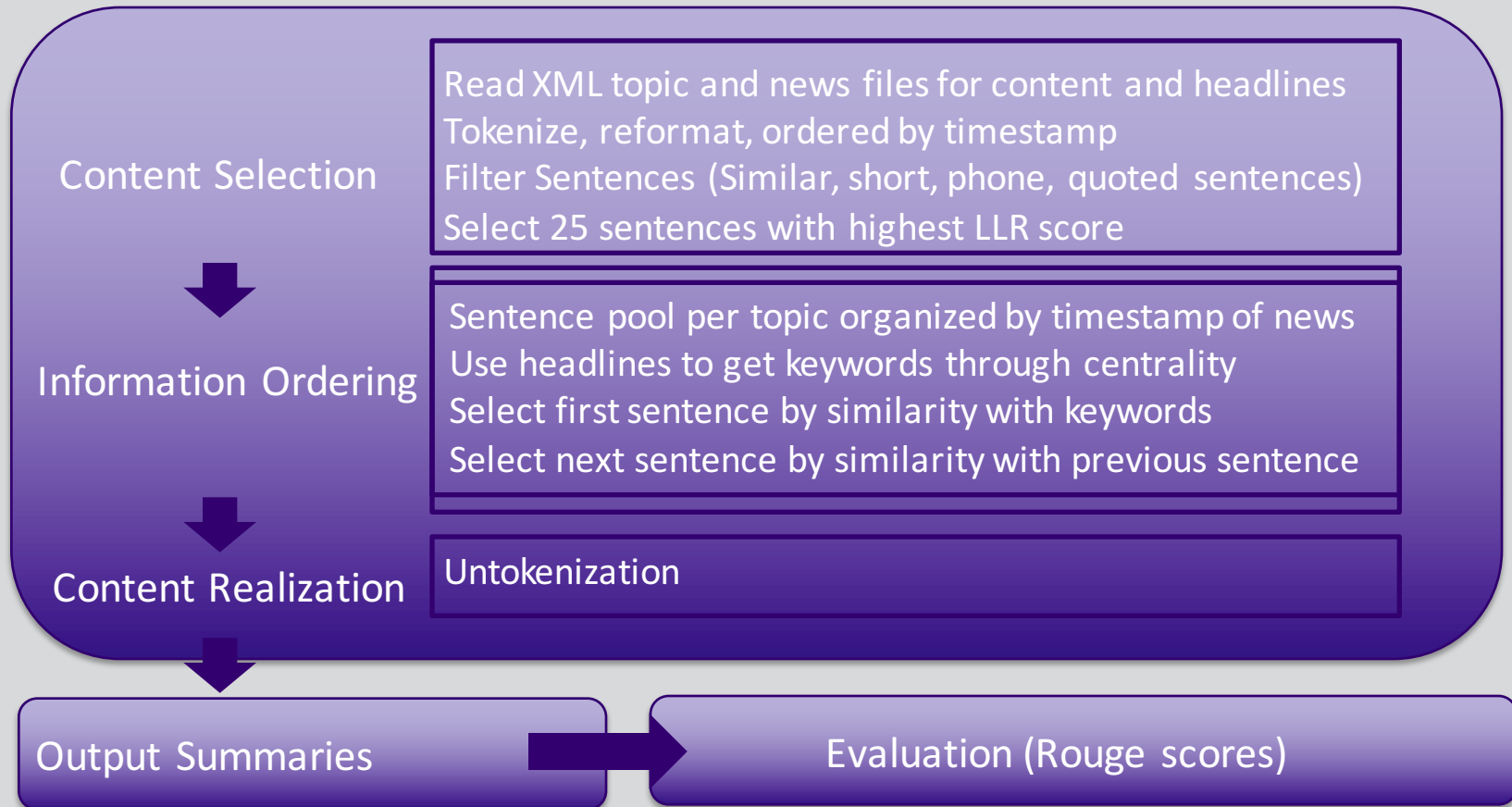
Summarization Task

Deliverable 3

LING 573 – Spring 2016

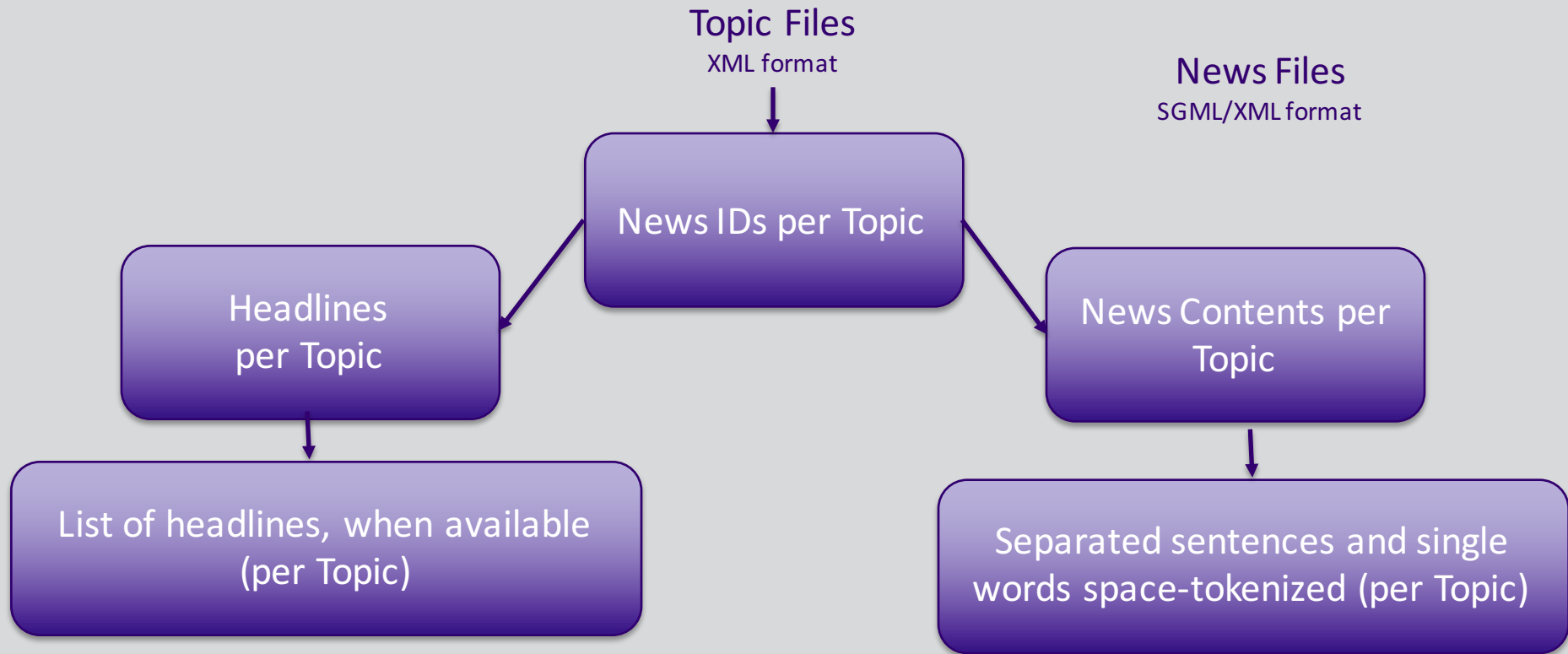


System Architecture – Diagram



W

Content Selection – Information Extraction



Content Selection - Log Likelihood Ratio

