Text summarization

Tutorial
ACM SIGIR
Sheffield, UK
July 25, 2004

Dragomir R. Radev

CLAIR: Computational Linguistics And Information Retrieval group
University of Michigan

radev@umich.edu
Part I
Introduction
Information overload

• The problem:
  – 4 Billion URLs indexed by Google
  – 200 TB of data on the Web [Lyman and Varian 03]

• Possible approaches:
  – information retrieval
  – document clustering
  – information extraction
  – visualization
  – question answering
  – text summarization
MILAN, Italy, April 18. A small airplane crashed into a government building in heart of Milan, setting the top floors on fire, Italian police reported. There were no immediate reports on casualties as rescue workers attempted to clear the area in the city's financial district. Few details of the crash were available, but news reports about it immediately set off fears that it might be a terrorist act akin to the Sept. 11 attacks in the United States. Those fears sent U.S. stocks tumbling to session lows in late morning trading.

Witnesses reported hearing a loud explosion from the 30-story office building, which houses the administrative offices of the local Lombardy region and sits next to the city's central train station. Italian state television said the crash put a hole in the 25th floor of the Pirelli building. News reports said smoke poured from the opening. Police and ambulances rushed to the building in downtown Milan. No further details were immediately available.
MILAN, Italy, April 18. A small airplane crashed into a government building in heart of Milan, setting the top floors on fire, Italian police reported. There were no immediate reports on casualties as rescue workers attempted to clear the area in the city's financial district. Few details of the crash were available, but news reports about it immediately set off fears that it might be a terrorist act akin to the Sept. 11 attacks in the United States. Those fears sent U.S. stocks tumbling to session lows in late morning trading.

Witnesses reported hearing a loud explosion from the 30-story office building, which houses the administrative offices of the local Lombardy region and sits next to the city's central train station. Italian state television said the crash put a hole in the 25th floor of the Pirelli building. News reports said smoke poured from the opening. Police and ambulances rushed to the building in downtown Milan. No further details were immediately available.
MILAN, Italy, April 18. A small airplane crashed into a government building in heart of Milan, setting the top floors on fire, Italian police reported. There were no immediate reports on casualties as rescue workers attempted to clear the area in the city's financial district. Few details were immediately available about it immediately set off fears that it might be a terrorist act akin to the Sept. 11 attacks in the United States. Those fears sent U.S. stocks tumbling to session lows in late morning trading.

Witnesses reported hearing a loud explosion from an office building, which houses the administrative offices of the local Lombardy region and sits next to the city's central train station. Italian state television said the crash put a hole in the 25th floor of the Pirelli building. News reports said smoke poured from the opening. Police and ambulances rushed to the building in downtown Milan. No further details were immediately available.
1. How many people were injured?
2. How many people were killed? (age, number, gender, description)
3. Was the pilot killed?
4. Where was the plane coming from?
5. Was it an accident (technical problem, illness, terrorist act)?
6. Who was the pilot? (age, number, gender, description)
7. When did the plane crash?
8. How tall is the Pirelli building?
9. Who was on the plane with the pilot?
10. Did the plane catch fire before hitting the building?
11. What was the weather like at the time of the crash?
12. When was the building built?
13. What direction was the plane flying?
14. How many people work in the building?
15. How many people were in the building at the time of the crash?
16. How many people were taken to the hospital?
17. What kind of aircraft was used?
Types of summaries

• Purpose
  – Indicative, informative, and critical summaries

• Form
  – Extracts (representative paragraphs/sentences/phrases)

• Dimensions
  – Single-document vs. multi-document

• Context
  – Query-specific vs. query-independent
Genres

- headlines
- outlines
- minutes
- biographies
- abridgments
- sound bites
- movie summaries
- chronologies, etc.

[Mani and Maybury 1999]
What does summarization involve?

- Three stages (typically)
  - content identification
  - conceptual organization
  - realization
BAGHDAD, Iraq (CNN) 6 July 2004 -- Three U.S. Marines have died in al Anbar Province west of Baghdad, the Coalition Public Information Center said Tuesday.

According to CPIC, "Two Marines assigned to [1st] Marine Expeditionary Force were killed in action and one Marine died of wounds received in action Monday in the Al Anbar Province while conducting security and stability operations."

Al Anbar Province -- a hotbed for Iraqi insurgents -- includes the restive cities of Ramadi and Fallujah and runs to the Syrian and Jordanian borders.

