

Text Summarization: News and Beyond

Kathleen McKeown
Department of Computer Science
Columbia University

Homework questions and information

- Data sets
 - Additional summary/document pairs added
 - Not homogeneous

Midterm Curve

- A: 72-93
 - A- 72-73
 - A+ 91-93
- B: 58-71
 - B- 58-59
 - B+ 70-71
- C: 47-57
 - C- 47-50
 - C+ 56-57
- D: 39-46

Questions (from Sparck Jones)

- Should we take the reader into account and how?
- “Similarly, the notion of a basic summary, i.e., one reflective of the source, makes hidden fact assumptions, for example that the subject knowledge of the output’s readers will be on a par with that of the readers for whom the source was intended. (p. 5)”
- Is the state of the art sufficiently mature to allow summarization from intermediate representations and still allow robust processing of domain independent material?

Text Summarization at Columbia

- Shallow analysis instead of information extraction
- Extraction of *phrases* rather than sentences
- Generation from surface representations in place of semantics

Problems with Sentence Extraction

- Extraneous phrases
 - “The five were apprehended along Interstate 95, *heading south in vehicles containing an array of gear including ...* ... authorities said.”
- Dangling noun phrases and pronouns
 - “The five”
- Misleading
 - Why would the media use this specific word (fundamentalists), so often with relation to Muslims?
*Most of them are radical Baptists, Lutheran and Presbyterian groups.

Cut and Paste in Professional Summarization

- Humans also reuse the input text to produce summaries
- But they “*cut and paste*” the input rather than simply extract
 - our automatic corpus analysis
 - 300 summaries, 1,642 sentences
 - 81% sentences were constructed by cutting and pasting
 - linguistic studies

Major Cut and Paste Operations

- (1) Sentence reduction



Major Cut and Paste Operations

- (1) Sentence reduction



Major Cut and Paste Operations

- (1) Sentence reduction



- (2) Sentence Combination

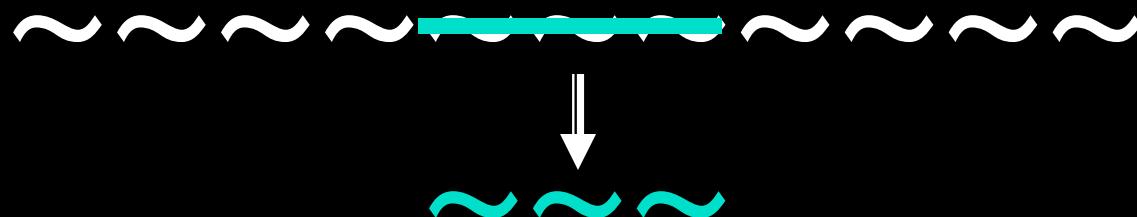


Major Cut and Paste Operations

- (3) Syntactic Transformation



- (4) Lexical paraphrasing



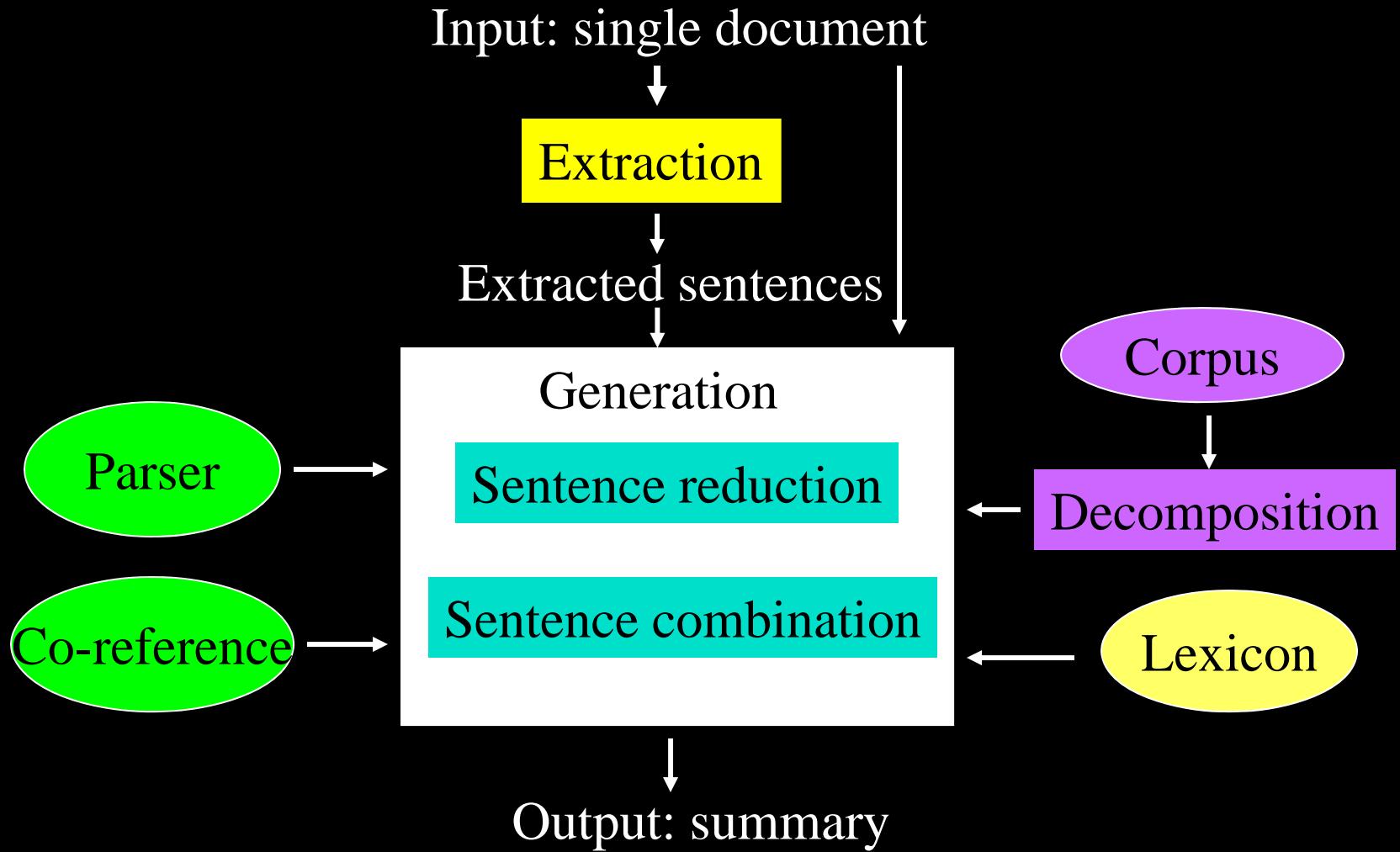
Summarization at Columbia

- News
- Email
- Meetings
- Journal articles
- Open-ended question-answering
 - What is a Loya Jurga?
 - Who is Mohammed Naeem Noor Khan?
 - What do people think of welfare reform?

Summarization at Columbia

- News
 - Single Document
 - Multi-document
- Email
- Meetings
 - Journal articles
 - Open-ended question-answering
 - What is a Loya Jurga?
 - Who is Al Sadr?
 - What do people think of welfare reform?

Cut and Paste Based Single Document Summarization -- System Architecture



(1) Decomposition of Human-written Summary Sentences

- Input:
 - a human-written summary sentence
 - the original document
- Decomposition analyzes how the summary sentence was constructed
- The need for decomposition
 - provide training and testing data for studying cut and paste operations

Sample Decomposition Output

Summary sentence:

Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc. and a member of the direct marketing association told the Communications Subcommittee of the Senate Commerce Committee that legislation to protect children's privacy on-line could destroy the spontaneous nature that makes the

Document sentences:

S1: A proposed new law that would require web publishers to obtain parental consent before collecting personal information from children could destroy the spontaneous nature that makes the internet unique, a member of the Direct Marketing Association told a Senate panel Thursday.

S2: Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc., said the association supported efforts to protect children on-line, but he...

S3: "For example, a child's e-mail address is necessary . . .," Sackler said in testimony to the Communications subcommittee of the Senate Commerce Committee.

S5: The subcommittee is considering the Children's Online Privacy Act, which was drafted...

A Sample Decomposition Output

Summary sentence:

Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc. and a member of the direct marketing association told the Communications Subcommittee of the Senate Commerce Committee that legislation to protect children's privacy on-line could destroy the spontaneous nature that makes the

Document sentences:

S1: A proposed new law that would require web publishers to obtain parental consent before collecting personal information from children could destroy the spontaneous nature that makes the internet unique, a member of the Direct Marketing Association told a Senate panel Thursday.