Obtain words tokens per sentence (excluding stopwords)

Used English Gigaword as background corpus

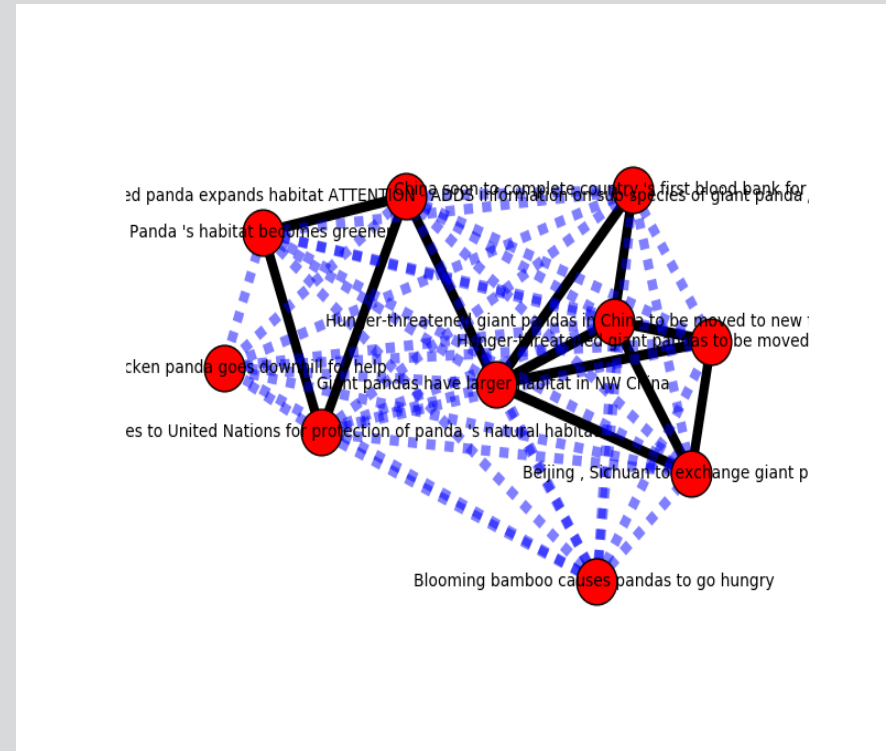
$$L(n, k, p) = p^k (1 - p)^{n-k} \rightarrow \text{weight}(w) = \begin{cases} 1, & \text{if } -2 \log \lambda > 10 \\ 0, & \text{otherwise} \end{cases}$$

