

Week 9

Thursday, August 11, 2016 15:18



9.1

09.01 Summarization NATURAL LANGUAGE PROCESSING **M**

NLP

09.01 Summarization NATURAL LANGUAGE PROCESSING **M**

Introduction to NLP

Summarization

Text Summarization

Health Benefits

- Eating a diet rich in vegetables and fruits as part of an overall healthy diet may reduce risk for heart disease, including heart attack and stroke.
- Eating a diet rich in some vegetables and fruits as part of an overall healthy diet may protect against certain types of cancers.
- Diets rich in foods containing fiber, such as some vegetables and fruits, may reduce the risk of heart disease, obesity, and type 2 diabetes.
- Eating vegetables and fruits rich in potassium as part of an overall healthy diet may lower blood pressure, and may also reduce the risk of developing kidney stones and help to decrease bone loss.
- Eating foods such as vegetables that are lower in calories per cup instead of some other higher-calorie food may be useful in helping to lower calorie intake.

Nutrients

- Most vegetables are naturally low in fat and calories. None have cholesterol. (Sauces or seasonings may add fat, calories, or cholesterol.)
- Vegetables are important sources of many nutrients, including potassium, dietary fiber, folate (folic acid), vitamin A, and vitamin C.
- Diets rich in potassium may help to maintain healthy blood pressure. Vegetable sources of potassium include sweet potatoes, white potatoes, white beans, tomato products (paste, sauce, and juice), beet greens, soybeans, lima beans, spinach, lentils, and kidney beans.
- Dietary fiber from vegetables, as part of an overall healthy diet, helps reduce blood cholesterol levels and may lower risk of heart disease. Fiber is important for proper bowel function. It helps reduce constipation and diverticulosis. Fiber-containing foods such as vegetables help provide a feeling of fullness with fewer calories.
- Folate (folic acid) helps the body form red blood cells. Women of childbearing age who may become pregnant should consume adequate folate from foods, and in addition 400 mcg of synthetic folic acid from fortified foods or supplements. This reduces the risk of neural tube defects, spina bifida, and anencephaly during fetal development.
- Vitamin A keeps eyes and skin healthy and helps to protect against infections.
- Vitamin C helps heal cuts and wounds and keeps teeth and gums healthy. Vitamin C aids in iron absorption.

Summary

Eating vegetables is healthy.

News Summarization



Book Summaries

- Cliff's notes → short description of large books.
- Book a minute (<http://www.rinkworks.com/bookaminute/>)
 - Which book is this?

(Some BOYS crash on an ISLAND.)

Ralph We need a fire. (*They make a fire. It goes out.*)

Ralph We need a fire. (*They make a fire. It goes out.*)

Ralph We need a fire.

Jack Forget the fire. Let's kill each other.

Other Boys Yeah! (*They do.*)

THE END

Lord of the Flies.

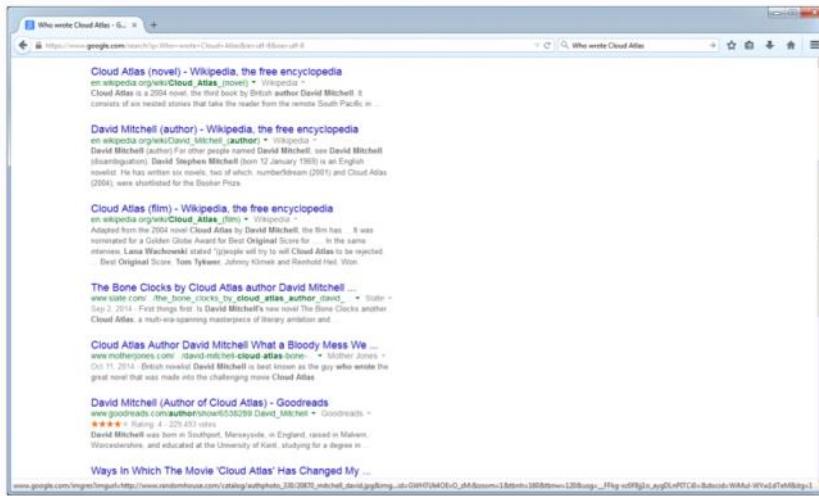
Ultra-Condensed by David J. Parker and Samuel Stoddard

Movie Summaries

- **Titanic**

- Beginning with genuine footage of the departure of the Titanic on its fateful voyage, this epic movie tells the events of that tragic night from the perspective of fictional survivor Rose. As an old lady of 100 she recounts her story of duty, love and disaster to a salvage crew searching for a lost gem.
- Winner of 11 Oscars, James Cameron's effects-driven blockbuster puts a human face on a tragedy of epic proportions by wedding the historical tale of the doomed ocean liner with a fictional romance between two of the ship's ill-fated passengers. Society girl Rose DeWitt Buquet and penniless artist Jack Dawson struggle to survive both the sinking ship -- and the wrath of Rose's wealthy fiancé.

Search Engine Snippets



Genres

- **headlines** very short sum
- outlines
- minutes
- biographies people
- abridgments → short version for specific audience
- sound bites
- movie summaries
- chronologies, etc.

[Mani and Maybury 1999]

Types of Summaries

- **Input**
- **Output**
- **Purpose**
 - Indicative, informative, and critical summaries
- **Form**
 - Extracts (representative paragraphs/sentences/phrases)
 - Abstracts: “a concise summary of the central subject matter of a document” [Paice90].
- **Dimensions**
 - Single-document vs. multi-document
- **Context**
 - Query-specific vs. generic

Stages



- **Three stages (typically)**

- content identification
- conceptual organization
- realization

input doc → determine what info to preserve

— additional issues —

↳ if
↳ they

*two sentences that go together should
be connected in some way.*

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 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.

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 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.

At least four previous air raids have targeted suspected Zarqawi safehouses in Fallujah.

Human Summarization and Abstracting

- What professional abstractors do
- Ashworth (1973):

• "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".

Difficult task

Extractive Summarization

- Selecting units of the original text
 - Usually sentences
 - No simplification
 - No rewriting
 - Baseline
 - Extract the first few sentences
- units usually
sentences.*
*2. 例句 1. 第二句
第一句*

NLP





NLP



Introduction to NLP

Summarization Techniques 1/3

Baxendale (1958)

First paper of
automated summarization
using first
reasonable approach

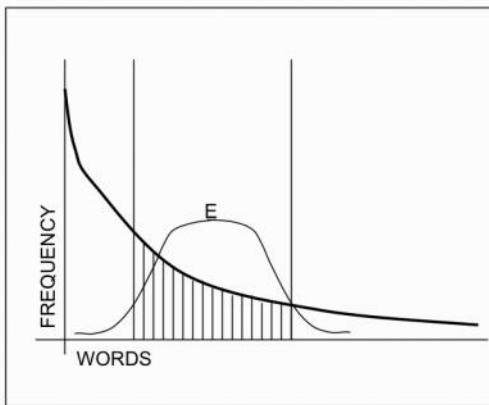
- Positional method

- Analysis of 200 paragraphs
- Pick the first and last sentences of the paragraph
- That's where the topic sentences are located
- Naïve but decent approach

Luhn (1958)

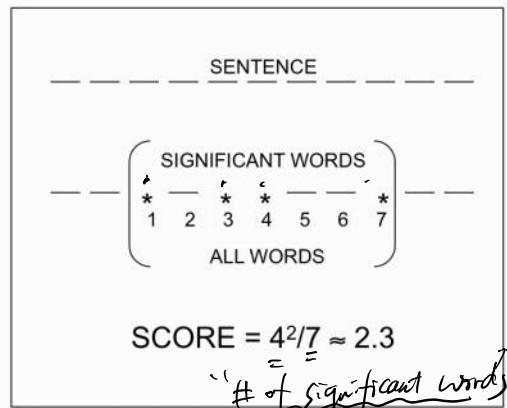
Automated Stemming

- Technical documents
- Stemming
- Stop words are removed
- Frequency of content terms



Luhn (1958)

- Sentence-level significance factor
- Look for concentrations of salient content terms



Concentration near each other $\xrightarrow{\text{is spreaded}}$

Edmundson (1969)

- Technical documents
- Position and frequency
- Cue words (bonus and stigma words)
 - Significant, hardly, impossible \rightarrow words that indicate this is an informative sentence
- Document structure
 - Is the sentence a title or heading or right under one of these
- Linear combination of the four features $\alpha_1C + \alpha_2K + \alpha_3T + \alpha_4L$ *Linear combination*

Frump (deJong 1979, 1982)

- Knowledge-based
- Slot-filling based on UPI news stories
- Based on 50 sketchy scripts — situations in the news
- Inputs matched to scripts based on manually selected keywords → hard to apply in different domains
- Difficult to port to other domains
- Missing scripts for many inputs

Frump

\$demonstration script

- The demonstrators arrive at the demonstration location.
- The demonstrators march.
- Police arrive on the scene.
- The demonstrators communicate with the target of the demonstration.
- The demonstrators attack the target of the demonstration.
- The demonstrators attack the police.