Meanwhile, officials said eight people died Monday in a U.S. air raid on a house in Fallujah that American commanders said was used to harbor Islamic militants.

A statement from interim Iraqi Prime Minister Ayad Allawi said his government's security forces provided "clear and compelling intelligence" that led to the raid.

A senior U.S. military official told CNN the target was a group of people suspected of planning suicide attacks using vehicles.

The strike was the latest in a series of raids on the city to target what U.S. military spokesmen have called safehouses for the network led by fugitive Islamic militant leader Abu Musab al-Zarqawi.

A statement from Allawi said: "The people of Iraq will not tolerate terrorist groups or those who collaborate with any other foreign fighters such as the Zarqawi network to continue their wicked ways."

"The sovereign nation of Iraq and our international partners are committed to stopping terrorism and will continue to hunt down these evil terrorists and weed them out, one by one. I call upon all Iraqis to close ranks and report to the authorities on the activities of these criminal cells."

American planes dropped two 1,000-pound bombs and four 500-pound bombs on the house about 7:15 p.m. (11:15 a.m. ET), according to a statement from the U.S.-led Multi-National Force-Iraq.

"This operation employed precision weapons and underscores the resolve of multinational forces and Iraqi security forces to jointly destroy terrorist networks in Iraq," a military statement said.

A doctor at Fallujah Hospital said the dead included four men, a woman and three children, some of them members of the same family. Another three people were wounded, the doctor said.

U.S. officials blame Zarqawi, who is believed to have links to al Qaeda, for numerous attacks on Iraqi and U.S. civilians and coalition troops.

At least four previous air raids have targeted suspected Zarqawi safehouses in Fallujah.
BAGHDAD, Iraq (CNN) 6 July 2004 -- Three U.S. Marines have died in al Anbar Province west of Baghdad, the Coalition Public Information Center said Tuesday. According to CPIC, "Two Marines assigned to [1st] Marine Expeditionary Force were killed in action and one Marine died of wounds received in action Monday in the Al Anbar Province while conducting security and stability operations."

Al Anbar Province -- a hotbed for Iraqi insurgents -- includes the restive cities of Ramadi and Fallujah and runs to the Syrian and Jordanian borders. Meanwhile, officials said eight people died Monday in a U.S. air raid on a house in Fallujah that American commanders said was used to harbor Islamic militants. A statement from interim Iraqi Prime Minister Ayad Allawi said his government's security forces provided "clear and compelling intelligence" that led to the raid. A senior U.S. military official told CNN the target was a group of people suspected of planning suicide attacks using vehicles. The strike was the latest in a series of raids on the city to target what U.S. military spokesmen have called safehouses for the network led by fugitive Islamic militant leader Abu Musab al-Zarqawi. A statement from Allawi said: "The people of Iraq will not tolerate terrorist groups or those who collaborate with any other foreign fighters such as the Zarqawi network to continue their wicked ways. "The sovereign nation of Iraq and our international partners are committed to stopping terrorism and will continue to hunt down these evil terrorists and weed them out, one by one. I call upon all Iraqis to close ranks and report to the authorities on the activities of these criminal cells."

American planes dropped two 1,000-pound bombs and four 500-pound bombs on the house about 7:15 p.m. (11:15 a.m. ET), according to a statement from the U.S.-led Multi-National Force-Iraq. "This operation employed precision weapons and underscores the resolve of multinational forces and Iraqi security forces to jointly destroy terrorist networks in Iraq," a military statement said. A doctor at Fallujah Hospital said the dead included four men, a woman and three children, some of them members of the same family. Another three people were wounded, the doctor said. U.S. officials blame Zarqawi, who is believed to have links to al Qaeda, for numerous attacks on Iraqi and U.S. civilians and coalition troops. At least four previous air raids have targeted suspected Zarqawi safehouses in Fallujah.
Outline

I  Introduction
II  Traditional approaches
III Multi-document summarization
IV Knowledge-rich techniques
V  Evaluation methods
VI  Recent approaches
VII Appendix
Part II
Traditional approaches
Human summarization and abstracting

• What professional abstractors do
• Ashworth:
  • “To take an original article, understand it and pack it neatly into a nutshell without loss of substance or clarity presents a challenge which many have felt worth taking up for the joys of achievement alone. These are the characteristics of an art form”.

www.si.umich.edu
Borko and Bernier 75

- The abstract and its use:
  - Abstracts promote current awareness
  - Abstracts save reading time
  - Abstracts facilitate selection
  - Abstracts facilitate literature searches
  - Abstracts improve indexing efficiency
  - Abstracts aid in the preparation of reviews
Cremmins 82, 96

