

The Tutorial Programme

Introduction
Definitions
Human Summarization
Automatic Summarization
Sentence Extraction
Superficial Methods
Learning to Extract
Cohesion Models
Discourse Models
Abstractive Models
Novel Techniques
Multi-document Summarization
Summarization Evaluation
Evaluation of Extracts
Pyramids Evaluation
SUMMAC Evaluation
DUC Evaluation
DUC 2004, ROUGE, and Basic Elements
SUMMAC Corpus
MEAD
Summarization in GATE
SUMMBANK
Other Corpora and Tools
Some Research Topics

Tutorial Organiser(s)

**Horacio Saggion
University of Sheffield**

Text Summarization: Resources and Evaluation

Horacio Saggion

Department of Computer Science

University of Sheffield

England, United Kingdom

<http://www.dcs.shef.ac.uk/~saggion>

Tutorial Outline

- Introduction
- Definitions
- Human summarization
- Automatic summarization
- Sentence extraction
- Superficial methods
- Learning to extract
- Cohesion models
- Discourse models
- Abstractive models
- Novel Techniques
- Multi-document summarization
- Summarization Evaluation
- Evaluation of extracts
- Pyramids evaluation
- SUMMAC evaluation
- DUC evaluation
- DUC 2004 & ROUGE & BE
- SUMMAC corpus
- MEAD
- Summarization in GATE
- SUMMBANK
- Other Corpora & Tools
- Some research topics

Renewed interest in the field

- Dagstuhl Meeting, 1993
- Association for Computational Linguistics ACL/EACL Workshop, Madrid, 1997
- AAAI Spring Symposium, Stanford, 1998
- SUMMAC '98 summarization evaluation
- Workshop on Automatic Summarization (WAS) ANLP/NAACL, Seattle, 2000.
- NAACL, Pittsburgh, 2001. Barcelona 2004.
- Document Understanding Conference (DUC) since 2000, summarization evaluation
- Multilingual Summarization Evaluation (MSE) since 2005, summarization evaluation
- Crossing Barriers in Text Summarization, RANLP 2005

The summary I want...

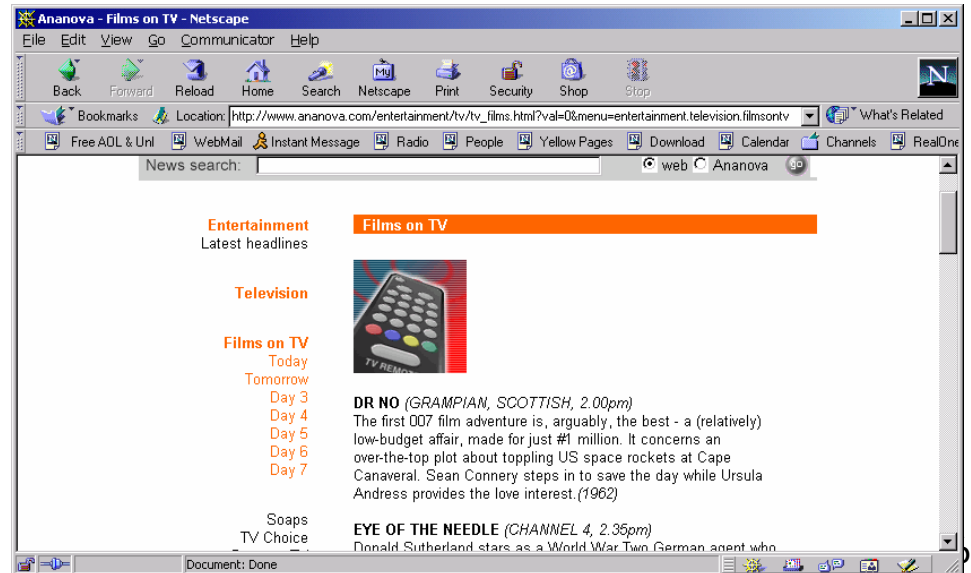
Margie was holding tightly to the string of her beautiful new balloon. Suddenly, a gust of wind caught it. The wind carried it into a tree. The balloon hit a branch and burst. Margie cried and cried.

Margie was sad when her balloon burst.

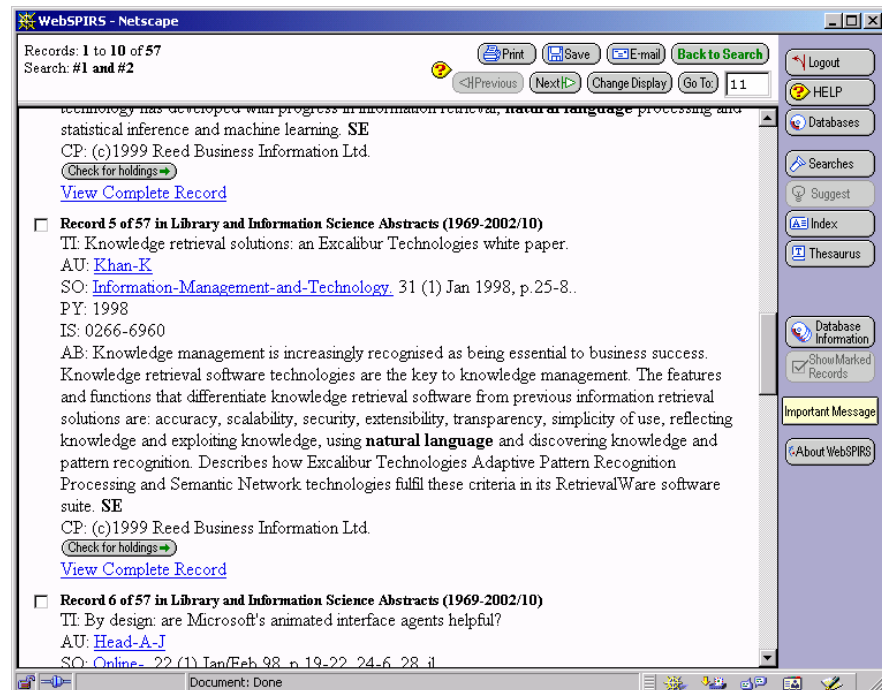
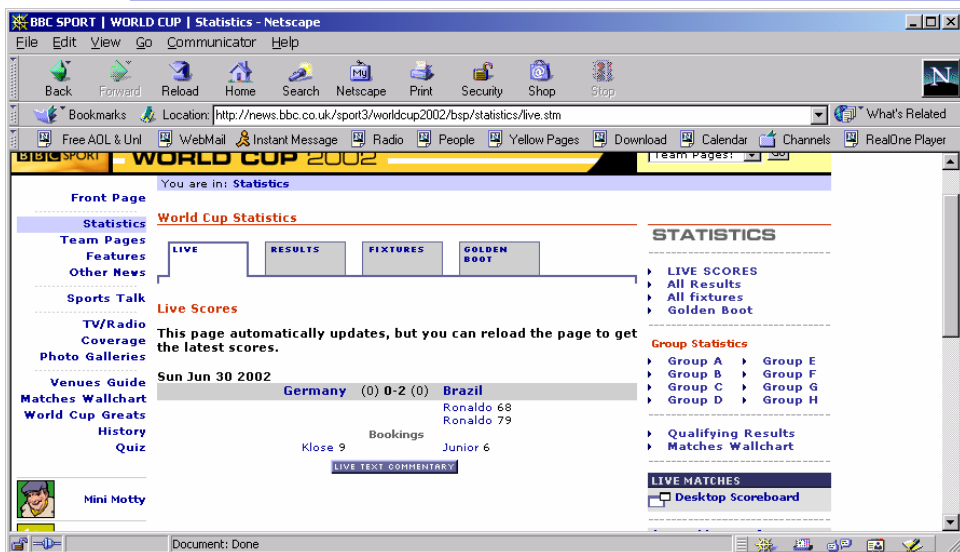
Summarization

- summary: brief but accurate representation of the contents of a document
- goal of summarization: take an information source, extract the most important content from it and present it to the user in a condensed form and in a manner sensitive to the user's needs.
 - compression: the amount of text to present or the length of the summary to the length of the source.
 - type of summary: indicative/informative
 - other parameters: topic/question

Surrounded by summaries!!!



Surrounded by summaries!!!



Some information ignored!

Alfred Hitchcock's landmark masterpiece of the macabre stars Anthony Perkins as the troubled Norman Bates, whose old dark house and adjoining motel are not the place to spend a quiet evening. No one knows that better than Marion Crane (Janet Leigh), the ill-fated traveller whose journey ends in the notorious "shower scene." First a private detective, then Marion's sister (Vera Miles) search for her, **the horror and suspense mount to a terrifying climax when the mysterious killer is finally revealed.**

Summary functions

- Direct functions
 - communicates substantial information;
 - keeps readers informed;
 - overcomes the language barrier;
- Indirect functions
 - classification;
 - indexing;

Typology

- Indicative
 - indicates types of information
 - “alerts”
- Informative
 - includes quantitative/qualitative information
 - “informs”
- Critic/evaluative
 - evaluates the content of the document

Indicative

The work of Consumer Advice Centres is examined. The information sources used to support this work are reviewed. The recent closure of many CACs has seriously affected the availability of consumer information and advice. The contribution that public libraries can make in enhancing the availability of consumer information and advice both to the public and other agencies involved in consumer information and advice, is discussed.

Informative

An examination of the work of Consumer Advice Centres and of the information sources and support activities that public libraries can offer. CACs have dealt with pre-shopping advice, education on consumers' rights and complaints about goods and services, advising the client and often obtaining expert assessment. They have drawn on a wide range of information sources including case records, trade literature, contact files and external links. The recent closure of many CACs has seriously affected the availability of consumer information and advice. Libraries can cooperate closely with advice agencies through local coordinating committees, shared premises, joint publicity referral and the sharing of professional expertise.

More on typology

- extract vs abstract
 - fragments from the document
 - newly re-written text
- generic vs query-based vs user-focused
 - all major topics equal coverage
 - based on a question “what are the causes of the war?”
 - users interested in chemistry
- for novice vs for expert
 - background
 - Just the new information

More on typology

- single-document vs multi-document
 - research paper
 - proceedings of a conference
- in textual form vs items vs tabular vs structured
 - paragraph
 - list of main points
 - numeric information in a table
 - with “headlines”
- in the language of the document vs in other language
 - monolingual
 - cross-lingual

Abstracting services

- Abstracting journals
 - not very popular today
- Abstracting databases
 - CD-ROM
 - Internet
- Mission
 - keep the scientific community informed
- LISA, CSA, ERIC, INSPEC, etc.

Professional abstracts

The screenshot shows a Netscape browser window titled "WebSPIRS - Netscape". The address bar is empty. The main content area displays search results for the query "distributed breadth first search". The results show one record from the "Library and Information Science Abstracts (1969-2002/10)" database. The record details include the title "Efficient distributed breadth-first search algorithm", author "Makki-S-A-M", source "Computer-Communications", volume and issue "19 (8)", date "Jul 96", page range "p.628-36", and reference count "refs..". The abstract (AB) describes a more efficient Distributed Breadth-First Search algorithm for asynchronous communication networks. The copyright (CP) is (c)1999 Reed Business Information Ltd. There are buttons for "Check for holdings" and "View Complete Record".

Records: 1 to 1 of 1
Search: **distributed breadth first search**

Print Save E-mail Back to Search

<HPrevious NextH> Change Display Go To: 1

Record 1 of 1 in Library and Information Science Abstracts (1969-2002/10)
TI: Efficient **distributed breadth-first search** algorithm.
AU: [Makki-S-A-M](#)
SO: [Computer-Communications](#). 19 (8) Jul 96, p.628-36. refs..
PY: 1996
IS: 0140-3664
AB: Presents a more efficient **Distributed Breadth-First Search** algorithm for an asynchronous communication network. Presents a model and gives an overview of related research. Analyzes the complexity of the algorithm, and gives some examples of performance on typical networks. SE
CP: (c)1999 Reed Business Information Ltd.
[Check for holdings](#) →
[View Complete Record](#)

Logout
HELP
Databases
Searches
Suggest
Index
Thesaurus
Database Information
Show Marked Records
Important Message
About WebSPIRS

Document: Done

Transformations during abstracting

Cremmins: The art of abstracting

| Source document | Abstract |
|--|---|
| There were significant positive associations between the concentration of the substance administered and mortality in rats and mice of both sexes. | Mortality in rats and mice of both sexes was dose related. |
| There was no convincing evidence to indicate that endrin ingestion induced any of the different types of tumors which were found in the treated animals. | No treatment related tumors were found in any of the animals. |

Abstractor's at work (Endres-Niggemeyer'95)

- systematic study of professional abstractors
- “speak-out-loud” protocols
- discovered operations during document condensation
 - use of document structure
 - top-down strategy + superficial features
 - cut-and-paste

Abstract's structure (Liddy'91)

- Identification of a text schema (grammar) of abstracts of empirical research
- Identification of lexical clues for predicting the structure
- From abstractors to a linguistic model
 - ERIC and PsycINFO abstractors as subjects of experimentation

Abstract's structure

- Three levels of information
 - proto-typical
 - hypothesis; subjects; conclusions; methods; references; objectives; results
 - typical
 - relation with other works; research topic; procedures; data collection; etc.
 - elaborated-structure
 - context; independent variable; dependent variable; materials; etc.
- Suggests that types of information can be identified based on "cue" words/expressions
- Many practical implications for IR systems

Finding source sentences (Saggion&Lapalme'02)

| Source document | Abstract |
|--|--|
| In this paper we have <u>presented</u> a more efficient distributed algorithm which constructs a breadth-first search tree in an asynchronous communication network. | <u>Presents</u> a more efficient distributed breadth-first search algorithm for an asynchronous communication network. |
| We <u>present</u> a model and <u>give</u> an overview of related research. | <u>Presents</u> a model and gives an overview of related research. |
| We <u>analyse</u> the the complexity of our algorithm and <u>give</u> some examples of performance on typical networks. | <u>Analyses</u> the complexity of the algorithm and <u>gives</u> some examples of performance on typical networks. |

Document structure for abstracting

| | |
|-----------------------|-----|
| Title | 2% |
| Author abstract | 15% |
| First section | 34% |
| Last section | 3% |
| Headings and captions | 33% |
| Other sections | 13% |

Automatic Summarization

- 50s-70s
 - Statistical techniques (scientific text)
- 80s
 - Artificial Intelligence (short texts, narrative, some news)
- 90s-
 - Hybrid systems (news, some scientific text)
- 00s-
 - Headline generation; multi-document summarization (much news, more diversity: law, medicine, e-mail, Web pages, etc.); hand-held devices; multimedia

Summarization steps

- Text interpretation
 - phrases; sentences; propositions; etc.
- Unit selection
 - some sentences; phrases; props; etc.
- Condensation
 - delete duplication, generalization
- Generation
 - text-text; propositions to text; information to text

Natural language processing

detecting syntactic structure for condensation

I: Solomon, a sophomore at Heritage School in Convers, is accused of opening fire on schoolmates.

O: Solomon is accused of opening fire on schoolmates.

meaning to support condensation

I: 25 people have been killed in an explosion in the Iraqi city of Basra.

O: Scores died in Iraq explosion

discourse interpretation/coreference

I: And as a conservative Wall Street veteran, Rubin brought market credibility to the Clinton administration.

O: Rubin brought market credibility to the Clinton administration.

I: Victoria de los Angeles died in a Madrid hospital today. She was the most acclaimed Spanish soprano of the century. She was 81.

O: Spanish soprano De los Angeles died at 81.

Summarization by sentence extraction

- extract
 - subset of sentence from the document
- easy to implement and robust
- how to discover what type of linguistic/semantic information contributes with the notion of relevance?
- how extracts should be evaluated?
 - create ideal extracts
 - need humans to assess sentence relevance

Evaluation of extracts

choosing sentences

| N | Human | System |
|---|-------|--------|
| 1 | + | + |
| 2 | - | + |
| | | |
| n | - | - |

contingency table

| | S | |
|---|----|----|
| H | + | - |
| + | TP | FN |
| - | FP | TN |

- precision

$$\frac{TP}{TP + FP}$$

- recall

$$\frac{TP}{TP + FN}$$

$$TP + FN + TN + FP = n$$

Evaluation of extracts (instance)

| N | Human | System |
|---|-------|--------|
| 1 | + | + |
| 2 | - | + |
| 3 | + | - |
| 4 | - | - |
| 5 | + | - |

| | S | |
|---|---|---|
| H | + | - |
| + | 1 | 2 |
| - | 1 | 1 |

- precision = $1/2$
- recall = $1/3$

Keyword method: Luhn'58

- words which are frequent in a document indicate the topic discussed
- stemming algorithm ("systems" = "system")
- ignore "stop words" (i.e. "the", "a", "for", "is")
- compute the distribution of each word in the document (tf)

Keyword method

- compute distribution of words in corpus (i.e., collection of texts)
- inverted document frequency

$$idf(term) = \log\left(\frac{NUMDOC}{NUMDOC(term)}\right)$$

$NUMDOC$ #docs in corpus

$NUMDOC(term)$ #docs where term occurs

Keyword method

- consider only those terms such that $tf \cdot idf > thr$
- identify clusters of keywords
 - $[X_i \ X_{i+1} \ \dots \ X_{i+n-1}]$
- compute weight
- normalize

$$\frac{\# \text{significant}(C)^2}{\# \text{words}(C)}$$

$$\text{weight}(t) = tf(t) \cdot idf(t)$$

$$\text{weight}(S) = \sum_{t \in S} \text{weight}(t)$$

Position: Edmundson'69

- Important sentences occur in specific positions
 - “lead-based” summary (Brandow'95)
 - inverse of position in document works well for the “news”

$$position(S_i) = (i)^{-1}$$

- Important information occurs in specific sections of the document (introduction/conclusion)

Position

- Extra points for sentences in specific sections
 - make a list of important sections
LIST= "introduction", "method", "conclusion",
"results", ...
- Position evidence (Baxendale'58)
 - first/last sentences in a paragraph are topical
 - give extra points to = initial | middle | final

Position

- Position depends on type of text!
- “Optimum Position Policy” (Lin & Hovy'97) method to learn “positions” which contain relevant information $OPP = \{ (p_1, s_2), (p_2, s_1), (p_1, s_1), \dots \}$
 - p_i = paragraph num; s_i = sentence num
 - “learning” method uses documents + abstracts + keywords provided by authors
 - average number of keywords in the sentence
 - 30% topic not mentioned in text
 - title contains 50% topics
 - title + 2 best positions 60% topics

Title method: Edmundson'69

- Hypothesis: title of document indicates its content
- therefore, words in title help find relevant content
- create a list of title words, remove "stop words"

$$title(S) = |TIT I S|$$

Cue method: Edmundson'69;Paice'81

- Important sentences contain cue words/indicative phrases
 - “The main aim of the present paper is to describe...” (IND)
 - “The purpose of this article is to review...” (IND)
 - “In this report, we outline...” (IND)
 - “Our investigation has shown that...” (INF)
- Some words are considered bonus others stigma
 - bonus: comparatives, superlatives, conclusive expressions, etc.
 - stigma: negatives, pronouns, etc.

Experimental combination (Edmundson'69)

- Contribution of 4 features
 - title, cue, keyword, position
 - Linear equation

$$Weight(S) = \alpha.Title(S) + \beta.Cue(S) + \gamma.Keyword(S) + \delta.Position(S)$$

- first the parameters are adjusted using training data

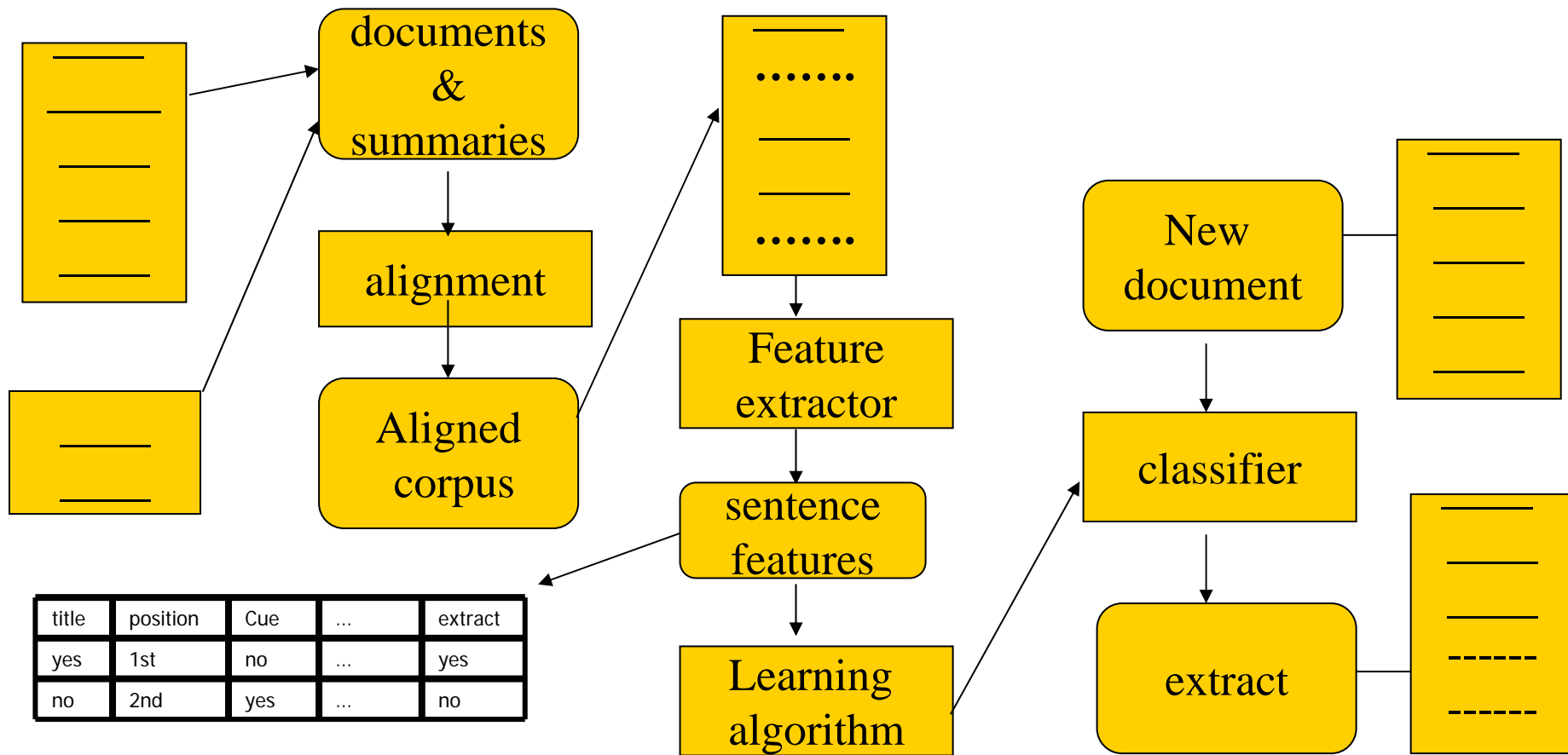
Experimental combination

- All possible combinations $4^2 - 1$ (=15 possibilities)
 - title + cue; title; cue; title + cue + keyword; etc.
- Produces summaries for test documents
- Evaluates co-selection (precision/recall)

Experimental combination

- Obtains the following results
 - best system
 - cue + title + position
 - individual features
 - position is best, then
 - cue
 - title
 - keyword

Learning to extract



Statistical combination

- method adopted by Kupiec&al'95
- need corpus of documents and extracts
 - professional abstracts
 - high cost
- alignment
 - program that identifies similar sentences
 - manual validation

Statistical combination

- length of sentence (true/false)

$$\textit{len}(S) > u_l$$

- cue (true/false)

$$(S_i \cap \textit{DIC}_{cue}) \neq \phi$$

or

$$\textit{heading}(S_{i-1}) \wedge (S_{i-1} \cap \textit{DIC}_{headings}) \neq \phi$$

Statistical combination

- position (discrete)
 - paragraph # $\{1, 2, \dots, 10\} \vee \{last, last-1, \dots, last-4\}$
 - in paragraph $\{initial, middle, final\}$
- keyword (true/false) $rank(S) > u_k$
- proper noun (true/false)
 - similar to keyword

Statistical combination

- combination

$$p(s \in E | f_1, \dots, f_n) = \frac{p(f_1, \dots, f_n | s \in E) \cdot p(s \in E)}{p(f_1, \dots, f_n)}$$

$$p(f_1, \dots, f_n | s \in E) = \prod p(f_i | s \in E)$$

$$p(f_1, \dots, f_n) = \prod p(f_i)$$

$$p(s \in E)$$

Statistical combination

- results for individual features
 - position
 - cue
 - length
 - keyword
 - proper name
- best combination
 - position+cue+length

Problems with extracts

- Lack of cohesion

source

A single-engine airplane crashed Tuesday into a ditch beside a dirt road on the outskirts of Albuquerque, killing all five people aboard, authorities said.

Four adults and one child died in the crash, which witnesses said occurred about 5 p.m., when it was raining, Albuquerque police Sgt. R.C. Porter said.

The airplane was attempting to land at nearby Coronado Airport, Porter said. It aborted its first attempt and was coming in for a second try when it crashed, he said...

extract

Four adults and one child died in the crash, which witnesses said occurred about 5 p.m., when it was raining, Albuquerque police Sgt. R.C. Porter said.

It aborted its first attempt and was coming in for a second try when it crashed, he said.

Problems with extracts

- Lack of coherence

source

Supermarket A announced a big profit for the third quarter of the year. The directory studies the creation of new jobs. Meanwhile, B's supermarket sales drop by 10% last month. The company is studying closing down some of its stores.

extract

Supermarket A announced a big profit for the third quarter of the year. The company is studying closing down some of its stores.