Very influential paper

G. DeJong (1979) FRUMP: Fast Reading Understanding and Memory Program

Paice (1990)

- Survey up to 1990
- Techniques that (mostly) failed
 - Syntactic criteria (Earl 1970)
 - Indicator phrases — Discourse structure phrases
- Problems with extracts
 - Lack of balance — e.g. only extract one side of a balanced discussion in a piece
 - Lack of cohesion

A ... B) ... C: He... cohesion

About scientific article: Paice (1990)

- Lack of balance
 - later approaches based on text rhetorical structure
 - Lack of cohesion
 - anaphoric reference
 - lexical or definite reference
 - rhetorical connectives
 - 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*
- flow to deal with lack of balance.*

Brandow et al. (1995)

- 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 (term frequency * inverse document frequency)
- sentence-based features:
 -
 -
 -
 -
- 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:
 - user"

about
news article

→ to discuss in
info retrieval section
high tf.idf.
⇒ more important
word

Kupiec et al. (1995)

- First trainable method
 - 20% extract
 - 188 documents from scientific journals
 - Naïve Bayes classifier
- New features
 - Sentence length ($|S| > 5$) exclude $|S| \leq 5$.
 - Presence of uppercase words (except common acronyms)
 - Thematic words
 - Set of 26 manually fixed phrases
 - Sentence position in paragraph

First use ML technique
Naïve Bayes method.

Kupiec et al. (1995)

- Uses Naïve Bayesian classifier

$$P(s \in S | F_1, F_2, \dots, F_k) = \frac{P(F_1, F_2, \dots, F_k | s \in S) P(s \in S)}{P(F_1, F_2, \dots, F_k)}$$

- Assuming statistical independence

$$P(s \in S | F_1, F_2, \dots, F_k) = \frac{\prod_{j=1}^k P(F_j | s \in S) P(s \in S)}{\prod_{j=1}^k P(F_j)}$$

Kupiec et al. (1995)

- Performance:

- For 25% summaries, 84% precision
- For smaller summaries, 74% improvement over Lead

Summons (McKeown & Radev 1995)

Input: a set of news story of the same topic.

- First work on multi-document summarization
- Approach
 - Knowledge-based
 - Information extraction (MUC templates)
 - Text generation

↓ How to generate sentences ...

Summons (McKeown & Radev 1995)

NICOSIA, Cyprus (AP) – Two bombs exploded near government ministries in Baghdad, but there was no immediate word of any casualties, Iraqi dissidents reported Friday. There was no independent confirmation of the claims by the Iraqi National Congress. Iraq's state-controlled media have not mentioned any bombings.

Multiple sources and disagreement

Explicit mentioning of “no information”.

Agreement
among diff.
sources

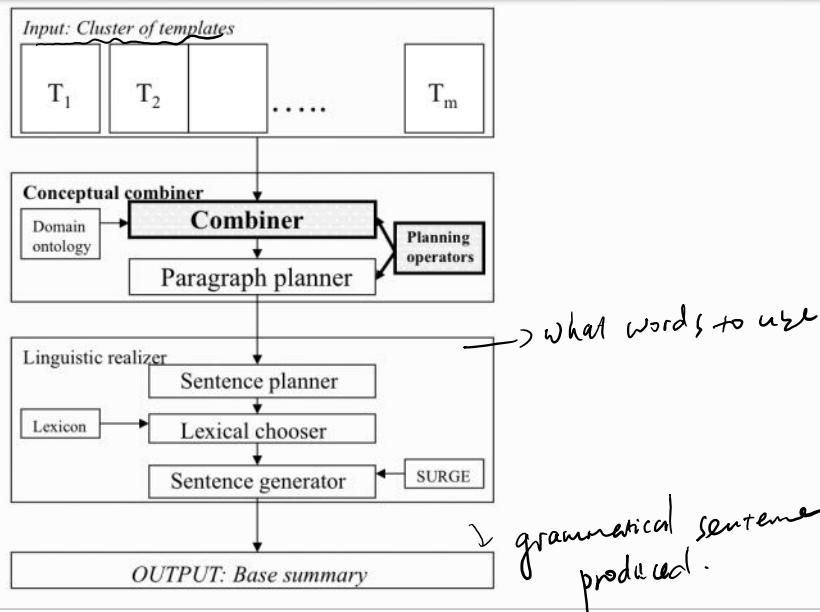
Summons

```

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      TERRORIST ACT
PERP: INCIDENT CATEGORY    "TERRORIST"
PERP: INDIVIDUAL ID
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             I: "1 CIVILIAN"
HUM TGT: FOREIGN NATION
HUM TGT: EFFECT OF INCIDENT DEATH: "1 CIVILIAN"
HUM TGT: TOTAL NUMBER

```

type of
events etc.





Summons

MESSAGE: ID SEC SOURCE: SOURCE SEC SOURCE: DATE PRIM SOURCE: SOURCE INCIDENT: DATE INCIDENT: LOCATION INCIDENT: TYPE HUM TGT: NUMBER PERP: ORGANIZATION ID	TST-REU-0001 Reuters March 3, 1996 11:30 March 3, 1996 Jerusalem Bombing "killed: 18" "wounded: 10"	MESSAGE: ID SEC SOURCE: SOURCE SEC SOURCE: DATE PRIM SOURCE: SOURCE INCIDENT: DATE INCIDENT: LOCATION INCIDENT: TYPE HUM TGT: NUMBER PERP: ORGANIZATION ID	TST-REU-0002 Reuters March 4, 1996 07:20 Israel Radio March 4, 1996 Tel Aviv Bombing "killed: at least 10" "wounded: more than 100"
MESSAGE: ID SEC SOURCE: SOURCE SEC SOURCE: DATE PRIM SOURCE: SOURCE INCIDENT: DATE INCIDENT: LOCATION INCIDENT: TYPE HUM TGT: NUMBER PERP: ORGANIZATION ID	TST-REU-0003 Reuters March 4, 1996 14:20 March 4, 1996 Tel Aviv Bombing "killed: at least 13" "wounded: more than 100" "Hamas"	MESSAGE: ID SEC SOURCE: SOURCE SEC SOURCE: DATE PRIM SOURCE: SOURCE INCIDENT: DATE INCIDENT: LOCATION INCIDENT: TYPE HUM TGT: NUMBER PERP: ORGANIZATION ID	TST-REU-0004 Reuters March 4, 1996 14:30 March 4, 1996 Tel Aviv Bombing "killed: at least 12" "wounded: 105"



Summons

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.

Summons

- 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...

Summons

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.

Summons

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.

Summons

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.

Add updated info.

Agreement

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

Summons

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.