- American National Standard for Writing Abstracts:
  - State the purpose, methods, results, and conclusions presented in the original document, either in that order or with an initial emphasis on results and conclusions.
  - Make the abstract as informative as the nature of the document will permit, so that readers may decide, quickly and accurately, whether they need to read the entire document.
  - Avoid including background information or citing the work of others in the abstract, unless the study is a replication or evaluation of their work.
Cremmins 82, 96

- Do not include information in the abstract that is not contained in the textual material being abstracted.
- Verify that all quantitative and qualitative information used in the abstract agrees with the information contained in the full text of the document.
- Use standard English and precise technical terms, and follow conventional grammar and punctuation rules.
- Give expanded versions of lesser known abbreviations and acronyms, and verbalize symbols that may be unfamiliar to readers of the abstract.
- Omit needless words, phrases, and sentences.
Cremmins 82, 96

- **Original version:**

  There were significant positive associations between the concentrations of the substance administered and mortality in rats and mice of both sexes.

  There was no convincing evidence to indicate that endrin ingestion induced and of the different types of tumors which were found in the treated animals.

- **Edited version:**

  Mortality in rats and mice of both sexes was dose related.

  No treatment-related tumors were found in any of the animals.
Morris et al. 92

- Reading comprehension of summaries
- 75% redundancy of English [Shannon 51]
- Compare manual abstracts, Edmundson-style extracts, and full documents
- Extracts containing 20% or 30% of original document are effective surrogates of original document
- Performance on 20% and 30% extracts is no different than informative abstracts
Luhn 58

- Very first work in automated summarization
- Computes measures of significance
- Words:
  - stemming
  - bag of words

Resolving power of significant words
Luhn 58

- Sentences: concentration of high-score words
- Cutoff values established in experiments with 100 human subjects

\[ \text{SCORE} = \frac{4^2}{7} \approx 2.3 \]
Edmundson 69

- **Cue method:**
  - stigma words ("hardly", "impossible")
  - bonus words ("significant")
- **Key method:**
  - similar to Luhn

- **Title method:**
  - title + headings
- **Location method:**
  - sentences under headings
  - sentences near beginning or end of document and/or paragraphs (also [Baxendale 58])
Edmundson 69

- Linear combination of four features:

  \[ \alpha_1 C + \alpha_2 K + \alpha_3 T + \alpha_4 L \]

- Manually labelled training corpus
- Key not important!
Paice 90

- Survey up to 1990
- Techniques that (mostly) failed:
  - syntactic criteria [Earl 70]
  - indicator phrases (“The purpose of this article is to review…”)

- Problems with extracts:
  - lack of balance
  - lack of cohesion
    - anaphoric reference
    - lexical or definite reference
    - rhetorical connectives
Paice 90

- **Lack of balance**
  - later approaches based on text rhetorical structure

- **Lack of cohesion**
  - recognition of anaphors [Liddy et al. 87]

- **Example: “that” is**
  - nonanaphoric if preceded by a research-verb (e.g., “demonstrat-”),
  - nonanaphoric if followed by a pronoun, article, quantifier, ...
  - external if no later than 10th word, else
  - internal
Brandow et al. 95

- ANES: commercial news from 41 publications
- “Lead” achieves acceptability of 90% vs. 74.4% for “intelligent” summaries

- 20,997 documents
- words selected based on tf*idf
- sentence-based features:
  - signature words
  - location
  - anaphora words
  - length of abstract
Brandow et al. 95

- Sentences with no signature words are included if between two selected sentences
- Evaluation done at 60, 150, and 250 word length

- Non-task-driven evaluation:
  “Most summaries judged less-than-perfect would not be detectable as such to a user”
Lin & Hovy 97

- Optimum position policy
- Measuring yield of each sentence position against keywords (signature words) from Ziff-Davis corpus

- Preferred order

[(T) (P2,S1) (P3,S1) (P2,S2) {(P4,S1) (P5,S1) (P3,S2)} {(P1,S1) (P6,S1) (P7,S1) (P1,S3) (P2,S3) ...}]
Kupiec et al. 95

- Extracts of roughly 20% of original text
- Feature set:
  - sentence length
    - |S| > 5
  - fixed phrases
    - 26 manually chosen
  - paragraph
    - sentence position in paragraph
  - thematic words
    - binary: whether sentence is included in manual extract
  - uppercase words
    - not common acronyms
- Corpus:
  - 188 document + summary pairs from scientific journals
Kupiec et al. 95