S2: Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc., said the association supported efforts to protect children on-line, but he...

S3: “For example, a child’s e-mail address is necessary ...,” Sackler said in testimony to the Communications subcommittee of the Senate Commerce Committee.

S5: The subcommittee is considering the Children’s Online Privacy Act, which was drafted...

A Sample Decomposition Output

Summary sentence:

Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc. and a member of the direct marketing association told the Communications Subcommittee of the Senate Commerce Committee that legislation to protect children's privacy on-line could destroy the spontaneous nature that makes the

Document sentences:

S1: A proposed new law that would require web publishers to obtain parental consent before collecting personal information from children could destroy the spontaneous nature that makes the internet unique, a member of the Direct Marketing Association told a Senate panel Thursday.

S2: Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc., said the association supported efforts to protect children on-line, but he...

S3: “For example, a child’s e-mail address is necessary ...,” Sackler said in testimony to the Communications subcommittee of the Senate Commerce Committee.

S5: The subcommittee is considering the Children’s Online Privacy Act, which was drafted...

A Sample Decomposition Output

Summary sentence:

Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc. and a member of the direct marketing association told the Communications Subcommittee of the Senate Commerce Committee that legislation to protect children's privacy on-line could destroy the spontaneous nature that makes the

Document sentences:

S1: A proposed new law that would require web publishers to obtain parental consent before collecting personal information from children could destroy the spontaneous nature that makes the internet unique, a member of the Direct Marketing Association told a Senate panel Thursday.

S2: Arthur B. Sackler, vice president for law and public policy of Time Warner Cable Inc., said the association supported efforts to protect children on-line, but he...

S3: “For example, a child’s e-mail address is necessary ...,” Sackler said in testimony to the Communications subcommittee of the Senate Commerce Committee.

S5: The subcommittee is considering the Children’s Online Privacy Act, which was drafted...

The Algorithm for Decomposition

- A Hidden Markov Model based solution
- Evaluations:
 - Human judgements
 - 50 summaries, 305 sentences
 - 93.8% of the sentences were decomposed correctly
 - Summary sentence alignment
 - Tested in a legal domain
- Details in (Jing&McKeown-SIGIR99)

(2) Sentence Reduction

- An example:

Original Sentence: When it arrives sometime next year in new TV sets, the V-chip will give parents a new and potentially revolutionary device to block out programs they don't want their children to see.

Reduction Program: The V-chip will give parents a new and potentially revolutionary device to block out programs they don't want their children to see.

Professional: The V-chip will give parents a device to block out programs they don't want their children to see.

The Algorithm for Sentence Reduction

- Preprocess: syntactic parsing
- Step 1: Use linguistic knowledge to decide what phrases MUST NOT be removed
- Step 2: Determine what phrases are most important in the local context
- Step 3: Compute the probabilities of humans removing a certain type of phrase
- Step 4: Make the final decision

Step 1: Use linguistic knowledge to decide what MUST NOT be removed

- Syntactic knowledge from a large-scale, reusable lexicon we have constructed
 - convince:
 - meaning 1: NP-PP :PVAL (“of”)
(E.g., “He convinced me of his innocence”)
 - NP-TO-INF-OC
(E.g., “He convinced me to go to the party”)
 - meaning 2: ...
- Required syntactic arguments are not removed

Step 2: Determining context importance based on lexical links

- Saudi Arabia on Tuesday decided to sign...
- **The official Saudi Press Agency reported that King Fahd made the decision during a cabinet meeting in Riyadh, the Saudi capital.**
- The meeting was called in response to ... the Saudi foreign minister, that the Kingdom...
- An account of the Cabinet discussions and decisions at the meeting...
- The agency...

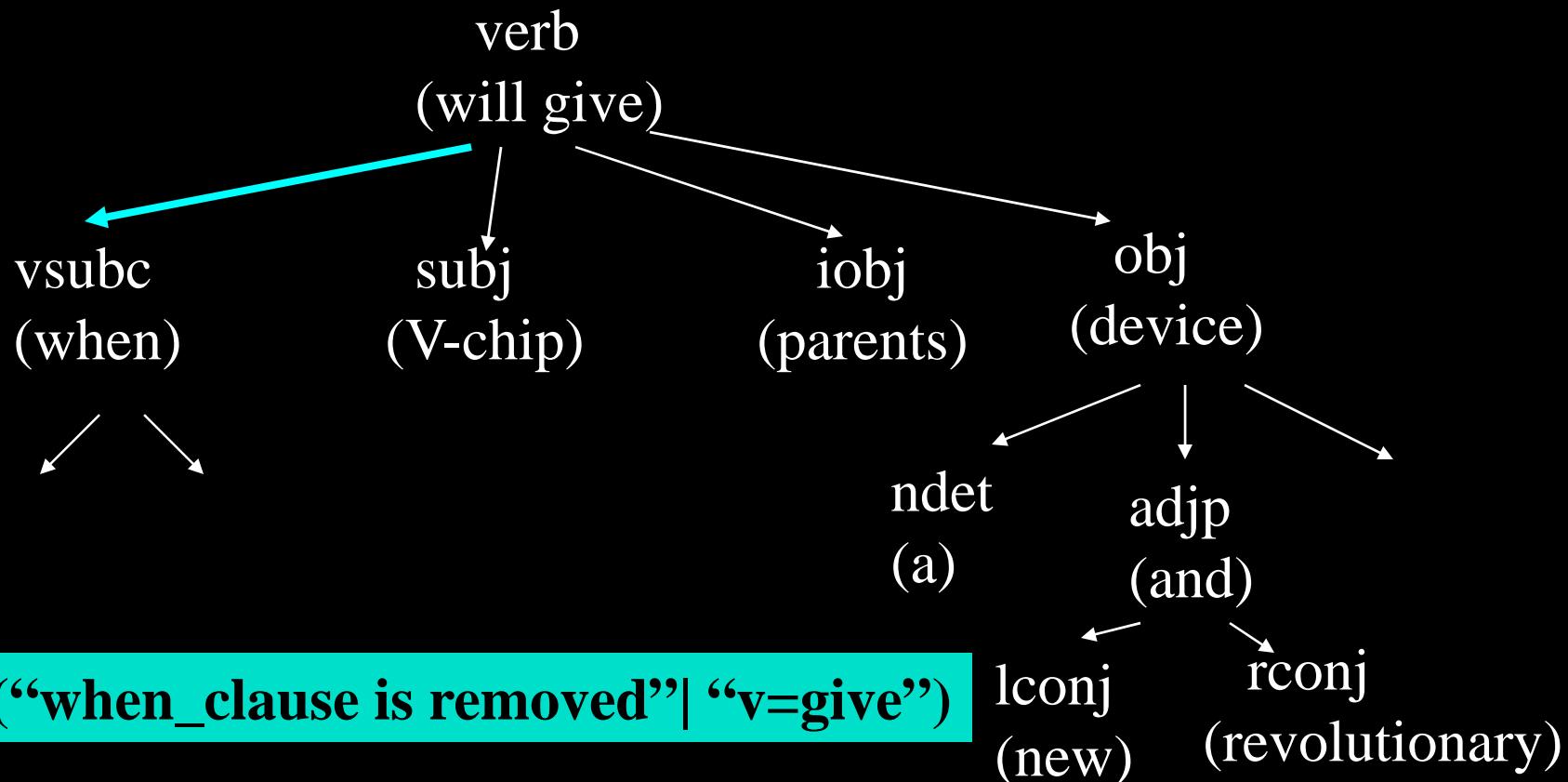
Step 2: Determining context importance based on lexical links

- Saudi Arabia on Tuesday decided to sign...
- **The official Saudi Press Agency reported that King Fahd made the decision during a cabinet meeting in Riyadh, the Saudi capital.**
- The meeting was called in response to ... the Saudi foreign minister, that the Kingdom...
- An account of the Cabinet discussions and decisions at the meeting...
- The agency...