Solution

- identification of document structure
- rules for the identification of anaphora
 - pronouns, logical and rhetorical connectives, and definite noun phrases
 - Corpus-based heuristics
- aggregation techniques
 - IF sentence contains anaphor THEN include preceding sentences
- anaphora resolution is more appropriate but
 - programs for anaphora resolution are far from perfect

Solution

- BLAB project (Johnson & Paice'93 and previous works by same group)
 - rules for identification: "that" is :
 - non-anaphoric if preceded by research-verb (e.g. "assume", "show", etc.)
 - non-anaphoric if followed by pronoun, article, quantifier, demonstrative,...
 - external if no later than 10th word of sentence
 - else: internal
 - selection (indicator) & rejection & aggregation rules; reported success: abstract > aggregation > extract

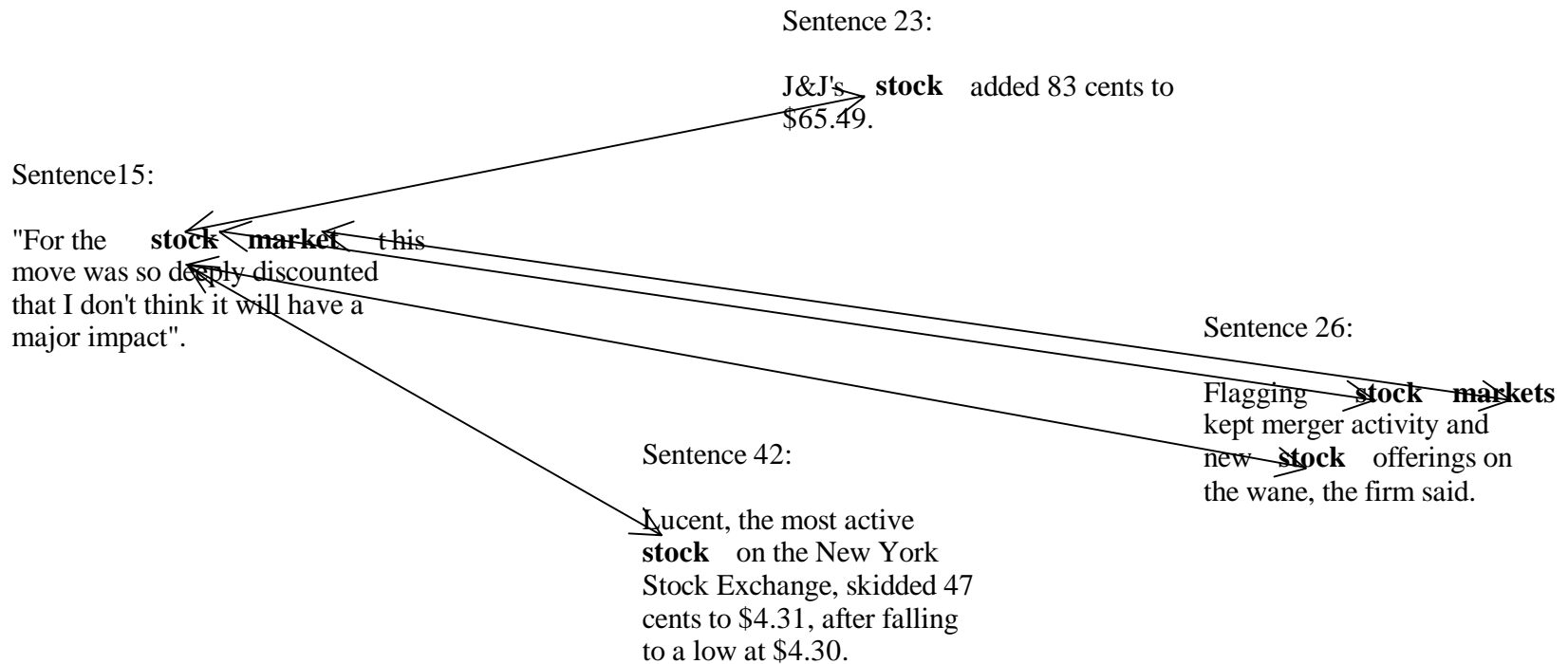
Cohesion analysis

- Repetition with identity
 - Adam bite the *apple*. The *apple* was not ripe enough.
- Repetition without identity
 - Adam ate the *apples*. He likes *apples*.
- Class/superclass
 - Adam ate the apple. He likes *fruit*.
- Systematic relation
 - He likes *green* apples. He does not like *red* ones.
- Non-systematic relation
 - Adam was three hours in the *garden*. He was *planting* an apple tree.

Telepattan system: (Bembrahim & Ahmad'95)

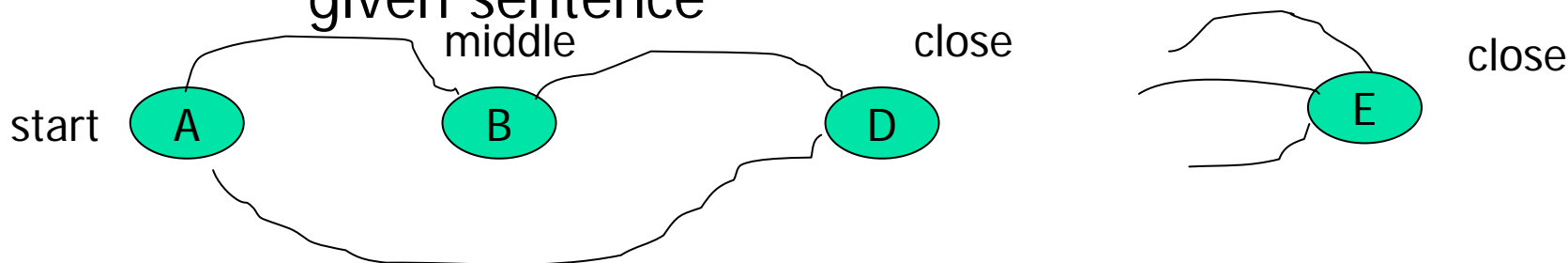
- Link two sentences if
 - they contain words related by repetition, synonymy, class/superclass (hypernymy), paraphrase
 - *destruct* ~ *destruction*
 - use thesaurus (i.e., related words)
- pruning
 - $\text{links}(s_i, s_j) > \text{thr} \Rightarrow \text{bond}(s_i, s_j)$

Telepattan system



Telepattan system

- Classify sentences as
 - start topic, middle topic, end of topic, according to the number of links
 - this is based on the number of links to and from a given sentence



- Summaries are obtained by extracting sentences that open-continue-end a topic

Lexical chains

- Lexical chain:
 - word sequence in a text where the words are related by one of the relations previously mentioned
- Use:
 - ambiguity resolution
 - identification of discourse structure

WordNet: a lexical database

- synonymy
 - dog, can
- hypernymy
 - dog, animal
- antonym
 - dog, cat
- meronymy (part/whole)
 - dog, leg

Search Word: bank

Searches for bank: Noun Verb

Senses:

The noun bank has 10 senses (first 9 from tagged texts)

1. depository financial institution, **bank**, banking concern, banking company -- (a financial institution)
2. **bank** -- (sloping land (especially the slope beside a body of water); "they pulled the canoe up on t
3. **bank** -- (a supply or stock held in reserve especially for future use (especially in emergencies))
4. **bank**, bank building -- (a building in which commercial banking is transacted; "the bank is on the
5. **bank** -- (an arrangement of similar objects in a row or in tiers; "he operated a bank of switches")
6. savings bank, coin bank, money box, **bank** -- (a container (usually with a slot in the top) for keep
7. **bank** -- (a long ridge or pile; "a huge bank of earth")
8. **bank** -- (the funds held by a gambling house or the dealer in some gambling games; "he tried to b
9. **bank**, cant, camber -- (a slope in the turn of a road or track; the outside is higher than the inside
10. **bank** -- (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning)

The verb bank has 7 senses (first 2 from tagged texts)

1. **bank** -- (tip laterally; of boats and aircraft)
2. **bank** -- (enclose with a bank; "bank roads")
3. **bank** -- (do business with a bank or keep an account at a bank; "Where do you bank in this town
4. **bank** -- (be in the banking business)
5. deposit, **bank** -- (put into a bank account)
6. **bank** -- (cover with ashes, of fires, to control the rate of burning)
7. trust, swear, rely, **bank** -- (have confidence or faith in; "We can trust in God"; "Rely on your frie

Search Word: bank

Redisplay Overview

Searches for bank: Noun Verb

Senses:

10 senses of bank

Sense 1

depository financial institution, bank, banking concern, banking company -- (a financial institution that accepts deposits and charn
=> financial institution, financial organization -- (an institution (public or private) that collects funds (from the public or other
=> institution, establishment -- (an organization founded and united for a specific purpose)
=> organization, organisation -- (a group of people who work together)
=> social group -- (people sharing some social relation)
=> group, grouping -- (any number of entities (members) considered as a unit)

Sense 2

bank -- (sloping land (especially the slope beside a body of water); "they pulled the canoe up on the bank"; "he sat on the bank of t
=> slope, incline, side -- (an elevated geological formation; "he climbed the steep slope"; "the house was built on the side of tl
=> geological formation, geology, formation -- (the geological features of the earth)
=> natural object -- (an object occurring naturally; not made by man)
=> object, physical object -- (a physical (tangible and visible) entity; "it was full of rackets, balls and other objects")
=> entity, something -- (anything having existence (living or nonliving))

Sense 3

bank -- (a supply or stock held in reserve especially for future use (especially in emergencies))
=> reserve, backlog, stockpile -- (something kept back or saved for future use or a special purpose)
=> accumulation -- ((finance) profits that are not paid out as dividends but are added to the capital base of the corporation

Extracts by lexical chains

- Barzilay & Elhadad'97; Silber & McCoy'02
- A chain C represents a "concept" in WordNet
 - *Financial institution* "bank"
 - *Place to sit down in the park* "bank"
 - *Sloppy land* "bank"
- A chain is a list of words, the order of the words is that of their occurrence in the text
- A noun N is inserted in C if N is related to C
 - relations used=identity; synonym; hypernym

Extracts by lexical chains

- Compute the contribution of N to C as follows
 - If C is empty consider the relation to be “repetition” (identity)
 - If not identify the last element M of the chain to which N is related
 - Compute distance between N and M in number of sentences (1 if N is the first word of chain)
 - Contribution of N is looked up in a table with entries given by type of relation and distance
 - e.g., hyper & distance=3 then contribution=0.5

Extracts by lexical chains

- After inserting all nouns in chains there is a second step
- For each noun, identify the chain where it most contributes; delete it from the other chains and adjust weights
- Select sentences that belong or are covered by "strong chains"

Extracts by lexical chains

- Strong chain:
 - $\text{weight}(C) > \text{thr}$
 - $\text{thr} = \text{average}(\text{weight}(Cs)) + 2 * \text{sd}(\text{weight}(Cs))$
- selection:
 - H1: select the first sentence that contains a member of a strong chain
 - H2: select the first sentence that contains a “representative” member of the chain
 - H3: identify a text segment where the chain is highly dense (density is the proportion of words in the segment that belong to the chain)

Information retrieval techniques (Salton&al'97)

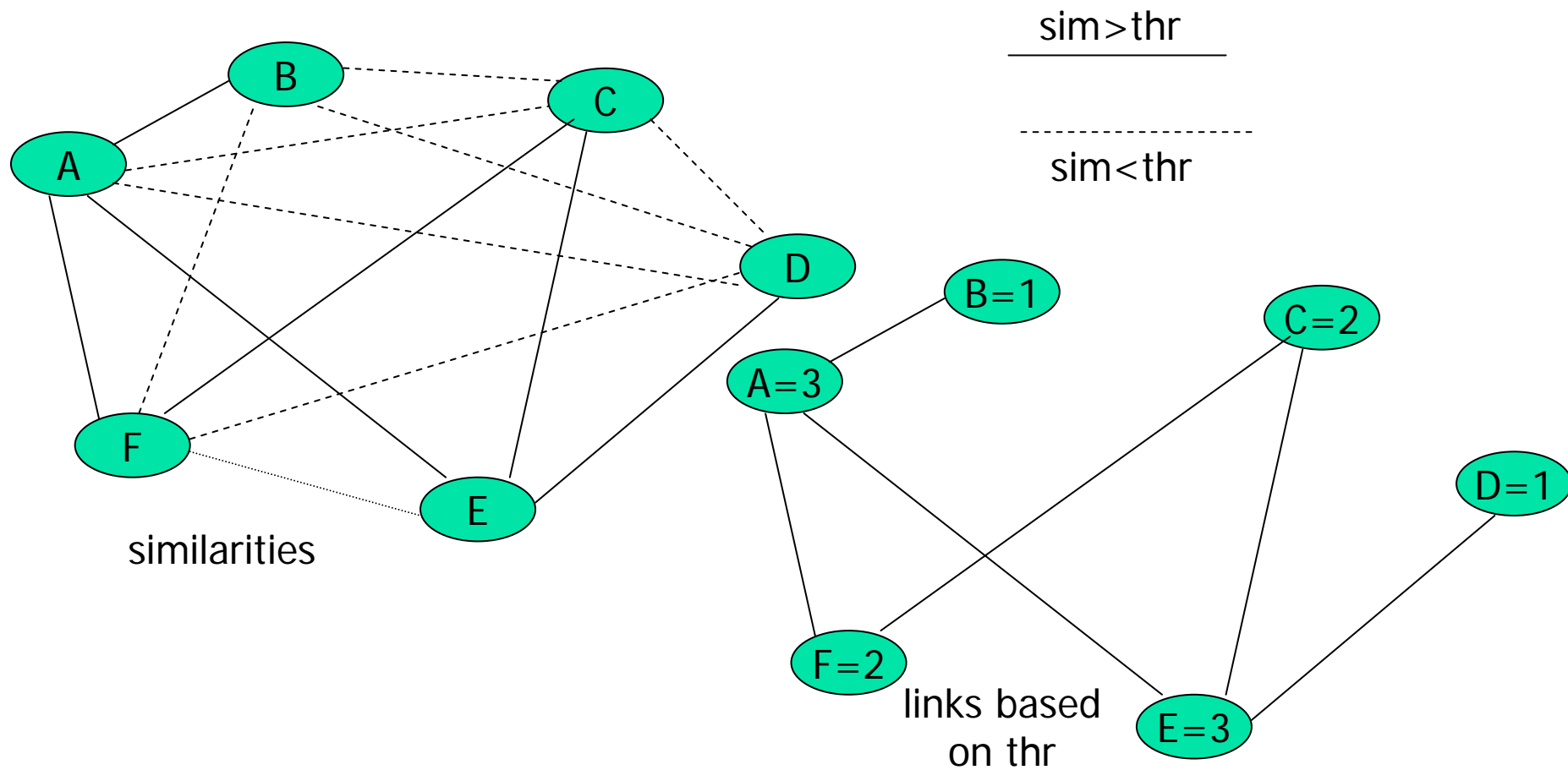
- Vector Space Model
 - each text unit represented as
- Similarity metric

$$D_i = (d_{i1}, \dots, d_{in})$$

$$\text{sim}(D_i, D_j) = \sum d_{ik} \cdot d_{jk}$$

- metric normalised to obtain 0-1 values
- Construct a graph of paragraphs.
Strength of link is the similarity metric
- Use threshold (thr) to decide upon similar paragraphs

Text relation map



Information retrieval techniques

- identify regions where paragraphs are well connected
- paragraph selection heuristics
 - bushy path
 - select paragraphs with many connections with other paragraphs and present them in text order
 - depth-first path
 - select one paragraph with many connections; select a connected paragraph (in text order) which is also well connected; continue
 - segmented bushy path
 - follow the bushy path strategy but locally including paragraphs from all “segments of text”: a bushy path is created for each segment

Information retrieval techniques

- Co-selection evaluation
 - because of low agreement across human annotators (~46%) new evaluation metrics were defined
 - optimistic scenario: select the human summary which gives best score
 - pessimistic scenario: select the human summary which gives worst score
 - union scenario: select the union of the human summaries
 - intersection scenario: select the overlap of human summaries

Rhetorical analysis

- Rhetorical Structure Theory (RST)
 - Mann & Thompson'88
- Descriptive theory of text organization
- Relations between two text spans
 - nucleus & satellite (hypotactic)
 - nucleus & nucleus (paratactic)
 - "IR techniques have been used in text summarization. For example, X used term frequency. Y used $tf \cdot idf$."

Rhetorical analysis

- relations are deduced by judgement of the reader
- texts are represented as trees, internal nodes are relations
- text segments are the leafs of the tree
 - (1) Apples are very cheap. (2) Eat apples!!!
 - (1) is an argument in favour of (2), then we can say that (1) motivates (2)
 - (2) seems more important than (1), and coincides with (2) being the nucleus of the motivation

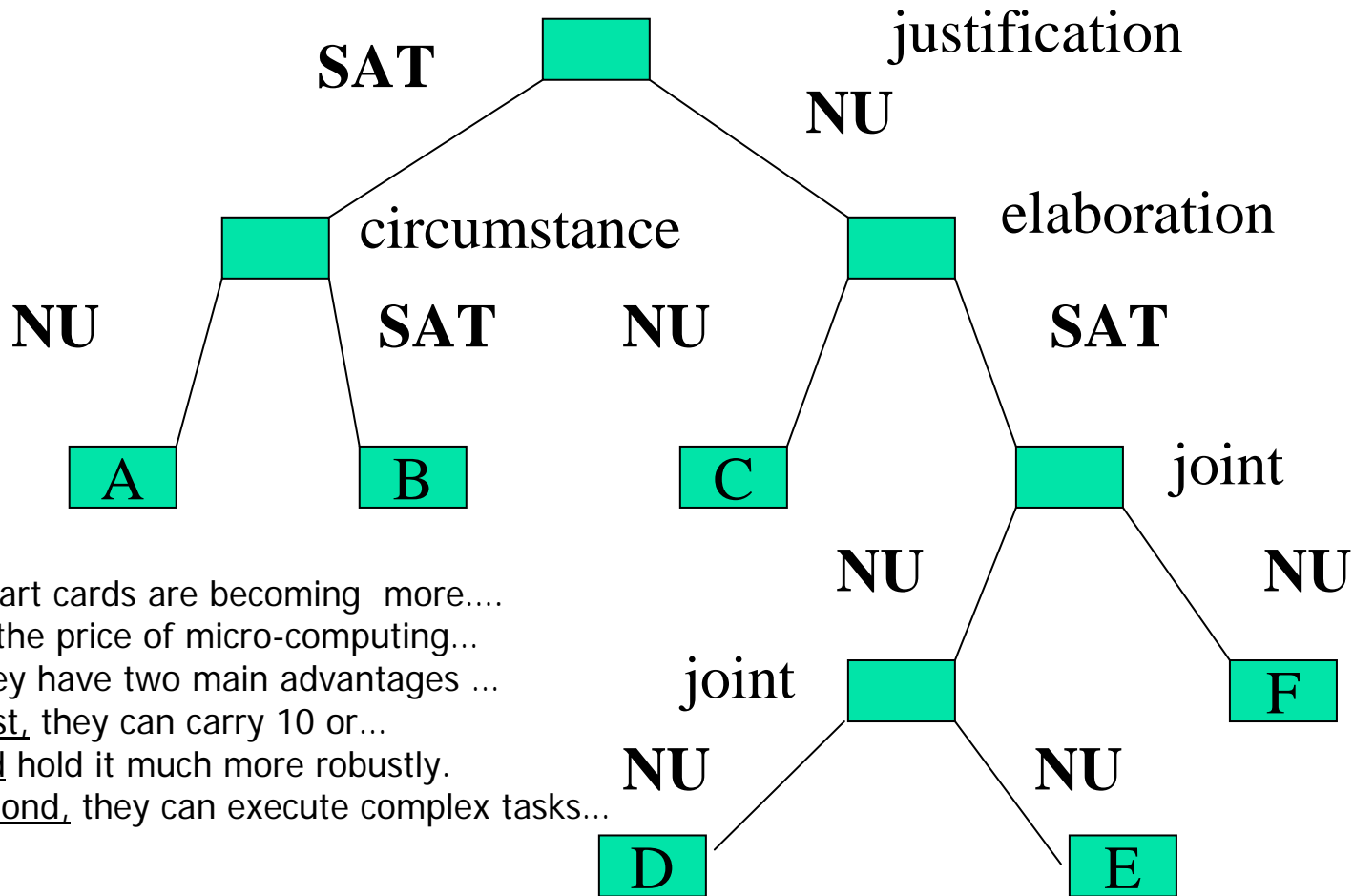
Rhetorical analysis

- Relations can be marked on the syntax
 - John went to sleep because he was tired.
 - Mary went to the cinema and Julie went to the theatre.
- RST authors say that markers are not necessary to identify a relation
- However all RTS analysers rely on markers
 - “however”, “therefore”, “and”, “as a consequence”, etc.
- strategy to obtain a complete tree
 - apply rhetorical parsing to “segments” (or paragraphs)
 - apply a cohesion measure (vocabulary overlap) to identify how to connect individual trees

Rhetorical analysis based summarization

- (A) Smart cards are becoming more attractive
- (B) as the price of micro-computing power and storage continues to drop.
- (C) They have two main advantages over magnetic strip cards.
- (D) First, they can carry 10 or even 100 times as much information
- (E) and hold it much more robustly.
- (F) Second, they can execute complex tasks in conjunction with a terminal.

Rhetorical tree



- (A) Smart cards are becoming more....
- (B) as the price of micro-computing...
- (C) They have two main advantages ...
- (D) First, they can carry 10 or...
- (E) and hold it much more robustly.
- (F) Second, they can execute complex tasks...

Penalty: Ono'94

Penalty

A=1

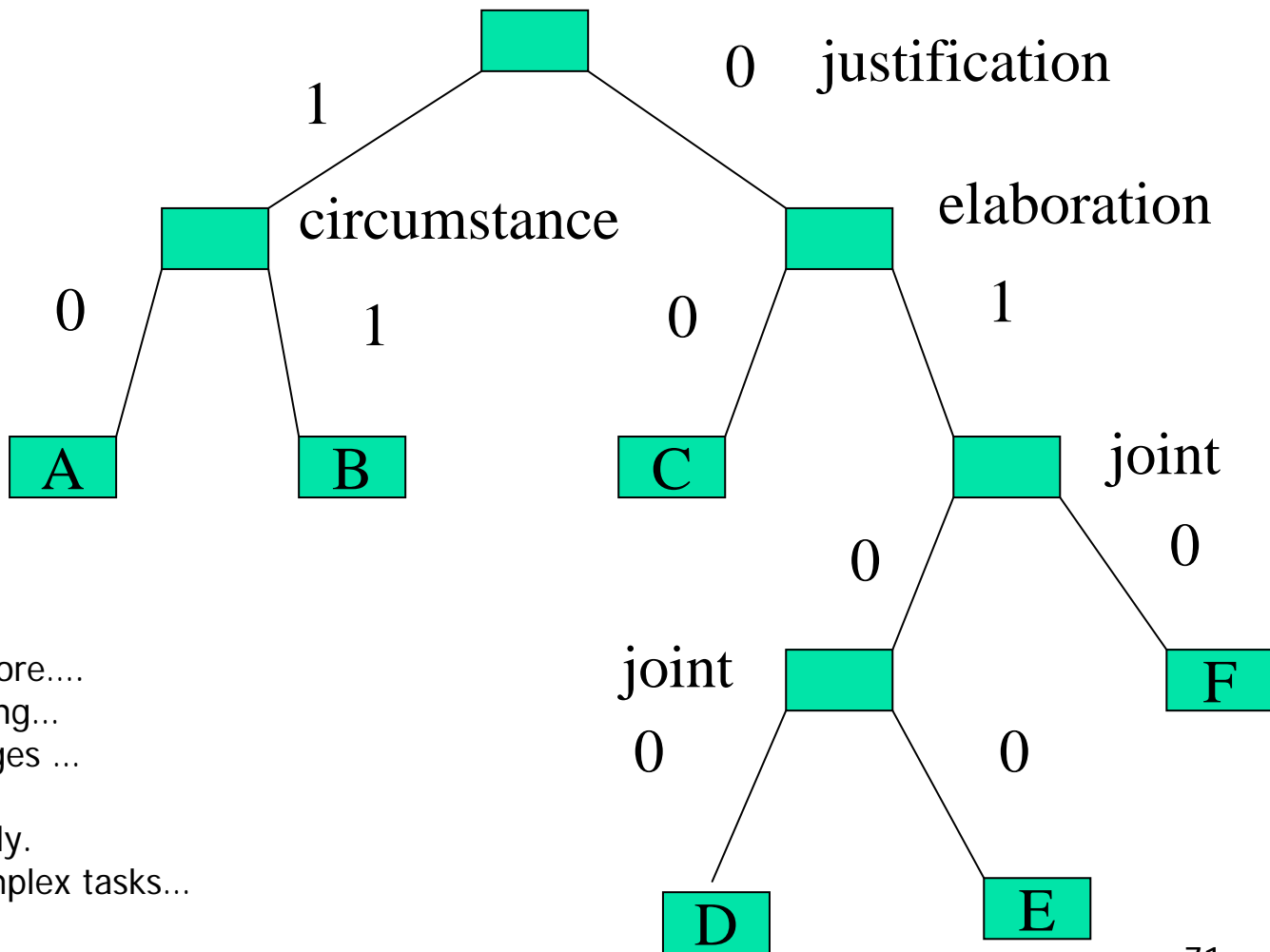
B=2

C=0

D=1

E=1

F=1



- (A) Smart cards are becoming more....
- (B) as the price of micro-computing...
- (C) They have two main advantages ...
- (D) First, they can carry 10 or...
- (E) and hold it much more robustly.
- (F) Second, they can execute complex tasks...

