Automatic Summarization

ANLP
Horacio Rodríguez

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Outline

- **Introduction**
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Introduction 1

- A summary is a reductive transformation of a source text into a summary text by extraction or generation
  - Sparck-Jones, 2001

Introduction 2

- A possible objetive

Illustration of the power of human abstracts

Mrs. Coolidge: *What did the preacher discuss in his sermon?*
President Coolidge: *Sin.*
Mrs. Coolidge: *What did he say?*
President Coolidge: *He said he was against it.*

- Bartlett's Quotations (via Graeme Hirst)

Introduction 3

- Locate the fragments of a text relevant (in general or given the user needs) and produce a summary of them
- Sum vs IE
  - IE
    - Structure to be extracted is predefined
      - "I know what I want, look for it"
  - Sum
    - A previous definition of what we want is not (always) provided
      - "What is interesting here"

What to read 1

- Tutorials
  - E.Hovy, D. Marcu (Coling-ACL, 1998)
  - Maybury, Mani (EACL-ACL, 2001)
  - Mani (RANLP, 2003)
  - Lin (ACL, 2004)
  - Saggion (LREC, 2006)
  - [Lloret, Palomar, 2012]
  - [Saggion, Poibeau, 2013]
  - Saggion (IJCNLP, 2013)
  - Saggion (RANLP, 2015)

What to read 2

- Surveys, Books, Chapters, Thesis
  - Hovy (1999), Hovy (2000), K. Sparck Jones (1998), Alonso et al, 2003, Afantenos et al, 2004
  - *R. Mitkov (ed.)*, **The Oxford handbook of Computational Linguistics**. Oxford University Press, 2004, Chapter 32 (E. Hovy)
  - I. Mani, M. Maybury (1999) (eds), Mani (2001)
  - Barzilay (master, 1997), Barzilay (2003), C-Y. Lin (1997),
  - D. R. Radev. (1999), Zechner (2001), Berger (2001),
  - Tucker (1999), Kan (2003), Alonso (2005),
  - Daumé III (2006), Nenkova (2006), Sundaram (2002)
  - M.Fuentes (2008), E.Lloret (2011)
  - J.Kumar (2012)

Material 1

- Groups.
  - U. Ottawa
  - U. Columbia
  - U. Michigan
  - U. Maryland
  - UPF
  - USC/ISI
  - CMU

Characteristics of summary 1

- Type
  - Indicative vs informative
  - Extract vs Abstract (vs gist)
  - Generic vs query based
  - Background vs just-the-news
  - Single document vs multidocument
  - general vs domain dependent
  - textual vs multimedia
- Input
  - domain, genre, form, size

Characteristics of summary 2

- Related disciplines
  - IE, IR, Q&A, Topic identification (TI), Document Classification (DC), Document clustering, Event (topic) detection and tracking (TDT), Text simplification, Paraphrasing
- Evaluation
- Aplications
  - biographies
  - Medical Reports summaries
  - E-mail summaries
    - Voice mails
  - Web pages summaries
  - News summaries
  - Headlines extraction
  - Support to IR
  - Meeting summaries

Characteristics of summary 3

- Current tasks involving research
  - Headline extraction
  - Cross-lingual MDS
  - Query-based
  - Definition Questions
  - Text simplification
  - Paraphrasing
  - (query based) information needs
    - Summarization, definitional Q&A, ...

Dimensions for AS description 1

- Sparck Jones (1998)
  - Input
  - Purpose
  - Output

Outline

- Introduction
- **Problems of AS**
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Input 1

- Input
  - Document structure
    - metadata
    - document organization
    - Multimedia integration
  - Domain
    - Domain dependent vs Open domain systems
    - Adaptable systems
  - Detail level
    - ordinary, specialized, restricted
  - Restricted or free language (sublanguages)
  - Media
    - text, video, meeting records, images, tables

Input 2

- Input
  - Scale
    - Document length, granularity
    - Meaning units, segmentation
    - Type
      - conventional (paragraph, sentence, clause, ...)
      - specific (Semantic Content Unit, SCU, nugget, factoid)
  - Genres
    - Medical reports, Medical articles, News, Radio fragments, meeting records, e-mails, Web pages, Voice mails (Koumpis, Renals, 2004)
  - Unit (SD vs MD)
  - Language ({mono,multi,cross}-lingual)