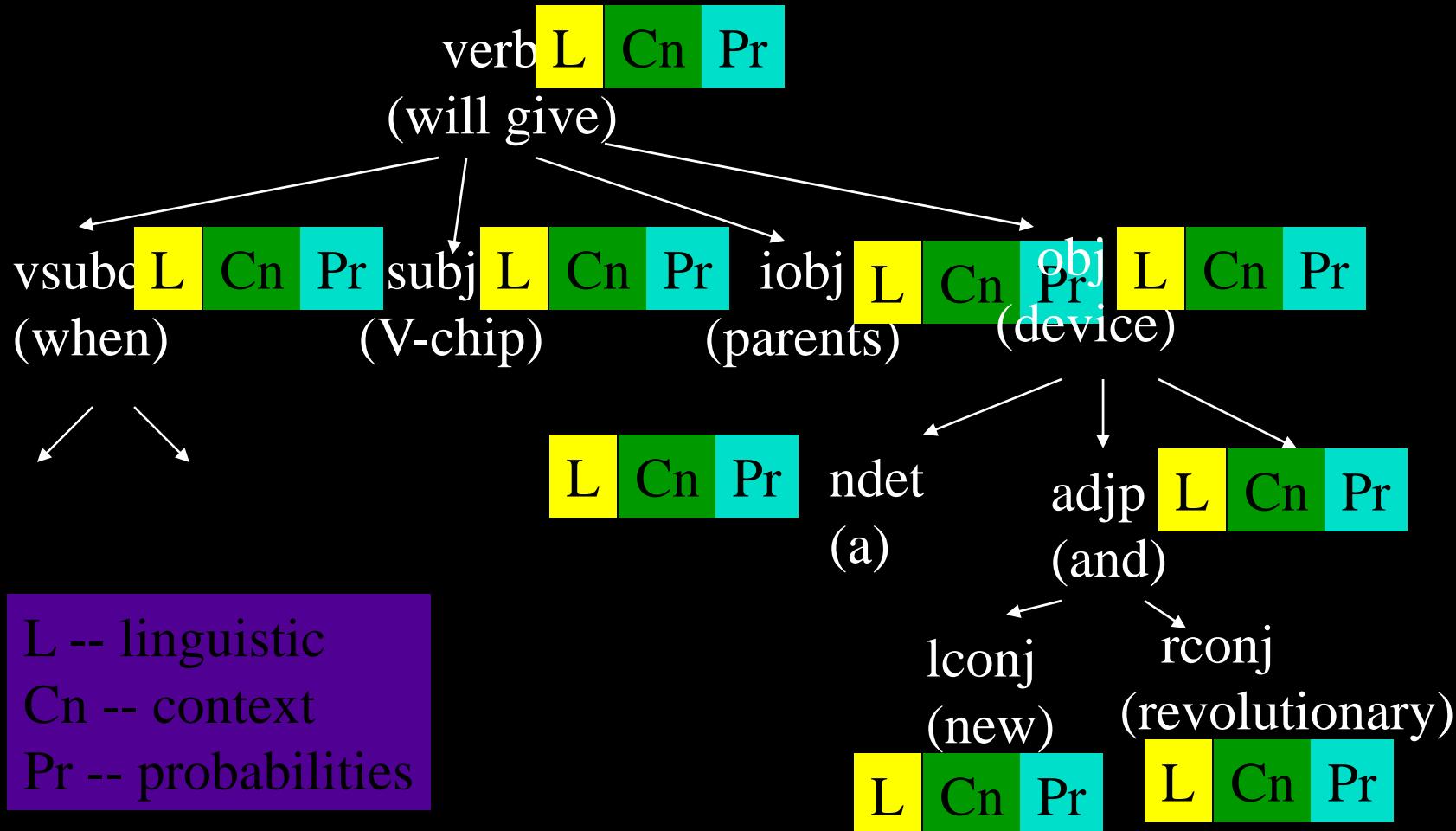
Step 2: Determining context importance based on lexical links

- Saudi Arabia on Tuesday decided to sign...
- **The official Saudi Press Agency reported that King Fahd made the decision during a cabinet meeting in Riyadh, the Saudi capital.**
 - The meeting was called in response to ... the Saudi foreign minister, that the Kingdom...
 - An account of the Cabinet discussions and decisions at the meeting...
 - The agency...

Step 3: Compute probabilities of humans removing a phrase



Step 4: Make the final decision



Evaluation of Reduction

- Success rate: 81.3%
 - 500 sentences reduced by humans
 - Baseline: 43.2% (remove all the clauses, prepositional phrases, to-infinitives,...)
- Reduction rate: 32.7%
 - Professionals: 41.8%
- Details in (Jing-ANLP00)

Multi-Document Summarization

Research Focus

- Monitor variety of online information sources
 - News, multilingual
 - Email
- Gather information on events across source and time
 - Same day, multiple sources
 - Across time
- Summarize
 - Highlighting similarities, new information, different perspectives, user specified interests in real-time

Our Approach

- Use a hybrid of statistical and linguistic knowledge
- Statistical analysis of multiple documents
 - Identify important new, contradictory information
- Information fusion and rule-driven content selection
- Generation of summary sentences
 - By re-using phrases
 - Automatic editing/rewriting summary

Newsblaster

Integrated in online environment for daily news updates

<http://newsblaster.cs.columbia.edu/>



Ani Nenkova



David Elson

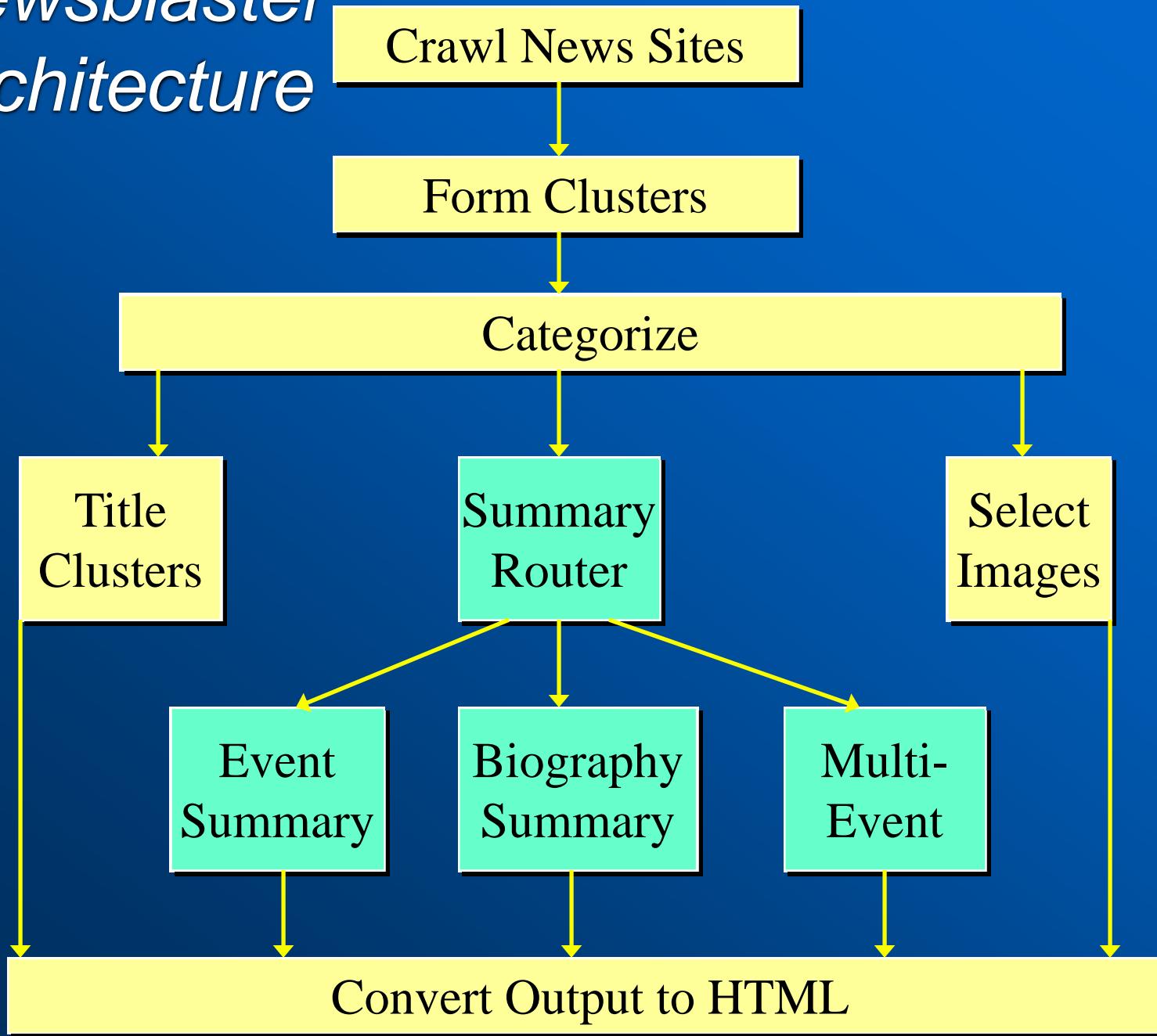
Newsblaster

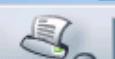
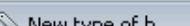
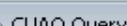
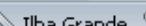
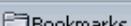
The screenshot shows a web browser window for the Columbia Newsblaster website. The URL in the address bar is <http://newsblaster.cs.columbia.edu/memoaries/-061.html>. The page title is "Columbia Newsblaster" and the subtitle is "Summarizing all the news on the Web". The date is Wednesday, January 29, 2003, and the last update was at 3:31 PM EST. On the left, there's a sidebar with links for U.S., World, Finance, Sci/Tech, Entertainment, Sports, View Today's Images, View Archives, Newsblaster.In Press, Academic Papers, and Old Interface. The main content area has a heading "Angry at Accord, Mobs Attack French Embassy in Ivory Coast". Below it is a large block of text about the situation in Ivory Coast, mentioning France's reinforcement of security measures and rebel attacks. There are also two small images: one of a crowd and another of a man's face.