RTS extract

(C) They have two main advantages over magnetic strip cards.

(A) Smart cards are becoming more attractive

(C) They have two main advantages over magnetic strip cards.

(D) First, they can carry 10 or even 100 times as much information

(E) and hold it much more robustly.

(F) Second, they can execute complex tasks in conjunction with a terminal.

(A) Smart cards are becoming more attractive

(B) as the price of micro-computing power and storage continues to drop.

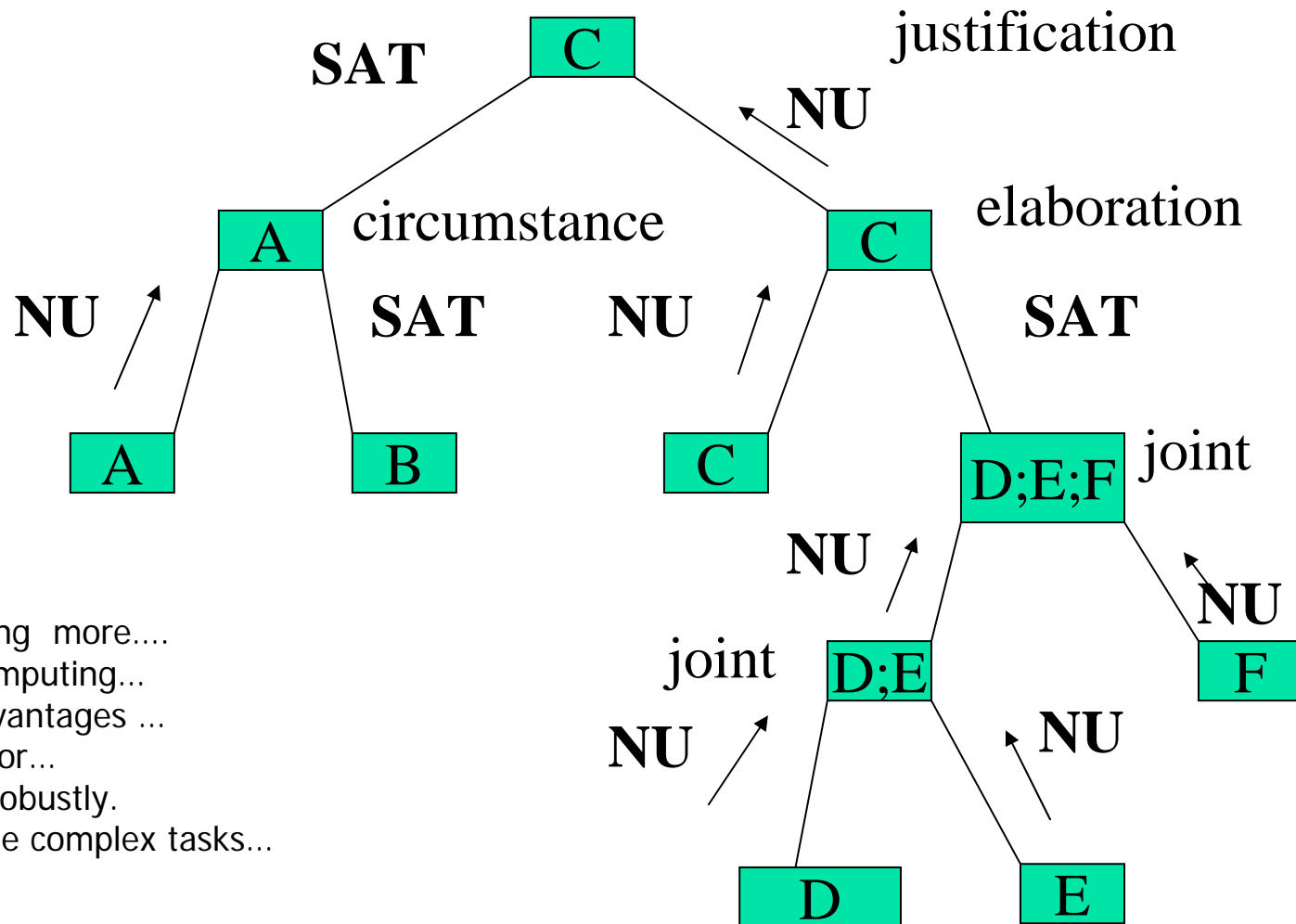
(C) They have two main advantages over magnetic strip cards.

(D) First, they can carry 10 or even 100 times as much information

(E) and hold it much more robustly.

(F) Second, they can execute complex tasks in conjunction with a terminal.

Promotion: Marcu'97



- (A) Smart cards are becoming more....
 (B) as the price of micro-computing...
 (C) They have two main advantages ...
 (D) First, they can carry 10 or...
 (E) and hold it much more robustly.
 (F) Second, they can execute complex tasks...

RST extract

(C) They have two main advantages over magnetic strip cards.

(A) Smart cards are becoming more attractive

(C) They have two main advantages over magnetic strip cards.

(A) Smart cards are becoming more attractive

(B) as the price of micro-computing power and storage continues to drop.

(C) They have two main advantages over magnetic strip cards.

(D) First, they can carry 10 or even 100 times as much information

(E) and hold it much more robustly.

(F) Second, they can execute complex tasks in conjunction with a terminal.

Observations

- Marcu showed that nucleus correlates with idea of centrality
- Compression can not be controlled
- No discrimination between relations
 - “elaboration” = “exemplification”
- Texts of interesting size untreatable
- RST is interpretative, therefore knowledge is needed

FRUMP (de Jong'82)

a small earthquake shook several Southern Illinois counties Monday night, the National Earthquake Information Service in Golden, Colo., reported. Spokesman Don Finley said the quake measured 3.2 on the Richter scale, "probably not enough to do any damage or cause any injuries." The quake occurred about 7:48 p.m. CST and was centered about 30 miles east of Mount Vernon, Finlay said. It was felt in Richland, Clay, Jasper, Effington, and Marion Counties.

There was an earthquake in Illinois with a 3.2 Richter scale.

FRUMP

- Knowledge structure = sketchy-scripts, adaptation of Shank & Abelson scripts (1977)
- sketchy-scripts contain only the relevant information of an event
- ~50 sketchy-scripts manually developed for FRUMP
- Interpretation is based on skimming

FRUMP

- When a key word is found one or more scripts are activated
- The activated scripts guide text interpretation, syntactic analysis is called on demand
- When more than one script is activated, heuristics decide which represents the correct interpretation
- Because the representation is language-independent, it can be used to generate summaries in various languages

FRUMP

- Evaluation: one day of processing text
- 368 stories
 - 100 not news articles
 - 147 not of the script type
 - 121 could be understood
 - for 29 FRUMP has scripts
 - only 11 were processed correctly + 2 almost correctly = 3% correct; on average 10% correct
- problems
 - incorrect variable binding
 - could not identify script
 - incorrect script used to interpret (no script)
 - incorrect script used to interpret (correct script present)

FRUMP

- 50 scripts is probably not enough for interpreting most stories
- knowledge was manually coded
- how to learn new scripts

Vatican City. The dead of the Pope shakes the world. He passed away...

Earthquake in the Vatican. One dead.

Information Extraction for Summarization

- Message Understanding Conferences (1987-1997)
- extract key information from a text
- automatic fill-in forms (i.e., for a database)
- idea of scenario/template
 - terrorist attacks; rocket/satellite launch; management succession; etc.
- characteristics of the problem
 - only a few parts of the text are relevant
 - only a few parts of the relevant sentences are relevant

Information Extraction

ALGIERS, May 22 (AFP) - At least 538 people were killed and 4,638 injured when a powerful earthquake struck northern Algeria late Wednesday, according to the latest official toll, with the number of casualties set to rise further ... The epicentre of the quake, which measured 5.2 on the Richter scale, was located at Thenia, about 60 kilometres (40 miles) east of Algiers, ...

| | |
|-----------|-----------------|
| DATE | 21/05/2003 |
| DEATH | 538 |
| INJURED | 4,638 |
| EPICENTER | Thenia, Algeria |
| INTENSITY | 5.2, Richter |

CBA: Concept-based Abstracting (Paice&Jones'93)

- Summaries in an specific domain, for example crop husbandry, contain specific concepts.
 - SPECIES (the crop in the study)
 - CULTIVAR (variety studied)
 - HIGH-LEVEL-PROPERTY (specific property studied of the cultivar, e.g. yield, growth)
 - PEST (the pest that attacks the cultivar)
 - AGENT (chemical or biological agent applied)
 - LOCALITY (where the study was conducted)
 - TIME (years of the study)
 - SOIL (description of the soil)

CBA

- Given a document in the domain, the objective is to instantiate with “well formed strings” each of the concepts
- CBA uses patterns which implement how the concepts are expressed in texts
 - “fertilized with *procymidane*” gives the pattern “fertilized with AGENT”
- Can be quite complex and involve several concepts
 - PEST is a ? pest of SPECIES
 - where ? matches a sequence of input tokens

CBA

- Each pattern has a weight
- Criteria for variable instantiation
 - Variable is inside pattern
 - Variable is on the edge of the pattern
- Criteria for candidate selection
 - all hypothesis' substrings are considered
 - decrease of SPECIES
 - effect of ? in SPECIES
 - count repetitions and weights
 - select one substring for each semantic role

CBA

- Canned-text based generation

this paper studies the effect of [AGENT] on the [HLP] of [SPECIES] OR this paper studies the effect of [METHOD] on the [HLP] of [SPECIES] when it is infested by [PEST].

....

Summary: This paper studies the effect of G. pallida on the yield of potato. An experiment in 1985 and 1986 at York was undertaken.

- evaluation

- central and peripheral concepts
- form of selected strings

- pattern acquisition can be done automatically

- informative summaries include verbatim “conclusive” sentences from document

Headline generation: Banko&a'00

- Generate a summary shorter than a sentence
 - Text: Acclaimed Spanish soprano de los Angeles dies in Madrid after a long illness.
 - Summary: de Los Angeles died

- Generate a sentence with pieces combined from different parts of the texts
 - Text: Spanish soprano de los Angeles dies. She was 81.
 - Summary: de Los Angeles dies at 81

- Method borrowed from statistical machine translation
 - model of word selection from the source
 - model of realization in the target language

Headline generation

- Content selection
 - how many and what words to select from document
- Content realization
 - how to put words in the appropriate sequence in the headline such that it looks ok
- training: 25K texts + headlines

Headline generation

■ Content selection

- What document features influence the words of the headline
- A possible feature: the words of the document
 - w is in summary & w is in document
- This feature can be computed as

$$p(w_i \in T | w_i \in D) = \frac{p(w_i \in D | w_i \in T) \cdot p(w_i \in T)}{p(w_i \in D)}$$

Headline generation

- Content selection
 - Other feature: how many words to select?

$$p(\text{len}(T)=n)$$

- Easiest solution is to use a fixed length per document type

Headline generation

- Surface realization
 - Compute the probability of observing $w_1 \dots w_n$

$$\prod p(w_i | w_1 \dots w_{i-1})$$

- 2-grams approximation

$$\prod p(w_i | w_{i-1})$$

Headline generation

- Model combination
 - we want the best sequence of words

$$p(w_1 \dots w_n) = \prod p(w_i \in T \mid w_i \in D) * \prod p(\text{len}(T) = n) * \prod p(w_i \mid w_1 \dots w_{i-1})$$

content model

realization model

Headline generation

- Search using the following formula (note the use logarithm)

$$\begin{aligned} \operatorname{argmax}_T & (\alpha \sum \log(p(w_i \in T | w_i \in D)) + \\ & \beta \log(p(\operatorname{len}(T) = n)) + \\ & \gamma \sum \log(w_i | w_{i-1})) \end{aligned}$$

- Viterbi algorithm can be used to find the best sequence

Headline generation

- One has to consider the problem of data sparseness
 - Words never seen
 - 2-grams never seen
- There are “smoothing” and “back-off” models to deal with the problems

Example

President Clinton met with his top Mideast adviser, including Secretary of State Madeleine Albright and U.S. peace envoy Dennis Ross, in preparation for a session with Israel Prime Minister Benjamin Netanyahu tomorrow. Palestinian leader Yasser Arafat is to meet with Clinton later this week. Published reports in Israel say Netanyahu will warn Clinton that Israel can't withdraw from more than nine percent of the West Bank in its next scheduled pullback, although Clinton wants 12-15 percent pullback.

- original title: *U.S. pushes for mideast peace*
- automatic title
 - *clinton*
 - *clinton wants*
 - *clinton netanyahu arafat*
 - *clinton to mideast peace*

Evaluation

- Compare automatic headline with original headline
 - Words in common
- Various lengths evaluated
 - 4 words give acceptable results (?) 1 out of 5 headlines contain all words of the original
- Grammaticality is an issue, however headlines have their own syntax
- Other features
 - POS & position

Novel Techniques: condensation

- Cut&Paste Summarization: Jing&McKeown'00
 - "HMM" for word alignment to answer the question: what document positions a word in the summary comes from?
 - a word in a summary sentence may come from different positions, not all of them are equally likely
 - given words $I_1 \dots I_n$ (in a summary sentence) the following probability table is needed:
 $P(I_{k+1} = \langle S2, W2 \rangle \mid I_k = \langle S1, W1 \rangle)$
 - they associate probabilities by hand following a number of heuristics
 - given a sentence summary, the alignment is computed using the Viterbi algorithm

Summary sentence:

(F0:S1 arthur b sackler vice president for law and public policy of time warner inc) (F1:S-1 and) (F2:S0 a member of the direct marketing association told) (F3:S2 the communications subcommittee of the senate commerce committee) (F4:S-1 that legislation) (F5:S1to protect) (F6:S4 children' s) (F7:S4 privacy) (F8:S4 online) (F9:S0 could destroy the spontaneous nature that makes the internet unique)

Source document sentences:

Sentence 0: a proposed new law that would require web publishers to obtain parental consent before collecting personal information from children (F9 could destroy the spontaneous nature that makes the internet unique) (F2 a member of the direct marketing association told) a senate panel thursday

Sentence 1: (F0 arthur b sackler vice president for law and public policy of time warner inc) said the association supported efforts (F5 to protect) children online but he urged lawmakers to find some middle ground that also allows for interactivity on the internet

Sentence 2: for example a child's e-mail address is necessary in order to respond to inquiries such as updates on mark mcguire's and sammy sosa's home run figures this year or updates of an online magazine sackler said in testimony to (F3 the communications subcommittee of the senate commerce committee)

Sentence 4: the subcommittee is considering the (F6 children's) (F8 online) (F7 privacy) protection act which was drafted on the recommendation of the federal trade commission

Novel Techniques: condensation

- Cut&Paste Summarization
 - Sentence reduction
 - a number of resources are used (lexicon, parser, etc.)
 - exploits connectivity of words in the document (each word is weighted)
 - uses a table of probabilities to learn when to remove a sentence component
 - final decision is based on probabilities, mandatory status, and local context
 - Rules for sentence combination were manually developed

Cut&Paste human examples

Example 1: add description for people or organization

Original Sentences:

Sentence 34: "We're trying to prove that there are big benefits to the patients by involving them more deeply in their treatment", said Paul Clayton, chairman of the dept. dealing with computerized medical information at Columbia.

Sentence 77: "The economic payoff from breaking into health care records is a lot less than for banks", said Clayton at Columbia.

Rewritten Sentences:

Combined: "The economic payoff from breaking into health care records is a lot less than for banks", said Paul Clayton, chairman of the dept. dealing with computerized medical information at Columbia.

Example 2: extract common elements

Original Sentences:

Sentence 8: but it also raises serious questions about the privacy of such highly personal information wafting about the digital world

Sentence 10: The issue thus fits squarely into the broader debate about privacy and security on the internet whether it involves protecting credit card numbers or keeping children from offensive information

Rewritten Sentences :

Combined: but it also raises the issue of privacy of such personal information and this issue hits the head on the nail in the broader debate about privacy and security on the internet.

Cut&Paste human examples

Example 3: reduce and join sentences by adding connectives or punctuations

Original Sentences:

Sentence 7: Officials said they doubted that Congressional approval would be needed for the changes, and they foresaw no barriers at the Federal level.

Sentence 8: States have wide control over the availability of methadone, however.

Rewritten Sentences :

Combined: Officials said they foresaw no barriers at the Federal level; however, States have wide control over the availability of methadone.

Example 4: reduce and change one sentence to a clause

Original Sentences:

Sentence 25: in GPI, you specify an RGB COLOR value with a 32-bit integer encoded as follows: 00000000* Red * Green * Blue The high 8 bits are set to 0.

Sentence 27: this encoding scheme can represent some 16 million colors

Rewritten Sentences :

Combined: GPI describes RGB colors as 32-bit integers that can describe 16 million colors

Novel Techniques: condensation

- Sentence condensation: Knight&Marcu'00
- probabilistic framework: noisy-channel model
- corpus: automatically collected <sentences, compressions>
- model explains how short sentences can be re-written
- a long sentence L can be generated from a short sentence S, two probabilities are needed
 - $P(L/S)$ and $P(S)$
 - the model seeks to maximize $P(L/S) \times P(S)$

Paraphrase

- Alignment based paraphrase: Barzilay&Lee'2003
- unsupervised approach to learn:
 - patterns in the data & equivalences among patterns
 - X injured Y people, Z seriously = Y were injured by X among them Z were in serious condition
 - learning is done over two different corpus which are comparable in content
- use a sentence clustering algorithm to group together sentences that describe similar events

Similar event descriptions

- **Cluster of similar sentences**

- **A Palestinian suicide bomber blew himself up in** a southern city Wednesday, **killing** two other **people and wounding** 27.
- **A suicide bomber blew himself up in** the settlement of Efrat, on Sunday, **killing** himself **and injuring** seven people.
- **A suicide bomber blew himself up in** the coastal resort of Netanya on Monday, **killing** three other **people and wounding** dozens more.

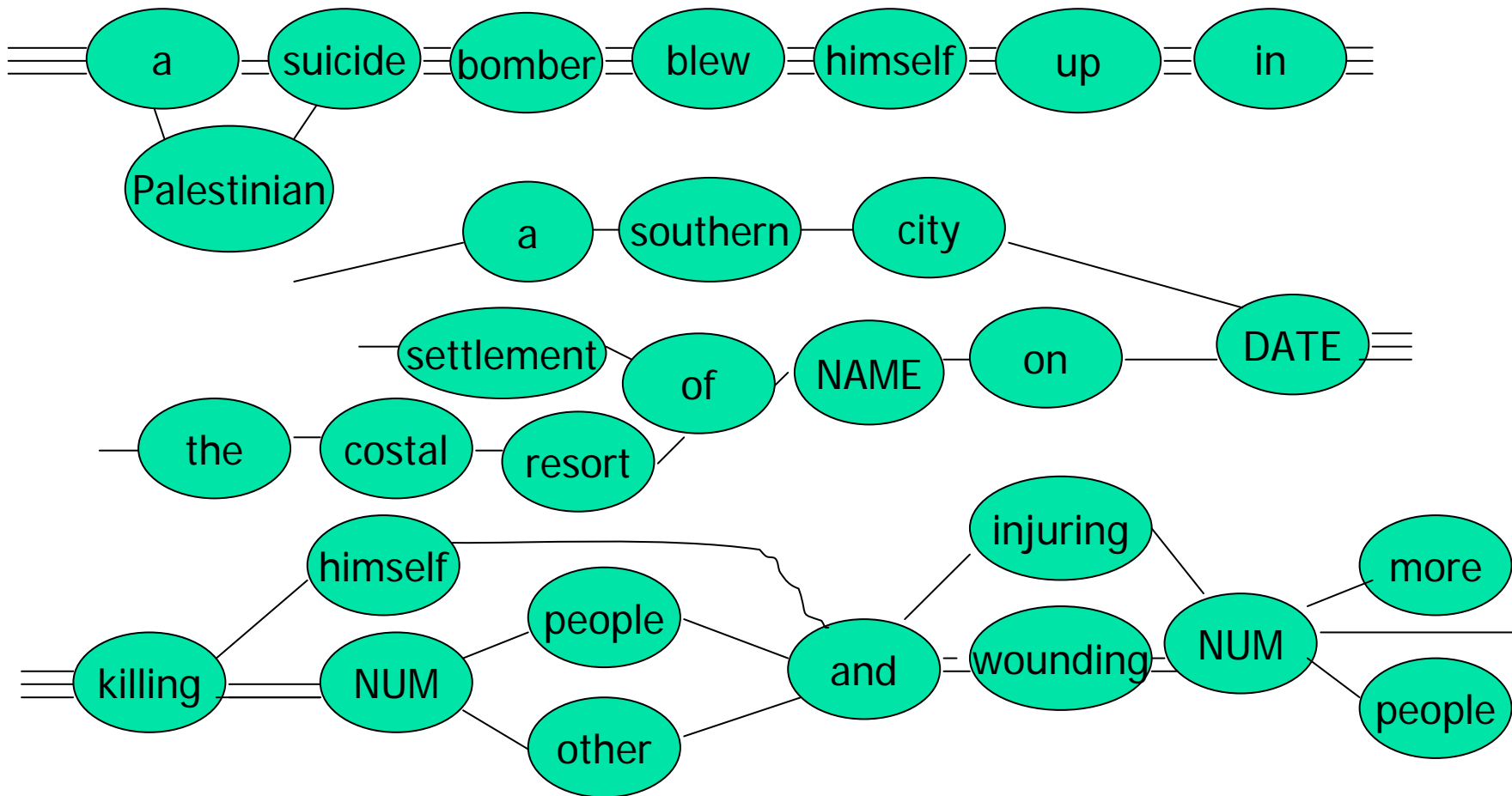
- **Variable substitution**

- **A Palestinian suicide bomber blew himself up in** a southern city DATE, **killing** NUM other **people and wounding** NUM.
- **A suicide bomber blew himself up in** the settlement of NAME, on DATE, **killing** himself **and injuring** NUM people.
- **A suicide bomber blew himself up in** the coastal resort of NAME on NAME, **killing** NUM other **people and wounding** dozens more.

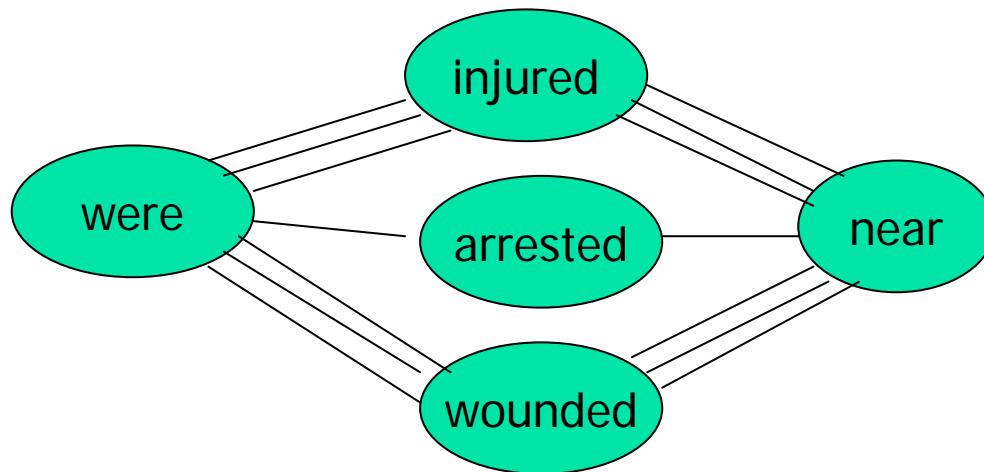
Paraphrase

- apply a multi-sequence alignment algorithm to represent paraphrases as lattices
- identify arguments (variable) as zones of great variability in the lattices
- generation of paraphrases can be done by matching against the lattices and generating as many paraphrases as paths in the lattice

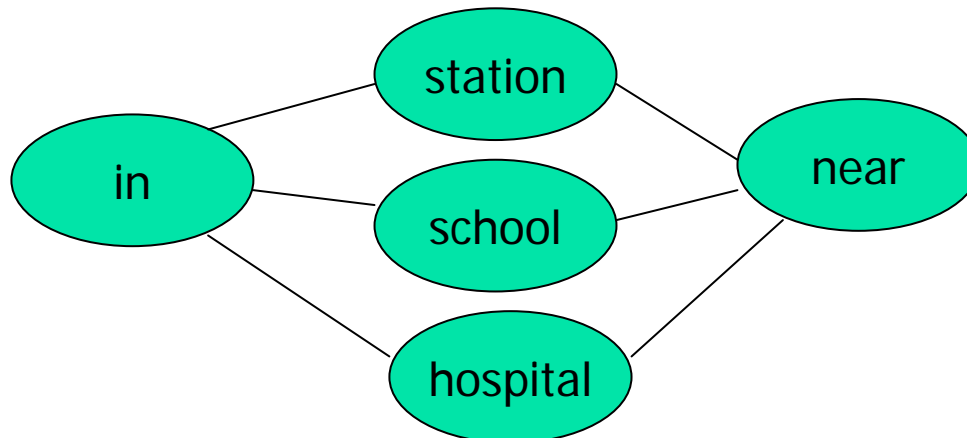
Lattices and backbones



Arguments or Synonyms?