NLP

Graph-based summarization tech



9.3



NLP

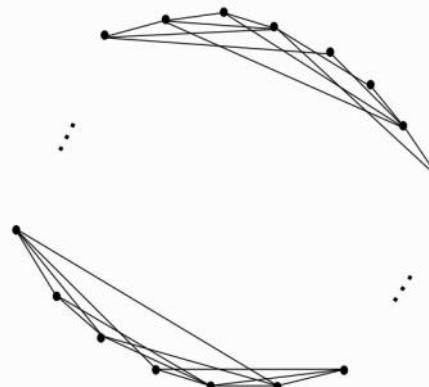


Introduction to NLP

*Summarization Techniques 1/3
(cont'd)*

Mitra/Allen/Salton

- Encyclopedia articles
- Semantic hyperlinks
 - Between pairs of paragraphs with lexical similarity above a threshold
 - Paths linking highly connected paragraphs are more likely to contain information central to the topic of the article



NLP

Introduction to NLP

Summarization Techniques 2/3

Mani/Bloedorn (1997, 1999)

- Graph-based method for identifying similarities and differences between documents
 - Single event or sequence of events
 - Content entities (nodes) and relations (edges)
 - Relation types
 - SAME, ADJACENT, ALPHA (WordNet, NetOwl), COREF
 - Spreading activation
- Nodes : entities
relation : edges.
- 1.25: "We are clear: the liberation of all our comrades, or we die with all the hostages," a rebel who did not give his name told a local radio station in a telephone call from inside the compound. ADJ
- 1.28: Many leaders of the Tupac Amaru which is smaller than Peru's Maoist Shining Path movement are in jail. 1.29 Its chief Victor Polay, was captured in June 1992 and is serving a life sentence, as is his lieutenant, Peter Cardenas. ALPHA
- 1.30: Other top commanders conceded defeat and surrendered in July 1993.
- identify most important nodes that should be highlighted.

Barzilay & M. Elhadad (1997)

- Lexical chains

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.

... with the longest lexical chain

Identify the longest lexical chain
pick the one that contains as many as
possible these chains.

Barzilay and M. Elhadad (1997)

- WordNet-based NLTK available
- Three types of relations:
 - extra-strong (repetitions)
 - strong (WordNet relations)
 - medium-strong (link between synsets is longer than one + some additional constraints)

Barzilay and M. Elhadad (1997)

- Scoring chains:
 - Length
 - Homogeneity index:
 $= 1 - \# \text{ distinct words in chain}$

$$\text{Score} = \text{Length} * \text{Homogeneity}$$

$$\text{Score} > \text{Average} + 2 * \text{st.dev.}$$

Marcu (1997-1999)

- Focuses on text coherence
- Based on Rhetorical Structure Theory (Mann and Thompson 1988)
 - Nucleus+Satellite
- Example: evidence
 - N+S combination increases reader's belief in N

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

they're in U Toronto

rhetorical
structure

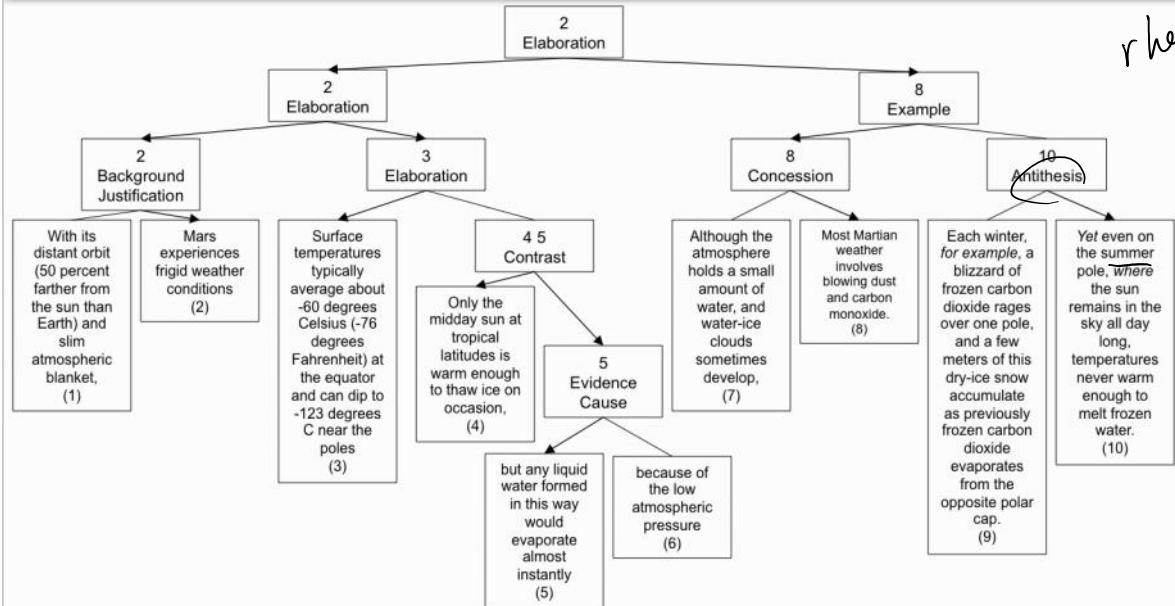
→ should go in sum

→ optimized - not in sum

Nuclear
+
Satellite

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)	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)
---	--	---	---	--	---	--	---	---	---

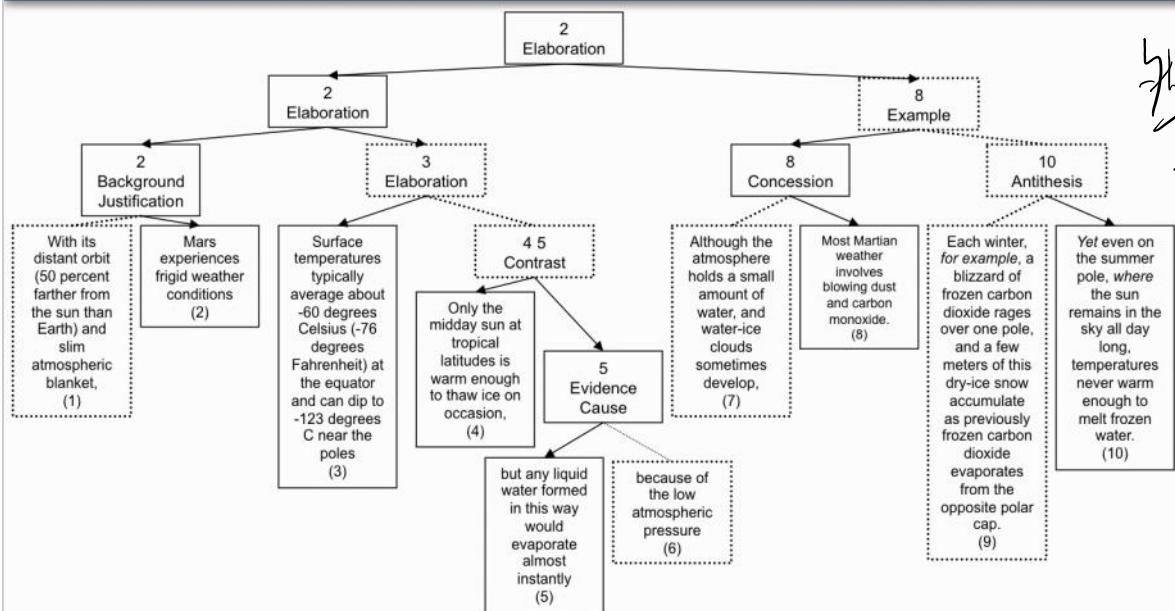
09.03 Summarization Techniques 2/3



M rhetorical
perched
tree.