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



Headline Content Selection: Method 1

- > Read headlines, remove stopwords, stem content words
- > Create a graph where sentences are nodes
- > Edges are equal to the cosine similarity between two sentences
- > Find graph center using degree centrality
- > Pass on headline at graph center

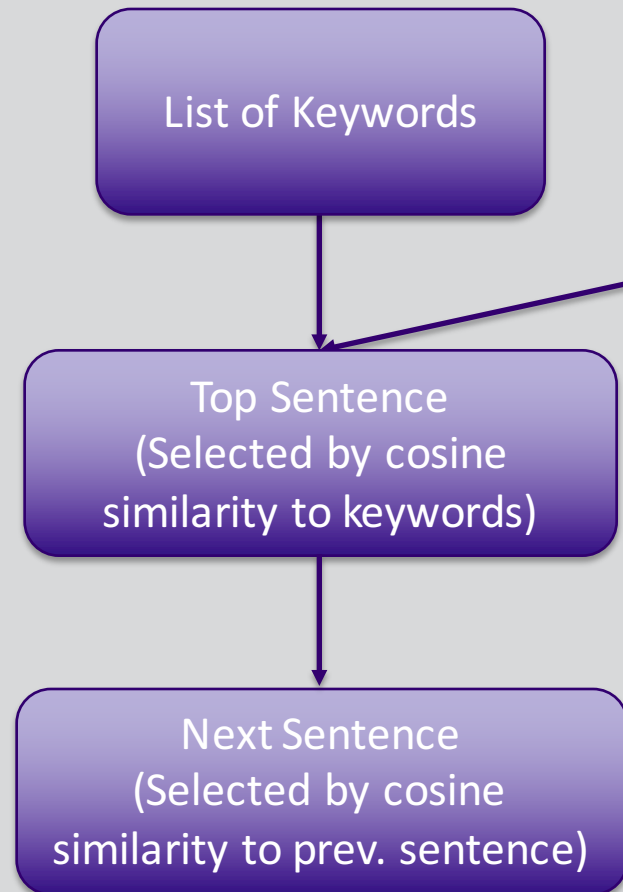


Headline Content Selection: Method 2

- > 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)
- > This method was more effective, and was used in the final version



Content Selection – Information Ordering



Next Sentence

- Must be >80 characters, <0.5 cosine similarity
- Prioritizes chronologically later sentences over earlier sentences
- At word limit, continues trying to add less and less similar sentences, case any fit

W

Evaluation of Results

Recall ROUGE scores for 1,2,3 and 4-ngrams.
Here average scores over all 46 summary topics:

	R-1	R-2	R-3	R-4
D2	0.18020	0.04338	0.01398	0.00575
D3	0.17798	0.04681	0.01615	0.00658



Output Analysis

D2 - Baseline

“Wolong is a famous giant panda habitat where the world-known China Conservation and Research Center of the Giant Panda is located.

"We have given priority to the afforestation in the habitats, especially nature reserves, for giant pandas," said Yang.

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.”

D3 – Improved System

“Wang said, Shaanxi has so far established 13 giant pandas protection zones and nature reserves focused on pandas' habitats.

In Sichuan and Shaanxi provinces, two other habitats of giant pandas, arrow bamboo was also found blooming.

However, a survey found that 12 percent of planted bamboo showed signs of blooming, rendering it inedible, China Daily said.

Zhang Kerong, director of the Baishuijiang State Nature Preserve, said that the preserve will intervene and help giant pandas find new food source.”

W

Output Analysis

D2 - Baseline

Fish and Wildlife Service is preparing to remove Yellowstone's grizzly bears from the endangered species list later this year.

Fish and Wildlife Service is poised to remove the park's renowned bears from the endangered species list.

"Once bears fill that habitat, the excess bears will probably wind up where they don't belong and are going to die.