<http://newsblaster.cs.columbia.edu/>

- Clustering articles into events
- Categorization by broad topic
- Multi-document summarization
- Generation of summary sentences
 - Fusion
 - Editing of references

Newsblaster Architecture





Newsblaster Archived Run

[Click here](#) to return to today's news.

Friday, November 14, 2003

Articles from 11/10/2003 to 11/13/2003

Last update: 12:19 AM EST

Search for:

Go

in summaries

[U.S.](#)
[World](#)
[Finance](#)
[Sci/Tech](#)
[Entertainment](#)
[Sports](#)

[View Today's Images](#)

[Back to Archive Index](#)

[About Newsblaster](#)

[About today's run](#)

[Newsblaster in Press](#)

[Academic Papers](#)



UK 'ready to send more troops to Iraq' (World, 23 articles)

A military spokesman said U.S. forces attacked three sites across the city, including a building used by insurgents on Wednesday to attack American soldiers with rockets. A suicide truck bomb exploded outside an Italian military police base here Wednesday, tearing off the facade of the three-story building and killing at least 26 people, including 12 Italian military police and a 10-day-old Iraqi baby.. U.S. troops mounted air and ground attacks in the Iraqi capital Thursday for a second straight night, targeting suspected insurgent positions around Baghdad, the U.S. command said.. The suicide bombing was the deadliest attack against the coalition since the occupation in Iraq began an insurgency that the top American general said numbers no more than 5,000 fighters.. General John Abizaid, head of the U.S. Central Command based in Tampa, Fla., said the fighters battling forces of the U.S.-led coalition number no more than 5,000 and appear to be organized at regional and local levels.. Soldiers arrested 18 people in connection with a deadly missile barrage last month that Deputy Defense Secretary Paul Wolfowitz narrowly escaped, officials said yesterday, as U.S. warplanes dropped bombs near the center of Iraqi resistance..

Other stories about Iraq, Iraqi and Baghdad:

- [U.S. allies rethinking role in post-war Iraq](#) (7 articles)
- [Handing over the keys in Iraq](#) (10 articles)

Top News

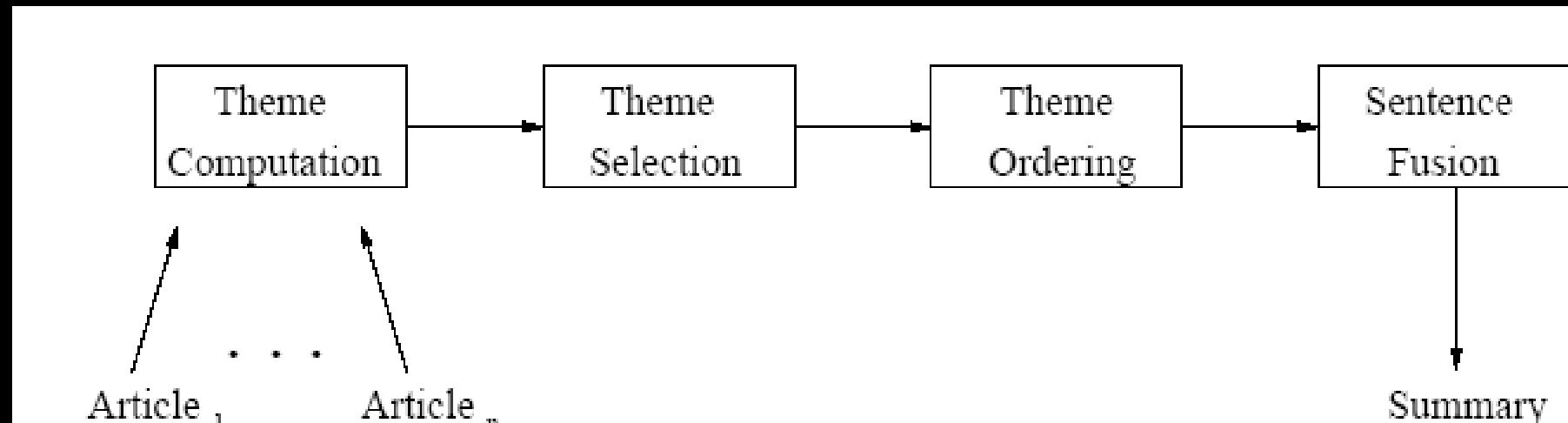
Tembec loses \$51.5 million in fourth quarter; lumber joint venture lagging (Finance, 8 articles)

Canadian Tire Corp. increased its third-quarter profit by 13.9 per cent as sales rose for everything from garden tools to car

New Palestinian Cabinet approved; PM pledges to end 'chaos' (World, 10 articles)

The Israeli and Palestinian prime ministers are expected to meet within 10 days in an effort to restart the peace process

Fusion



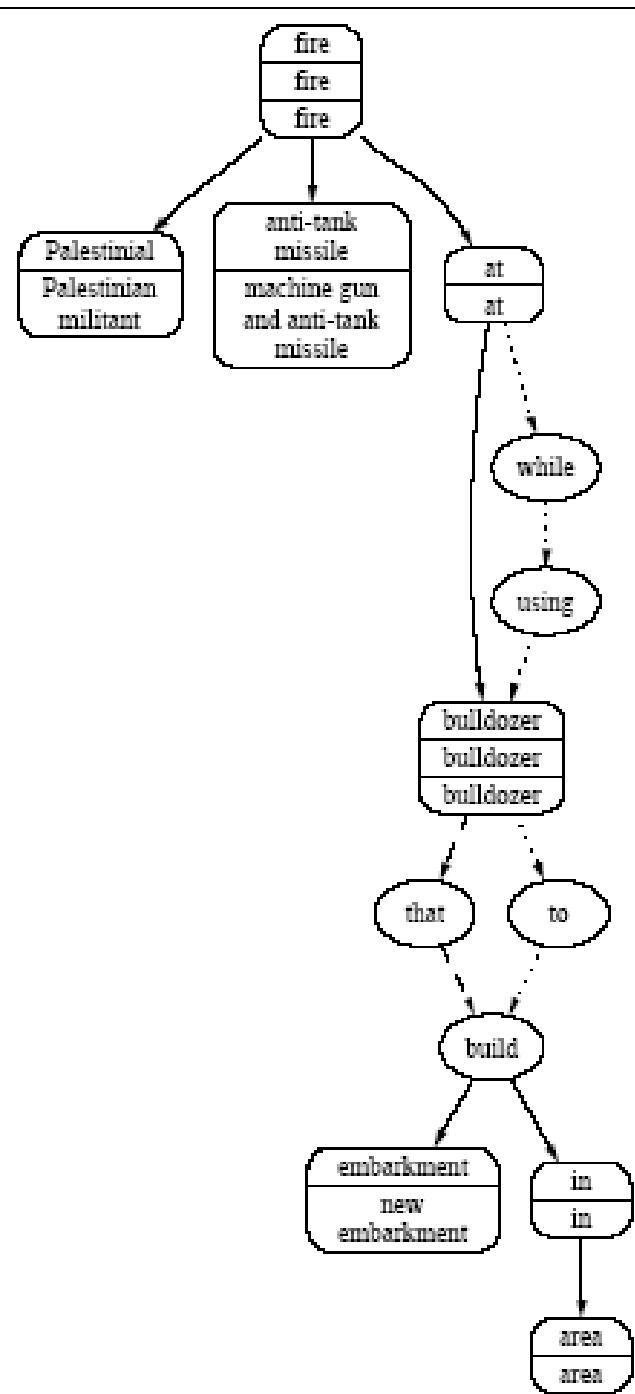
- | |
|--|
| 1. IDF Spokeswoman did not confirm this, but said the Palestinians fired an anti-tank missile at a bulldozer. |
| 2. The clash erupted when Palestinian militants fired machine-guns and anti-tank missiles at a bulldozer that was building an embankment in the area to better protect Israeli forces. |
| 3. The army expressed “regret at the loss of innocent lives” but a senior commander said troops had shot in self-defense after being fired at while using bulldozers to build a new embankment at an army base in the area.
fusion sentence: Palestinians fired an anti-tank missile at a bulldozer. |