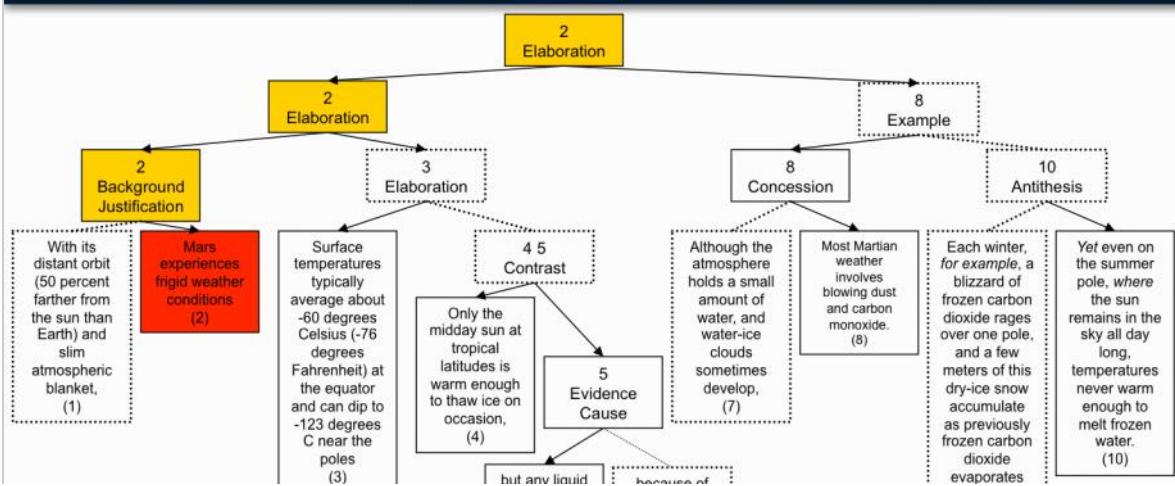
even on
summer
where
the sun
shines in the
all day
long,
temperatures
are very warm
though to
frozen
water.
(10)

09.03 Summarization Techniques 2/3

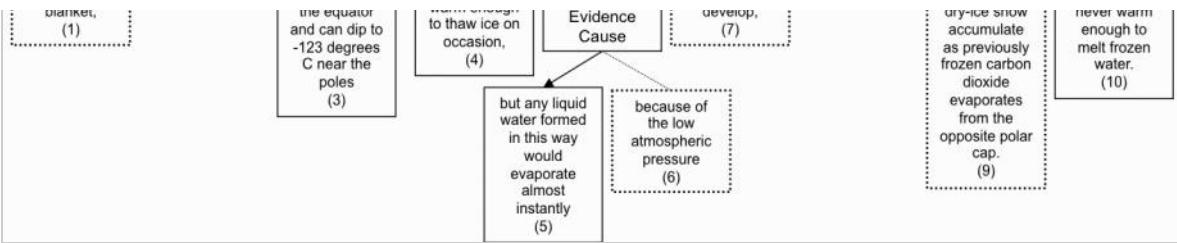


4. 7. 2018
2. 重慶 3. 8

09.03 Summarization Techniques 2/3

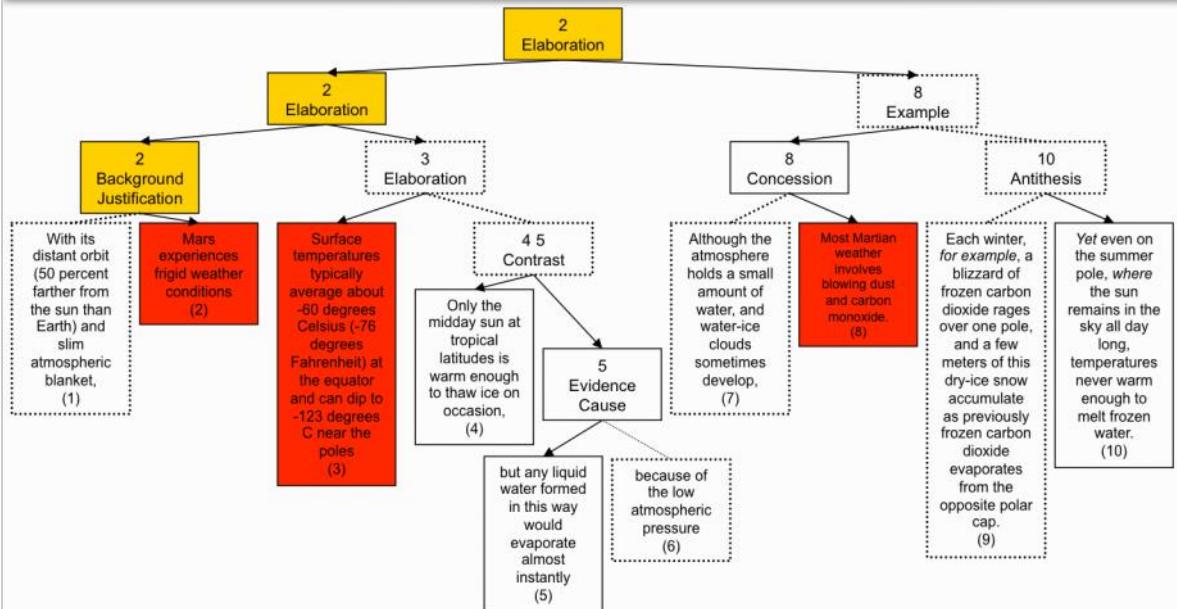


satellite



09.03 Summarization Techniques 2/3

NATURAL LANGUAGE PROCESSING



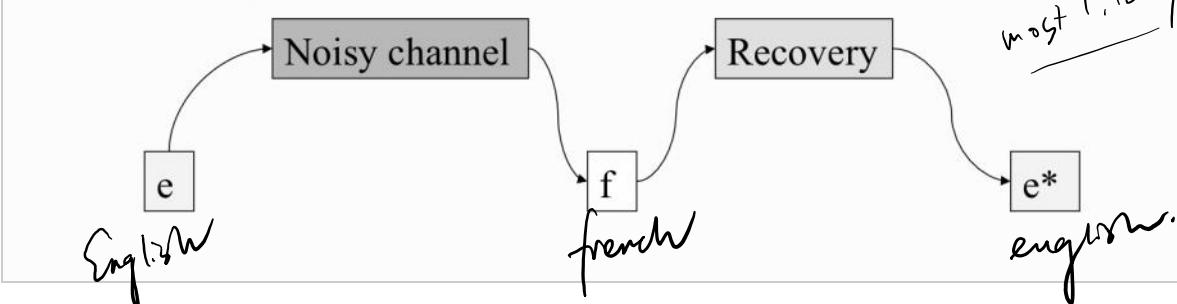
09.03 Summarization Techniques 2/3

NATURAL LANGUAGE PROCESSING



Noisy Channel Models

- Source/target language
- Coding process



Berger and Mittal 2000

- Source language
 - full document
- Target language
 - Summary
- Gisting (OCELOT)

$$g^* = \underset{g}{\operatorname{argmax}} p(g|d) = \underset{g}{\operatorname{argmax}} p(g) \cdot p(d|g)$$

$= g_{\text{gist}}$

→ consider
 headline gen
 consider the doc in
 one language
 & headline is a
 shorter sum (sum)
 of the doc

Berger & Mittal 2000

- Training on 100K summary+document pairs
- Testing on 1046 pairs
- Use Viterbi-type search
- Evaluation: word overlap (0.2–0.4)
- No word ordering

Viterbi decoding to identify
 most likely.

Berger & Mittal 2000

Sample output:

Audubon society atlanta area savannah georgia chatham
and local birding savannah keepers chapter of the audubon
georgia and leasing

headline produced from
fa (only doc)
I include word
but still preserve grammar
structure.