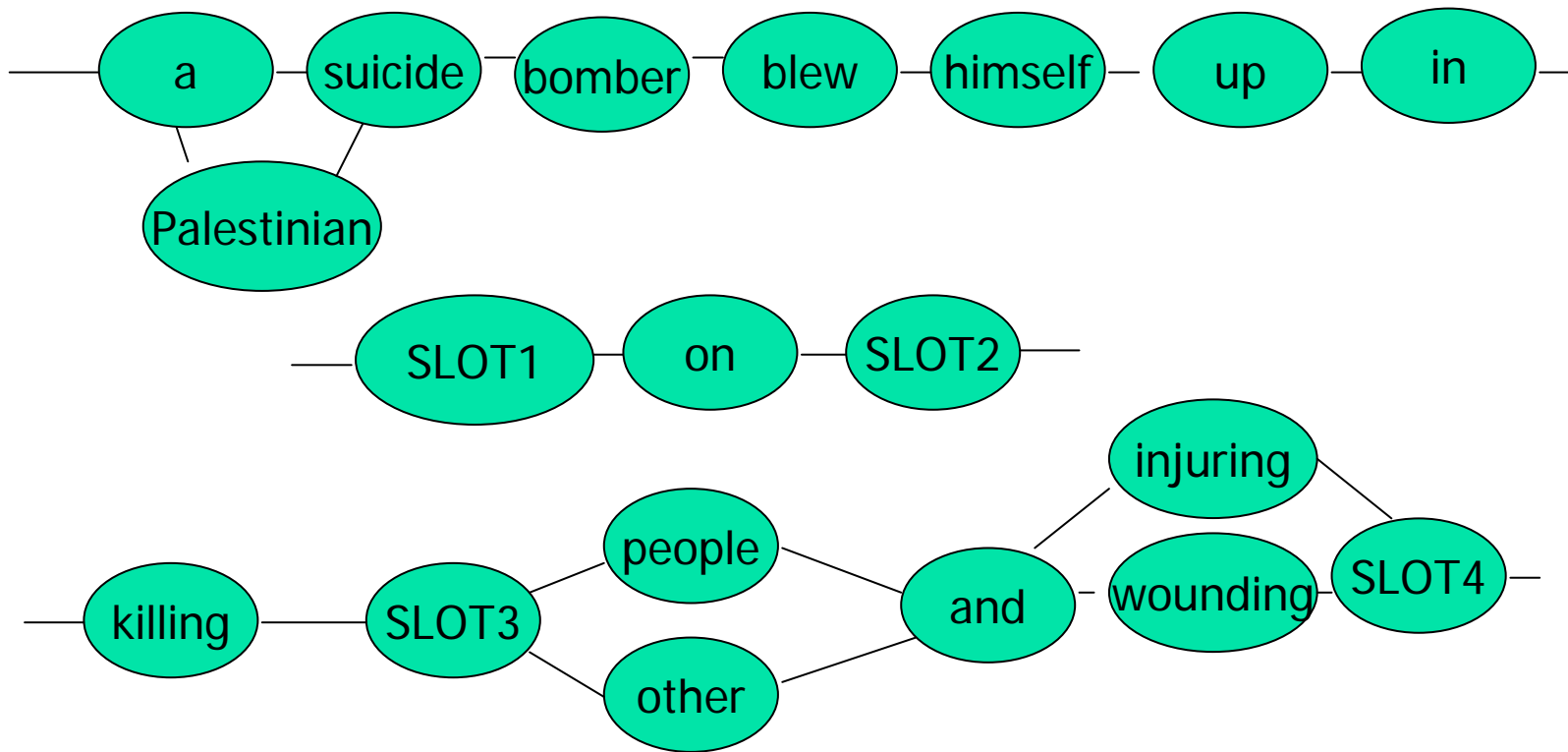


keep words



replace by
arguments

Patterns induced



Generating paraphrases

- finding equivalent patterns
 - X injured Y people, Z seriously = Y were injured by X among them Z were in serious condition
- exploit the corpus
 - equivalent patterns will have similar arguments/slots in the corpus
 - given two clusters from where the patterns were derived identify sentences “published” on the same date & topic
 - compare the arguments in the pattern variables
 - patterns are equivalent if overlap of word in arguments > thr

Multi-document summarization

- motivation
 - I want a summary of all major political events in the UK from May 2001 to June 2001
 - search on the Web or in a closed collection can return thousands of hits
 - none of them has all the answers we need

Multi-document summarization

- professional abstractors
 - conference proceedings o journals
- journal editors
 - introduction
- government analysts
 - organization and people profiles
- academics
 - summary of state of the art

Multi-document summarization

- definition
 - Brief representation of the contents of a set of “related” documents (by event, event type, group, or terms, etc) where important tasks are redundancy elimination and identification and expression of differences between sources

Multi-document summarization

- Redundancy of information
 - the destruction of Rome by the Barbarians in 410....
 - Rome was destroyed by Barbarians.
 - Barbarians destroyed Rome in the V Century
 - In 410, Rome was destroyed. The Barbarians were responsible.
- fragmentary information
 - D1="earthquake in Turkey"; D2="measured 6.5"
- contradictory information
 - D1="killed 3"; D2="killed 4"
- relations between documents
 - inter-document-coreference
 - D1="Tony Blair visited Bush"; D2="UK Prime Minister visited Bush"

Similarity metrics

- text fragments (sentences, paragraphs, etc.) represented in a vector space model OR as bags of words and use set operations to compare them
- can be “normalized” (stemming, lemmatised, etc)
- stop words can be removed
- weights can be term frequencies or tf*idf...

$$D_i = (d_{i1}, \dots, d_{in})$$

$$\text{sim}(D_i, D_j) = \sum_k d_{ik} \cdot d_{jk} \quad \cos(D_i, D_j) = \frac{\sum_k (d_{ik} \cdot d_{jk})}{\sqrt{\sum_k (d_{ik})^2 \sum_k (d_{jk})^2}}$$

Morphological techniques

- IR techniques: a query is the input to the system
- Goldstein&al'00. Maximal Marginal Relevance
 - a formula is used allowing the inclusion of sentences relevant to the query but different from those already in the summary

Q = query

R = list of documents

D_k = k - document in list

S = subset of R already scanned

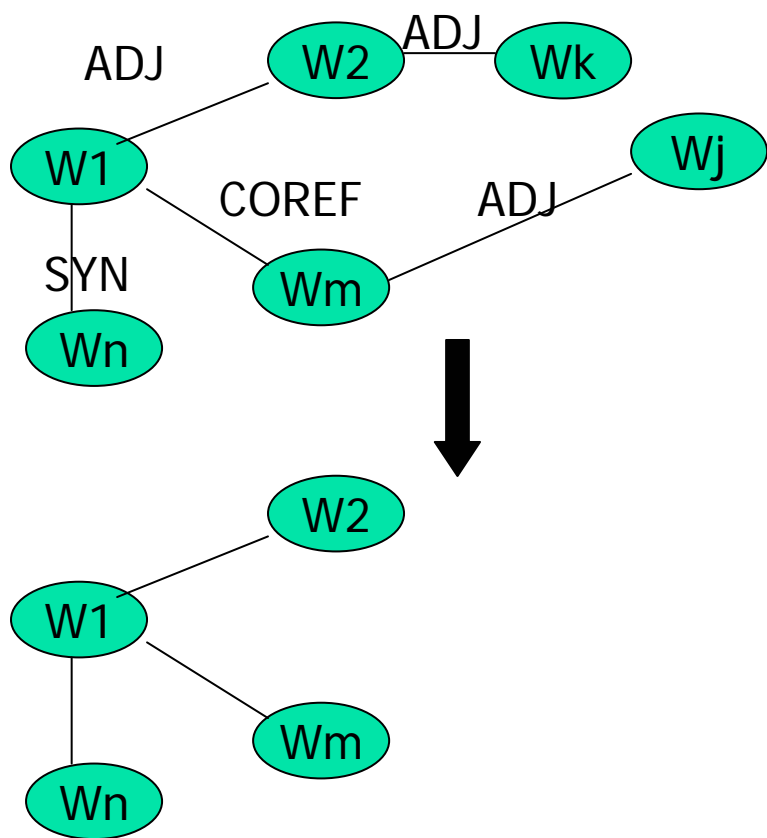
$$MMR(Q, R, S) = \arg \max_{D_i \in R \setminus S} (\lambda sim_1(D_i, Q) +$$

$$(\lambda - 1) \max_{D_j \in S} sim_2(D_i, D_j))$$

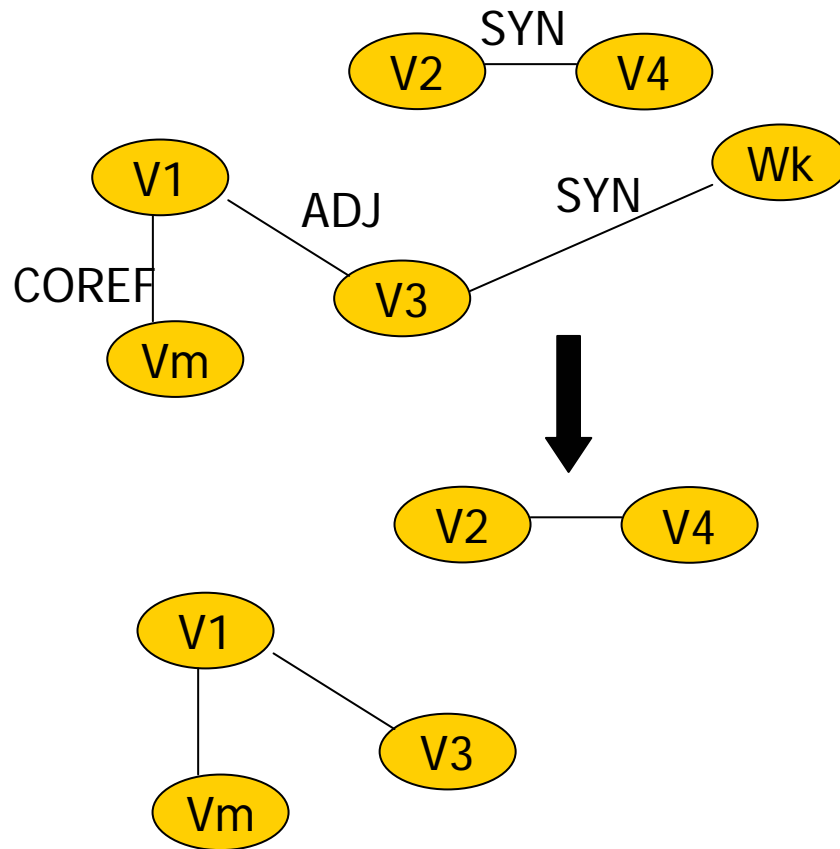
Morphological techniques

- Mani & Bloedor'99. Graphs representing text structure
 - proximity (ADJ), coreference (COREF), synonym (SYN)
 - link words by relations (create a graph)
 - identify regions in graph related to query (input to the system)
 - identification of common terms
 - identification of different terms
 - use common words & different words to select sentences from the texts

Cohesion graph



DOC1



DOC2

Sentence ordering

- important for both single and multi-document summarization (Barzilay, Elhadad, McKeown'02)
- some strategies
 - Majority order
 - Chronological order
 - Combination
- probabilistic model (Lapata'03)
 - the model learns order constraints in a particular domain
 - the main component is a probability table
 - $P(S_i|S_{i-1})$ for sentences S
 - the representation of each sentence is a set of features for
 - verbs, nouns, and dependencies

Semantic techniques

- Knowledge-based summarization in SUMMONS (Radev & McKeown'98)
- Conceptual summarization
 - reduction of content
- Linguistic summarization
 - conciseness

SUMMONS

- corpus of summaries
 - strategies for content selection
 - summarization lexicon
- summarization from a template knowledge base
- planning operators for content selection
 - 8 operators
- linguistic generation
 - generating summarization phrases
 - generating descriptions

Example summary

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.

Input

- correct templates sorted by date
- templates which refer to the same event are grouped together
- primary and secondary sources are added to the initial set of templates

Input

MESSAGE: ID
 SECSOURCE: SOURCE
 SECSOURCE: DATE
 PRIMSOURCE: 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
 SECSOURCE: SOURCE
 SECSOURCE: DATE
 PRIMSOURCE: 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
 SECSOURCE: SOURCE
 SECSOURCE: DATE
 PRIMSOURCE: 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
 SECSOURCE: SOURCE
 SECSOURCE: DATE
 PRIMSOURCE: 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"

Operators

- Change of perspective

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.*

- Contradiction

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.*

Logical operators

- Contradiction operator: given templates T1 & T2

```
T1.LOC == T2.LOC &&
```

```
T1.TIME < T2.TIME && ...
```

```
T1.SRC2 != T2.SRC2 =>
```

```
    apply contradiction "with-new-account" to  
    T1,T2
```

- templates have weights which are reduced when combined
- the combined template has its weights boosted
- ideally the combined resulting template will be used for generating the final summary

Text Summarization Evaluation

- Identify when a particular algorithm can be used commercially
- Identify the contribution of a system component to the overall performance
- Adjust system parameters
- Objective framework to compare own work with work of colleagues

Text Summarization Evaluation

- Expensive because requires the construction of standard sets of data and evaluation metrics
- May involve human judgement
- There is disagreement among judges
- Automatic evaluation would be ideal but not always possible

Intrinsic Evaluation

- Summary evaluated on its own or comparing it with the source
 - Is the text cohesive and coherent?
 - Does it contain the main topics of the document?
 - Are important topics omitted?

Extrinsic Evaluation

- Evaluation in an specific task
 - Can the summary be used instead of the document?
 - Can the document be classified by reading the summary?
 - Can we answer questions by reading the summary?

Evaluation metrics

- extracts
 - automatic vs. human
- precision
 - Ratio of correct summary sentences
- recall
 - Ratio of relevant sentences included in summary

Evaluation of extracts

| | System | |
|-------|--------|----|
| Human | + | - |
| + | TP | FN |
| - | FP | TN |

- F-score (F)
- Accuracy (A)

- precision (P)

$$\frac{TP}{TP+FP}$$

- recall (R)

$$\frac{TP}{TP + FN}$$

$$\frac{(\beta^2 + 1)P.R}{\beta^2 P + R}$$

$$\frac{TP + TN}{TP + FP + FP + FN}$$

Evaluation of extracts

- Relative utility (fuzzy) (Radev&al'00)
 - each sentence has a degree of "belonging to a summary"
 - $H = \{(S_1, 10), (S_2, 7), \dots, (S_n, 1)\}$
 - $A = \{S_2, S_5, S_n\} \Rightarrow \text{val}(S_2) + \text{val}(S_5) + \text{val}(S_n)$
 - Normalize dividing by maximum

Other metrics

- Content based metrics
 - “The president visited China” vs “The visit of the President to China”
 - overlap
 - Based on set n-gram intersection
 - Fine grained metrics than combine different sets of n-grams can be used
 - cosine in Vector Space Model
 - Longest subsequence
 - Minimal number of deletions/insertions needed to obtain two identical chains
 - Do they really measure semantic content?

- We will see ROUGE adopted by DUC

Pyramids

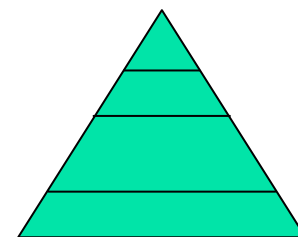
- Human evaluation of content: Nenkova & Passonneau (2004)
- based on the distribution of content in a pool of summaries
- Summarization Content Units (SCU):
 - fragments from summaries
 - identification of similar fragments across summaries

Pyramids

- SCU have
 - id, a weight, a NL description, and a set of contributors
 - similar to Teufel & van Halterer (2003)
- SCU1 ($w=4$)
 - A1 - two Libyans indicted
 - B1 - two Libyans indicted
 - C1 - two Libyans accused
 - D2 – two Libyans suspects were indicted

Pyramids

- a “pyramid” of SCUs of height n is created for n gold standard summaries
- each SCU in tier T_i in the pyramid has weight i
- with highly weighted SCU on top of the pyramid
- the best summary is one which contains all units of level n , then all units from $n-1, \dots$
- if D_i is the number of SCU in a summary which appear in T_i for summary D , then the weight of the summary is:



$w=n$

$w=n-1$

$w=1$

$$D = \sum_{i=1}^n i * D_i$$

Pyramids score

- let X be the total number of units in a summary
- it is shown that more than 4 ideal summaries are required to produce reliable rankings

$$Max = \sum_{i=j+1}^n i * |T_i| + j * (X - \sum_{i=j+1}^n |T_i|)$$

$$j = \max_i \left(\sum_{t=i}^n |T_t| \geq X \right)$$

$$Score = D / Max$$

DucView v. 1.1 - Annotating Peer [Close] [Maximize] [Minimize]

File Edit Options Help

On the European Community, for example, Mr Kinnock's opposition to British membership was less that the EC was a 'capitalist club' and more that the hated Tories' favoured it.

When the referendum was held in 1978 just 18 per cent voted for a Welsh assembly.

Today's British government plays into the nationalist hand by a staggering lack of vision in adapting constitutionally as it has done over 150 years,' he said.

The government plans to set up 21 new authorities to replace eight counties and 37 districts in Wales.

A promise of a plebiscite on British participation in a single currency union would add extra assurance.

The Welsh nationalists would probably be willing to provide ad hoc support to a Labour minority government, but without the sort of unconditional guarantees required for stable rule.

Yet with only days to the election to go, Mr Kinnock's personality remains at the very centre of the British political plot.

The prime minister also rejected a call for a Welsh parliament by Mr Ieuan Jones, the Plaid Cymru MP for Ynys Mon.

Both Johns - Major and Smith - are against referenda.

Facts with Labour and the Tories are equally dangerous.

| Add Contributor | Remove | Order |
|---|--------|-------|
| <ul style="list-style-type: none"> ④ Wales currently has eight counties <ul style="list-style-type: none"> eight counties ③ Blair has set out proposals for devolution ③ Differences in opinion among Wales countrymen fall along language lines ③ Implementation of the local government reform will take place in April 1995 ③ In 1992, the British prime minister rejected a call for a Welsh parliament <ul style="list-style-type: none"> The prime minister also rejected a call for a Welsh parliament ③ Plaid Cymru wants to bypass devolution ③ Shadow elections to the new unitary authorities will be held as early as next year <ul style="list-style-type: none"> shadow elections to new unitary authorities as early as next year ③ The British government is pressing ahead with restructuring plans ③ The government's restructuring plan calls for 21 new authorities <ul style="list-style-type: none"> The government plans to set up 21 new authorities to replace ③ Tony Blair is the Labour party leader ③ Welsh devolution has popular support ② Blair has set off a constitutional battle with the Tories ② Conservatives called the present arrangement the best for the UK ② Conservatives want to maintain a strong Welsh Office with a cabinet minister ② Implementation of the white paper was postponed until at least 1995 ② In the 1979 referendum on devolution, the vote was four to one against ② In Wales, Welsh speakers tend to favor full independence ② John Major is Prime Minister ② Major reminded the Welsh nationalists in 1992 of popular disapproval of devolution ② Only 18 percent of Welsh favored devolution in 1978 <ul style="list-style-type: none"> When the referendum was held in 1978 just 18 per cent voted for a Welsh assembly ② Plaid Cymru has only 3 or 4 representatives in Parliament ② Plaid Cymru is seeking support from the Labour party ② Support for devolution crystallized by 1994 | | |

< >

In 1992, the British prime minister, John Major, rejected a call for a Welsh parliament with devolved powers by a Welsh member of parliament. Major reminded him that results of an earlier referendum, in 1978, showed that only 18 percent of Welsh voters favored a parliament. In November 1993, the British Tory government pressed ahead with plans to reform the structure of local government in Wales by holding shadow elections to new unitary authorities as early as 1994. It intended that these councils should take over in April 1995. This was in lieu of granting demands for an independent Welsh

Loaded C:\mydata\duc-2005\633\633.M.250.G.6.pan

SUMMAC evaluation

- System independent evaluation
- high scale
- basically extrinsic
- 16 systems
- summaries in tasks carried out by defence analysis of the American government

SUMMAC

- “ad hoc” task
 - indicative summaries
 - system receives a document + a topic and has to produce a topic-based
 - analyst has to classify the document in two categories
 - Document deals with topic
 - Document does not deal with topic

SUMMAC

- Categorization task
 - generic summaries
 - given n categories and a summary, the analyst has to classify the document in one of the n categories or none of them
 - one wants to measure whether summaries reduce classification time without losing classification accuracy

SUMMAC

- Experimental conditions
 - text: full-document; fixed-length summary; variable-length summary; default summary (baseline)
 - technology: each of the participants
 - consistency: 51 analysts

SUMMAC

- data
 - “ad hoc”: 20 topics each with 50 documents
 - categorization: 10 topics each with 100 documents (5 categories)

SUMMAC

- Results “ad hoc” task
 - Variable length summaries take less time to classify by a factor of 2 (33.12 sec/doc vs. 58.89 sec/doc with full-text)
 - Classification accuracy reduced but not significantly

SUMMAC

- Results of categorization task
 - only significant differences in time between 10% length summaries and full-documents
 - no difference in classification accuracy
 - many FN observed (automatic summaries lack many relevant topics)
- 3 groups of systems observed
- ad hoc: pair-wise human agreement 69%; 53% 3-way; 16% unanimous

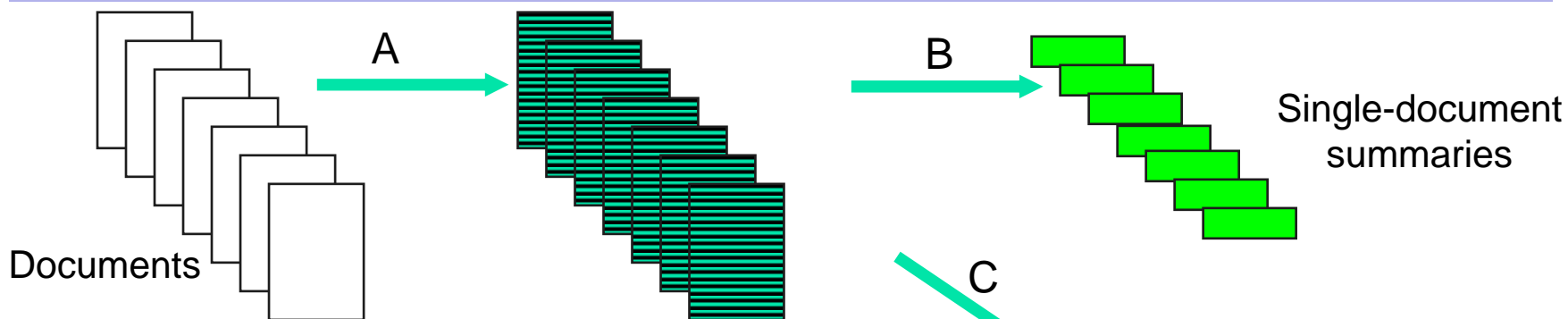
DUC experience

- National Institute of Standards and Technology (NIST)
- further progress in summarization and enable researchers participate in large-scale experiments
- Document Understanding Conference
 - 2000-2006
- Call begin of the year, data released in ~May

DUC 2001

- Task 1
 - given a document, create a generic summary of the document (100 words)
 - 30 sets of ~10 documents each
- Task 2
 - given a set of documents, create summaries of the set (400, 200, 100, 50 words)
 - 30 sets of ~ 10 documents each

Human summary creation

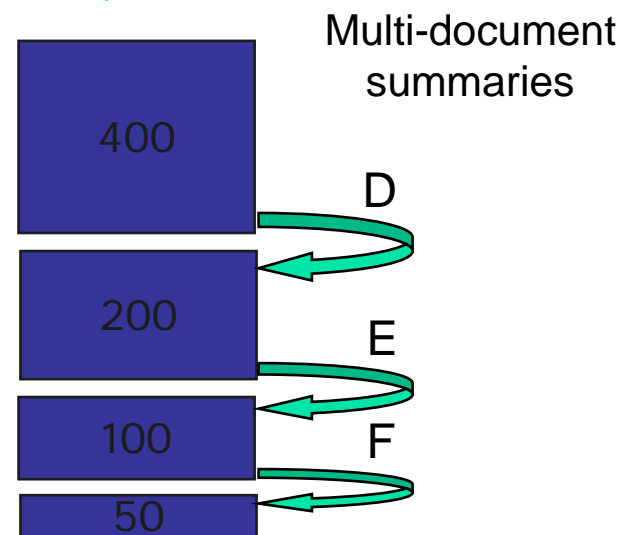


A: Read hardcopy of documents.