Chris Servheen, grizzly bear recovery coordinator for the Fish and Wildlife Service, said he also supported taking bears off the list.

D3 – Improved System

If adopted, the proposal would apply only to bears outside Yellowstone and Grand Teton national parks.

It plotted, rather feebly, to assassinate local officials and then go to war with the National Guard and NATO.

But National Wildlife Federation senior wildlife biologist Sterling Miller, who spent 21 years studying grizzlies in Alaska, said the time has come to take Yellowstone's bears off the list.

Less than 20 species have been delisted since the law was signed by President Richard Nixon was signed in 1973.

W

Output Analysis

D2 - Baseline

Athens radio quoted eyewitnesses as saying that the plane was being accompanied by two Greek fighter jets when it went down.

After losing contact with the Athens airport's control tower, the Greek air force immediately sent two F-16 fighter jets to lead the airliner.

Greek radio and television stations reported that the air force pilots in the two fighter jets did not see any movement in the cockpit of the airliner before the crash and it was unclear if the two pilots were in their seats.

D3 – Improved System

European Union (EU) Transport Commissioner Jacques Barrot expressed his condolences to victims following a Cypriot airliner crashed into a hill in Greece earlier on Sunday.

In Greece, Prime Minister Costas Karamanlis returned to Athens from a holiday on the Aegean island of Tinos. In Cyprus, President Tassos Papadopoulos declared three days of official mourning. Gerard Feldzer, the head of France's Air and Space Museum, said many questions remained unanswered.

The Boeing 737-300 was due to fly onto Prague, Czech Republic, after stopping in Athens.



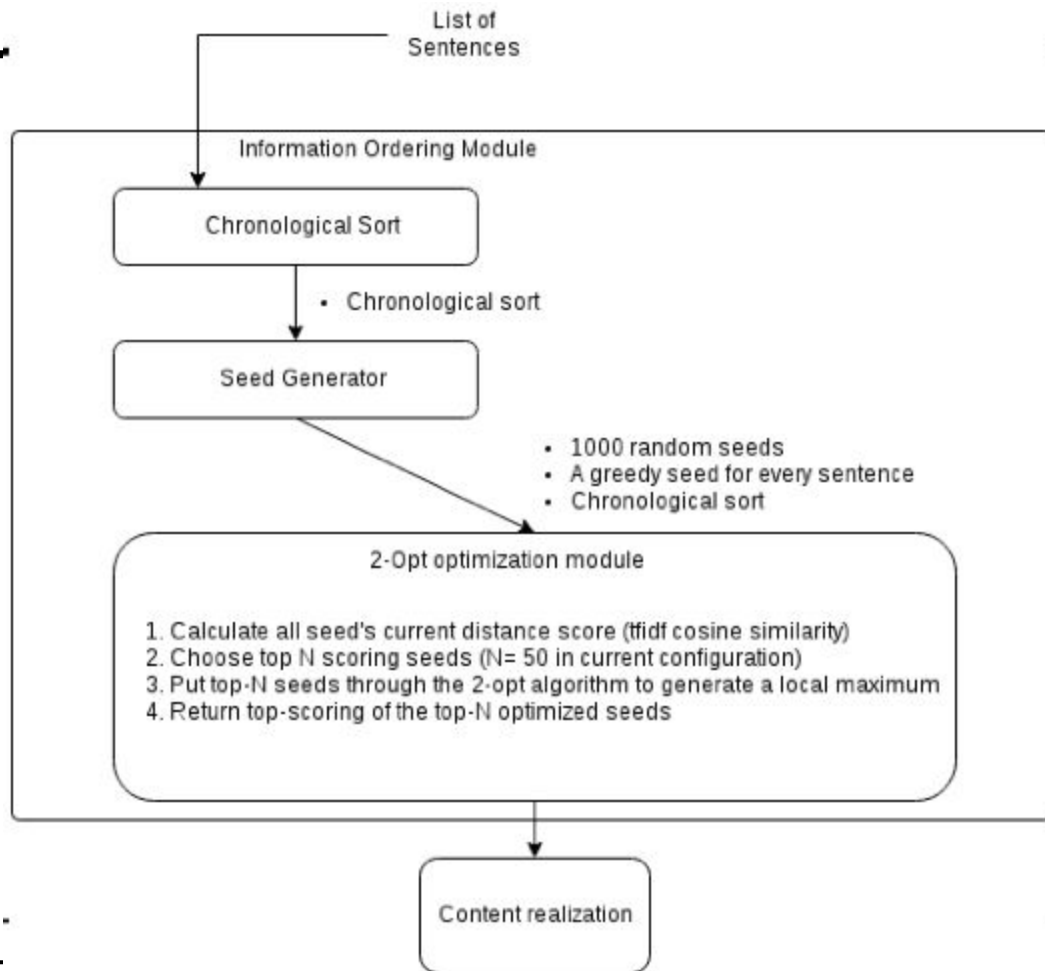
Ling 573 - Multi- document Summarization System : D3