Theme Computation

- Input: A set of related documents
- Output: Sets of sentences that “mean” the same thing
- Algorithm
 - Compute similarity across sentences using the **Cosine Metric**
 - Can compare word overlap or phrase overlap
 - (PLACEHOLDER: IR vector space model)

Sentence Fusion Computation

- Common information identification
 - Alignment of constituents in parsed theme sentences: *only some subtrees match*
 - Bottom-up local multi-sequence alignment
 - Similarity depends on
 - ◆ Word/paraphrase similarity
 - ◆ Tree structure similarity
- Fusion lattice computation
 - Choose a basis sentence
 - Add subtrees from fusion not present in basis
 - Add alternative verbalizations
 - Remove subtrees from basis not present in fusion
- Lattice linearization
 - Generate all possible sentences from the fusion lattice
 - Score sentences using statistical language model





Newsbaster Archived Run

[Click here to return to today's news.](#)

Thursday, June 24, 2004

Articles from 06/21/2004 to 06/24/2004

Last update: 9:48 AM EST

Search for:

in summaries

[U.S.](#)
[World](#)
[Finance](#)
[Entertainment](#)
[Sports](#)

[View Today's Images](#)

[Back to Archive Index](#)

[About Newsblaster](#)

[About today's run](#)

[Newsblaster in Press](#)

[Academic Papers](#)

Mattie Stepanek: Child poet battled muscular dystrophy

Summary from multiple countries, from articles in English

Mattie Stepanek the child poet whose inspirational verse made him best ([article 3](#)) selling writer and an advocate for muscular dystrophy research ([article 4](#)) died yesterday from complications of the disease. ([article 3](#)) His mother has ([article 3](#)) a milder adult onset form of the disease ([article 6](#)) and his three older siblings died of it in early childhood. ([article 3](#)) Within weeks the book reached the top of the New York Times best seller list the Arizona based Mda said. ([article 4](#)) Mattie began ([article 3](#)) writing poetry at age 3 partly as salve for his grief over brother's death from the same disease. ([article 7](#))



Other summaries about this story:

- [Summary from United States, from articles in English](#) (6 articles) [[compare](#)]
- [Summary from the United Kingdom, from articles in English](#) (1 articles) [[compare](#)]

Event tracking:

- [Track this story's development in time](#)

Story keywords

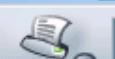
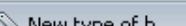
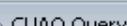
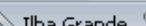
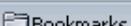
Mattie, Heartsongs, Stepanek, Dystrophy, Muscular

[Search articles](#)

Internet

Tracking Across Days

- Users want to follow a story across time and watch it unfold
- Network model for connecting clusters across days
 - Separately cluster events from today's news
 - Connect new clusters with yesterday's news
 - Allows for forking and merging of stories
- Interface for viewing connections
- Summaries that update a user on what's new
 - Statistical metrics to identify differences between article pairs
 - Uses learned model of features
 - Identifies differences at clause and paragraph levels



Newsblaster Archived Run

[Click here](#) to return to today's news.

Friday, November 14, 2003

Articles from 11/10/2003 to 11/13/2003

Last update: 12:19 AM EST

Search for:

Go

in summaries

[U.S.](#)
[World](#)
[Finance](#)
[Sci/Tech](#)
[Entertainment](#)
[Sports](#)

[View Today's Images](#)

[Back to Archive Index](#)

[About Newsblaster](#)

[About today's run](#)

[Newsblaster in Press](#)

[Academic Papers](#)



UK 'ready to send more troops to Iraq' (World, 23 articles)

A military spokesman said U.S. forces attacked three sites across the city, including a building used by insurgents on Wednesday to attack American soldiers with rockets. A suicide truck bomb exploded outside an Italian military police base here Wednesday, tearing off the facade of the three-story building and killing at least 26 people, including 12 Italian military police and a 10-day-old Iraqi baby.. U.S. troops mounted air and ground attacks in the Iraqi capital Thursday for a second straight night, targeting suspected insurgent positions around Baghdad, the U.S. command said.. The suicide bombing was the deadliest attack against the coalition since the occupation in Iraq began an insurgency that the top American general said numbers no more than 5,000 fighters.. General John Abizaid, head of the U.S. Central Command based in Tampa, Fla., said the fighters battling forces of the U.S.-led coalition number no more than 5,000 and appear to be organized at regional and local levels.. Soldiers arrested 18 people in connection with a deadly missile barrage last month that Deputy Defense Secretary Paul Wolfowitz narrowly escaped, officials said yesterday, as U.S. warplanes dropped bombs near the center of Iraqi resistance..

Other stories about Iraq, Iraqi and Baghdad:

- [U.S. allies rethinking role in post-war Iraq](#) (7 articles)
- [Handing over the keys in Iraq](#) (10 articles)

Top News

Tembec loses \$51.5 million in fourth quarter; lumber joint venture lagging (Finance, 8 articles)

Canadian Tire Corp. increased its third-quarter profit by 13.9 per cent as sales rose for everything from garden tools to car

New Palestinian Cabinet approved; PM pledges to end 'chaos' (World, 10 articles)

The Israeli and Palestinian prime ministers are expected to meet within 10 days in an effort to restart the peace process



Newsblaster Archived Run

[Click here](#) to return to today's news.

Friday, November 14, 2003

Articles from 11/10/2003 to 11/13/2003

Last update: 12:19 AM EST

Search for:

Go

in summaries

[U.S.](#)
[World](#)
[Finance](#)
[Sci/Tech](#)
[Entertainment](#)
[Sports](#)

[View Today's
Images](#)

U.S. allies rethinking role in post-war Iraq

Summary from multiple countries, from articles in English

[See Summary with New Information Only](#)

French papers report on the trial of 37 people accused of embezzling money from former state oil giant Elf, while a nationwide rail strike preoccupies the Austrian press.. Il Giorno reports "Italians massacred in Nasiriya", while L'Unita speaks of "the most serious attack since World War II" on Italian armed forces, which it says were "mired in an Iraqi inferno".. Germany's Der Tagesspiegel hopes Prime Minister Silvio Berlusconi will resist pressure to pull the Italian troops out.. A further 20 Italians and 59 Iraqis were wounded in the attack in which a suicide squad driving a petrol tanker rammed Italian police headquarters in Nasiriya, a Shia Muslim town in the south of the country.. That idyllic picture came crashing down yesterday, when insurgents blew up the Italian military headquarters in Nasiriya, killing at least 18 Italians and eight others and



Tracking this event across days (click [here](#) to return)

11/12/2003

[Bremer Will Return to Iraq to Plan Accelerated Transition](#) [\[recenter\]](#)

11/13/2003

|||||

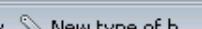
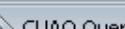
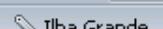
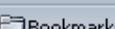
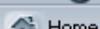
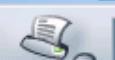
[Italy debates pulling out of Iraq](#)
[\[recenter\]](#)

11/14/2003

[U.S. allies rethinking role in post-war Iraq](#)

[Truck bomb strikes Italian base in deadliest attack against U.S. ally in Iraq](#) [\[recenter\]](#)

|||||



Update Summary

Italian officials confirmed on Wednesday evening the deaths of 16 Italian military and police personnel , two Italian civilians and nine Iraqis.

Some politicians asked for Italy 's contingent to be immediately pulled out , and recalled that the opposition , as well as the vast majority of Italians , had always opposed the deployment of troops in post-war Iraq.

That idyllic picture came crashing down yesterday , when insurgents blew up the Italian military headquarters in Nasiriyah , killing at least 18 Italians and eight others and plunging the European country into shocked debate.

Although the carabinieri police and the army 's crack Bersaglieri infantrymen , distinctive in helmets topped by plumes of black feathers , have frequently participated in peacekeeping missions , it has always been under the clear guise of a United Nations , European Union or NATO mandate.