Purpose

Situation
- Complete Aplications
- Components of wider systems
  - MT, IR, QA
- Task-driven vs General

- Audience
  - Background vs just-the-news
- Use
  - Retrieving
  - Previewing

Output

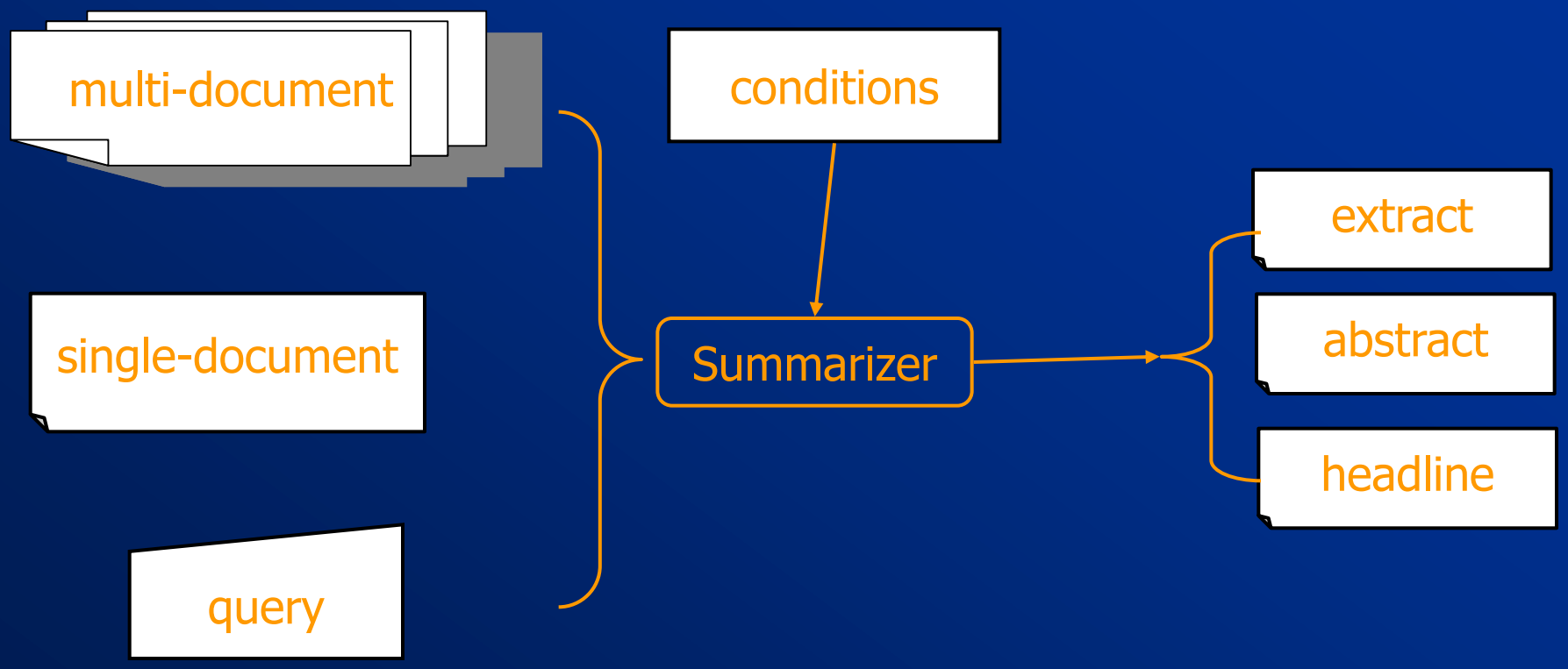
- Content
  - generic vs user-focused vs query-driven
- Format
  - Plain text, list of ítems, structured
- Style
  - informative, indicative, aggregative, critic
- Production
  - extractive vs abstractive
  - use of NLG (U. Columbia)
  - Relation with the source
  - Substitution, linked, included (highlighted text)
- Size
  - Compression degree
  - Headline, ultra-summarization, subtitle generation