Carbonell and Goldstein (1998)

- Maximal marginal relevance
- Greedy selection method
- Query-based summaries
- Law of diminishing returns

C = doc collection
Q = user query
R = IR(C, Q, θ)
S = already retrieved documents
Sim = similarity metric used

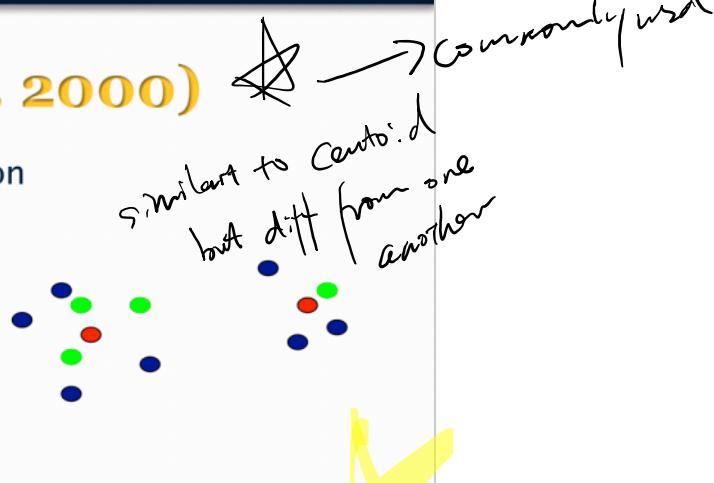
$$\text{MMR} = \underset{D_i \in S}{\operatorname{argmax}} [\lambda \underset{j \in R}{\text{Sim}_1(D_j, Q)} - (1-\lambda) \max_{D_j \in S} \underset{i \in S}{\text{Sim}_2(D_i, D_j)}]$$

don't want to get too many doc similar to each other

HL: penalize similarity
with already
Selected
document

Mead (Radev et al. 2000)

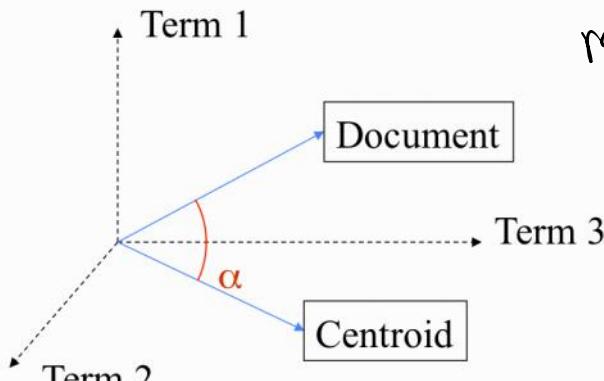
- Salience-based extractive summarization
- In 6 languages
- Centroid-based method
- Single and multi document
- Vector space model
- Features: position, length, centroid
- Cross-document structure theory
- Reranker – similar to MMR



- features, position, length, centroid
- Cross-document structure theory
- Reranker – similar to MMR
- Open source library (www.summarization.com/mead)



Mead



ref. a doc as
vector

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

Mead

- **Input**
 - Cluster of d documents with n sentences (compression rate = r)
- **Output**
 - $(n * r)$ sentences from the cluster with the highest scores

$$SCORE(s) = \sum_i (w_c C_i + w_p P_i + w_f F_i)$$

NewsInEssence (Radev et al. 2001)

- Web-based multi-document news summarization system

Top Left Screenshot: Interactive Multi-source News Summarization. Shows a news item about a Florida fireworks blast on July 2, 2003, with a summary and a link to the full article.

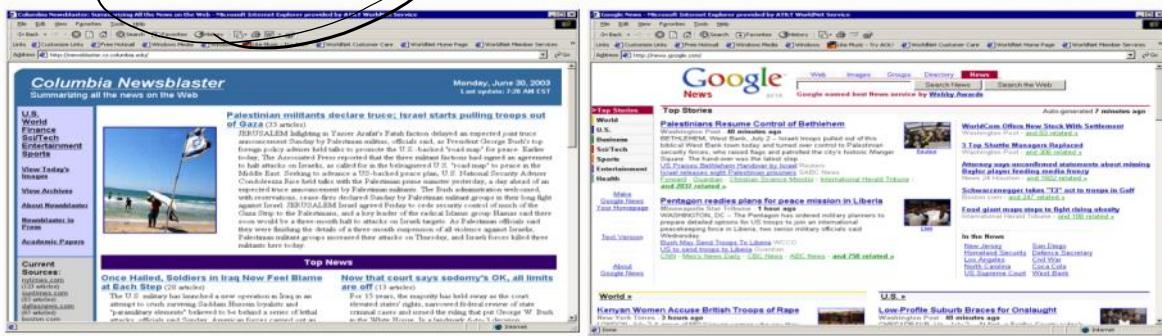
Top Right Screenshot: Interactive Multi-source News Summarization. Shows a news item about a Swedish hijack attempt on July 2, 2003, with a summary and a link to the full article.

Bottom Left Screenshot: Interactive Multi-source News Summarization. Shows a news item about a Florida fireworks blast on July 2, 2003, with a summary and a link to the full article.

Bottom Right Screenshot: Interactive Multi-source News Summarization. Shows a news item about a Swedish hijack attempt on July 2, 2003, with a summary and a link to the full article.

Other News Summarization Systems

- Newsblaster (McKeown et al. 2002)
- Google News



NLP



9.4



NLP



Introduction to NLP

Summarization Techniques 3/3

Conroy and O'Leary (2001)

- Using Hidden Markov Models
- Takes into account the local dependencies between sentences
- Features
 - Position, number of terms, similarity to document terms
- HMM alternates between summary and non-summary states

(local dependence)
Don't want to assume independent sentences

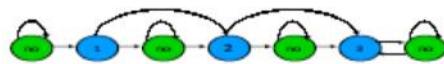


Figure 1: Summary Extraction Markov Model to Extract 2 Lead Sentences and Additional Supporting Sentences

Osborne (2002)

- Don't assume feature independence
- Use maxent (log-linear) models
- Better than Naïve Bayes
- Features
 - Sentence length
 - Sentence position
 - Inside introduction
 - Inside conclusion

Maxent. (log-linear)
Better than Naïve Bayes

Lexrank (Erkan and Radev 2004)

- Single and multi-document summarization
- Lexical Centrality
 - Represent text as graph
 - Graph centrality
 - Graph clustering
 - Random walks

Random Walk.
 for multiple doc summ.
 Is a sentence more likely
 visited → the sentence
 ⇒ more likely
 to be summ

doc1 sent1

Lexrank

- 1 (d1s1) Iraqi Vice President Taha Yassin Ramadan announced today, Sunday, that Iraq refuses to back down from its decision to stop cooperating with disarmament inspectors before its demands are met.
- 2 (d2s1) Iraqi Vice president Taha Yassin Ramadan announced today, Thursday, that Iraq rejects cooperating with the United Nations except on the issue of lifting the blockade imposed upon it since the year 1990.
- 3 (d2s2) Ramadan told reporters in Baghdad that "Iraq cannot deal positively with whoever represents the Security Council unless there was a clear stance on the issue of lifting the blockade off of it."
- 4 (d2s3) Baghdad had decided late last October to completely cease cooperating with the inspectors of the United Nations Special Commission (UNSCOM), in charge of disarming Iraq's weapons, and whose work became very limited since the fifth of August, and announced it will not resume its cooperation with the Commission even if it were subjected to a military operation.
- 5 (d3s1) The Russian Foreign Minister, Igor Ivanov, warned today, Wednesday against using force against Iraq, which will destroy, according to him, seven years of difficult diplomatic work and will complicate the regional situation in the area.
- 6 (d3s2) Ivanov contended that carrying out air strikes against Iraq, who refuses to cooperate with the United Nations inspectors, "will end the tremendous work achieved by the international group during the past seven years and will complicate the situation in the region."
- 7 (d3s3) Nevertheless, Ivanov stressed that Baghdad must resume working with the Special Commission in charge of disarming the Iraqi weapons of mass destruction (UNSCOM).
- 8 (d4s1) The Special Representative of the United Nations Secretary-General in Baghdad, Prakash Shah, announced today, Wednesday, after meeting with the Iraqi Deputy Prime Minister Tariq Aziz, that Iraq refuses to back down from its decision to cut off cooperation with the disarmament inspectors.
- 9 (d5s1) British Prime Minister Tony Blair said today, Sunday, that the crisis between the international community and Iraq "did not end" and that Britain is still "ready, prepared, and able to strike Iraq."
- 10 (d5s2) In a gathering with the press held at the Prime Minister's office, Blair contended that the crisis with Iraq "will not end until Iraq has absolutely and unconditionally respected its commitments" towards the United Nations.
- 11 (d5s3) A spokesman for Tony Blair had indicated that the British Prime Minister gave permission to British Air Force Tornado planes stationed in Kuwait to join the aerial bombardment against Iraq.