After the bombing , Italy 's Interior Ministry placed the nation on special alert , fearing an attack at home.

In the past couple of years , a number of people alleged to have ties to al-Qaida have been arrested in Italy.

Although the Carabinieri police and the army 's crack Bersaglieri infantrymen , distinctive in helmets topped by plumes of black feathers , have frequently participated in peacekeeping missions , it has always been under the clear guise of a United Nations , European Union or NATO mandate.

Tracking this event across days (click [here](#) to return)

11/12/2003

[Bremer Will Return to Iraq to Plan Accelerated Transition](#) [\[recenter\]](#)

11/13/2003

|||||

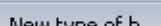
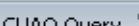
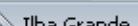
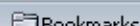
[Italy debates pulling out of Iraq](#)
[\[recenter\]](#)

11/14/2003

[U.S. allies rethinking role in post-war Iraq](#)

[Truck bomb strikes Italian base in deadliest attack against U.S. ally in Iraq](#) [\[recenter\]](#)

|||||



Home

Bookmarks

Ilha Grande

Model

CUAQ Query

New type of b...

Newsblaster Archived Run

[Click here](#) to return to today's news.

Thursday, November 13, 2003

Articles from 11/10/2003 to 11/13/2003

Last update: 5:02 PM EST

Search for:

Go

in summaries

[U.S.](#)
[World](#)
[Finance](#)
[Sci/Tech](#)
[Sports](#)

[View Today's
Images](#)

Italy debates pulling out of Iraq

Summary from multiple countries, from articles in multiple languages

[See Summary with New Information Only](#)

U.S. military forces have launched strikes aimed at cracking down on Iraqi insurgents following Wednesday's bombing that killed 26 people in Nasiriya and continuing guerrilla attacks against coalition forces.. U.S. troops attacked targets in western and southern Baghdad overnight in Operation Iron Hammer, said Capt. David Gercken, a spokesman for the U.S. Army's 1st Armored Division.. A suicide truck bomb exploded outside an Italian military police base here Wednesday, tearing off the facade of the three-story building and killing at least 26 people, including 12 Italian military police and a 10-day-old Iraqi baby.. The bombing, which injured at least 100 people, was the deadliest attack to date on international troops who have come to Iraq

at the request of President George W. Bush to aid in the U.S.-led reconstruction efforts. Japan has



Tracking this event across days (click [here](#) to return)

10/20/2003

[Chaos reigns as Saddam's plan unfolds](#) [\[recenter\]](#)

11/12/2003

[Bremer Will Return to Iraq to Plan Accelerated Transition](#) [\[recenter\]](#)

11/13/2003

[Italy debates pulling out of Iraq](#)

11/14/2003

[U.S. allies rethinking role in post-war Iraq](#) [\[recenter\]](#)

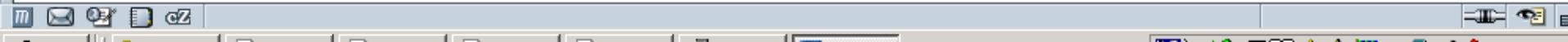
[Chaos reigns as Saddam's plan unfolds](#) [\[recenter\]](#)

[Truck bomb strikes Italian base in deadliest attack against U.S. ally in Iraq](#) [\[recenter\]](#)

|||||

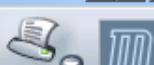
[UK 'ready to send more troops to Iraq'](#) [\[recenter\]](#)

|||||



Different Perspectives

- Hierarchical clustering
 - Each event cluster is divided into clusters by country
- Different perspectives can be viewed side by side
- Experimenting with update summarizer to identify key differences between sets of stories



Columbia Newsblaster

Summarizing all the news on the Web

Wednesday, November 12, 2003

Articles from 0/0/0000 to 11/10/2003

Last update: 12:35 AM EST

Search for:

Go

in summaries

U.S.
World
Finance
Sci/Tech
Entertainment
Sports

[View Today's Images](#)

[View Archive](#)

[About Newsblaster](#)

[About today's run](#)

[Newsblaster in Press](#)

[Academic Papers](#)

[Article Sources:](#)

U.S. pledges to help Saudi war on terror after weekend attacks

Summary from multiple countries, from articles in multiple languages

Saudi security officials are deploying thousands of troops to the city of Mecca because of concerns that terrorists may be planning new attacks during the Muslim holy month of Ramadan, Saudi government sources told CNN.. U.S. to close embassy in Sudan, officials say The attack occurred a day after the United States said it was shutting its embassy and consulates in Saudi Arabia, citing intelligence of an imminent terrorist attack.. In May, 35 people were killed in suicide attacks on a Western compound in Riyadh, and analysts believe the latest attacks bear the hallmarks an al-Qaeda operation.. The officials said the attack started with two to three gunmen standing high atop the khaki desert cliffs facing the gated complex and raining bullets on the guards.. Three explosions rocked a residential compound in the Saudi capital last night, killing at least two people and wounding 86, in what a government official said was a suicide car bombing.. The attacks " were very similar in nature to the East African bombings one U.S. official said, referring to the 1998 bombings of the U.S. Embassies in Kenya and Tanzania that killed 231 people, including 12 Americans..



TOP500 neu
 Die größten deutschen Unternehmen
 Jetzt auch als Download!



Other summaries about this story:

- [Summary from the United Kingdom, from articles in English](#) (15 articles) [[compare](#)]
- [Summary from Canada, from articles in English](#) (4 articles) [[compare](#)]
- [Summary from multiple countries, from articles in English](#) (38 articles) [[compare](#)]
- [Summary from Germany, from articles in German](#) (5 articles) [[compare](#)]
- [Summary from Spain, from articles in Spanish](#) (1 articles) [[compare](#)]



Columbia Newsblaster

Summarizing all the news on the Web

Wednesday, November 12, 2003

Articles from 0/0/0000 to 11/10/2003

Last update: 12:35 AM EST

Search for:

Go

in summaries

[U.S.](#)
[World](#)
[Finance](#)
[Sci/Tech](#)
[Entertainment](#)
[Sports](#)

[View Today's Images](#)[View Archive](#)[About Newsblaster](#)[About today's run](#)[Newsblaster in Press](#)[Academic Papers](#)[Article Sources:](#)

U.S. pledges to help Saudi war on terror after weekend attacks

[Summary from multiple countries, from articles in multiple languages](#)

Saudi security officials are deploying thousands of troops to the city of Mecca because of concerns that terrorists may be planning new attacks during the Muslim holy month of Ramadan, Saudi government sources told CNN.. U.S. to close embassy in Sudan, officials say The attack occurred a day after the United States said it was shutting its embassy and consulates in Saudi Arabia, citing intelligence of an imminent terrorist attack.. In May, 35 people were killed in suicide attacks on a Western compound in Riyadh, and analysts believe the latest attacks bear the hallmarks an al-Qaeda operation.. The officials said the attack started with two to three gunmen standing high atop the khaki desert cliffs facing the gated complex and raining bullets on the guards.. Three explosions rocked a residential compound in the Saudi capital last night, killing at least two people and wounding 86, in what a government official said was a suicide car bombing.. The attacks " were very similar in nature to the East African bombings one U.S. official said, referring to the 1998 bombings of the U.S. Embassies in Kenya and Tanzania that killed 231 people including 12

[Summary from Germany, from articles in German](#)

After the devastating notice in Riyadh the US government with Saudi Arabia in the fight against the international terror wants to co-operate more strongly.. The authorities make the terrorist organization El Kaida responsible for the notice for Riyadh/Cairo - after the devastating bomb attack on a foreigner housing development in Riyadh the number of the victims increased to 17.. How the Saudi Arabian press agency SPA in the Sunday evening reported, in the rubble of the completely destroyed block of flats 6 further corpses were discovered.. Saudi Arabia has likewise the network of Osama is made shop for the notice of Sunday on a housing estate of foreigners responsible.. The ruler family explained, a goal of the teuflischen terrorists was the destabilization of the kingdom 17 humans had died.. Authorities prepare safety precautions in the diplomat quarter of Riyadh on after the blood bath again strengthened the authorities according to data of eye-witnesses.. With a devastating suicide

Home Bookmarks Ilha Grande Model CUAQ Query New type of b...