- **Uses Bayesian classifier:**

\[
P(s \in S \mid F_1, F_2, \ldots F_k) = \frac{P(F_1, F_2, \ldots F_k \mid s \in S)P(s \in S)}{P(F_1, F_2, \ldots F_k)}
\]

- **Assuming statistical independence:**

\[
P(s \in S \mid F_1, F_2, \ldots F_k) = \frac{\prod_{j=1}^{k} P(F_j \mid s \in S)P(s \in S)}{\prod_{j=1}^{k} P(F_j)}
\]
Kupiec et al. 95

• Performance:
  – For 25% summaries, 84% precision
  – For smaller summaries, 74% improvement over Lead
Salton et al. 97

- Document analysis based on semantic hyperlinks (among pairs of paragraphs related by a lexical similarity significantly higher than random)

- Bushy paths (or paths connecting highly connected paragraphs) are more likely to contain information central to the topic of the article
Salton et al. 97
**Salton et al. 97**

**Overlap between manual extracts: 46%**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Optimistic</th>
<th>Pessimistic</th>
<th>Intersection</th>
<th>Union</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global bushy</td>
<td>45.60%</td>
<td>30.74%</td>
<td>47.33%</td>
<td>55.16%</td>
</tr>
<tr>
<td>Global depth-first</td>
<td>43.98%</td>
<td>27.76%</td>
<td>42.33%</td>
<td>52.48%</td>
</tr>
<tr>
<td>Segmented bushy Random</td>
<td>45.48%</td>
<td>26.37%</td>
<td>38.17%</td>
<td>52.95%</td>
</tr>
<tr>
<td>Random</td>
<td>39.16%</td>
<td>22.07%</td>
<td>38.47%</td>
<td>44.24%</td>
</tr>
</tbody>
</table>
Marcu 97-99

- Based on RST (nucleus+satellite relations)
- Text coherence
- 70% precision and recall in matching the most important units in a text

- Example: evidence
  [The truth is that the pressure to smoke in junior high is greater than it will be any other time of one’s life:] [we know that 3,000 teens start smoking each day.]

- N+S combination increases R’s belief in N [Mann and Thompson 88]
2 Elaboration

2 Elaboration

2 Background Justification

With its distant orbit (50 percent farther from the sun than Earth) and slim atmospheric blanket,
(1)

Mars experiences frigid weather conditions
(2)

Surface temperatures typically average about -60 degrees Celsius (-76 degrees Fahrenheit) at the equator and can dip to -123 degrees C near the poles
(3)

Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion,
(4)

but any liquid water formed in this way would evaporate almost instantly
(5)

because of the low atmospheric pressure
(6)

4 5 Contrast

5 Evidence Cause

3 Elaboration

4 5 Contrast

Although the atmosphere holds a small amount of water, and water-ice clouds sometimes develop,
(7)

Most Martian weather involves blowing dust and carbon monoxide.
(8)

Each winter, for example, a blizzard of frozen carbon dioxide rages over one pole, and a few meters of this dry-ice snow accumulate as previously frozen carbon dioxide evaporates from the opposite polar cap.
(9)

Yet even on the summer pole, where the sun remains in the sky all day long, temperatures never warm enough to melt frozen water.
(10)

8 Concession

8 Example

10 Antithesis
Barzilay and Elhadad 97

• Lexical chains [Stairmand 96]

Mr. Kenny is the person that invented the anesthetic machine which uses micro-computers to control the rate at which an anesthetic is pumped into the blood. Such machines are nothing new. But his device uses two micro-computers to achieve much closer monitoring of the pump feeding the anesthetic into the patient.
Barzilay and Elhadad 97

- WordNet-based
- three types of relations:
  - extra-strong (repetitions)
  - strong (WordNet relations)
  - medium-strong (link between synsets is longer than one + some additional constraints)
Barzilay and Elhadad 97

- Scoring chains:
  - Length
  - Homogeneity index:
    = 1 - # distinct words in chain

Score = Length * Homogeneity

Score > Average + 2 * st.dev.
Osborne 02

- Maxent (loglinear) model – no independence assumptions
- Features: word pairs, sentence length, sentence position, discourse features (e.g., whether sentence follows the “Introduction”, etc.)
- Maxent outperforms Naïve Bayes
Part III
Multi-document summarization
Mani & Bloedorn 97,99