strongly connected in the graph

Lexrank

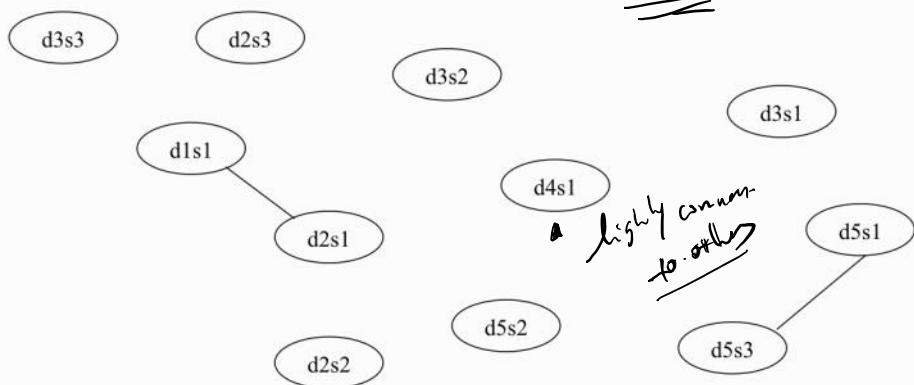
	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.45	0.02	0.17	0.03	0.22	0.03	0.28	0.06	0.06	0.00
2	0.45	1.00	0.16	0.27	0.03	0.19	0.03	0.21	0.03	0.15	0.00
3	0.02	0.16	1.00	0.03	0.00	0.01	0.03	0.04	0.00	0.01	0.00
4	0.17	0.27	0.03	1.00	0.01	0.16	0.28	0.17	0.00	0.09	0.01
5	0.03	0.03	0.00	0.01	1.00	0.29	0.05	0.15	0.20	0.04	0.18
6	0.22	0.19	0.01	0.16	0.29	1.00	0.05	0.29	0.04	0.20	0.03
7	0.03	0.03	0.03	0.28	0.05	0.05	1.00	0.06	0.00	0.00	0.01
8	0.28	0.21	0.04	0.17	0.15	0.29	0.06	1.00	0.25	0.20	0.17
9	0.06	0.03	0.00	0.00	0.20	0.04	0.00	0.25	1.00	0.26	0.38
10	0.06	0.15	0.01	0.09	0.04	0.20	0.00	0.20	0.26	1.00	0.12
11	0.00	0.00	0.00	0.01	0.18	0.03	0.01	0.17	0.38	0.12	1.00

Lexrank

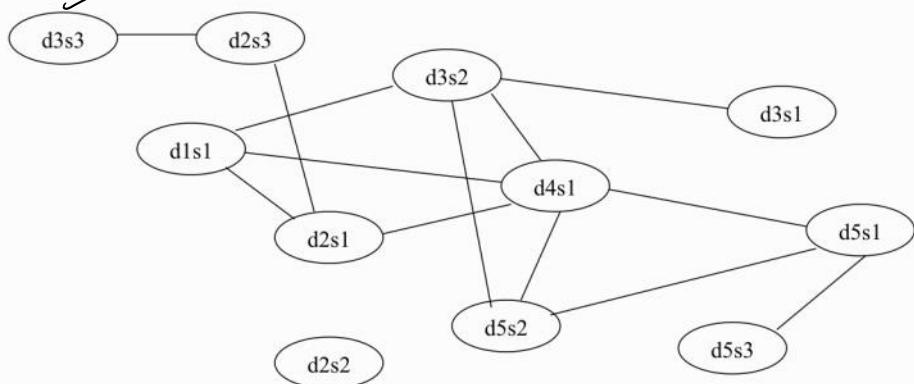
	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.45	0.02	0.17	0.03	0.22	0.03	0.28	0.06	0.06	0.00
2	0.45	1.00	0.16	0.27	0.03	0.19	0.03	0.21	0.03	0.15	0.00
3	0.02	0.16	1.00	0.03	0.00	0.01	0.03	0.04	0.00	0.01	0.00
4	0.17	0.27	0.03	1.00	0.01	0.16	0.28	0.17	0.00	0.09	0.01
5	0.03	0.03	0.00	0.01	1.00	0.29	0.05	0.15	0.20	0.04	0.18
6	0.22	0.19	0.01	0.16	0.29	1.00	0.05	0.29	0.04	0.20	0.03
7	0.03	0.03	0.03	0.28	0.05	0.05	1.00	0.06	0.00	0.00	0.01
8	0.28	0.21	0.04	0.17	0.15	0.29	0.06	1.00	0.25	0.20	0.17
9	0.06	0.03	0.00	0.00	0.20	0.04	0.00	0.25	1.00	0.26	0.38
10	0.06	0.15	0.01	0.09	0.04	0.20	0.00	0.20	0.26	1.00	0.12
11	0.00	0.00	0.00	0.01	0.18	0.03	0.01	0.17	0.38	0.12	1.00

Cosine Centrality ($t=0.3$)

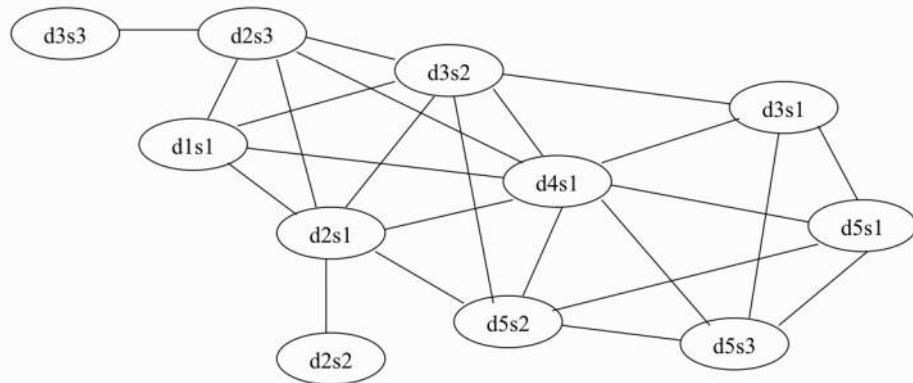
interesting



Cosine Centrality ($t=0.2$)

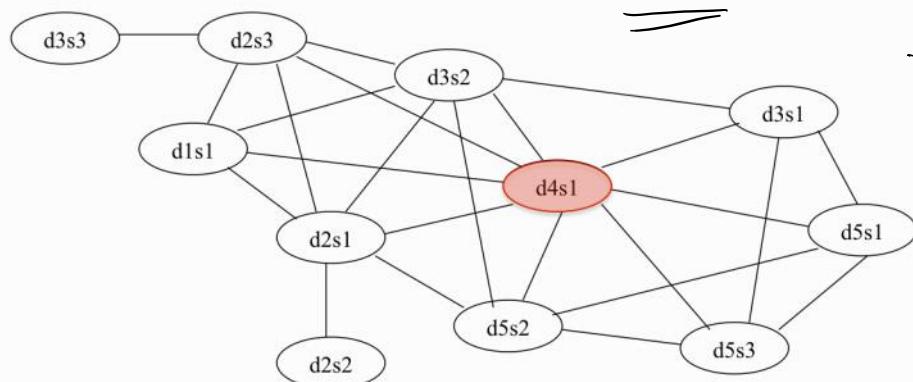


Cosine Centrality ($t=0.1$)



Sentences vote for the most central sentence!

Cosine Centrality ($t=0.1$)



Sentences vote for the most central sentence!