Approaches to AS

- Top-Down
  - query-driven
  - User needs
    - Some information
  - system
    - Specific needs for guiding the search
    - examples :
      - templates with slots with semantic features
      - list of relevant terms
  - Similar to IE

- Bottom-up
  - text-driven
  - User needs
    - Everything important
  - system
    - metrics over the generic importance (strategy)
    - Examples:
      - Connectivity degree in a semantic graph
      - frequency of (co-) occurrence of terms
  - Similar to IR

Basic schema



Tasks

- Extracting the most relevant fragments
  - sentences
  - paragraphs
  - pasages
- ranking of these fragments by relevance trying to avoid redundancy
- Summary production
  - Saggion
    - text to text
    - propositions to text
    - information to text

Analizing

Transforming

Sinthesizing

Outline

- Introduction
- Problems of AS
- **Basic Architectures**
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Techniques used 1

- Processing level
  - Superficial
  - Entity
  - Discourse level
- Techniques used
  - Linguistic
  - Statistical
  - ML
  - IE
  - Text mapping and MT
  - Graph based
  - Combination

Techniques used 2

Superficial level

- Frequency of terms
- Position
  - Lead method
  - Title-based method
  - OPP (Optimum Position Policy)
    - Lin, Hovy, 1997
- Bias
  - Presence of terms from the title and headings
    - Conroy, Schlesinger, 2001, 2002
- cue words & phrases

Techniques used 3

Entity Level

- Build a representation of the document from their entities and relations
  - Words, MWT, word n-grams, NE
- Lexical Similarity
  - Distribucional or semantic
- Proximity
- Logical relations
  - Agreement, contradiction, entailment
- Semantic relations
  - Predicate-argument

Techniques used 4

Entity Level

- Cohesion
  - Word Coocurrence
  - Coreference
  - Local saliency
  - Lexical chains
    - Barzilay, 1997
    - Fuentes, Rodriguez, 2002
  - MMR
    - Carbonell, Goldstein, 1998
  - Graph based methods

Techniques used 5

Discoursive level

- Format and organization of the document
- Thread of topics
  - Topic Signatures
    - Hovy, Lin, 1999
    - Harabagiu, Lacatusu, 2005
- Topic Detection
  - Hovy, Lin, 1999,
  - Hovy, 2000
  - Radev et al, 2000,2001
- Rethorical Structure
  - Marcu, 1997
  - Alonso, 2005

Techniques used 6

Linguistic techniques

- Sentence compression, reduction or simplification
  - Saggion 2013, 2015
  - Cohn, Lapata, 2009

Techniques used 7

Statistical techniques

- Language models
  - Berger, 2001,
  - Berger, Mittal, 2000
- Bayesian models
  - Kupiec et al, 1995
  - Schlesinger et al, 2001
- HMM
- ME, Logistic regression
  - Conroy et al, 2001
- K-mixtures, combinations
  - Lloret et al 2009
  - Chandra et al, 2011
  - Bhatia et al, 2012

Techniques used 8

ML

- Decision trees
  - Copeck et al, 2002
- ILP
  - Knight, Marcu, 2000, Tzoukerman et al, 2001
- SVM
  - Hirao et al, 2002
- Perceptron
  - Fisher, Roark, 2005, 2006
- Evolutionary algorithms
  - Binwahlan et al, 2009, Suanmali et al, 2011, Nandhini et al, 2013

Techniques used 9

IE

- Kan, McKeown, 1999, Harabagiu, Lacatusu, 2002, Daumé, 2002, Lal, Rueger, 2002

Text mapping and MT

- Banko et al, 1999, Daumé III, Marcu, 2006

Techniques used 10

Graphs, Matrices, Tensors

- Graphs of lexical connexion
  - (adjacency, grammatical relations, similarity, coreference)
  - Mani, Bloedorn, 1999
  - Leskovec et al, 2005
  - Lin, H., et al 2009
  - Kumaresh, N., & Ramakrishnan, 2012
- Random Walk algorithms
  - LexRank - Erkan, Radev, 2004
  - TextRank - Mihalcea, Tarau, 2004
- Bigraphs
  - Zha, 2002