- Summarizing differences and similarities across documents
- Single event or a sequence of events

- Text segments are aligned
- Evaluation: TREC relevance judgments
- Significant reduction in time with no significant loss of accuracy
Carbonell & Goldstein 98

- Maximal Marginal Relevance (MMR)
- Query-based summaries
- Law of diminishing returns

\[ \text{MMR} = \arg\max_{D_i \in R \setminus S} \lambda \left( \text{Sim}_1(D_i, Q) - (1-\lambda) \max_{D_i \in S} \text{Sim}_2(D_i, D_j) \right) \]

- \( C \) = doc collection
- \( Q \) = user query
- \( R = IR(C, Q, \theta) \)
- \( S \) = already retrieved documents
- \( \text{Sim} \) = similarity metric used
Radev et al. 00

- MEAD
- Centroid-based
- Based on sentence utility

- Topic detection and tracking initiative [Allen et al. 98, Wayne 98]
ARTICLE 18853: ALGIERS, May 20 (AFP)

1. Eighteen decapitated bodies have been found in a mass grave in northern Algeria, press reports said Thursday, adding that two shepherds were murdered earlier this week.

2. Security forces found the mass grave on Wednesday at Chbila, near Djelfa, 275 kilometers (170 miles) south of the capital.

3. It contained the bodies of people killed last year during a wedding ceremony, according to Le Quotidien Liberte.

4. The victims included women, children and old men.

5. Most of them had been decapitated and their heads thrown on a road, reported the Es Sahafa.

6. Another mass grave containing the bodies of around 10 people was discovered recently near Algiers, in the Eucalyptus district.

7. The two shepherds were killed Monday evening by a group of nine armed Islamists near the Moulai Slissen forest.

8. After being injured in a hail of automatic weapons fire, the pair were finished off with machete blows before being decapitated, Le Quotidien d'Oran reported.

9. Seven people, six of them children, were killed and two injured Wednesday by armed Islamists near Medea, 120 kilometers (75 miles) south of Algiers, security forces said.

10. The same day a parcel bomb explosion injured 17 people in Algiers itself.

11. Since early March, violence linked to armed Islamists has claimed more than 500 lives, according to press tallies.

ARTICLE 18854: ALGIERS, May 20 (UPI)

1. Algerian newspapers have reported that 18 decapitated bodies have been found by authorities in the south of the country.

2. Police found the “decapitated bodies of women, children and old men, with their heads thrown on a road” near the town of Jelfa, 275 kilometers (170 miles) south of the capital Algiers.

3. In another incident on Wednesday, seven people -- including six children -- were killed by terrorists, Algerian security forces said.

4. Extremist Muslim militants were responsible for the slaughter of the seven people in the province of Medea, 120 kilometers (74 miles) south of Algiers.

5. The killers also kidnapped three girls during the same attack, authorities said, and one of the girls was found wounded on a nearby road.

6. Meanwhile, the Algerian daily Le Matin today quoted Interior Minister Abdul Malik Silal as saying that “terrorism has not been eradicated, but the movement of the terrorists has significantly declined.”

7. Algerian violence has claimed the lives of more than 70,000 people since the army cancelled the 1992 general elections that Islamic parties were likely to win.

8. Mainstream Islamic groups, most of which are banned in the country, insist their members are not responsible for the violence against civilians.

9. Some Muslim groups have blamed the army, while others accuse “foreign elements conspiring against Algeria.”
Vector-based representation