B: Create a 100-word softcopy summary for each document **using the document author's perspective.**

C: Create a 400-word softcopy multi-document summary of all 10 documents **written as a report for a contemporary adult newspaper reader.**

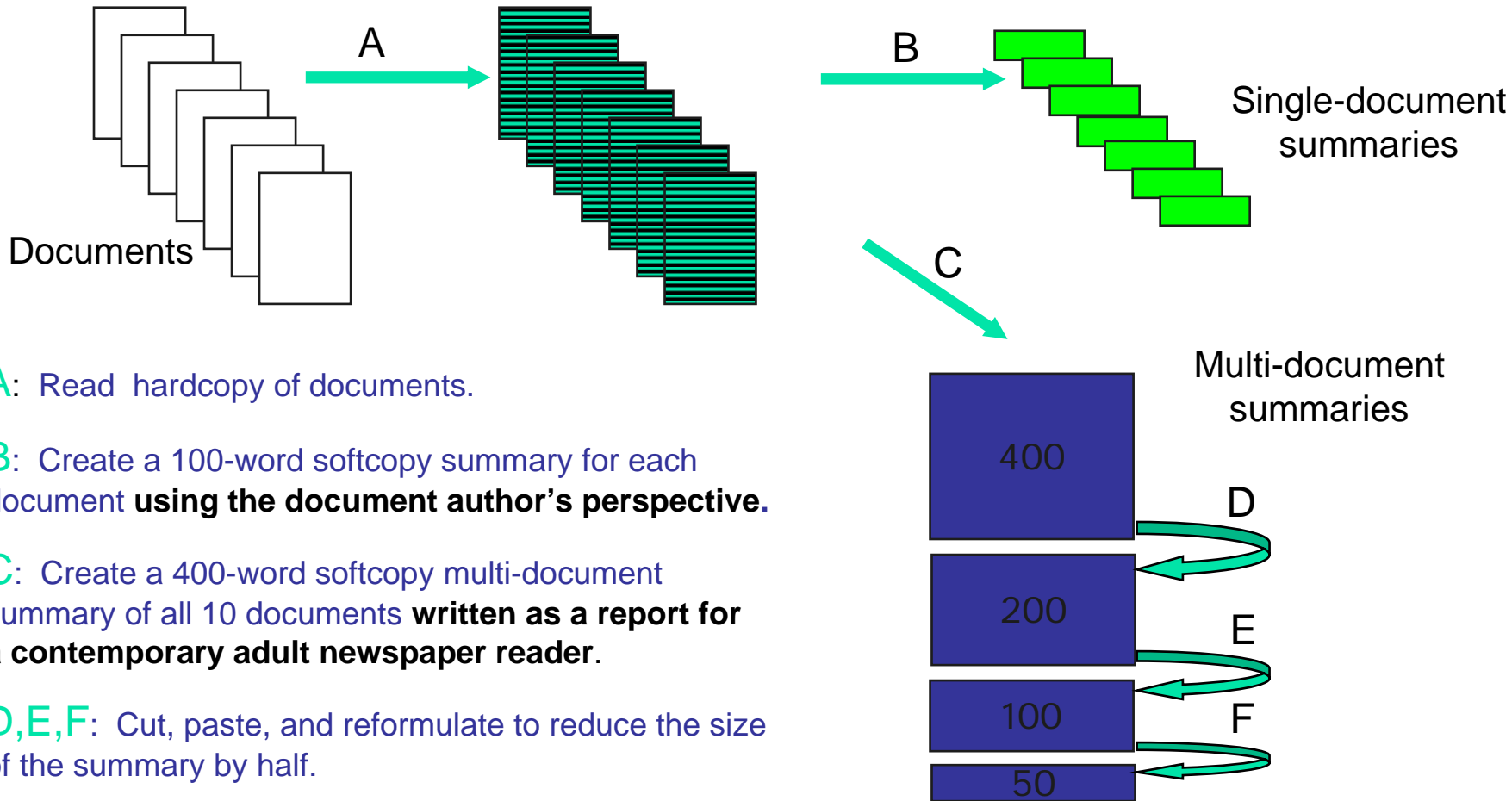
D,E,F: Cut, paste, and reformulate to reduce the size of the summary by half.



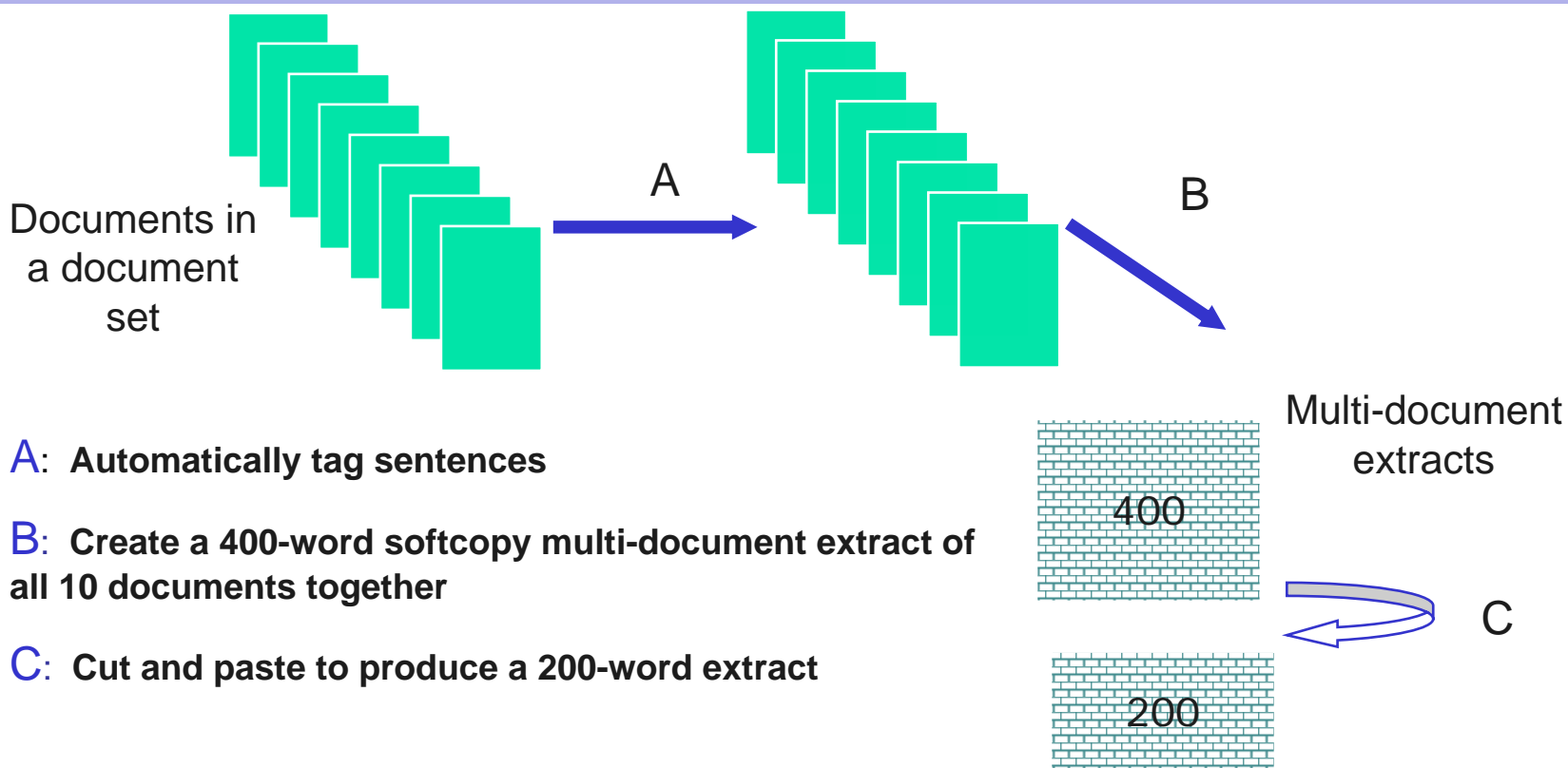
DUC 2002

- Task 1
 - given a document, create a generic summary of the document (100 words)
 - 60 sets of ~10 documents each
- Task 2
 - given a set of documents, create summaries of the set (400, 200, 100, 50 words)
 - given a set of documents, create two extracts (400, 200 words)
 - 60 sets of ~ 10 documents each

Human summary creation



Manual extract creation



A: Automatically tag sentences

B: Create a 400-word softcopy multi-document extract of all 10 documents together

C: Cut and paste to produce a 200-word extract

DUC 2003

- Task 1
 - 10 words single-document summary
- Task 2
 - 100 word multi-document summary of cluster related by an event
- Task 3
 - given a cluster and a viewpoint, 100 word multi-document summary of cluster
- Task 4
 - given a cluster and a question, 100 word multi-document summary of cluster

Viewpoints & Topics & Questions

Viewpoint:

Forty years after poor parenting was thought to be the cause of schizophrenia, researchers are working in many diverse areas to refine the causes and treatments of this disease and enable early diagnosis.

Topic:

30042 - PanAm Lockerbie Bombing Trial

Seminal Event

WHAT: Kofi Annan visits Libya to appeal for surrender of PanAm bombing suspects

WHERE: Tripoli, Libya

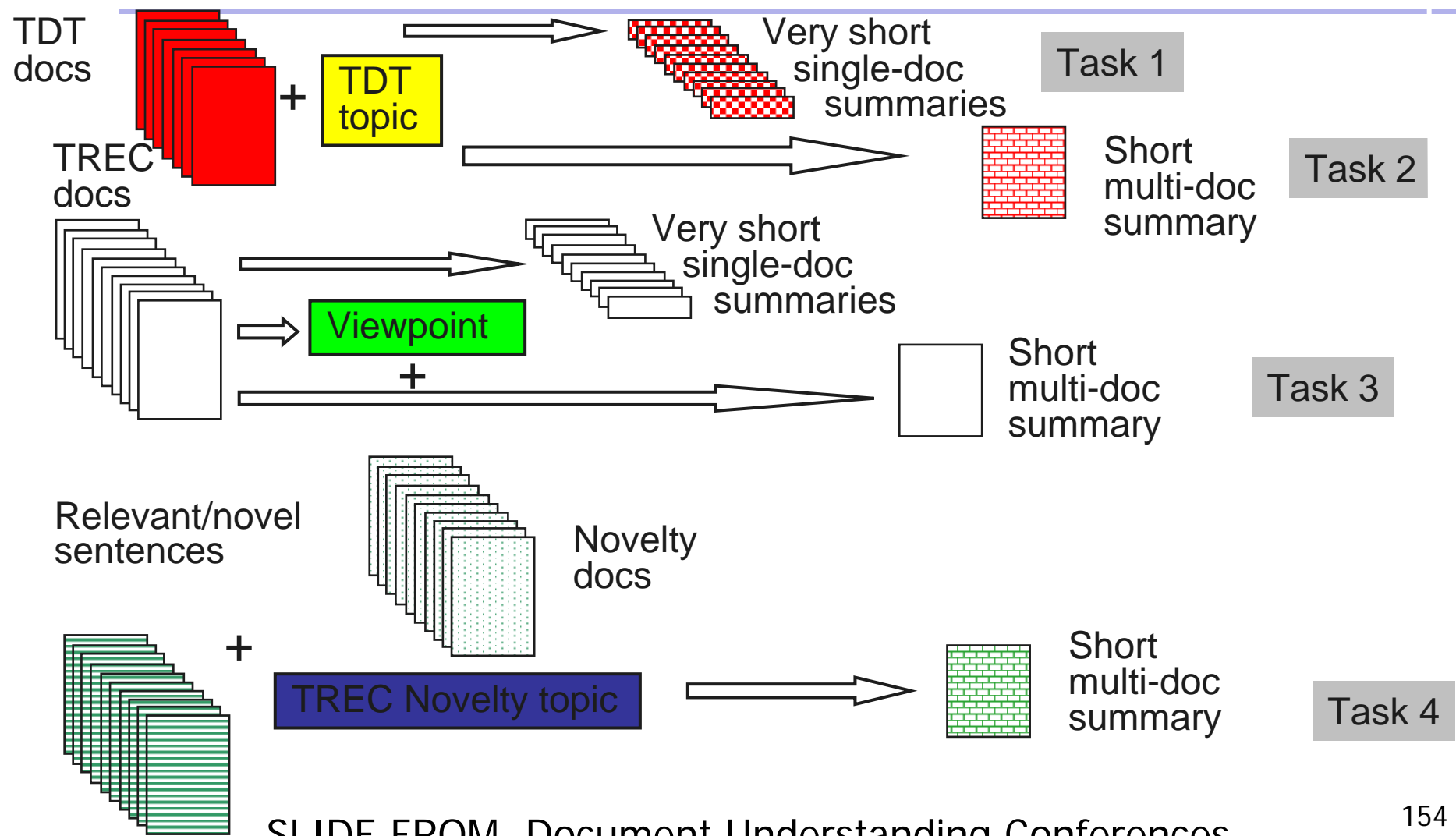
WHO: U.N. Secretary-General Kofi Annan; Libyan leader Moammar Gadhafi

WHEN: December, 1998

Question:

What are the advantages of growing plants in water or some substance other than soil?

Manual abstract creation



DUC 2004

- Tasks for 2004
 - Task 1: very short summary
 - Task 2: short summary of cluster of documents
 - Task 3: very short cross-lingual summary
 - Task 4: short cross-lingual summary of document cluster
 - Task 5: short person profile
- Very short (VS) summary ≤ 75 bytes
- Short (S) summary ≤ 665 bytes
- Each participant may submit up to 3 runs

DUC 2004 - Data

- 50 TDT English news clusters (tasks 1 & 2) from AP and NYT sources
 - 10 docs/topic
 - Manual S and VS summaries
- 24 TDT Arabic news clusters (tasks 3 & 4) from France Press
 - 13 topics as before and 12 new topics
 - 10 docs/topic
 - Related English documents available
 - IBM and ISI machine translation systems
 - S and VS summaries created from manual translations
- 50 TREC English news clusters from NYT, AP, XIE
 - Each cluster with documents which contribute to answering "Who is X?"
 - 10 docs/topic
 - Manual S summaries created

DUC 2004 - Tasks

- Task 1
 - VS summary of each document in a cluster
 - Baseline = first 75 bytes of document
 - Evaluation = ROUGE
- Task 2
 - S summary of a document cluster
 - Baseline = first 665 bytes of most recent document
 - Evaluation = ROUGE

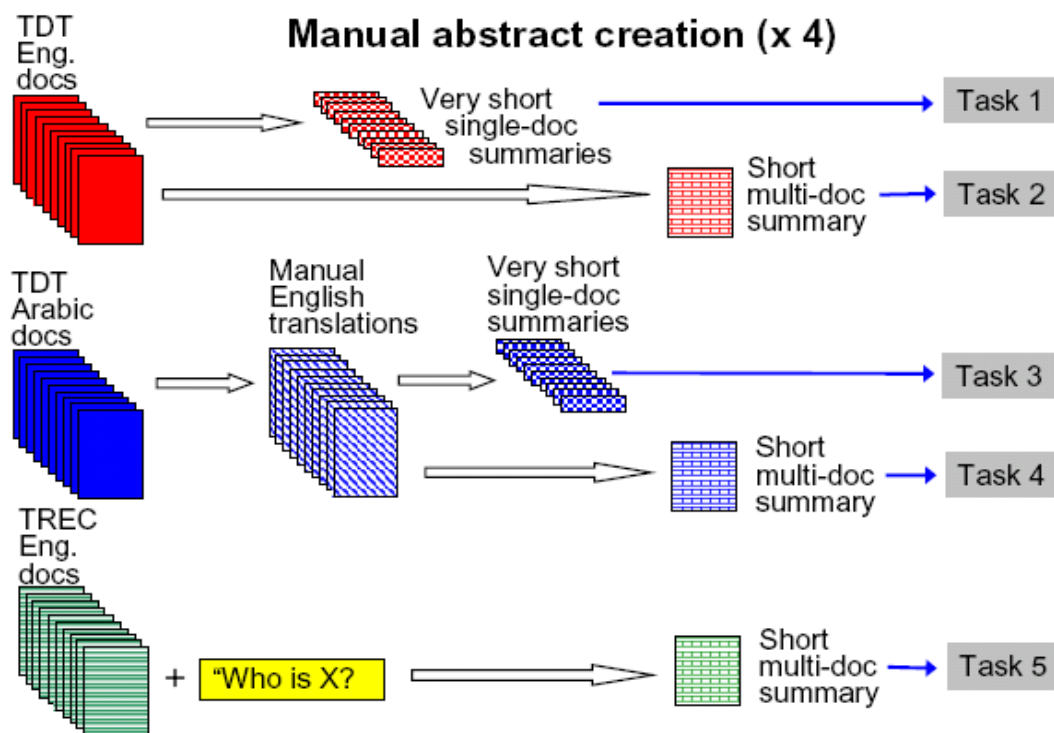
DUC 2004 - Tasks

- Task 3
 - VS summary of each translated document
 - Use: automatic translations; manual translations; automatic translations + related English documents
 - Baseline = first 75 bytes of best translation
 - Evaluation = ROUGE
- Task 4
 - S summary of a document cluster
 - Use: same as for task 3
 - Baseline = first 665 bytes of most recent best translated document
 - Evaluation = ROUGE

DUC 2004 - Tasks

- Task 5
 - S summary of document cluster + “Who is X?”
 - Evaluation = using Summary Evaluation Environment (SEE): quality & coverage; ROUGE

Summary of tasks



DUC 2004 – Human Evaluation

- Human summaries segmented in Model Units (MUs)
- Submitted summaries segmented in Peer Units (PUs)
- For each MU
 - Mark all PUs sharing content with the MU
 - Indicates whether the PUs express 0%, 20%, 40%, 60%, 80%, 100% of MU
 - For all non-marked PU indicate whether 0%, 20%, ... 100% of PUs are related but needn't to be in summary

Summary evaluation environment (SEE)

SEE - OUTPUT.D076.M.200.B.E.E.19

File Options Help

Peer Summary Path /nlpir/duc/duc2002/eval/peer5/D076.M.200.B.19.html Prev Summary Pair

Model Summary Path /nlpir/duc/duc2002/eval/models/D076.M.200.B.E.html Next Summary Pair

| Peer Summary | Model Summary |
|--|--|
| [1] Margaret Thatcher will be seen with Winston Churchill as the greatest British prime minister of the last 50 years. [2] She was elected in 1979, the first female prime minister in Europe, and won re-election in 1983 and in 1987, when she said she planned to "go on and on". [3] Earlier this year, Mrs. Thatcher overtook Liberal Lord Asquith's 1908-1916 tenure as prime minister to become Britain's longest continuously serving prime minister of the 20th century. [4] Margaret Thatcher set the example of what a woman could achieve in British society, but her critics say she did little else to help women along. [5] She led her party to victory in three elections, steered it through the war with Argentina to reclaim the Falklands, faced down the miners union in a long strike | [1] Prime Minister Margaret Thatcher, the Iron Lady of British politics, resigned Thursday. [2] Serving for over 11 years, longer than any prime minister in the 20th Century. [3] the announcement of her resignation took the world by surprise. [4] Mrs. Thatcher was the first woman prime minister in Great Britain [5] and is credited with reviving the faltering British economy in the early '80s. [6] Former President Reagan had nothing but praise for Mrs. Thatcher. [7] While he was still in office, the two shared a special relationship, [8] calling each other Margaret and Ronnie and often appearing together at international gatherings. [9] The relationship with American cooled with the coming of the Bush administration, but had improved in recent months. [10] Soviet President |

Quality Judgment 1 Quality Judgment 2 Content Unmarked Peer Units

Serving for over 11 years, longer than any prime minister in the 20th Century, Prev Next

Unit Coverage 3

The marked PUs, taken together, express:

100% 80% 60% 40% 20% 0%

of the meaning expressed by the current model unit.

0 of 12 quality questions judged (at 5 of 5 summary p... file://nlpir/duc/duc2002/eval/peer5/D076.M.200.B.19.html#3

DUC 2004 – Questions

- 7 quality questions
- 1) Does the summary build from sentence to sentence to a coherent body of information about the topic?
 - A. Very coherently
 - B. Somewhat coherently
 - C. Neutral as to coherence
 - D. Not so coherently
 - E. Incoherent
- 2) If you were editing the summary to make it more concise and to the point, how much useless, confusing or repetitive text would you remove from the existing summary?
 - A. None
 - B. A little
 - C. Some
 - D. A lot
 - E. Most of the text

DUC 2004 - Questions

- Read summary and answer the question
- Responsiveness (Task 5)
 - Given a question “Who is X” and a summary
 - Grade the summary according to how responsive it is to the question
 - 0 (worst) - 4 (best)

ROUGE package

- Recall-Oriented Understudy for Gisting Evaluation
- Developed by Chin-Yew Lin at ISI (see DUC 2004 paper)
- Compares quality of a summary by comparison with ideal(s) summaries
- Metrics count the number of overlapping units

ROUGE package

- ROUGE-N: N-gram co-occurrence statistics is a recall oriented metric

S1- Police killed the gunman

S2- Police kill the gunman

S3- The gunman kill police

S2=S3

ROUGE package

- ROUGE-L: Based on longest common subsequence

S1- Police killed the gunman

S2- Police kill the gunman

S3- The gunman kill police

S2 better than S3

ROUGE package

- ROUGE-W: weighted longest common subsequence, favours consecutive matches

X - A B C D E F G

Y1 - A B C D H I K

Y2 - A H B K C I D

Y1 better than Y2

ROUGE package

- ROUGE-S: Skip-bigram recall metric
- Arbitrary in-sequence bigrams are computed
 - S1 - police killed the gunman
 - S2 - police kill the gunman
 - S3 - the gunman kill police
 - S4 - the gunman police killed

S2 better than S4 better than S3
- ROUGE-SU adds unigrams to ROUGE-S

ROUGE package

- Co-relation with human judgment
- Experiments on DUC 2000-2003 data
- 17 ROUGE metrics tested
- Pearson's correlation coefficients computed

ROUGE Results

- ROUGE-S4, S9, and ROUGE-W1.2 were the best in 100 words single doc task, but were statistically indistinguishable from most other ROUGE metrics.
- ROUGE-1, ROUGE-L, ROUGE-SU4, ROUGE-SU9, and ROUGE-W1.2 worked very well in 10 words headline like task (Pearson's $\rho \sim 97\%$).
- ROUGE-1, 2, and ROUGE-SU* were the best in 100 words multi-doc task but were statistically equivalent to other ROUGE-S and SU metrics.
- ROUGE-1, 2, ROUGE-S, and SU worked well in other multi-doc tasks.

Basic Elements: going “semantics”

- BE (Hovy, Lin, Zhou'05)
 - head of a major syntactic structure (noun, verb, adjective, adverbial phrase)
 - relation between head-BE and single dependent
- Example
 - two Libyans were indicted for the Lockerbie bombing in 1991
 - lybians|two|nn (HM)
 - indicted|libyans|obj (HMR)
 - bombing|lockerbie|nn
 - indicted|bombing|for
 - bombing|1991|nn

Basic elements

- break ideal and system summaries in units
 - use parser + a set of rules
 - Charniak parser + CYL rules = BE-L
 - Minipar + JF rules = BE-F
 - each unit receives one point per summary where it is observed, for example
- match units in system summaries against units in ideal summaries obtaining scores
 - lexical identity; lemma identity; synonymy; etc.
- combine scores
 - sum up individual scores for BE in system summaries
- more work is needed

DUC 2004 – Some systems

- Task 1
 - TOPIARY (Zajic&al'04)
 - University of Maryland; BBN
 - Sentence compression from parse tree
 - Unsupervised Topic Discovery (UTD): statistical technique to associate meaningful names to topics
 - Combination of both techniques
 - MEAD (Erkan&Radev'04)
 - University of Michigan
 - Centroid + Position + Length
 - Select one sentence as S summary

DUC 2004 – Some systems

- Task 2
 - CLASSY (Conroy&al'04)
 - IDA/Center for Computing Sciences; Department of Defence; University of Maryland
 - HMM with summary and non-summary states
 - Observation input = topic signatures
 - Co-reference resolution
 - Sentence simplification
 - Cluster Relevance & Redundancy Removal (Saggion&Gaizauskas'04)
 - University of Sheffield
 - Sentence cluster similarity + sentence lead document similarity + absolute position
 - N-gram based redundancy detection

DUC 2004 – Some systems

- Task 3
 - LAKHAS (Douzidia&Lapalme'04)
 - Universite de Montreal
 - Summarize from Arabic documents, then translates
 - Sentence scoring = lead + title + cue + tf*idf
 - Sentence reduction = name substitution; word removal; phrase removal; etc.
 - After translation with Ajeeb (commercial system) good results
 - After translation with ISI best system

DUC 2004 – Some systems

- Task 5
 - Lite-GISTexter (Lacatusu&al'04)
 - Language Computer Corporation
 - Syntactic structure
 - entity in appositive construction ("X, a ...")
 - entity subject of copula ("X is the...")
 - sentence containing key are scored by syntactic features

DUC 2005

```
<TOPIC ID="d324e" GRANULARITY="specific">
```

How have relations between Argentina and Great Britain developed since the 1982 war over the Falkland Islands? Have diplomatic, economic, and military relations been restored? Do differences remain over the status of the Falkland Islands?

```
</TOPIC>
```

```
<TOPIC ID="d332h" GRANULARITY="general">
```

What kinds of non-tax crimes have lead to tax evasion prosecutions (failure to file, inaccurate filing), instead of or in addition to prosecution for the non-tax crimes themselves?

```
</TOPIC>
```

- Topic based summarization
 - given a set of documents and a topic description, generate a 250 words summary
- Evaluation
 - ROUGE
 - Pyramid

Single-document summary (DUC)

```
<SUM DOCSET="d04" TYPE="PERDOC" SIZE="100" DOCREF="FT923  
6455" SELECTOR="A" SUMMARIZER="A">
```

US cities along the Gulf of Mexico from Alabama to eastern Texas were on alert last night as Hurricane Andrew headed west after hitting southern Florida leaving at least eight dead, causing severe property damage, and leaving 1.2 million homes without electricity. Gusts of up to 165 mph were recorded. It is the fiercest hurricane to hit the US in decades. As Andrew moved across the Gulf there was concern that it might hit New Orleans, which would be particularly susceptible to flooding, or smash into the concentrated offshore oil facilities. President Bush authorized federal disaster assistance for the affected areas.</SUM>

Multi-document summaries (DUC)

```
<SUM DOCSET="d04" TYPE="MULTI" SIZE="50" DOCREF="FT923-5267 FT923-6110 FT923-6455 FT923-5835 FT923-5089 FT923-5797 FT923-6038" SELECTOR="A" SUMMARIZER="A">
```

Damage in South Florida from Hurricane Andrew in August 1992 cost the insurance industry about \$8 billion making it the most costly disaster in the US up to that time. There were fifteen deaths and in Dade County alone 250,000 were left homeless.</SUM>

```
<SUM DOCSET="d04" TYPE="MULTI" SIZE="100" DOCREF="FT923-5267 FT923-6110 FT923-6455 FT923-5835 FT923-5089 FT923-5089 FT923-5797 FT923-6038" SELECTOR="A" SUMMARIZER="A">
```

Hurricane Andrew which hit the Florida coast south of Miami in late August 1992 was at the time the most expensive disaster in US history. Andrew's damage in Florida cost the insurance industry about \$8 billion. There were fifteen deaths, severe property damage, 1.2 million homes were left without electricity, and in Dade county alone 250,000 were left homeless. Early efforts at relief were marked by wrangling between state and federal officials and frustrating delays, but the White House soon stepped in, dispatching troops to the area and committing the federal government to rebuilding and funding an effective relief effort.</SUM>

Extracts (DUC)

```
<SUM DOCSET="d061" TYPE="MULTI-E" SIZE="200"
DOCREF="AP880911-0016 AP880912-0137 AP880912-0095 AP880915-0003 AP880916-0060
WSJ880912-0064" SELECTOR="J" SUMMARIZER="B">
<s docid="WSJ880912-0064" num="18" wdcnt="15"> Tropical Storm Gilbert formed in the
eastern Caribbean and strengthened into a hurricane Saturday night.</s>
<s docid="AP880912-0137" num="22" wdcnt="13"> Gilbert reached Jamaica after skirting
southern Puerto Rico, Haiti and the Dominican Republic.</s>
<s docid="AP880915-0003" num="13" wdcnt="33"> Hurricane Gilbert, one of the
strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled
thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts
of Cancun and Cozumel.</s>
<s docid="AP880915-0003" num="44" wdcnt="21"> The Mexican National Weather Service
reported winds gusting as high as 218 mph earlier Wednesday with sustained winds
of 179 mph.</s>
```

Other evaluations

- Multilingual Summarization Evaluation (MSE) 2005
 - basically task 4 of DUC 2004
 - Arabic/English multi-document summarization
 - human evaluation with pyramids
 - automatic evaluation with ROUGE
- MSE 2006 underway
 - automatic evaluation with ROUGE

Other evaluations

- Text Summarization Challenge (TSC)
 - Summarization in Japan
 - Two tasks in TSC-2
 - A: generic single document summarization
 - B: topic based multi-document summarization
 - Evaluation
 - summaries ranked by content & readability
 - summaries scored in function of a revision based evaluation metric

SUMMAC Corpus

- Categorization & ad-hoc tasks
 - documents with relevance judgements
- 2000 full text sources
- each sentence annotated with information as to which summarization system selected that sentence
- suggested use:
 - train to behave as a summarizer which will select sentence chosen by most summarizers

Annotated Sentences

The screenshot shows the GATE 3.0 build 1846 interface. The main window displays a document with several sentences highlighted in red, indicating they are annotated. The interface includes a menu bar (File, Options, Tools, Help), a toolbar, and a left-hand pane with a tree view showing 'GATE', 'Applications', 'Language Resources', and 'Processing Resources'. The main area is divided into a table of annotations and a text editor.

| Type | Set | Start | End | Features |
|------|------------------|-------|-----|---|
| S | Original markups | 4 | 189 | {sys_adhoc_best=[5,6,8,14,16]} |
| S | Original markups | 194 | 309 | {sys_adhoc_best=[5]} |
| S | Original markups | 310 | 509 | {sys_adhoc_best=[5,8,14]} |
| S | Original markups | 514 | 683 | {sys_adhoc_best=[6,11,12,14,16], sys_adhoc_fixed=[11,12,13,14]} |
| S | Original markups | 684 | 894 | {sys_adhoc_best=[11], sys_adhoc_fixed=[11,15]} |

Below the table, the text editor shows the following text with red highlights:

In a courtroom filled with 100 lawyers, a bankruptcy judge on Friday began trying to sort out how to handle the complex, historic Public Service Company of New Hampshire reorganization.