AKTUELLES

Montag, 17. November 2003 Berlin, 04:19 Uhr

DIE WELT

suche

BILDER DES TAGES

 ...es weihnachtet bald

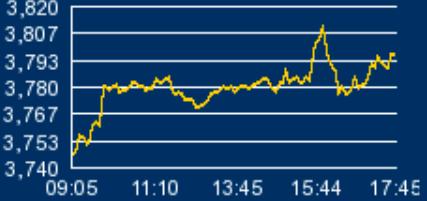
news TICKER Themen heute

- 04:03 Paris geht Machtübergabe im Irak zu langsam
- 03:58 Nach «Queen Mary 2»-Drama Ermittlungen unter Hochdruck
- 03:54 Regierungsbildung in Katalonien offen
- 03:47 Kranker Luther Vandross Doppelgewinner bei American Music Awards
- 03:45 SPD-Parteidag in Bochum beginnt

→ weitere aktuelle Meldungen

BREAKING NEWS per SMS

dax INTRADAY



3,820
3,807
3,793
3,780
3,767
3,753
3,740
09:05 11:10 13:45 15:44 17:45
14.11.2003 17:45 Uhr: 3797,4

EDA EDITION DEUTSCHE MEDIEN

Start Resear... PI-me... email.ppt Summ... search... disco.c... USA: E...

Multilingual Summarization

- Given a set of documents on the same event
- Some documents are in English
- Some documents are translated from other languages

Issues for Multilingual Summarization

- Problem: Translated text is errorful
- Exploit information available during summarization
 - Similar documents in cluster
- Replace translated sentences with similar English
- Edit translated text
 - Replace named entities with extractions from similar English

Multilingual Redundancy

BAGDAD. - A total of 21 prisoners has been died and a hundred more hurt by firings from mortar in the jail of Abu Gharib (to 20 kilometers to the west of Bagdad), according to has informed general into the U.S.A. Marco Kimmitt.

Spanish

Bagdad in the Iraqi capital Aufstaendi attacked Bagdad on Tuesday a prison with mortars and **killed after USA gifts 22 prisoners**. Further 92 passengers of the Abu Ghraib prison were hurt, communicated a spokeswoman of the American armed forces.

German

The Iraqi being stationed US military shot on the 20th, the same day to the allied forces detention facility which is in アブグアブグレイブighdad west approximately 20 kilometers, mortar 12 shot and you were packed, **22 Iraqi human prisoners died**, it announced that nearly 100 people were injured.

Japanese

BAGHDAD, Iraq – Insurgents fired 12 mortars into Baghdad's Abu Ghraib prison Tuesday, **killing 22 detainees** and injuring 92, U.S. military officials said.

English

Multilingual Redundancy

BAGDAD. - A total of 21 prisoners has been died and a hundred more hurt by firings from *mortar in the jail of Abu Ghraib* (to 20 kilometers to the west of Bagdad), according to has informed general into the U.S.A. Marco Kimmitt.

Spanish

Bagdad in the Iraqi capital Aufstaendi attacked Bagdad on Tuesday *a prison with mortars* and *killed after USA gifts 22 prisoners*. Further 92 passengers of the Abu Ghraib prison were hurt, communicated a spokeswoman of the American armed forces.

German

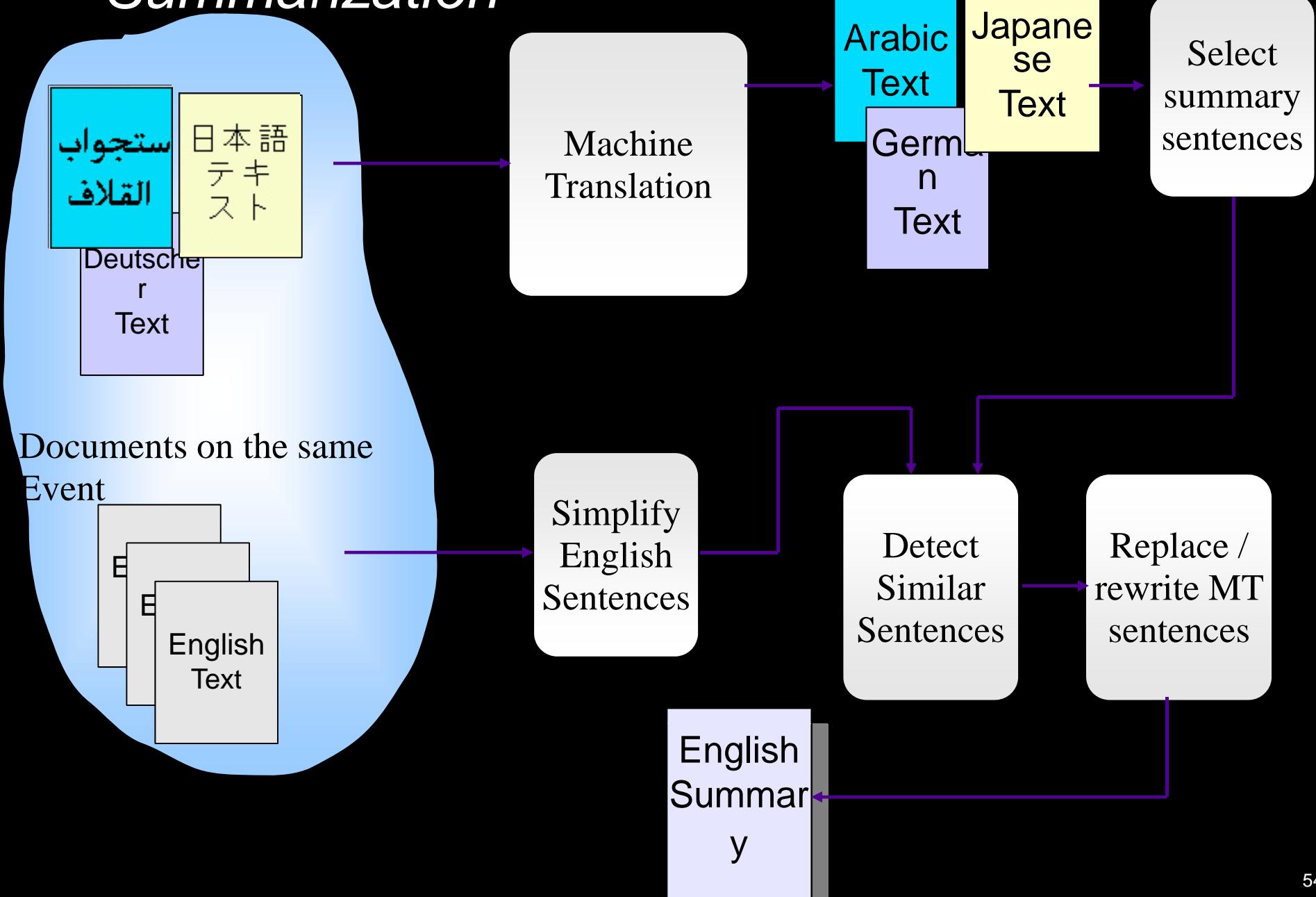
The Iraqi being stationed US military shot on the 20th, the same day to the allied forces detention facility which is in アブグアブグレイブighdad west approximately 20 kilometers, *mortar 12 shot* and you were packed, *22 Iraqi human prisoners died*, it announced that nearly 100 people were injured.

Japanese

BAGHDAD, Iraq – *Insurgents fired 12 mortars into Baghdad's Abu Ghraib prison* Tuesday, *killing 22 detainees* and injuring 92, U.S. military officials said.

English

Multilingual Similarity-based Summarization



Sentence 1

Iraqi President Saddam Hussein that the government of Iraq over 24 years in a "black" near the port of the northern Iraq after nearly eight months of pursuit was considered the largest in history .

Similarity 0.27: Ousted Iraqi President Saddam Hussein is in custody following his dramatic capture by US forces in Iraq.