Martin Horn, William Lane, Ryan Lish, Spencer Morris

Improvements in content selection

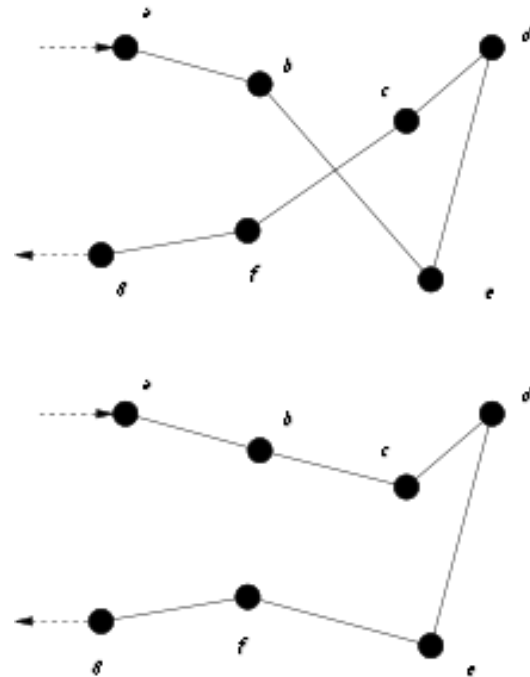
- Focus LexRank
 - Incorporated focus to previous centrality score
 - $\text{ratio} * \text{Focus_Score} + (1 - \text{ratio}) * \text{Centrality_Score}$
 - Sentence length and position
 - Total score: combo of focus LexRank, length score, position score
 - IDF using Gigaword
 - More regexes
 - Remove location headers, whitespace
-

Information Ordering



Information Ordering: 2-opt algorithm

- G. A. CROES (1958). *A method for solving traveling salesman problems*. Operations Res. 6 (1958) , pp., 791-812.
- Basically, you loop through the existing path and “uncross” paths, checking to see if you’ve made an improvement.
 - a. In our case, we are actually “crossing” as many paths as possible to get the maximum path, since cosine similarity is a similarity measure, not a distance measure



Results

ROUGE average recall scores for D2 and D3

	D2	D3
R-1	0.1717	0.2211
R-2	0.0402	0.0552
R-3	0.0115	0.0184
R-4	0.0031	0.0068

Successes

- Improved scores due to focused LexRank
 - Sentence length weighting allows more content per summary
 - ~5 sentences per summary instead of 2
 - Improved cohesion from information ordering
-

D1017-A.M.100.D.1

A hurricane warning was issued north to Brunswick late Monday afternoon and a watch was extended as far up the coast as Savannah.

“This hurricane,” said Jerry Jarrell, director of the National Hurricane Center, “could be catastrophic.”

“There are different tracks that have been projected and we don't come out unscathed on any of them,” Heller said.

Wherever Hurricane Floyd hits, the federal agency responsible for emergencies says it won't be as tardy as it was after Hurricane Andrew, seven years ago.

Hurricane Floyd got stronger and headed toward the Bahamas Saturday, packing 110 mph winds and leaving weather pundits wondering whether it will hit South Florida this week.

Issues

- Ordering could still be improved
 - Optimization limited by cosine similarity
 - Still some redundancy
 - Lack of co-reference resolution damages readability
-

D1002-A.M.100.A.1

It was the first time ever that a New York City police officer has been indicted for murder though a few have faced manslaughter charges.

The cheering began as the parents of Amadou Diallo left the prosecutor's office. They left the courtroom to the cheers of nearly 200 other officers, gathered outside the Bronx County Building.

Officers Kenneth Boss, Sean Carroll, Edward McMellon and Richard Murphy pleaded innocent in a Bronx courtroom to second-degree murder.

The arraignment and the subsequent comments by the officers' lawyers offered a glimpse of the defense they will provide.

D1028-A.M.100.E.1

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** 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** 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:

- G. A. CROES (1958). *A method for solving traveling salesman problems*. Operations Res. 6 (1958) , pp., 791-812.
 - John M. Conroy, Judith D. Schlesinger, Dianne P. O'Leary, Jade Goldstein. 2006. Back to Basics: CLASSY 2006.
-