Techniques used 11

Combination

- Voting democratic or weighted
- ML
  - Classifiers
  - SVM, DT, NB,
  - Rankers
  - Stacked
  - Ensemble methods

Outline

- Introduction
- Problems of AS
- Basic Architectures
- **Evaluations**
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

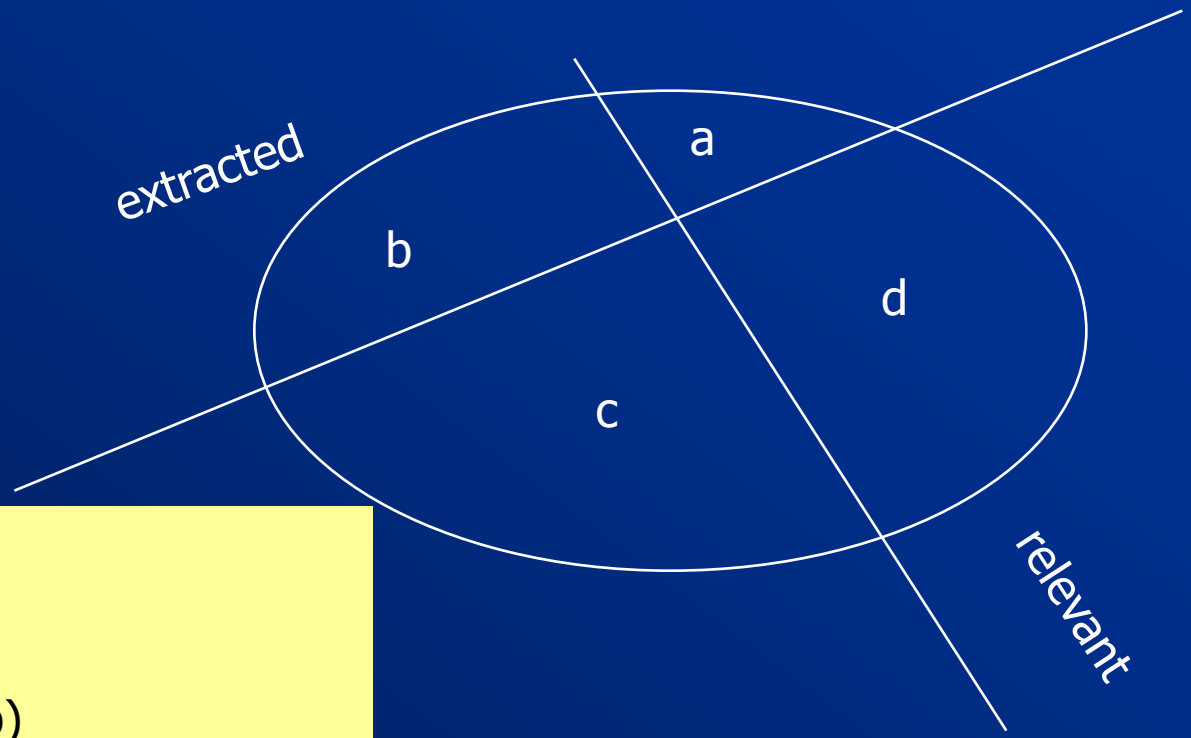
Evaluation 1

- Process complex, controversial and costly
- **Basic Problem**
  - Very low agreement between human annotators
- Types of evaluation
  - intrinsic
    - Summary is evaluated per se or by comparing it to the source
      - coherence, cohesion
      - **informativity**
      - omisions
  - extrinsic
    - The system is evaluated against an (external) specific task
      - Can the summary be used instead of the document?
        - » DC, Q&A

Evaluation 2

- Standard Metrics for evaluating extracts
  - Precision
  - Recall
  - Accuracy
  - F-score
- Specific Metrics
  - Based on ranking
    - RU (Relative Utility) (Radev)
  - Based on content
    - Overlap (n-gram, cosine, longest subsequence)
    - ROUGE (Lin, Hovy)
    - Basic Elements (Lin, Hovy)
    - Pyramid (Nenkova, Passonneau)
  - Based on combination
    - QARLA (Amigó)