Similarity 0.07: Saddam Hussein, the former president of Iraq, has been captured and *is being held by US forces in the country.*

Similarity 0.04: *Coalition authorities have said that the former Iraqi president could be tried at a war crimes tribunal, with Iraqi judges presiding and international legal experts acting as advisers.*

Sentence Simplification

- Machine translated sentences long and ungrammatical
- Use sentence simplification on English sentences to reduce to approximately “one fact” per sentence
- Use Arabic sentences to find most similar simple sentences
- Present multiple high similarity sentences

Simplification Examples

- '*Operation Red Dawn*', which led to the capture of Saddam Hussein, followed crucial information from a member of a family close to the former Iraqi leader.
 - 'Operation Red Dawn' followed crucial information from a member of a family close to the former Iraqi leader.
 - *Operation Red Dawn led to the capture of Saddam Hussein.*
- Saddam Hussein had been the object of intensive searches by US-led forces in Iraq but previous attempts to locate him had proved unsuccessful.
 - Saddam Hussein had been the object of intensive searches by US-led forces in Iraq.
 - *But previous attempts to locate him had proved unsuccessful.*

Rewrite proper and common nouns to remove MT errors

(Siddharthan and McKeown 05)

- Use redundancy in input to summarization and multiple translations to build attribute value matrices (AVMs)
 - Record country, role, description for all people
 - Record name variants
- Use generation grammar with semantic categories (role, organization, location) to re-order phrases for fluent output

the representative of Iraq in the United Nations Nizar Hamdoon

+

representative of Iraq of the United Nations Nizar HAMDOON

↓

name	Nizar Hamdoon
role	representative
country	Iraq (<i>arg1</i>)
organization	United Nations (<i>arg2</i>)

↓

Iraqi United Nations representative Nizar Hamdoon

Evaluation

- DUC (Document Understanding Conference): run by NIST yearly
- Manual creation of topics (sets of documents)
- 2-7 human written summaries per topic
- How well does a system generated summary cover the information in a human summary?
- Metrics
 - Rouge
 - Pyramid

Rouge

- ROUGE version 1.2.1:
 - Publicly available at: <http://www.isi.edu/~cyl/ROUGE>
 - Version 1.2.1 includes:
 - ROUGE-N - n-gram-based co-occurrence statistics
 - ROUGE-L - longest common subsequence-based (LCS) co-occurrence statistics
 - ROUGE-W - LCS-based co-occurrence statistics favoring consecutive LCSes
- Measures recall
 - Rouge-1: How many unigrams in the human summary did the system summary find?
 - Rouge-2: How many bigrams?

Pyramids

- Uses multiple human summaries
 - Previous data indicated 5 needed for score stability
- Information is ranked by its importance
- Allows for multiple good summaries
- A pyramid is created from the human summaries
 - Elements of the pyramid are content units
 - System summaries are scored by comparison with the pyramid

Summarization Content Units

- Near-paraphrases from different human summaries
- Clause or less
- Avoids explicit semantic representation
- Emerges from analysis of human summaries

SCU: A cable car caught fire

(Weight = 4)

- A. The cause of the fire was unknown.
- B. A cable car caught fire just after entering a mountainside tunnel in an alpine resort in Kaprun, Austria on the morning of November 11, 2000.
- C. A cable car pulling skiers and snowboarders to the Kitzsteinhorn resort, located 60 miles south of Salzburg in the Austrian Alps, caught fire inside a mountain tunnel, killing approximately 170 people.
- D. On November 10, 2000, a cable car filled to capacity caught on fire, trapping 180 passengers inside the Kitzsteinhorn mountain, located in the town of Kaprun, 50 miles south of Salzburg in the central Austrian Alps.

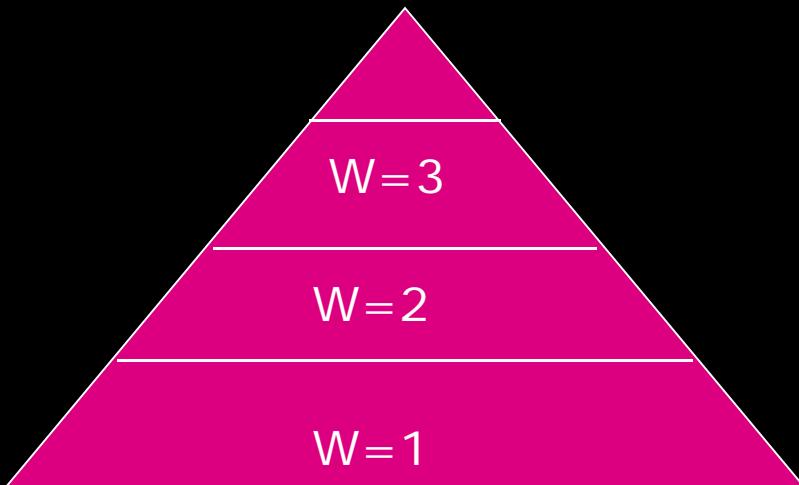
SCU: The cause of the fire is unknown (Weight = 1)

- A. The cause of the fire was unknown.
- B. A cable car caught fire just after entering a mountainside tunnel in an alpine resort in Kaprun, Austria on the morning of November 11, 2000.
- C. A cable car pulling skiers and snowboarders to the Kitzsteinhorn resort, located 60 miles south of Salzburg in the Austrian Alps, caught fire inside a mountain tunnel, killing approximately 170 people.
- D. On November 10, 2000, a cable car filled to capacity caught on fire, trapping 180 passengers inside the Kitzsteinhorn mountain, located in the town of Kaprun, 50 miles south of Salzburg in the central Austrian Alps.

SCU: The accident happened in the Austrian Alps (Weight = 3)

- A. The cause of the fire was unknown.
- B. A cable car caught fire just after entering a mountainside tunnel in an alpine resort in Kaprun, Austria on the morning of November 11, 2000.
- C. A cable car pulling skiers and snowboarders to the Kitzsteinhorn resort, located 60 miles south of Salzburg in the Austrian Alps, caught fire inside a mountain tunnel, killing approximately 170 people.
- D. On November 10, 2000, a cable car filled to capacity caught on fire, trapping 180 passengers inside the Kitzsteinhorn mountain, located in the town of Kaprun, 50 miles south of Salzburg in the central Austrian Alps.

Idealized representation



- Tiers of differentially weighted SCUs
- Top: few SCUs, high weight
- Bottom: many SCUs, low weight

Pyramid Score

$$\text{SCORE} = D/\text{MAX}$$

D: Sum of the weights of the SCUs in a summary

MAX: Sum of the weights of the SCUs in a ideally informative summary

Measures the proportion of good information in the summary: precision

User Study: Objectives

- Does multi-document summarization help?
 - Do summaries help the user find information needed to perform a report writing task?
 - Do users use information from summaries in gathering their facts?
 - Do summaries increase user satisfaction with the online news system?
 - Do users create better quality reports with summaries?
 - How do full multi-document summaries compare with minimal 1-sentence summaries such as Google News?

User Study: Design

- Four parallel news systems
 - *Source documents only*; no summaries
 - *Minimal single sentence summaries* (Google News)
 - *Newsblaster summaries*
 - *Human summaries*
- All groups write reports given four scenarios
 - A task similar to analysts
 - Can only use Newsblaster for research
 - Time-restricted

User Study: Execution

- 4 scenarios
 - 4 event clusters each
 - 2 directly relevant, 2 peripherally relevant
 - Average 10 documents/cluster
- 45 participants
 - Balance between liberal arts, engineering
 - 138 reports
- Exit survey
 - Multiple-choice and open-ended questions
- Usage tracking
 - Each click logged, on or off-site

“Geneva” Prompt

- The conflict between Israel and the Palestinians has been difficult for government negotiators to settle. Most recently, implementation of the “road map for peace”, a diplomatic effort sponsored by
 - Who participated in the negotiations that produced the Geneva Accord?
 - Apart from direct participants, who supported the Geneva Accord preparations and how?
 - What has the response been to the Geneva Accord by the Palestinians?

Measuring Effectiveness

- Score report content and compare across summary conditions
- Compare user satisfaction per summary condition
- Comparing where subjects took report content from

Summary Level	Pyramid Score
Level 1 (documents only)	0.3354
Level 2 (one sentence summary)	0.3757
Level 3 (System-X summary)	0.4269
Level 4 (Human summary)	0.4027

Table 2: Mean Pyramid Scores on Reports, Scenario 1 (Geneva Accords) excluded.

User Satisfaction

- More effective than a web search with Newsblaster
 - Not true with documents only or single-sentence summaries
- Easier to complete the task with summaries than with documents only
- Enough time with summaries than documents only
- Summaries helped most
 - 5% single sentence summaries
 - 24% Newsblaster summaries
 - 43% human summaries

User Study: Conclusions

- Summaries measurably improve a news browser's effectiveness for research
- Users are more satisfied with Newsblaster summaries than single-sentence summaries like those of Google News
- Users want search
 - Not included in evaluation

Questions (from Sparck Jones)

- Should we take the reader into account and how?
- Need more power than text extraction and more flexibility than fact extraction (p. 4)
- “Similarly, the notion of a basic summary, i.e., one reflective of the source, makes hidden fact assumptions, for example that the subject knowledge of the output’s readers will be on a par with that of the readers for whom the source was intended. (p. 5)”
- Is the state of the art sufficiently mature to allow summarization from intermediate representations and still allow robust processing of domain independent material?
- Evaluation: gold standard vs. user study? Difficulty of evaluation?