Judge James Yacos reached no immediate decision on a 'procedural mechanism' for the case after a 2(-hour hearing. But he approved several preliminary motions, including the utility's request to retain a law firm to consider a U.S. Supreme Court challenge of a state law that bars advance charges for power plants.

Public Service, choked by debt from its \$2.1 billion investment in the stalled Seabrook nuclear power plant, sought refuge from creditors in bankruptcy court on Jan. 28. Two days earlier, the state Supreme Court upheld the state law that bars the utility from charging ratepayers for Seabrook unless the reactor surmounts evacuation-planning hurdles and wins a commercial license.

New Hampshire's largest power company, with 360,000 customers,

The right-hand pane shows 'Original markups' with a checked box for 'S' and an unchecked box for 'doc'. A 'New' button is visible at the bottom right of the pane.

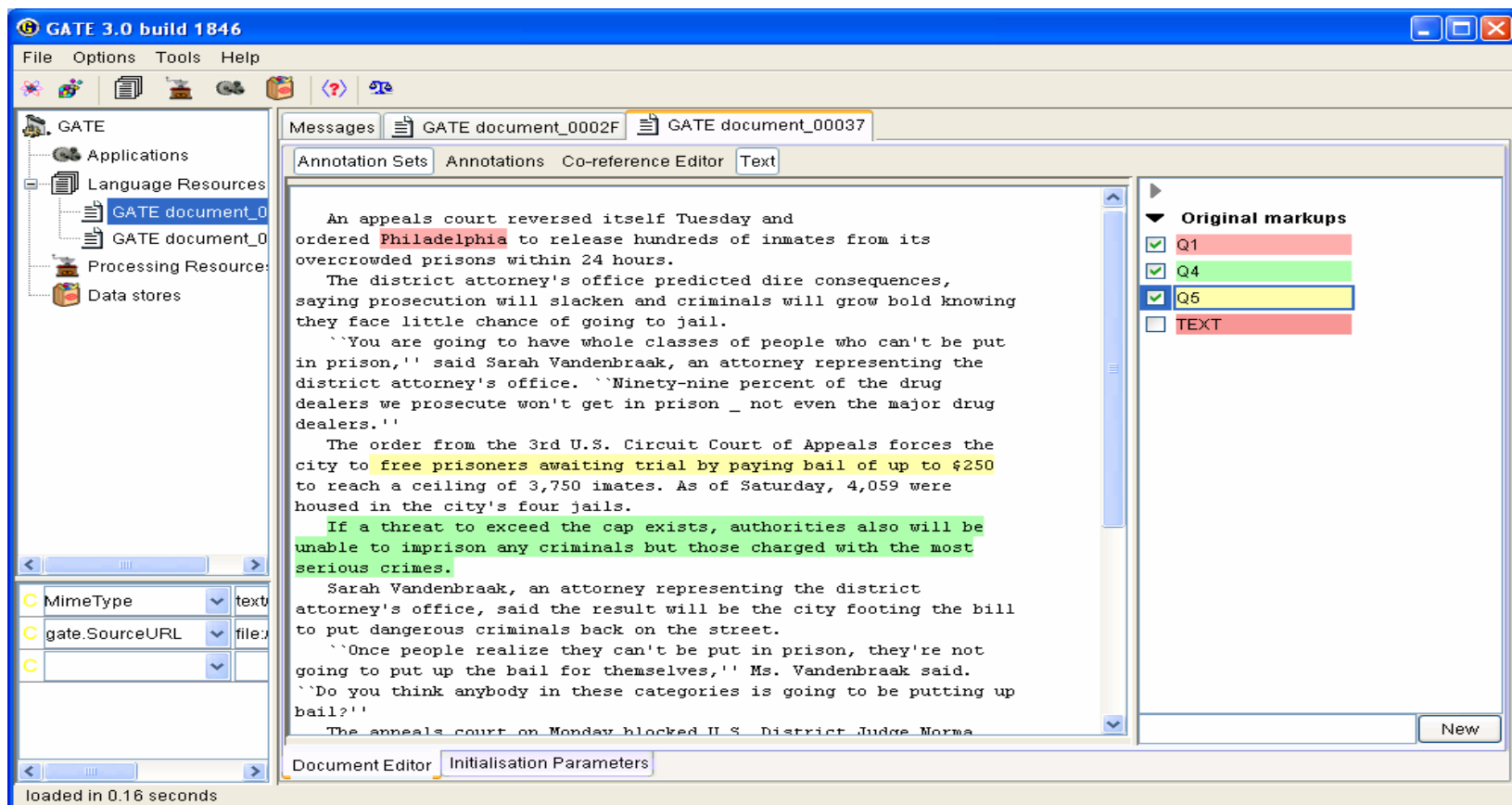
SUMMAC Q&A

- Topic descriptions
- Questions per topic
- Documents per topic
- Answer keys
- Model summaries

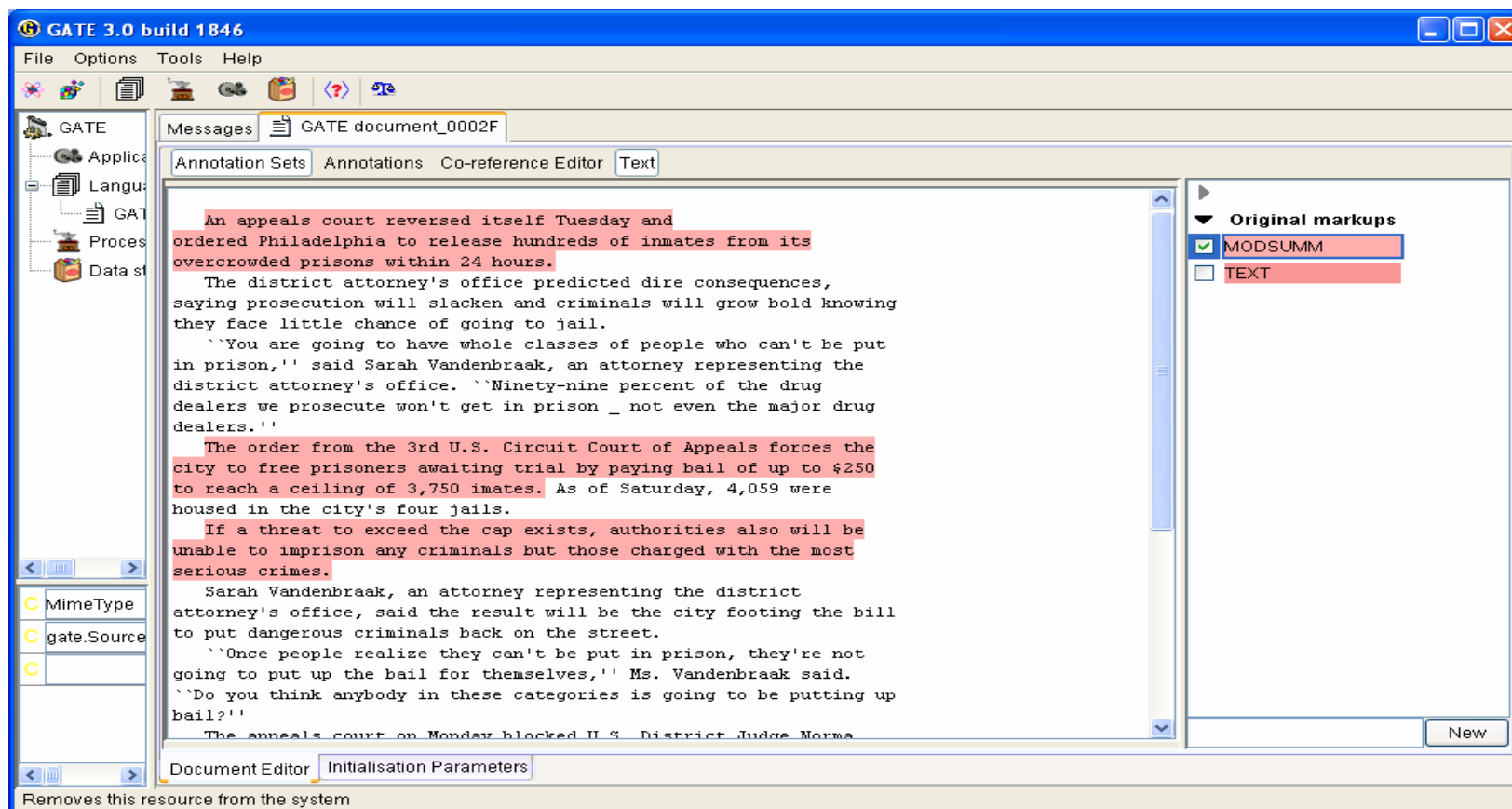
SUMMAC Q&A: Topics and Questions

- **Topic 151: “Coping with overcrowded prisons”**
 1. What are name and/or location of the correction facilities where the reported overcrowding exists?
 2. What negative experiences have there been at the overcrowded facilities (whether or not they are thought to have been caused by the overcrowding)?
 3. What measures have been taken/planned/recommended (etc.) to accommodate more inmates at penal facilities, e.g., doubling up, new construction?
 4. What measures have been taken/planned/recommended (etc.) to reduce the number of new inmates, e.g., moratoriums on admission, alternative penalties, programs to reduce crime/recidivism?
 5. What measures have been taken/planned/recommended (etc.) to reduce the number of existing inmates at an overcrowded facility, e.g., granting early release, transferring to un-crowded facilities?

Q&A Keys



Model Q&A Summaries



Summac tools

- Sentence alignment tool
 - sentence-similarity program
 - measures the similarity between each sentence in the summary with each sentence in the full document

MEAD

- Dragomir Radev and others at University of Michigan
- publicly available toolkit for multi-lingual summarization and evaluation
- implements different algorithms: position-based, centroid-based, $it * idf$, query-based summarization
- implements evaluation methods: co-selection, relative-utility, content-based metrics

MEAD

- Perl & XML-related Perl modules
- runs on POSIX-conforming operating systems
- English and Chinese
- summarizes single documents and clusters of documents

MEAD

- compression = words or sentences; percent or absolute
- output = console or specific file
- ready-made summarizers
 - lead-based
 - random

MEAD architecture

- configuration files
- feature computation scripts
- classifiers
- re-rankers

Configuration file

```
<MEAD-CONFIG TARGET='GA3' LANG='ENG' CLUSTER-PATH='/clair4/mead/data/GA3'  
  DATA-DIRECTORY='/clair4/mead/data/GA3/docsent'>  
  
  <FEATURE-SET BASE-DIRECTORY='/clair4/mead/data/GA3/feature/'>  
    <FEATURE NAME='Centroid'  
SCRIPT='/clair4/mead/bin/feature-scripts/Centroid.pl HK-WORD-enidf ENG' />  
    <FEATURE NAME='Position'  
SCRIPT='/clair4/mead/bin/feature-scripts/Position.pl' />  
    <FEATURE NAME='Length'  
SCRIPT='/clair4/mead/bin/feature-scripts/Length.pl' />  
  </FEATURE-SET>  
  
  <CLASSIFIER COMMAND-LINE='/clair4/mead/bin/default-classifier.pl \  
    Centroid 1 Position 1 Length 9' SYSTEM='MEADORIG' RUN='10/09' />  
  
  <RERANKER COMMAND-LINE='/clair4/mead/bin/default-reranker.pl MEAD-cosine 0.7' />  
  
  <COMPRESSION BASIS='sentences' PERCENT='20' />  
  
</MEAD-CONFIG>
```

clusters & sentences

```
<?xml version='1.0'?>
<!DOCTYPE CLUSTER SYSTEM '/clair4/mead/dtd/cluster.dtd'>

<CLUSTER LANG='ENG'>
  <D DID='41' />
  <D DID='81' />
  <D DID='87' />
</CLUSTER>
```

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE DOCSSENT SYSTEM '/clair4/mead/dtd/docsent.dtd'>

<DOCSSENT DID='41' LANG='ENG'>
<BODY>
<HEADLINE>
<S PAR="1" RSNT="1" SNO="1">Egyptians Suffer Second Air
Tragedy in a Year </S>
</HEADLINE>
<TEXT>
<S PAR='2' RSNT='1' SNO='2'>CAIRO, Egypt -- The crash of a
Gulf Air flight that killed 143 people in Bahrain is a disturbing
deja vu for Egyptians: It is the second plane crash within a
year to devastate this Arab country.</S>
<S PAR='2' RSNT='2' SNO='3'>Sixty-three Egyptians were on
board the Airbus A320, which crashed into shallow Persian Gulf
waters Wednesday night after circling and trying to land in
Bahrain.</S>
```


extract & summary

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE EXTRACT SYSTEM '/clair/tools/mead/dtd/extract.dtd'>

<EXTRACT QID='GA3' LANG='ENG' COMPRESSION='7'
SYSTEM='MEADORIG' RUN='Sun Oct 13 11:01:19 2002'>
<S ORDER='1' DID='41' SNO='2' />
<S ORDER='2' DID='41' SNO='3' />
<S ORDER='3' DID='41' SNO='11' />
<S ORDER='4' DID='81' SNO='3' />
<S ORDER='5' DID='81' SNO='7' />
<S ORDER='6' DID='87' SNO='2' />
<S ORDER='7' DID='87' SNO='3' />
</EXTRACT>
```

[1]The Disaster Relief Fund Advisory Committee has approved a grant of \$3 million to Hong Kong Red Cross for emergency relief for flood victims in Jiangxi, Hunan and Hubei, the Mainland.

[2]Together with the earlier grant of \$3 million to World Vision Hong Kong, the Advisory Committee has so far approved \$6 million from the Disaster Relief Fund for relief projects to assist the victims affected by the recent floods in the Mainland.

Mead at work

- Mead computes sentence features (real-valued)
 - position, length, centroid, etc.
 - similarity with first, is longest sentence, various query-based features
- Mead combines features
- Mead re-rank sentences to avoid repetition

Summarization with GATE

- GATE (<http://gate.ac.uk>)
 - General Architecture for Text Engineering
 - Processing & Language Resources
 - Documents follow the TIPTSTER
- Text Summarization in GATE (Saggion'02)
 - processing resources compute feature-values for each sentence in a document
 - features are stored in documents
 - feature-values are combined to score sentences
 - need gate + summarization jar file + creole.xml

Summarization with GATE

- implemented in JAVA
- platform independent
 - Windows, Unix, Linux
- is a Java library which can be used to create summarization applications
- summarization applications
 - single document summarization: English, Swedish, Latvian, Finnish, Spanish
 - multi-document summarization: centroid-based
 - 2nd position in DUC 2004 (task 2)
 - cross-lingual summarization: (English, Arabic)

Functions

- sentence identification
- NE recognition & coreference resolution
- summarization components
 - position, keyword, title, query
 - Vector Space Model for content analysis
 - similarity metrics implemented
- evaluation of extracts is possible with GATE AnnotationDiff tool
- evaluation of abstracts is possible with an implementation of BLUE (Pastra&Saggion'03)

Units represented in a VSM

- linear feature combination
- text fragment represented as $\langle \text{term}, \text{tf} \cdot \text{idf} \rangle$
- cosine used as one metric to measure similarity

$$\cos(v_i, v_j) = \frac{\sum_k (t_{ik} \cdot t_{jk})}{\sqrt{\sum_k (t_{ik})^2 \sum_k (t_{jk})^2}}$$

Gate 2.1_01 build 1191

File Options Tools Help

Messages | SUMMARIZER | SIMPLE SUMMARIZER | WSJ911023-0158.xml_00064 | WSJ900706-0092.xml_0004C

Text Annotations Annotation Sets Coreference Print

Gate

- Applications
- SUMMARIZER
- Language Resources
- Processing Resources
- SIMPLE SUMMARIZER
- INCLUDE TEXT PAIR
- COREFERENCE SCORER
- ORTHO MATCHER
- NE TRANSDUCER
- LOOK UP
- TERM FREQ SCORER
- STATISTIC COHESION
- SENT/SENT SIMILARITY
- DOC/SENT SIMILARITY
- CORPUS STATISTICS
- SPLITTER

WASHINGTON — In searching for someone to care for her three children, Bonnie Fay of Alexandria, Va., turned to Mother's Helper Agency, which billed itself in national advertising as "serving America's most prestigious families." But the nanny sent by the agency turned out to be something less than Mary Poppins.

Ms. Fay says she found marijuana in the room of the woman she hired three years ago as a live-in nanny. And she fired the woman after she admitted stealing money from Ms. Fay to silence a boyfriend who threatened to tell the nanny's parents about a past abortion. Mother's Helper then demanded an additional \$100 to find a replacement nanny, Ms. Fay says.

The agency says it doesn't know whether Ms. Fay's allegations are true, and it asked for the extra fee because the 60-day replacement period was past. Ms. Fay sued the agency and won a judgment of \$1,323 for costs incurred, including fees paid to the agency. Her advice to other working parents: consumer beware.

In a nation of ever more two-paycheck couples, nannies have become a hot commodity. Agencies have sprung up to supply them for fees of \$400 to \$2,500. Many of the agencies do a conscientious job. But some dissatisfied customers tell of unchecked references, deficient screening, broken contracts and unrefunded fees.

Annotations Editor | Features Editor | Initialisation Parameters

SUMMARIZER run in 74.497 seconds

- Date
- FirstPerson
- Identifier
- JobTitle
- Lead
- Location
- Lookup
- Money
- Organization
- Person
- Sentence
- SpaceToken
- Split
- Temp
- Title
- Token
- sent_token_tf
- sentence_vector
- Original markups annotations
- SUMMARY 20 annotations
 - Summary

Training the summarizer

- GATE incorporates ML functionalities through WEKA
- training and testing modes are available
 - annotate sentences selected by humans as keys
 - annotate sentences with feature-values
 - learn model
 - use model for creating extracts of new documents

Resources: SummBank

- Johns Hopkins Summer Workshop 2001
- Language Data Consortium (LDC)
- Drago Radev, Simone Teufel, Wai Lam, Horacio Saggion
- Development & implementation of resources for experimentation in text summarization
- <http://www.summarization.com>

SummBank

- Hong Kong News Corpus
- formatted in XML
- 40 topics/themes identified by LDC
- creation of a list of relevant documents for each topic
- 10 documents selected for each topic = clusters

SummBank

- 3 judges evaluate each sentence in each document
- relevance judgements associated to each sentence (relative utility)
- these are values between 0-10 representing how relevant is the sentence to the theme of the cluster
- they also created multi-document summaries at different compression rates (50 words, 100 words, etc.)

C:\development\resources\summarization\resources\jhu-clusters\551\19980731_003.bis.xml - Microsoft Inte...