Lexrank (Advanced Material)

- Square connectivity matrix
- Directed vs. undirected
- An eigenvalue for a square matrix A is a scalar λ such that there exists a vector $x \neq 0$ such that $Ax = \lambda x$
- The normalized eigenvector associated with the largest λ is called the principal eigenvector of A
- A matrix is called a stochastic matrix when the sum of entries in each row sum to 1 and none is negative. All stochastic matrices have a principal eigenvector

How
lexrank
works
eigenvalue ✓

All stochastic
matrix
has eigen
vector

Lexrank (Advanced Material)

- The connectivity matrix used in PageRank [Page & al. 1998] is irreducible [Langville & Meyer 2003]
- An iterative method (power method) can be used to compute the principal eigenvector
- That eigenvector corresponds to the stationary value of the Markov stochastic process described by the connectivity matrix
- This is also equivalent to performing a random walk on the matrix

PageRank

PageRank

Markov Stochastic matrix

Lexrank (Advanced Material)

- The stationary value of the Markov stochastic matrix can be computed using an iterative power method:

$$p = E^T p \rightarrow \text{vector of values that conv. to stationary}$$

$$(I - E^T)p = 0$$
- PageRank adds an extra twist to deal with dead-end pages. With a probability $1-\varepsilon$, a random starting point is chosen. This has a natural interpretation in the case of Web page ranking

$$p(v) = \frac{1-\varepsilon}{n} + \varepsilon \sum_{u \in pr[v]} \frac{p(u)}{|su[u]|}$$

su = successor nodes
pr = predecessor nodes



$$p(v) = \frac{1 - \varepsilon}{n} + \varepsilon \sum_{u \in pr[v]} \frac{p(u)}{|su[u]|}$$

su = successor nodes
pr = predecessor nodes

- Eigenvector centrality: the paths in the random walk are weighted by the centrality of the nodes that the path connects

*A
use
many years.*

Gong and Liu (2001)

- Using Latent Semantic Analysis (LSA)
- Single and multi-document
- Not using WordNet
- Each document is represented as a word by sentence matrix (row=word, column=sentence)
- TF*IDF weights in the matrix
- SVD: $A = USV^T$
- The rows of V^T are independent topics
- Select sentences that cover these independent topics

*1st paper using
Latent Semantic Analysis*

*Pick sentence that
cover
topics.*

NLP



9.5

09.05 Summarization Evaluation

NATURAL LANGUAGE
PROCESSING



NLP

09.05 Summarization Evaluation

NATURAL LANGUAGE
PROCESSING



Introduction to NLP

Summarization Evaluation

Evaluation Criteria

- Appropriate length
- Fidelity *faithful*.
- Salience
- Grammaticality
- Non-redundancy
- Referential well-formedness
- Structure and coherence *structure-property
coherent*

Ideal Evaluation

Information content

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

$$R = \text{Retention Ratio} = \frac{i(S)}{i(D)}$$

$$R > C$$

*What part of
original doc. is sum.
→ the info. in sum.
want $\frac{R}{C}$.
??*

Types Of Evaluation Methods

- Extrinsic techniques (task-based)
 - Can you make the same decision using the summary as with the full text, but in less time
- Intrinsic techniques
 - Comparing summaries against gold standards

Precision and Recall

	Relevant	Non-relevant
System: relevant	A	B
System: non-relevant	C	D

$$\text{Precision: } P = \frac{A}{A + B}$$

$$\text{Recall: } R = \frac{A}{A + C}$$

Comparable to doc retrieval.

a special case of this is

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

which is geometric mean

high score
for recall
for length

Rouge (Lin and Hovy 2003)

- Very convenient for prototyping
- Based on Papineni et al's BLEU (used for MT) → *high score for high precision*
- R stands for Recall
- ROUGE-n is a measure of n-gram overlap between a summary and a set of reference summaries
- ROUGE-L uses longest common subsequence instead of n-gram overlap
- Correlated with manual evaluations when averaged over many examples
- It can be easily gamed → *but you need to have human generated gold-standard summaries*

Relative Utility (Radev et al. 2000)

	Ideal	System 1	System 2
S1	+	+	-
S2	+	+	+
S3	-	-	-
S4	-	-	+
S5	-	-	-
S6	-	-	-
S7	-	-	-
S8	-	-	-
S9	-	-	-
S10	-	-	-

but maybe
these are
other "ideal"

	Ideal	System 1	System 2
S1	10(+)	10(+)	5
S2	8(+)	9(+)	8(+)
S3	2	3	4
S4	7	6	9(+)

assign utility score
for each sentence. ↗ not
for bad.

$$U(\text{system}, \text{ideal}) = \% \text{ of ideal utility covered by system summary}$$

Relative Utility

$$\text{RU} =$$

17

	Judge1	Judge2	Judge3
Sentence 1	10	10	5
Sentence 2	8	9	8
Sentence 3	2	3	4
Sentence 4	5	6	9

Relative Utility

$$\text{RU} = \frac{13}{17}$$

$$= 0.765$$

Ask judges to
Score the
Sentences?

	Judge1	Judge2	Judge3
Sentence 1	10	10	5
Sentence 2	8	9	8
Sentence 3	2	3	4
Sentence 4	5	6	9

Normalized System Performance

	Judge 1	Judge 2	Judge 3	Average
Judge 1	1.000	1.000	0.765	0.883
Judge 2	1.000	1.000	0.765	0.883
Judge 3	0.722	0.789	1.000	0.756

will both pick sent 82

6 pairs +
relative weights

what's interjudge
agreement

$$D = \frac{(S-R)}{(J-R)}$$

Normalized system performance

System performance

Random performance

Interjudge agreement

This you don't
expect sys to
perform between
them judges

Random Performance

average of all

$$\frac{n!}{(n(1-r))! (r^n)!}$$

systems

$$D = \frac{(S-R)}{(J-R)}$$

$\frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}}$
 $\frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}}$
 $\frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}}$

{12} {13} {14} {23} {24} {34}

$D = \frac{1}{\sqrt{n}}$
 $S = R$
 $\Rightarrow D = 1$
 $S = R$
 $\Rightarrow D = 0$
 $\Rightarrow D = 1$
 $\Rightarrow D = 0$

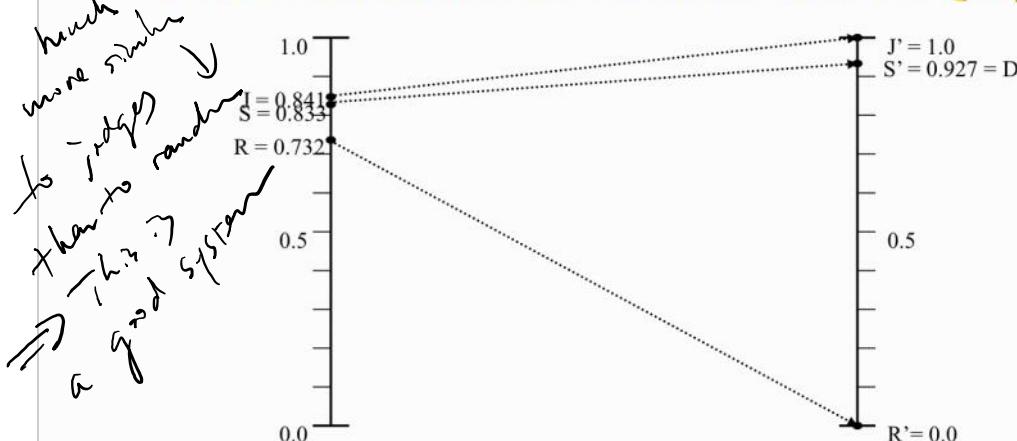
{12} {13} {14} {23} {24} {34}

Example

$$D \{14\} = \frac{(S-R)}{(J-R)} = \frac{0.833 - 0.732}{0.841 - 0.732} = 0.927$$

$$D \{24\} = 0.963$$

Normalized Evaluation of {14}



09.05 Summarization Evaluation

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 Chbika, 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 Moulay Sissen 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 Djelfa, 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."

09.05 Summarization Evaluation

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 Chbika, 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 Moulay Sissen 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.

same content

=> part include

both

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 Djelfa, 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."

Subsumption and Equivalence

- *Subsumption*: If the information content of sentence a (denoted as $I(a)$) is contained within sentence b, then a becomes informationally redundant and the content of b is said to subsume that of a:

$$I(a) \subset I(b) \quad I(a) \subset I(b) \Rightarrow \text{skip. } a$$

- *Equivalence*: If $I(a) \subset I(b) \wedge I(b) \subset I(a)$

Subsumption

- (1) John Doe was found guilty of the murder.
- (2) The court found John Doe guilty of the murder of Jane Doe last August and sentenced him to life.