Evaluation 3



extracted = a + b
relevant = a + d
recall = a / (a + d)
precisión = a / (a + b)
accuracy = (a + c) / (a + b + c + d)

$$F = \frac{(\beta^2 + 1) \cdot p \cdot r}{\beta^2 \cdot p + r}$$

$$F_1 = \frac{2 \cdot p \cdot r}{p + r}$$

Relative utility

- Radev, Tam, 2003
- N judges assign utility scores to each sentence of each cluster, Best scored e sentences are an extract of size e.

ROUGE

- Recall Oriented Understudy for Gisting Evaluation
- C.Y. Lin, E. Hovy, 2003
- Comparison with a gold standard (one or more judges)
- Overlapping units
  - ROUGE-n: n-gram co-occurrence statistics
  - ROUGE-L: Longest common substring.
  - ROUGE-W: Weighted longest common substring.
  - ROUGE-Sn: Skip-bigram co-occurrence statistics without gap length limit and with maximum gap lengths of n words.
  - ROUGE-SUn: Extension of skip-bigram.

Pyramids 1

- Nenkova & Passonneau (2004)
- Evaluates content and quality
- Pool of human summary as gold standard
- Summarization Content Units (SCU)
  - An SCU is a set of contributors expressing a unique content identied by a common label
- Starts from a set of manually built summaries
- A pyramid model that exploits the intersection and frequency of SCU is built.
- Informativity of a new summary against the pyramid model
- Matching of SCUs with the summary
  - Add weights of SCUs (Obs)
  - Compute Obs/Max
    - 2 scores original (precision) and modified (recall), depending on the value of Max

Pyramids 2

[*Label*:*The Concorde crossed the Atlantic in less than 4 hours*]

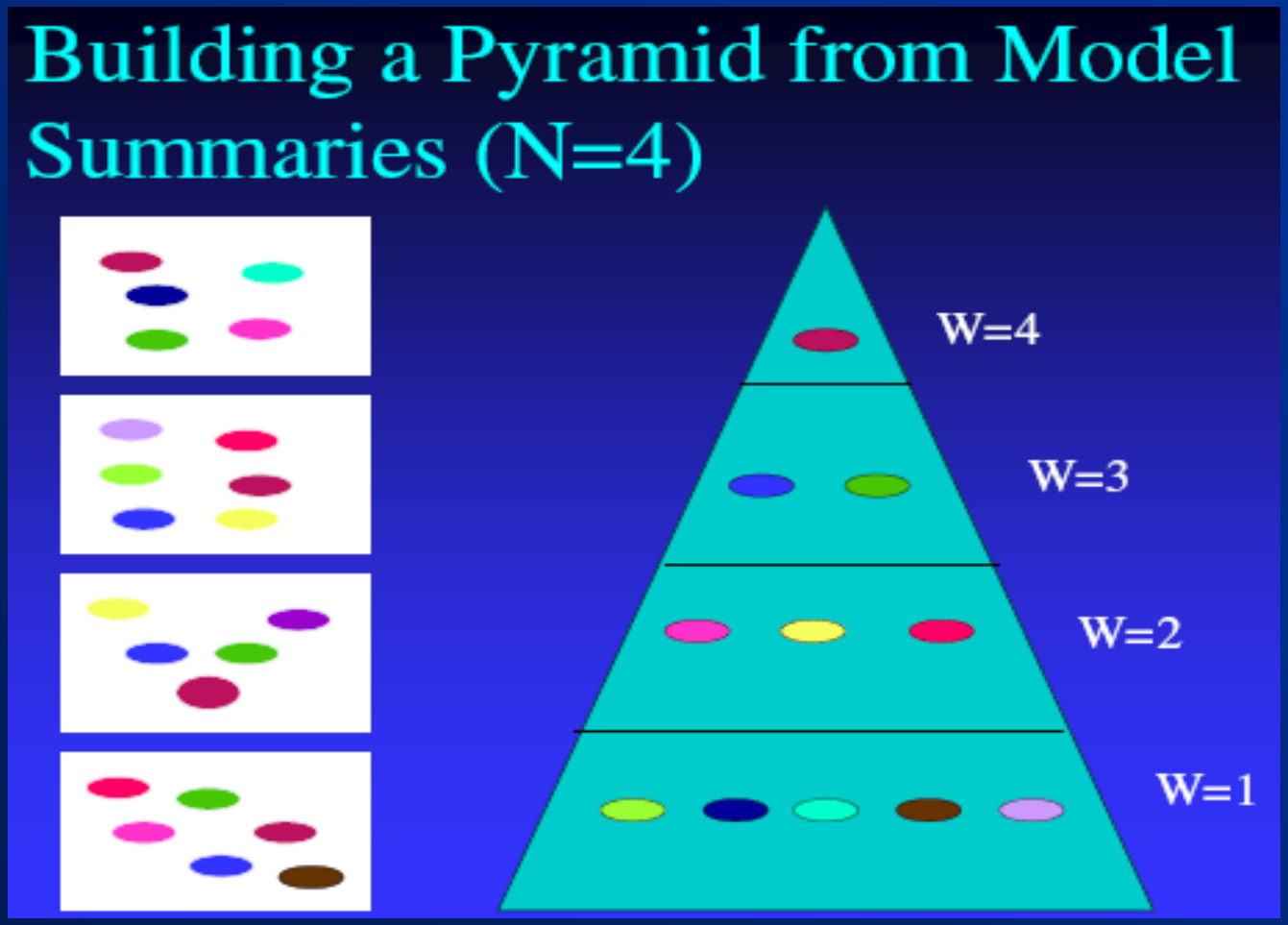
Sum1 < making the transatlantic flight in 3 and ½ hrs >