Term 1

Term 2

Term 3

Document

Centroid

$\alpha$
Vector-based matching

- The cosine measure

\[
\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}
\]
CIDR

\[ \text{sim} \geq T \]  
\[ \text{sim} < T \]
## Centroids

<table>
<thead>
<tr>
<th>C 00022 (N=44)</th>
<th>C 00026 (N=10)</th>
<th>C 00025 (N=19)</th>
<th>C 10007 (N=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>diana 1.93</td>
<td>universe 1.50</td>
<td>albanians 3.00</td>
<td>crashes 1.00</td>
</tr>
<tr>
<td>princess 1.52</td>
<td>expansion 1.00</td>
<td></td>
<td>safety 0.55</td>
</tr>
<tr>
<td></td>
<td>bang 0.90</td>
<td></td>
<td>transportat 0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>drivers 0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>board 0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>flight 0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>buckle 0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pittsburgh 0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>graduating 0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>automobile 0.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C 00035 (N=22)</th>
<th>C 10062 (N=161)</th>
<th>C 00008 (N=113)</th>
</tr>
</thead>
<tbody>
<tr>
<td>airlines 1.45</td>
<td>microsoft 3.24</td>
<td>space 1.98</td>
</tr>
<tr>
<td>finnair 0.45</td>
<td>justice 0.93</td>
<td>shuttle 1.17</td>
</tr>
<tr>
<td></td>
<td>departmen 0.88</td>
<td>station 0.75</td>
</tr>
<tr>
<td></td>
<td>windows 0.98</td>
<td>nasa 0.51</td>
</tr>
<tr>
<td></td>
<td>corp 0.61</td>
<td>columbia 0.37</td>
</tr>
<tr>
<td></td>
<td>software 0.57</td>
<td>mission 0.33</td>
</tr>
<tr>
<td></td>
<td>ellison 0.07</td>
<td>mir 0.30</td>
</tr>
<tr>
<td></td>
<td>hatch 0.06</td>
<td>astronaut 0.14</td>
</tr>
<tr>
<td></td>
<td>netscape 0.04</td>
<td>steering 0.11</td>
</tr>
<tr>
<td></td>
<td>metcalfe 0.02</td>
<td>safely 0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C 00031 (N=34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>el 1.85</td>
</tr>
<tr>
<td>nino 1.56</td>
</tr>
</tbody>
</table>
MEAD

```
  |   |
  |   |
  |   |
  |   |
  |   |
  |   |
  |   |
...
```

```
  |   |
  |   |
  |   |
  |   |
  |   |
  |   |
  |   |
...
```
MEAD

- **INPUT:** Cluster of $d$ documents with $n$ sentences (compression rate $= r$)
- **OUTPUT:** $(n \times r)$ sentences from the cluster with the highest values of SCORE

$\text{SCORE} (s) = \sum_i (w_c C_i + w_p P_i + w_r F_i)$
[Barzilay et al. 99]

- Theme intersection (paraphrases)
- Identifying common phrases across multiple sentences:
  - evaluated on 39 sentence-level predicate-argument structures
  - 74% of p-a structures automatically identified
Other multi-document approaches

- Reformulation [McKeown et al. 99, McKeown et al. 02]
- Generation by Selection and Repair [DiMarco et al. 97]
Part IV
Knowledge-rich approaches
Overview

• Schank and Abelson 77
  – scripts
• DeJong 79
  – FRUMP (slot-filling from UPI news)
• Graesser 81
  – Ratio of inferred propositions to these explicitly stated is 8:1
• Young & Hayes 85
  – banking telexes
### MESSAGE: ID
TST3-MUC4-0010

### MESSAGE: TEMPLATE
2

### INCIDENT: DATE
30 OCT 89

### INCIDENT: LOCATION
EL SALVADOR

### INCIDENT: TYPE
ATTACK

### INCIDENT: STAGE OF EXECUTION
ACCOMPLISHED

### INCIDENT: INSTRUMENT ID

### INCIDENT: INSTRUMENT TYPE

### PERP: INCIDENT CATEGORY
TERRORIST ACT

### PERP: INDIVIDUAL ID
"TERRORIST"

### PERP: ORGANIZATION ID
"THE FMLN"

### PERP: ORG. CONFIDENCE
REPORTED: "THE FMLN"

### PHYS TGT: ID

### PHYS TGT: TYPE

### PHYS TGT: NUMBER

### PHYS TGT: FOREIGN NATION

### PHYS TGT: EFFECT OF INCIDENT

### PHYS TGT: TOTAL NUMBER

### HUM TGT: NAME

### HUM TGT: DESCRIPTION
"1 CIVILIAN"

### HUM TGT: TYPE
CIVILIAN: "1 CIVILIAN"

### HUM TGT: NUMBER
1: "1 CIVILIAN"

### HUM TGT: FOREIGN NATION

### HUM TGT: EFFECT OF INCIDENT
DEATH: "1 CIVILIAN"

### HUM TGT: TOTAL NUMBER
Generating text from templates

On October 30, 1989, one civilian was killed in a reported FMLN attack in El Salvador.
### Excerpts from four articles