Subsumption

	Article 1	Article 2	Article 3
S1	10	→ 10	5
S2	8	9	8
S3	2	3	→ 4
S4	7	6	→ 9

3 articles
 want to pick
 2 sentences
 $\underline{S1A1} \underline{S1A2}$
 What
 $S1A1 \subset S1A2$?
 Then don't want
 to pick $S1A1$

Subsumption (Cont'd)

$$SCORE(s) = \sum_i (w_c C_i + w_p P_i + w_f F_i) \quad \underbrace{w_R R_s}_{\text{to count the redundant sentences.}}$$

$$R_s = \underbrace{\text{cross-sentence word overlap}}$$

$$R_s = \frac{2 * (\# \text{ overlapping words})}{(\# \text{ words in sentence 1} + \# \text{ words in sentence 2})}$$

$$w_R = \underbrace{\text{Max}_s (SCORE(s))}$$

Pyramid

- Nenkova and Passonneau 2004
- Based on Semantic Content Units (SCU)
- Used for multi-document summarization
- Different surface realizations with equivalent meanings

*in 2 diff ways
to say the same
thing*

Pyramid

Varia^r of relevant knowledge base

4 points

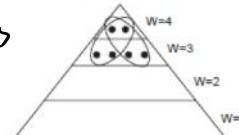


Figure 2: Two of six optimal summaries with 4 SCUs

A1 In 1998 two Libyans indicted in 1991 for the Lockerbie bombing were still in Libya.

B1 Two Libyans were indicted in 1991 for blowing up a Pan Am jumbo jet over Lockerbie, Scotland in 1988.

C1 Two Libyans, accused by the United States and Britain of bombing a New York bound Pan Am jet over Lockerbie, Scotland in 1988, killing 270 people, for 10 years were harbored by Libya who claimed the suspects could not get a fair trial in America or Britain.

D2 Two Libyan suspects were indicted in 1991.

SCU1 (w=4): two Libyans were officially accused of the Lockerbie bombing

A1 [two Libyans]1 [indicted]1

B1 [Two Libyans were indicted]1

C1 [Two Libyans,]1 [accused]1

D2 [Two Libyan suspects were indicted]1

SCU2 (w=3): the indictment of the two Lockerbie suspects was in 1991

A1 [in 1991]2

B1 [in 1991]2

D2 [in 1991.]2

Available Corpora

- DUC and TAC corpora
 - <http://duc.nist.gov>
 - <http://tac.nist.gov>
- SummBank corpus
 - <http://www.summarization.com/summbank>
- SUMMAC corpus

→ include Chinese

- <http://www.summarization.com/summbank>
- SUMMAC corpus
- NY Times corpus (from LDC)

*Newspaper
non extractive
summarization*

NLP



9.6

NLP

Introduction to NLP

Sentence Simplification

*Going beyond
sentence simplification*

*challenge
in text
summarization*

Sentence Simplification

- Removing some parts of sentences
 - Quotes
 - Appositives
 - Adjectives and adverbs
 - Embedded clauses
 - Attribution clauses
- Applications
 - Subtitling
 - Headline generation
 - Mobile devices
 - Applications for the visually impaired

Part of sentence
Part of sentence
Part of sentence

Knight and Marcu 2000

- Use structured (syntactic) information
- Two approaches
 - Noisy channel
 - Decision based

$P_{exp}(NP \rightarrow DT\ NN \mid NP \rightarrow DT\ NN) = 0.8678$
 $P_{exp}(NP \rightarrow DT\ JJ\ NN \mid NP \rightarrow DT\ NN) = 0.0287$
 $P_{exp}(NP \rightarrow DT\ NNP\ NN \mid NP \rightarrow DT\ NN) = 0.0230$
 $P_{exp}(NP \rightarrow DT\ JJS\ NN \mid NP \rightarrow DT\ NN) = 0.0115$
 $P_{exp}(NP \rightarrow DT\ NNP\ CD\ NN \mid NP \rightarrow DT\ NN) = 0.0057$

Knight and Marcu

- The documentation is typical of Epson quality, excellent.
 - Documentation is excellent.
- All of our design goals were achieved and the delivered performance matches the speed of the underlying device.
 - All design goals were achieved.
- Although the modules themselves may be physically and/or electronically incompatible, the cable-specific jacks on them provide industry-standard connections.
 - Cable-specific jacks provide industry-standard connections.
- Beyond the basic level, the operations of the three products vary widely.
 - The operations of the three products vary widely.
- Arborscan is reliable and worked accurately in testing, but it produces very large dxf files.
 - Arborscan produces very large dxf files.
- Many debugging features, including user-defined break points and variable-watching and message-watching windows, have been added.
 - Many debugging features have been added.

English Wikipedia

- Anthony Charles Lynton Blair (born 6 May 1953)^[1] is a former British Labour Party politician who served as the Prime Minister of the United Kingdom from 2 May 1997 to 27 June 2007. He was the Member of Parliament (MP) for Sedgefield from 1983 to 2007 and Leader of the Labour Party from 1994 to 2007. He resigned from all of these positions in June 2007.
- Tony Blair was elected Leader of the Labour Party in the leadership election of July 1994, following the sudden death of his predecessor, John Smith. Under his leadership, the party adopted the term "New Labour"^[2] and moved away from its traditional left wing position towards the centre ground.^{[3][4]} Blair subsequently led Labour to a landslide victory in the 1997 general election. At 43 years old, he became the youngest Prime Minister since Lord Liverpool in 1812. In the first years of the New Labour government, Blair's government implemented a number of 1997 manifesto pledges, introducing the minimum wage, Human Rights Act and Freedom of Information Act, and carrying out regional devolution, establishing the Scottish Parliament, the National Assembly for Wales, and the Northern Ireland Assembly.
- Blair's role as Prime Minister was particularly visible in foreign and security policy, including in Northern Ireland, where he was involved in the 1998 Good Friday Agreement. From the start of the War on Terror in 2001, Blair strongly supported the foreign policy of US President George W. Bush, notably by participating in the 2001 invasion of Afghanistan and 2003 invasion of Iraq. Blair is the Labour Party's longest-serving Prime Minister, the only person to have led the Labour Party to three consecutive general election victories, and the only Labour Prime Minister to serve consecutive terms more than one of which was at least four years long.
- He was succeeded as Leader of the Labour Party on 24 June 2007 and as Prime Minister on 27 June 2007 by Gordon Brown.^[5] On the day he resigned as Prime Minister, he was appointed the official Envoy of the Quartet on the Middle East. In May 2008, Blair launched his Tony Blair Faith Foundation.^[6] This was followed in July 2009 by the launching of the Faith and Globalisation Initiative with Yale University in the USA, Durham University in the UK and the National University of Singapore in Asia to deliver a postgraduate programme in partnership with the Foundation.^{[7][8]}

Simple English Wikipedia

- **Anthony Charles Lyton Blair**, usually called **Tony Blair**, is a former [Prime Minister](#) of the [United Kingdom](#). He was born in [Edinburgh](#). He was Prime Minister from May 1997 until June 2007 and was succeeded by the former [Chancellor of the Exchequer](#), [Gordon Brown](#).
- As leader of the Labour party, he won three general [elections](#) in the UK, in [1997](#), [2001](#) and [2005](#). He is married to [Cherie Booth](#). They met on the top deck of a [double-decker bus](#) in 1975. Cherie claimed "It was a double-decker and we went upstairs. It was completely empty and by the time we got off we knew each other better than when we'd got on. And even better the next morning. He was a very good-looking young man, tall and slim, yet broad in the shoulders. A really strong body."^[1]. Cherie Blair is a [lawyer](#), who graduated from the [London School of Economics](#) with a first class honours degree. Blair himself left [Oxford University](#) with a second class degree. They have four children, Euan, Nicky, Kathryn, and Leo. There was a controversy over Blair sending his eldest son Euan to a [grant-maintained](#) school. As a result of this, [Alastair Campbell](#) discovered Blair "standing stark naked reading the [Daily Mail](#)"^[2].
- He attributes his success in politics to a pair of [lucky brogues](#) which he wore for every single [Prime Ministers Questions](#) of his leadership. He claimed that "cheap shoes are a false economy".^[3]

NLP