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media Recycle Bin Mail Print Send To Favorites

Address C:\development\resources\summarization\resources\jhu-clusters\551\19980731_003.bis.xml Go Links >>

```

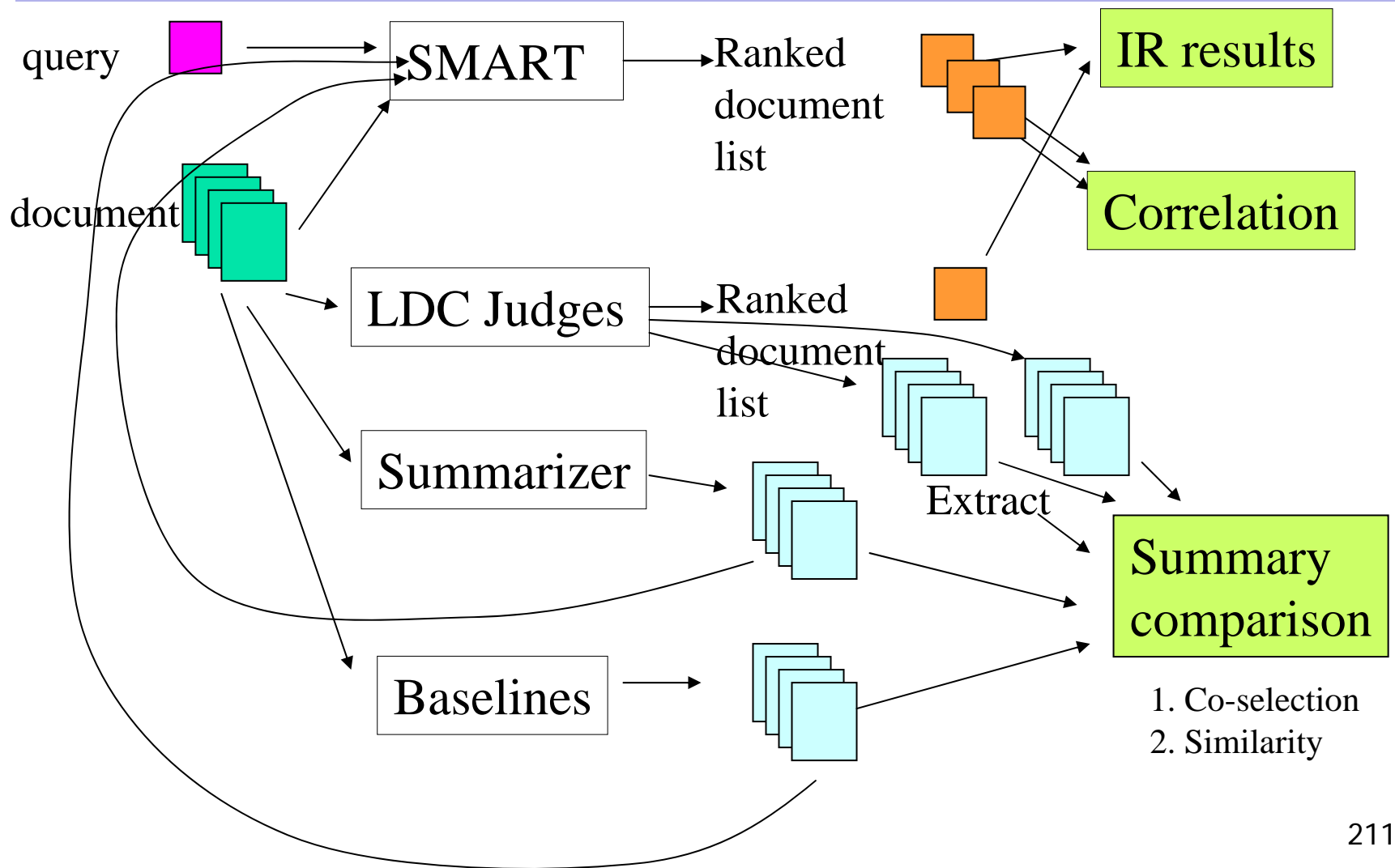
<!DOCTYPE DOCTYPE (View Source for full doctype...)>
- <DOCTYPE CLUSTER="551" QUERY="Natural disaster victims aided" DID="D-19980731_003.e"
  DOCNO="4334" LANG="ENG" CORR-DOC="D-19980731_006.c">
- <BODY>
- <HEADLINE>
  <S PAR="1" RSNT="1" SNO="1" JUDGE3="pfried" UTILITY3="6" JUDGE2="jtyson" UTILITY2="10"
    JUDGE1="ahester" UTILITY1="10">Aid for flood victims in the Mainland</S>
</HEADLINE>
- <TEXT>
  <S PAR="2" RSNT="1" SNO="2" JUDGE3="pfried" UTILITY3="10" JUDGE2="jtyson"
    UTILITY2="10" JUDGE1="ahester" UTILITY1="6">The Disaster Relief Fund Advisory
    Committee has approved a grant of $3 million to Hong Kong Red Cross for emergency
    relief for flood victims in Jiangxi, Hunan and Hubei, the Mainland.</S>
  <S PAR="3" RSNT="1" SNO="3" JUDGE3="pfried" UTILITY3="10" JUDGE2="jtyson" UTILITY2="9"
    JUDGE1="ahester" UTILITY1="6">Together with the earlier grant of $3 million to World
    Vision Hong Kong, the Advisory Committee has so far approved $6 million from the
    Disaster Relief Fund for relief projects to assist the victims affected by the recent
    floods in the Mainland.</S>
  <S PAR="3" RSNT="2" SNO="4" JUDGE3="pfried" UTILITY3="9" JUDGE2="jtyson" UTILITY2="3"
    JUDGE1="ahester" UTILITY1="8">The Committee hopes that the grants can help to
    provide some immediate relief to the victims.</S>
  <S PAR="4" RSNT="1" SNO="5" JUDGE3="pfried" UTILITY3="7" JUDGE2="jtyson" UTILITY2="6"
    JUDGE1="ahester" UTILITY1="7">To ensure that the money will be used for the purpose
    designated, the Government has required Hong Kong Red Cross to submit an
    evaluation report and audited accounts on the use of the grant after the project has
  
```

Done My Computer

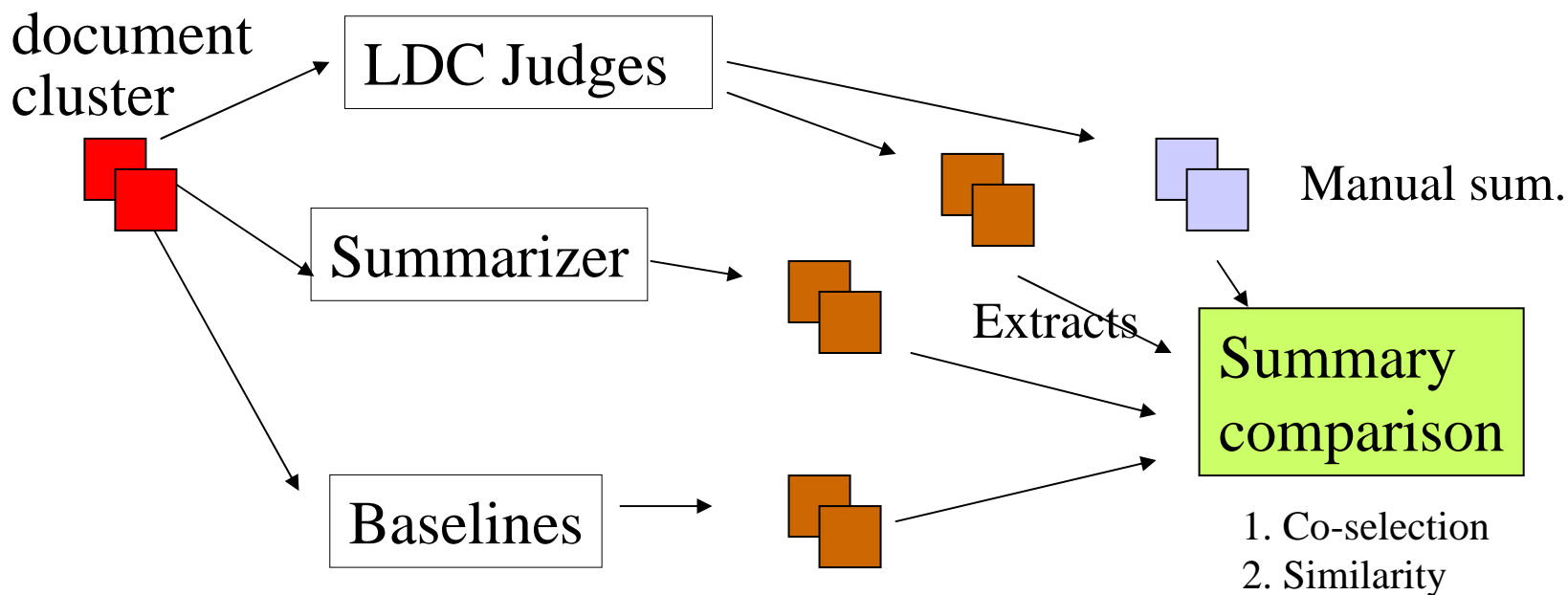
SummBank

- extracts were created for all documents
- implementation of evaluation metrics
 - co-selection
 - content-based
 - rank correlation in IR context

Single document evaluation



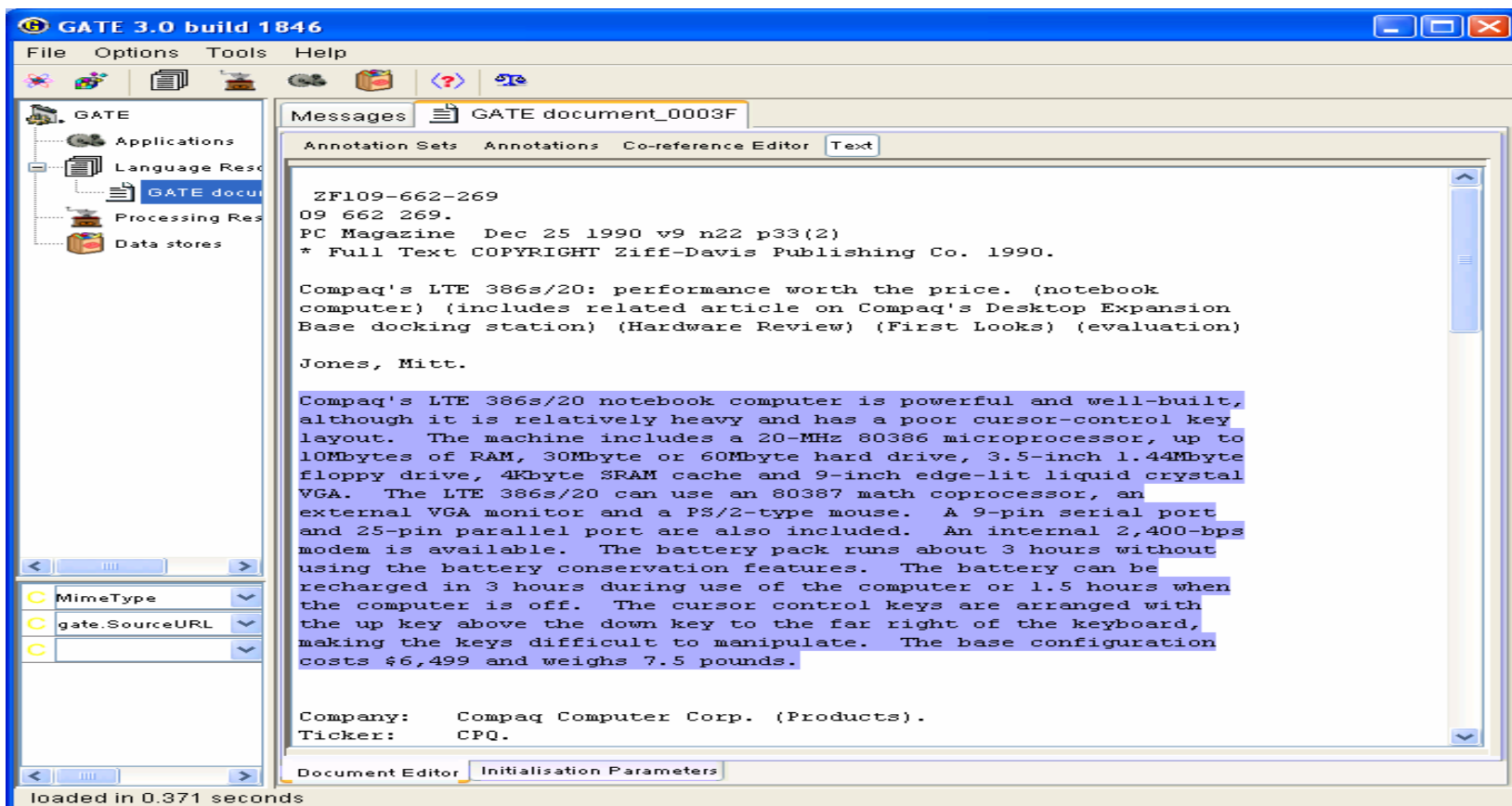
Multi-document evaluation



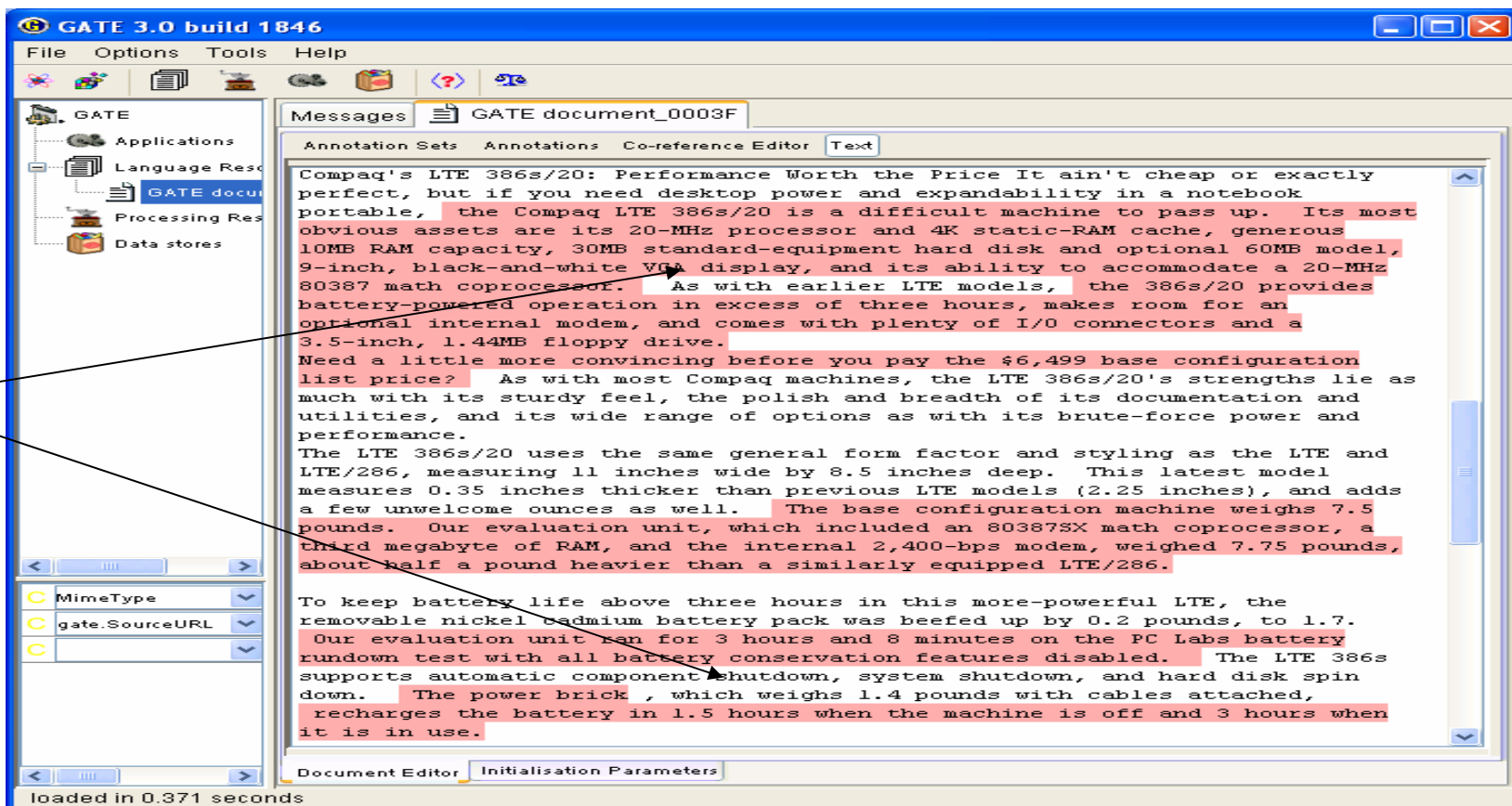
Ziff-Davis Corpus for Summarization

- Each document contains the DOC, DOCNO, and TEXT fields, etc.
- The SUMMARY field contains a summary of the full text within the TEXT field.
- The TEXT has been marked with ideal extracts at the clause level.

Document Summary



Clause Extract



The extracts

- Marcus'99
- Greedy-based clause rejection algorithm
 - clauses obtained by segmentation
 - "best" set of clauses
 - reject sentence such that the resulting extract is closer to the ideal summary

Uses of the corpus

- Study of sentence compression
 - following Knight & Marcu'01
- Study of sentence combination
 - following Jing&McKeown'00

Other corpora

- SumTime-Meteo (Sripada&Reiter'05)
 - University of Aberdeen
 - (<http://www.siggen.org/>)
 - weather data to text
- KTH eXtract Corpus (Dalianis&Hassel'01)
 - Stockholm University and KTH
 - news articles (Swedish & Danish)
 - various sentence extracts per document

Other corpora

- University of Woverhampton
- CAST (Computer-Aided Summarisation Tool) Project (Hasler&Orasan&Mitkov'03)
- newswire texts + popular science
- annotated with:
 - essential sentences
 - unessential fragments in those sentences
 - links between sentences when one is needed for the understanding of the other

Text Reuse in METER

- University of Sheffield
- Texts from the Press Association and British news paper reports
 - 1,700 texts
 - texts are topic-related
 - newspaper texts can be: wholly derived; partially derived; or non-derived
 - marked-up with SGML and TEI
 - two domains: law/courts and showbiz

Types of re-use

- rewriting
 - re-arranging order or positions
 - replacing words by synonyms or substitutable terms
 - deleting parts
 - change inflection, voice, etc.
- at word/string level
 - verbatim
 - rewrite
 - new

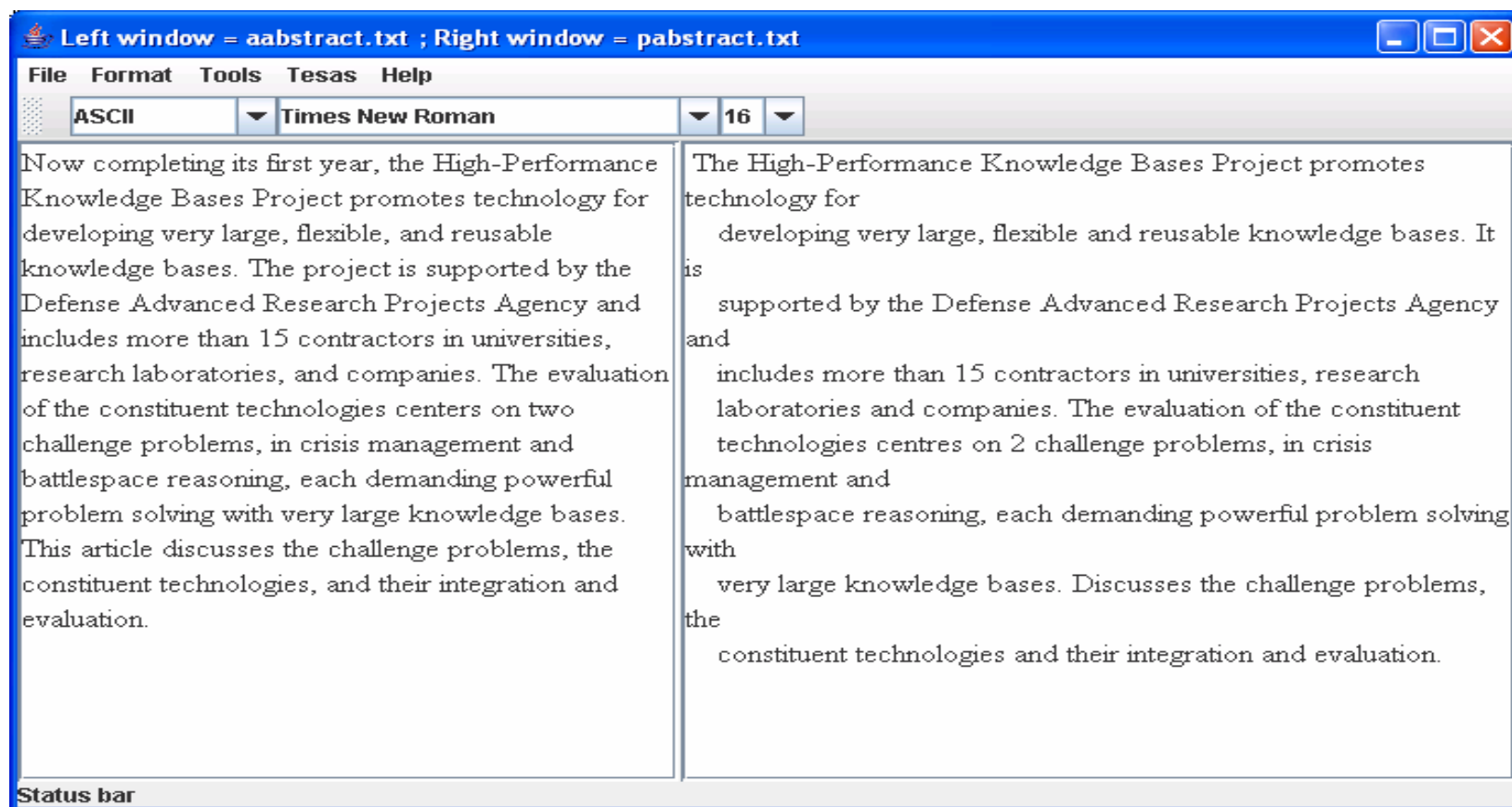
Original (PA) *A drink-driver who ran into the Queen Mother's official Daimler was fined £700 and banned from driving for two years.*

Rewrite (The Sun) *A DRUNK driver who ploughed into the Queen Mother's limo was fined £700 and banned for two years yesterday.*

Rewrite (The Mirror) *A BOOZY driver who smashed into the Queen Mums's chauffeur-driven Daimler minutes after she had been dropped off was banned for two years and fined £700 yesterday.*

Rewrite (Daily Star) *A DRUNK driver who crashed into the back of the Queen Mum's limo was banned for two years yesterday.*

Texas Tool for Sentence Alignment



Texas Tool for Sentence Alignment

The screenshot shows a software window titled "TESAS Report Package [1]" with several tabs: "Text Output", "Scores Analysis", "Weighted Score", "Alignment Table", and "Summary Report". Below the tabs are two buttons: "Save Left Window" and "Save Right Window". The main area is split into two panes. The left pane shows candidate source text with three sentences, each enclosed in an XML-style tag: `<s n="1">`, `<s n="2">`, and `<s n="3">`. The right pane shows suggested derived text with three corresponding sentences, each enclosed in a tag: `<s n="1" corresp="1" score="0.989">`, `<s n="2" corresp="2" score="0.964">`, and `<s n="3" corresp="3" score="0.989">`. A blue banner at the bottom of the window provides a detailed explanation of the alignment tags.

TESAS Report Package [1]

Text Output Scores Analysis Weighted Score Alignment Table Summary Report

Save Left Window Save Right Window

`<body>`

`<s n="1">`Now completing its first year, the High-Performance Knowledge Bases Project promotes technology for developing very large, flexible, and reusable knowledge bases.`</s>`

`<s n="2">`The project is supported by the Defense Advanced Research Projects Agency and includes more than 15 contractors in universities, research laboratories, and companies.`</s>`

`<s n="3">`The evaluation of the constituent technologies centers on two challenge problems, in crisis management and battlespace reasoning, each demanding powerful problem solving with very large knowledge bases.`</s>`

`<body>`

`<s n="1" corresp="1" score="0.989">`The High-Performance Knowledge Bases Project promotes technology for developing very large, flexible and reusable knowledge bases.`</s>`

`<s n="2" corresp="2" score="0.964">`It is supported by the Defense Advanced Research Projects Agency and includes more than 15 contractors in universities, research laboratories and companies.`</s>`

`<s n="3" corresp="3" score="0.989">`The evaluation of the constituent technologies centres on 2 challenge problems, in crisis management and battlespace reasoning, each demanding powerful problem solving with very large knowledge bases.`</s>`

This window displays a suggested derived text (right) and its candidate source text (left). The sentences are delimited and given sequential numbers. For each sentence in the candidate derived text, the position of suggested source sentence(s) is given with tag: `<s n corresp="{location}">`. If no match, zero is given for the location.

Texas Tool for Sentence Alignment

TESAS Report Package [1]

| Text Output | Scores Analysis | Weighted Score | Alignment Table | Summary Report |
|---|------------------|----------------|---|---|
| | Similarity Score | | Suggested derived Sentence | Suggested Source Sentence(s) |
| PSNG=0.951 WS=0.989 | | | developing very large, flexible and reusable knowledge bases. | Project promotes technology for developing very large, flexible, and reusable knowledge bases. |
| PSD=0.961 PS=0.938 PSNG=1.0 WS=0.964 | | | 2) It is supported by the Defense Advanced Research Projects Agency and includes more than 15 contractors in universities, research laboratories and companies. | 2) The project is supported by the Defense Advanced Research Projects Agency and includes more than 15 contractors in universities, research laboratories, and companies. |
| PSD=1.0 PS=1.0 PSNG=0.889 WS=0.989 | | | 3) The evaluation of the constituent technologies centres on 2 challenge problems, in crisis management and battlespace reasoning, each demanding powerful problem solving with very large knowledge bases. | 3) The evaluation of the constituent technologies centers on two challenge problems, in crisis management and battlespace reasoning, each demanding powerful problem solving with very large knowledge bases. |

In this table, the numbers assigned to sentences denote locations of the sentences in the the suggested derived and source texts. The n-grams shared by each pair of aligned sentences are highlighted with green, and shared single words (including those undergone inflectional and letter-case changes; function words are excluded) are highlighted with red colour.

Research topics

- “adaptive summarization”
 - create a system that adapts itself to a new topic (Learning FRUMP)
- machine translation techniques for summarization
 - going beyond headline generation
- abstraction operations
 - linguistic condensation, generalisation, etc. (more than headlines)

Research topics

- text types
 - Legal texts; Science; Medical texts
 - Imaginative works (narrative, films, etc.)
- profile creation
 - organizations, people, etc.
- multimedia summarization/presentation
 - digital libraries; meetings

Research topics

- Crossing the sentence barrier
 - coreference to support merging
- Identifying “nuggets” instead of sentences & combine them in a cohesive, well-formed summary
- Crossing the language barrier with summaries
 - you obtain summaries in your own language for news available in a language you don't understand

Some links

- <http://www.summarization.com>
- <http://duc.nist.gov>
- <http://www.newsinsence.com>
- <http://www.clsp.jhu.edu/ws2001/groups/asmd>
- <http://www.cs.columbia.edu/~jing/summarization.html>
- <http://www.shef.ac.uk/~saggion>
- <http://www.csi.uottawa.ca/~swan/summarization>

Thanks!