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JERUSALEM - A Muslim suicide bomber blew apart 18 people on a Jerusalem bus and wounded 10 in a mirror-image of an attack one week ago. The carnage could rob Israel's Prime Minister Shimon Peres of the May 29 election victory he needs to pursue Middle East peacemaking. Peres declared all-out war on Hamas but his tough talk did little to impress stunned residents of Jerusalem who said the election would turn on the issue of personal security.</td>
</tr>
<tr>
<td>2</td>
<td>JERUSALEM - A bomb at a busy Tel Aviv shopping mall killed at least 10 people and wounded 30, Israel radio said quoting police. Army radio said the blast was apparently caused by a suicide bomber. Police said there were many wounded.</td>
</tr>
<tr>
<td>3</td>
<td>A bomb blast ripped through the commercial heart of Tel Aviv Monday, killing at least 13 people and wounding more than 100. Israeli police say an Islamic suicide bomber blew himself up outside a crowded shopping mall. It was the fourth deadly bombing in Israel in nine days. The Islamic fundamentalist group Hamas claimed responsibility for the attacks, which have killed at least 54 people. Hamas is intent on stopping the Middle East peace process. President Clinton joined the voices of international condemnation after the latest attack. He said the &quot;forces of terror shall not triumph&quot; over peacemaking efforts.</td>
</tr>
<tr>
<td>4</td>
<td>TEL AVIV (Reuters) - A Muslim suicide bomber killed at least 12 people and wounded 105, including children, outside a crowded Tel Aviv shopping mall Monday, police said. Sunday, a Hamas suicide bomber killed 18 people on a Jerusalem bus. Hamas has now killed at least 56 people in four attacks in nine days. The windows of stores lining both sides of Dizengoff Street were shattered, the charred skeletons of cars lay in the street, the sidewalks were strewn with blood. The last attack on Dizengoff was in October 1994 when a Hamas suicide bomber killed 22 people on a bus.</td>
</tr>
</tbody>
</table>
Four templates

1. MESSAGE: ID
   TST-REU-0001
   SECSOURCE: SOURCE
   Reuters
   SECSOURCE: DATE
   March 3, 1996 11:30
   INCIDENT: SOURCE
   March 3, 1996
   INCIDENT: LOCATION
   Jerusalem
   INCIDENT: TYPE
   Bombing
   HUM TGT: NUMBER
   “killed: 18”
   “wounded: 10”
   PERP: ORGANIZATION ID

2. MESSAGE: ID
   TST-REU-0002
   SECSOURCE: SOURCE
   Reuters
   SECSOURCE: DATE
   March 4, 1996 07:20
   INCIDENT: SOURCE
   Israel Radio
   INCIDENT: DATE
   March 4, 1996
   INCIDENT: LOCATION
   Tel Aviv
   INCIDENT: TYPE
   Bombing
   HUM TGT: NUMBER
   “killed: at least 10”
   “wounded: more than 100”
   PERP: ORGANIZATION ID

3. MESSAGE: ID
   TST-REU-0003
   SECSOURCE: SOURCE
   Reuters
   SECSOURCE: DATE
   March 4, 1996 14:20
   INCIDENT: SOURCE
   March 4, 1996
   INCIDENT: LOCATION
   Tel Aviv
   INCIDENT: TYPE
   Bombing
   HUM TGT: NUMBER
   “killed: at least 13”
   “wounded: more than 100”
   “Hamas”
   PERP: ORGANIZATION ID

4. MESSAGE: ID
   TST-REU-0004
   SECSOURCE: SOURCE
   Reuters
   SECSOURCE: DATE
   March 4, 1996 14:30
   INCIDENT: SOURCE
   March 4, 1996
   INCIDENT: LOCATION
   Tel Aviv
   INCIDENT: TYPE
   Bombing
   HUM TGT: NUMBER
   “killed: at least 12”
   “wounded: 105”
   PERP: ORGANIZATION ID
Fluent summary with comparisons

Reuters reported that 18 people were killed on Sunday in a bombing in Jerusalem. The next day, a bomb in Tel Aviv killed at least 10 people and wounded 30 according to Israel radio. Reuters reported that at least 12 people were killed and 105 wounded in the second incident. Later the same day, Reuters reported that Hamas has claimed responsibility for the act.

(OUTPUT OF SUMMONS)
Operators

• If there are two templates
  AND
  the location is the same
  AND
  the time of the second template is after the time of the first template
  AND
  the source of the first template is different from the source of the second template
  AND
  at least one slot differs
  THEN
  combine the templates using the contradiction operator...
Operators: Change of Perspective

Change of perspective

Precondition:
The same source reports a change in a small number of slots

March 4th, Reuters reported that a bomb in Tel Aviv killed at least 10 people and wounded 30. Later the same day, Reuters reported that exactly 12 people were actually killed and 105 wounded.
Operators: Contradiction

Contradiction

Precondition:
Different sources report contradictory values for a small number of slots

The afternoon of February 26, 1993, Reuters reported that a suspected bomb killed at least six people in the World Trade Center. However, Associated Press announced that exactly five people were killed in the blast.
Operators: Refinement and Agreement

Refinement
On Monday morning, Reuters announced that a suicide bomber killed at least 10 people in Tel Aviv. In the afternoon, Reuters reported that Hamas claimed responsibility for the act.