Horacio Saggion

saggion@dcs.shef.ac.uk

International meetings

- 1993 Summarizing Text for Intelligent Communication, Dagstuhl
- 1997 Summarization Workshop, ACL, Madrid
- 1998 AAI Intelligent Text Summarization, Spring Symposium, Stanford
- 1998 SUMMAC evaluation
- 1998 RIFRA Workshop, Sfax
- 2000 Workshop on Automatic Summarization (WAS), Seattle. 2001 (New Orleans). 2002 (Philadelphia). 2003 (Edmonton). 2004 (Barcelona)...
- 2005 Crossing Barriers in Text Summarization, RANLP, Bulgaria
- 2001-2006 Document Understanding Conference
- 2005-2006 Multilingual Summarization Evaluation

Tutorial materials

- COLING/ACL 1998 (Hovy & Marcu)
- IJCAI 1999 (Hahn & Mani)
- SIGIR 2000/2004 (Radev)
- IJCNLP 2005 (Lin)
- ESSLLI 2005 (Saggion)

References

- [1] AAAI. *Intelligent Text Summarization Symposium, AAAI 1998 Spring Symposium Series. March 23-25*, Stanford, USA, March 23-25 1998.
- [2] ACL. *Workshop on Automatic Summarization, ANLP-NAACL2000, April 30*, Seattle, Washington, USA, April 30 2000.
- [3] ACL/EACL. *Workshop on Intelligent Scalable Text Summarization, ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization*, 11 July 1997, Madrid, Spain, 1997.
- [4] Richard Alterman. A Dictionary Based on Concept Coherence. *Artificial Intelligence*, 25:153–186, 1985.
- [5] Richard Alterman. Text Summarization. In S.C. Shapiro, editor, *Encyclopedia of Artificial Intelligence*, volume 2, pages 1579–1587. John Wiley & Sons, Inc., 1992.
- [6] Richard Alterman and Lawrence A. Bookman. Some Computational Experiments in Summarization. *Discourse Processes*, 13:143–174, 1990.
- [7] ANSI. *Writing Abstracts*. American National Standards Institute, 1979.
- [8] M. Banko, V. Mittal, and M. Witbrock. Headline generation based on statistical translation.
- [9] Regina Barzilay and Michael Elhadad. Using Lexical Chains for Text Summarization. In *Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization*, pages 10–17, Madrid, Spain, July 1997.
- [10] Regina Barzilay, Kathleen McKeown, and Michael Elhadad. Information Fusion in the Context of Multi-Document Summarization. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 550–557, Maryland, USA, 20-26 June 1999.
- [11] Barzilay, R. and Lee, L. Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization. In *Proceedings of HLT-NAACL 2004*, 2004.
- [12] P.B. Baxendale. Man-made Index for Technical Literature - an experiment. *IBM J. Res. Dev.*, 2(4):354–361, 1958.
- [13] M. Benbrahim and K. Ahmad. Text Summarisation: the Role of Lexical Cohesion Analysis. *The New Review of Document & Text Management*, pages 321–335, 1995.
- [14] Mohamed Benbrahim and Kurshid Ahmad. Text Summarisation: the Role of Lexical Cohesion Analysis. *The New Review of Document & Text Management*, pages 321–335, 1995.
- [15] William J. Black. Knowledge based abstracting. *Online Review*, 14(5):327–340, 1990.
- [16] H. Borko and C. Bernier. *Abstracting Concepts and Methods*. Academic Press, 1975.
- [17] Jaime G. Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Research and Development in Information Retrieval*, pages 335–336, 1998.
- [18] Paul Clough, Robert Gaizauskas, Scott Piao, and Yorick Wilks. METER: MEasuring TExt Reuse. In *Proceedings of the ACL. Association for Computational Linguistics*, July 2002.
- [19] J.M. Conroy, J.D. Schlesinger, J. Goldstein, and D.P. O'Leary. Left-Brain/Right-Brain Multi-Document Summarization. In *Proceedings of the Document Understanding Conference*, 2004.

- [20] Eduard T. Cremmins. *The Art of Abstracting*. ISI PRESS, 1982.
- [21] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications. In *ACL 2002*, 2002.
- [22] *Summarizing Text for Intelligent Communication*, Dagstuhl, Germany, 1993.
- [23] H. Dalianis, M. Hassel, K. de Smedt, A. Liseth, T.C. Lech, and J. Wedekind. Porting and evaluation of automatic summarization. In *Nordisk Sprogteknologi*, pages 107–121, 2004.
- [24] Gerald DeJong. An Overview of the FRUMP System. In W.G. Lehnert and M.H. Ringle, editors, *Strategies for Natural Language Processing*, pages 149–176. Lawrence Erlbaum Associates, Publishers, 1982.
- [25] R.L. Donaway, K.W. Drummey, and L.A. Mather. A Comparison of Rankings Produced by Summarization Evaluation Measures. In *Proceedings of the Workshop on Automatic Summarization, ANLP-NAACL2000*, pages 69–78. Association for Computational Linguistics, 30 April 2000 2000.
- [26] F.S. Douzidia and G. Lapalme. Lakhas, an Arabic Summarization System. In *Proceedings of the Document Understanding Conference*, 2004.
- [27] Dragomir Radev, Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda elebi, Stanko Dimitrov, Elliott Drabek, Ali Hakim, Wai Lam, Danyu Liu, Jahna Otterbacher, Hong Qi, Horacio Saggion, Simone Teufel, Michael Topper, Adam Winkel, and Zhang Zhu. Mead - a platform for multidocument multilingual text summarization. In *Proceedings of LREC 2004*, May 2004.
- [28] H.P. Edmundson. New Methods in Automatic Extracting. *Journal of the Association for Computing Machinery*, 16(2):264–285, April 1969.
- [29] Brigitte Endres-Niggemeyer. SimSum: an empirically founded simulation of summarizing. *Information Processing & Management*, 36:659–682, 2000.
- [30] Brigitte Endres-Niggemeyer, E. Maier, and A. Sigel. How to Implement a Naturalistic Model of Abstracting: Four Core Working Steps of an Expert Abstractor. *Information Processing & Management*, 31(5):631–674, 1995.
- [31] Brigitte Endres-Niggemeyer, W. Waumans, and H. Yamashita. Modelling Summary Writing by Introspection: A Small-Scale Demonstrative Study. *Text*, 11(4):523–552, 1991.
- [32] G. Erkan and D. Radev. The University of Michigan at DUC 2004. In *Proceedings of the Document Understanding Conference*, 2004.
- [33] D.K. Evans, J.L. Klavans, and K.R. McKeown. Columbia Newsblaster: Multilingual News Summarization on the Web. In *Proceedings of NAACL/HLT*, 2004.
- [34] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, 1998.
- [35] Jade Goldstein, Mark Kantrowitz, Vibhu O. Mittal, and Jaime G. Carbonell. Summarizing text documents: Sentence selection and evaluation metrics. In *Research and Development in Information Retrieval*, pages 121–128, 1999.
- [36] Jade Goldstein, Vibhu O. Mittal, Jaime G. Carbonell, and Mark Kantrowitz. Multi-document summarization by sentence extraction. In *Proceedings of ANLP/NAACL workshop on Automatic Summarization*, Seattle, WA, April 2000.
- [37] Udo Hahn and U. Reimer. Knowledge-Based Text Summarization: Saliency and Generalization Operators for Knowledge Base Abstraction. In I. Mani and M.T. Maybury, editors, *Advances in Automatic Text Summarization*, pages 215–232. The MIT Press, 1999.

- [38] Laura Hasler, Constantin Orăsan, and Ruslan Mitkov. Building better corpora for summarisation. In *Proceedings of Corpus Linguistics 2003*, pages 309 – 319, Lancaster, UK, March 2003.
- [39] E. Hovy and C-Y. Lin. Automated Text Summarization in SUMMARIST. In I. Mani and M.T. Maybury, editors, *Advances in Automatic Text Summarization*, pages 81–94. The MIT Press, 1999.
- [40] *Information Processing & Management Special Issue on Text Summarization*, volume 31. Pergamon, 1995.
- [41] Hongyan Jing. Sentence Reduction for Automatic Text Summarization. In *Proceedings of the 6th Applied Natural Language Processing Conference*, pages 310–315, Seattle, Washington, USA, April 29 - May 4 2000.
- [42] Hongyan Jing and Kathleen McKeown. The Decomposition of Human-Written Summary Sentences. In M. Hearst, Gey. F., and R. Tong, editors, *Proceedings of SIGIR'99. 22nd International Conference on Research and Development in Information Retrieval*, pages 129–136, University of California, Beekely, August 1999.
- [43] Hongyan Jing and Kathleen McKeown. Cut and Paste Based Text Summarization. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 178–185, Seattle, Washington, USA, April 29 - May 4 2000.
- [44] Hongyan Jing, Kathleen McKeown, Regina Barzilay, and Michael Elhadad. Summarization Evaluation Methods: Experiments and Analysis. In *Intelligent Text Summarization. Papers from the 1998 AAAI Spring Symposium. Technical Report SS-98-06*, pages 60–68, Standford (CA), USA, March 23-25 1998. The AAAI Press.
- [45] Paul A. Jones and Chris D. Paice. A 'select and generate' approach to to automatic abstracting. In A.M. McEnry and C.D. Paice, editors, *Proceedings of the 14th British Computer Society Information Retrieval Colloquium*, pages 151–154. Springer Verlag, 1992.
- [46] D.E. Kieras. A model of reader strategy for abstracting main ideas from simple technical prose. *Text*, 2(1-3):47–81, 1982.
- [47] Walter Kintsch and Teun A. van Dijk. Comment on se rappelle et on résume des histoires. *Langages*, 40:98–116, Décembre 1975.
- [48] Walter A. Kintsch and Teun A. van Dijk. Towards a model of text comprehension and production. *Psychological Review*, 85:235–246, 1978.
- [49] Kevin Knight and Daniel Marcu. Statistics-based summarization - step one: Sentence compression. In *Proceedings of the 17th National Conference of the American Association for Artificial Intelligence. AAAI*, July 30 - August 3 2000.
- [50] Julian Kupiec, Jan Pedersen, and Francine Chen. A Trainable Document Summarizer. In *Proc. of the 18th ACM-SIGIR Conference*, pages 68–73, 1995.
- [51] F. Lacatusu, L. Hick, S. Harabagiu, and L. Nezd. Lite-GISTexter at DUC2004. In *Proceedings of DUC 2004*. NIST, 2004.
- [52] F. Lacatusu, A. Hickl, S. Harabagiu, and L. Nezda. Lite-GISTexter at DUC2004. In *Proceedings of the Document Understanding Conference*, 2004.
- [53] M. Lapata. Probabilistic text structuring: Experiments with sentence ordering, 2003.
- [54] W. Lehnert and C. Loiselle. An Introduction to Plot Units. In D. Waltz, editor, *Advances in Natural Language Processing*, pages 88–111. Lawrence Erlbaum, Hillsdale, N.J., 1989.

- [55] Wendy Lehnert. Plot Units and Narrative Summarization. *Cognitive Science*, 5:293–331, 1981.
- [56] Wendy G. Lehnert. Narrative Complexity Based on Summarization Algorithms. In B.G. Bara and G. Guida, editors, *Computational Models of Natural Language Processing*, pages 247–259. Elsevier Science Publisher B.V., North-Holland, 1984.
- [57] Alessandro Lenci, Roberto Bartolini, Nicoletta Calzolari, Ana Agua, Stephan Busemann, Emmanuel Cartier, Karine Chevreau, and Jos Coch. Multilingual Summarization by Integrating Linguistic Resources in the MLIS-MUSI Project. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC'02)*, May 29–31, Las Palmas, Canary Islands, Spain, 2002.
- [58] Elizabeth D. Liddy. The Discourse-Level Structure of Empirical Abstracts: An Exploratory Study. *Information Processing & Management*, 27(1):55–81, 1991.
- [59] C. Lin and E. Hovy. Identifying Topics by Position. In *Fifth Conference on Applied Natural Language Processing*, pages 283–290. Association for Computational Linguistics, 31 March - 3 April 1997.
- [60] C-Y. Lin. Knowledge-Based Automatic Topic Identification. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics. 26-30 June 1995, MIT, Cambridge, Massachusetts, USA*, pages 308–310. ACL, 1995.
- [61] Lin.C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries. In *Proceedings of the Workshop on Text Summarization*, Barcelona, 2004. ACL.
- [62] Hans P. Luhn. The Automatic Creation of Literature Abstracts. *IBM Journal of Research Development*, 2(2):159–165, 1958.
- [63] Robert E. Maizell, Julian F. Smith, and Tibor E.R. Singer. *Abstracting Scientific and Technical Literature*. Wiley-Interscience, A Division of John Wiley & Son, Inv., 1971.
- [64] I. Mani and M.T. Maybury, editors. *Advances in Automatic Text Summarization*. The MIT Press, 1999.
- [65] Inderjeet Mani. *Automatic Text Summarization*. John Benjamins Publishing Company, 2001.
- [66] Inderjeet Mani and Eric Bloedorn. Summarizing similarities and differences among related documents. *Information Retrieval*, 1(1):35–67, 1999.
- [67] Inderjeet Mani, Barbara Gates, and Eric Bloedorn. Improving Summaries by Revising Them. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 558–565, Maryland, USA, 20–26 June 1999.
- [68] Inderjeet Mani, David House, Gary Klein, Lynette Hirshman, Leo Obrst, Thérèse Firmin, Michael Chrzanowski, and Beth Sundheim. The TIPSTER SUMMAC Text Summarization Evaluation. Technical report, The Mitre Corporation, 1998.
- [69] W.C. Mann and S.A. Thompson. Rhetorical Structure Theory: towards a functional theory of text organization. *Text*, 8(3):243–281, 1988.
- [70] D. Marcu. *Encyclopedia of Library & Information Science*, chapter Automatic Abstracting, pages 245–256. Miriam Drake, 2003.
- [71] Daniel Marcu. From Discourse Structures to Text Summaries. In *The Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization*, pages 82–88, Madrid, Spain, July 11 1997.
- [72] Daniel Marcu. *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. PhD thesis, Department of Computer Science, University of Toronto, 1997.

- [73] Daniel Marcu. The automatic construction of large-scale corpora for summarization research. In M. Hearst, Gey. F., and R. Tong, editors, *Proceedings of SIGIR'99. 22nd International Conference on Research and Development in Information Retrieval*, pages 137–144, University of California, Beekely, August 1999.
- [74] K. McKeown, R. Barzilay, J. Chen, D. Eldon, D. Evans, J. Klavans, A. Nenkova, B. Schiffman, and S. Sigelman. Columbia's Newsblaster: New Features and Future Directions. In *NAACL-HLT'03 Demo*, 2003.
- [75] Kathleen McKeown, D. Jordan, and Hatzivassiloglou V. Generating patient-specific summaries of on-line literature. In *Intelligent Text Summarization. Papers from the 1998 AAAI Spring Symposium. Technical Report SS-98-06*, pages 34–43, Stanford (CA), USA, March 23-25 1998. The AAAI Press.
- [76] Kathleen McKeown, Judith Klavans, Vasileios Hatzivassiloglou, Regina Barzilay, and Eleazar Eskin. Towards multidocument summarization by reformulation: Progress and prospects. In *AAAI/IAAI*, pages 453–460, 1999.
- [77] Kathleen R. McKeown and Dragomir R. Radev. Generating summaries of multiple news articles. In *Proceedings, 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 74–82, Seattle, Washington, July 1995.
- [78] Kathleen R. McKeown, J. Robin, and K. Kukich. Generating concise natural language summaries. *Information Processing & Management*, 31(5):702–733, 1995.
- [79] K.R. McKeown, J. Robin, and K. Kukich. Generating concise natural language summaries. *Information Processing & Management*, 31(5):702–733, 1995.
- [80] S. Miike, E. Itoh, K. Ono, and K. Sumita. A Full-text Retrieval System with A Dynamic Abstract Generation Function. In W.B. Croft and C.J. van Rijsbergen, editors, *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 152–161, July 3-6, Dublin, Ireland, July 3-6 1994.
- [81] J.-L. Minel. *Filtrage sémantique. Du résumé automatique à la fouille de textes*. Editions Hermès, 2003.
- [82] J-L. Minel, J-P. Desclés, E. Cartier, G. Crispino, S.B. Hazez, and A. Jackiewicz. Résumé automatique par filtrage sémantique d'informations dans des textes. *TSI*, X(X/2000):1–23, 2000.
- [83] J-L. Minel, S. Nugier, and G. Piat. Comment Apprécier la Qualité des Résumés Automatiques de Textes? Les Exemples des Protocoles FAN et MLUCE et leurs Résultats sur SERAPHIN. In *1ères Journées Scientifiques et Techniques du Réseau Francophone de l'Ingénierie de la Langue de l'AUFELF-UREF.*, pages 227–232, 15-16 avril 1997.
- [84] Ani Nenkova and Rebecca Passonneau. Evaluating Content Selection in Summarization: The Pyramid Method. In *Proceedings of NAACL-HLT 2004*, 2004.
- [85] NIST. *Proceedings of the Document Understanding Conference*, September 13 2001.
- [86] Michael P. Oakes and Chris D. Paice. The Automatic Generation of Templates for Automatic Abstracting. In *21st BCS IRSG Colloquium on IR*, Glasgow, 1999.
- [87] Michael P. Oakes and Chris D. Paice. Term extraction for automatic abstracting. In D. Bourigault, C. Jacquemin, and M-C. L'Homme, editors, *Recent Advances in Computational Terminology*, volume 2 of *Natural Language Processing*, chapter 17, pages 353–370. John Benjamins Publishing Company, 2001.
- [88] Kenji Ono, Kazuo Sumita, and Seiji Miike. Abstract Generation Based on Rhetorical Structure Extraction. In *Proceedings of the International Conference on Computational Linguistics*, pages 344–348, 1994.

- [89] Chris D. Paice. The Automatic Generation of Literary Abstracts: An Approach based on Identification of Self-indicating Phrases. In O.R. Norman, S.E. Robertson, C.J. van Rijsbergen, and P.W. Williams, editors, *Information Retrieval Research*, London: Butterworth, 1981.
- [90] Chris D. Paice. Constructing Literature Abstracts by Computer: Technics and Prospects. *Information Processing & Management*, 26(1):171–186, 1990.
- [91] Chris D. Paice, William J. Black, Frances C. Johnson, and A.P. Neal. Automatic Abstracting. Technical Report R&D Report 6166, British Library, 1994.
- [92] Chris D. Paice and Paul A. Jones. The Identification of Important Concepts in Highly Structured Technical Papers. In R. Korfhage, E. Rasmussen, and P. Willett, editors, *Proc. of the 16th ACM-SIGIR Conference*, pages 69–78, 1993.
- [93] Chris D. Paice and Michael P. Oakes. A Concept-Based Method for Automatic Abstracting. Technical Report Research Report 27, Library and Information Commission, 1999.
- [94] K. Pastra and H. Saggion. Colouring summaries Bleu. In *Proceedings of Evaluation Initiatives in Natural Language Processing*, Budapest, Hungary, 14 April 2003. EACL.
- [95] M. Pinto Molina. Documentary Abstracting: Towards a Methodological Model. *Journal of the American Society for Information Science*, 46(3):225–234, April 1995.
- [96] J.J. Pollock and A. Zamora. Automatic abstracting research at Chemical Abstracts Service. *Journal of Chemical Information and Computer Sciences*, (15):226–233, 1975.
- [97] Dragomir R. Radev, Hongyan Jing, and Malgorzata Budzikowska. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *ANLP/NAACL Workshop on Summarization*, Seattle, WA, April 2000.
- [98] Dragomir R. Radev and Kathleen R. McKeown. Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, 24(3):469–500, September 1998.
- [99] Lisa F. Rau, Paul S. Jacobs, and Uri Zernik. Information Extraction and Text Summarization using Linguistic Knowledge Acquisition. *Information Processing & Management*, 25(4):419–428, 1989.
- [100] J.B. Reiser, J.B. Black, and W. Lehnert. Thematic knowledge structures in the understanding and generation of narratives. *Discourse Processes*, (8):357–389, 1985.
- [101] *RIFRA'98. Rencontre Internationale sur l'extraction le Filtrage et le Résumé Automatique. Novembre 11-14*, Sfax, Tunisie, Novembre 11-14 1998.
- [102] Jennifer Rowley. *Abstracting and Indexing*. Clive Bingley, London, 1982.
- [103] J.E. Rush, R. Salvador, and A. Zamora. Automatic Abstracting and Indexing. Production of Indicative Abstracts by Application of Contextual Inference and Syntactic Coherence Criteria. *Journal of the American Society for Information Science*, pages 260–274, July-August 1971.
- [104] H. Saggion. Shallow-based Robust Summarization. In *Automatic Summarization: Solutions and Perspectives*, ATALA, December, 14 2002.
- [105] H. Saggion and G. Lapalme. Concept Identification and Presentation in the Context of Technical Text Summarization. In *Proceedings of the Workshop on Automatic Summarization. ANLP-NAACL2000*, Seattle, WA, USA, 30 April 2000. Association for Computational Linguistics.
- [106] H. Saggion and G. Lapalme. Generating Indicative-Informative Summaries with SumUM. *Computational Linguistics*, 2002.

- [107] H. Saggion, D. Radev, S. Teufel, L. Wai, and S. Strassel. Developing Infrastructure for the Evaluation of Single and Multi-document Summarization Systems in a Cross-lingual Environment. In *LREC 2002*, pages 747–754, Las Palmas, Gran Canaria, Spain, 2002.
- [108] Horacio Saggion. Using Linguistic Knowledge in Automatic Abstracting. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 596–601, Maryland, USA, June 1999.
- [109] Horacio Saggion. *Génération automatique de résumés par analyse sélective*. PhD thesis, Département d’informatique et de recherche opérationnelle. Faculté des arts et des sciences. Université de Montréal, Août 2000.
- [110] Horacio Saggion and Robert Gaizauskas. Multi-document summarization by cluster/profile relevance and redundancy removal. In *Proceedings of the Document Understanding Conference 2004*. NIST, 2004.
- [111] Horacio Saggion and Guy Lapalme. The Generation of Abstracts by Selective Analysis. In *Intelligent Text Summarization. Papers from the 1998 AAAI Spring Symposium. Technical Report SS-98-06*, pages 137–139, Standford (CA), USA, March 23-25 1998. The AAAI Press.
- [112] Horacio Saggion and Guy Lapalme. Where does Information come from? Corpus Analysis for Automatic Abstracting. In *Rencontre Internationale sur l’Extraction le Filtrage et le Résumé Automatique. RIFRA’98*, pages 72–83, Sfax, Tunisie, Novembre 11-14 1998.
- [113] Horacio Saggion and Guy Lapalme. Summary Generation and Evaluation in SumUM. In *Advances in Artificial Intelligence. International Joint Conference: 7th Ibero-American Conference on Artificial Intelligence and 15th Brazilian Symposium on Artificial Intelligence. IBERAMIA-SBIA 2000.*, volume 1952 of *Lecture Notes in Artificial Intelligence*, pages 329–38, Berlin, Germany, 2000. Springer-Verlag.
- [114] Saggion, H. and Bontcheva, K. and Cunningham, H. Generic and Query-based Summarization. In *European Conference of the Association for Computational Linguistics (EACL) Research Notes and Demos*, Budapest, Hungary, 12-17 April 2003. EACL.
- [115] Gerald Salton, J. Allan, and Amit Singhal. Automatic text decomposition and structuring. *Information Processing & Management*, 32(2):127–138, 1996.
- [116] Gerald Salton, Amit Singhal, Mandar Mitra, and Chris Buckley. Automatic Text Structuring and Summarization. *Information Processing & Management*, 33(2):193–207, 1997.
- [117] B. Schiffman, I. Mani, and K.J. Concepcion. Producing Biographical Summaries: Combining Linguistic Knowledge with Corpus Statistics. In *Proceedings of EACL-ACL*, 2001.
- [118] Karen Sparck Jones. Discourse Modelling for Automatic Summarising. Technical Report 290, University of Cambridge, Computer Laboratory, February 1993.
- [119] Karen Sparck Jones. What Might Be in a Summary? In K. Knorz and Womser-Hacker, editors, *Information Retrieval 93: Von der Modellierung zur Anwendung*, 1993.
- [120] Karen Sparck Jones. Automatic Summarizing: Factors and Directions. In I. Mani and M. Maybury, editors, *Advances in Automatic Text Summarization*. MIT Press, Cambridge MA, 1999.
- [121] Karen Sparck Jones and Brigitte Endres-Niggemeyer. Automatic Summarizing. *Information Processing & Management*, 31(5):625–630, 1995.
- [122] Karen Sparck Jones and Julia R. Galliers. *Evaluating Natural Language Processing Systems: An Analysis and Review*. Number 1083 in *Lecture Notes in Artificial Intelligence*. Springer, 1995.

- [123] T. Strzalkowski, J. Wang, and Wise B. A Robust Practical Text Summarization. In *Intelligent Text Summarization Symposium (Working Notes)*, pages 26–33, Stanford (CA), USA, March 23-25 1998.
- [124] John I. Tait. *Automatic Summarising of English Texts*. PhD thesis, University of Cambridge, Computer Laboratory, December 1982.
- [125] S. Teufel and M. Moens. Sentence Extraction and Rhetorical Classification for Flexible Abstracts. In *Intelligent Text Summarization. Papers from the 1998 AAAI Spring Symposium. Technical Report SS-98-06*, pages 16–25, Stanford (CA), USA, March 23-25 1998. The AAAI Press.
- [126] Simone Teufel. Meta-Discourse Markers and Problem-Structuring in Scientific Texts. In M. Stede, L. Wanner, and E. Hovy, editors, *Proceedings of the Workshop on Discourse Relations and Discourse Markers, COLING-ACL'98*, pages 43–49, 15th August 1998.
- [127] Simone Teufel and Marc Moens. Argumentative classification of extracted sentences as a first step towards flexible abstracting. In I. Mani and M.T. Maybury, editors, *Advances in Automatic Text Summarization*, pages 155–171. The MIT Press, 1999.
- [128] Translingual Information Detection, Extraction and Summarization (TIDES) Program. <http://www.darpa.mil/ito/research/tides/index.html>, August 2000.
- [129] Anastasios Tombros, Mark Sanderson, and Phil Gray. Advantages of Query Biased Summaries in Information Retrieval. In *Intelligent Text Summarization. Papers from the 1998 AAAI Spring Symposium. Technical Report SS-98-06*, pages 34–43, Stanford (CA), USA, March 23-25 1998. The AAAI Press.
- [130] Teun A. van Dijk. Recalling and Summarizing Complex Discourse. In M.A. Just and Carpenters, editors, *Cognitive Processes in Comprehension*, 1977.
- [131] Teun A. van Dijk and Walter Kintsch. *Strategies of Discourse Comprehension*. Academic Press, Inc., 1983.
- [132] D. Zajic, B. Dorr, and R. Schwartz. BBN/UMD at DUC-2004: Topiary. In *Proceedings of the Document Understanding Conference*, 2004.