Agreement
The morning of March 1st 1994, both UPI and Reuters reported that a man was kidnapped in the Bronx.
Operators: Generalization

Generalization

According to UPI, three terrorists were arrested in Medellín last Tuesday. Reuters announced that the police arrested two drug traffickers in Bogotá the next day.

A total of five criminals were arrested in Colombia last week.
Other conceptual methods

- Operator-based transformations using terminological knowledge representation [Reimer and Hahn 97]
- Topic interpretation [Hovy and Lin 98]
Part V
Evaluation techniques
Ideal evaluation

Information content

Compression Ratio = $\frac{|S|}{|D|}$

Retention Ratio = $\frac{i(S)}{i(D)}$
Overview of techniques

- Extrinsic techniques (task-based)
- Intrinsic techniques
Hovy 98

- Can you recreate what’s in the original?
  - the Shannon Game [Shannon 1947–50].
  - but often only some of it is really important.
- Measure info retention (number of keystrokes):
  - 3 groups of subjects, each must recreate text:
    - group 1 sees original text before starting.
    - group 2 sees summary of original text before starting.
    - group 3 sees nothing before starting.
- Results (# of keystrokes; two different paragraphs):

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>approx. 10</td>
<td>approx. 150</td>
<td>approx. 1100</td>
</tr>
</tbody>
</table>
Hovy 98

• Burning questions:
  1. How do different evaluation methods compare for each type of summary?
  2. How do different summary types fare under different methods?
  3. How much does the evaluator affect things?
  4. Is there a preferred evaluation method?

• Small Experiment
  – 2 texts, 7 groups.

• Results:
  – No difference!
  – As other experiment...
  – ? Extract is best?

<table>
<thead>
<tr>
<th></th>
<th>Shannon</th>
<th>Q&amp;A</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>1</td>
<td>1</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>Abstract</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
<td>3</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>Just-the-News</td>
<td></td>
<td>3</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extract</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular</td>
<td>1</td>
<td>2</td>
<td>1 1 1 1</td>
</tr>
<tr>
<td>Keywords</td>
<td>2</td>
<td>4</td>
<td>1 1 1</td>
</tr>
<tr>
<td>Random</td>
<td>3</td>
<td>3</td>
<td>1 1 1</td>
</tr>
<tr>
<td>No Text</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

1-2: 50%  1-2: 30%
2-3: 50%  2-3: 20%
3-4: 20%  4-5: 100%
Precision and Recall

<table>
<thead>
<tr>
<th></th>
<th>Relevant</th>
<th>Non-relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>System: relevant</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>System: non-relevant</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>
Precision and Recall

Precision: \( P = \frac{A}{A + B} \)

Recall: \( R = \frac{A}{A + C} \)

\( F = \frac{2PR}{(P + R)} \)
Jing et al. 98

- Small experiment with 40 articles
- When summary length is given, humans are pretty consistent in selecting the same sentences
- Percent agreement
- Different systems achieved maximum performance at different summary lengths
- Human agreement higher for longer summaries
SUMMAC [Mani et al. 98]

- 16 participants
- 3 tasks:
  - ad hoc: indicative, user-focused summaries
  - categorization: generic summaries, five categories
  - question-answering
- 20 TREC topics
- 50 documents per topic (short ones are omitted)
SUMMAC [Mani et al. 98]

- Participants submit a fixed-length summary limited to 10% and a “best” summary, not limited in length.

- variable-length summaries are as accurate as full text
- over 80% of summaries are intelligible
- technologies perform similarly
Goldstein et al. 99

- Reuters, LA Times
- Manual summaries
- Summary length rather than summarization ratio is typically fixed
- Normalized version of $R$ & $F$

$$R' = \frac{A}{\min (A + B, A + C)}$$

$$F' = \frac{2PR'}{(P + R')$$
Conclusion

• Summarization is coming of age
• For general domains: sentence extraction
• Strong focus on evaluation
• New challenges: language modeling, multilingual summaries, summarization of email, spoken document summarization

www.summarization.com