Sum2 < The Concorde could make the flight in between New York and London or Paris in less than four hours>

Sum3 < completing its journey from London to New York in about 3 hours, 30 minutes >

Sum4 < took less than 4 hrs to cross the Atlantic >

Pyramids 3



DUC 1

- Document Understanding Conference
- Organized by NIST from 2001
- Tasks 2001 and 2002
  - SDS
    - fully automatic summarization of a single newswire/newspaper document in 100 words
  - MDS
    - given multiple newswire/newspaper documents on a single subject, to produce summaries of diferent lengths (400, 200, 100, 50 words in DUC 2001 and 200, 100, 50, 10 words in DUC 2002)

DUC 2

- Tasks 2003
  - Headline extraction
    - very short, 10-word, summaries from a single document.
  - Short (100-word) summaries focused by events
  - Given a cluster of documents produce short summaries focused by viewpoints
    - The viewpoint description was supposed to be a natural language string no larger than a sentence. It will describe the important facet(s) of the cluster the NIST assessor has decided to include in the short summary
  - Given a cluster of documents produce short summaries that answer a question
    - Given each document cluster, a question, and the set of sentences in each document deemed relevant to the question, the task consisted in the creation of short summaries of the cluster that answers the question.

DUC 3

- Tasks 2004
  - Task 1: very short (≤ 75 bytes) summary
  - Task 2: short (≤ 665 bytes) 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

DUC 4

- Tasks 2005 and2006
  - to synthesize from a set of 25-50 documents a brief, well-organized, fluent answer to a need for information that cannot be met by just stating a name, date, quantity, etc.
  - This task models real-world complex question answering.
  - Size: 250 words

DUC 5

<topic>
<num> d301i </num>
<title> International Organized Crime </title>
<narr>
Identify and describe types of organized crime that crosses borders or involves more than one country. Name the countries involved. Also identify the perpetrators involved with each type of crime, including both individuals and organizations if possible.
</narr>
<granularity> specific </granularity>
</topic>

<topic>
<num> d426a </num>
<title> Law enforcement with dogs </title>
<narr>
What sorts of law-enforcement tasks are dogs being used for worldwide? What law enforcement agencies are using dogs? What breeds of dogs are being used?
</narr>
<granularity> general </granularity>
</topic>

Table 4.11: DUC 2005 complex natural language questions.

M. Fuentes (2008)

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- **Single-document Systems**
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

ML

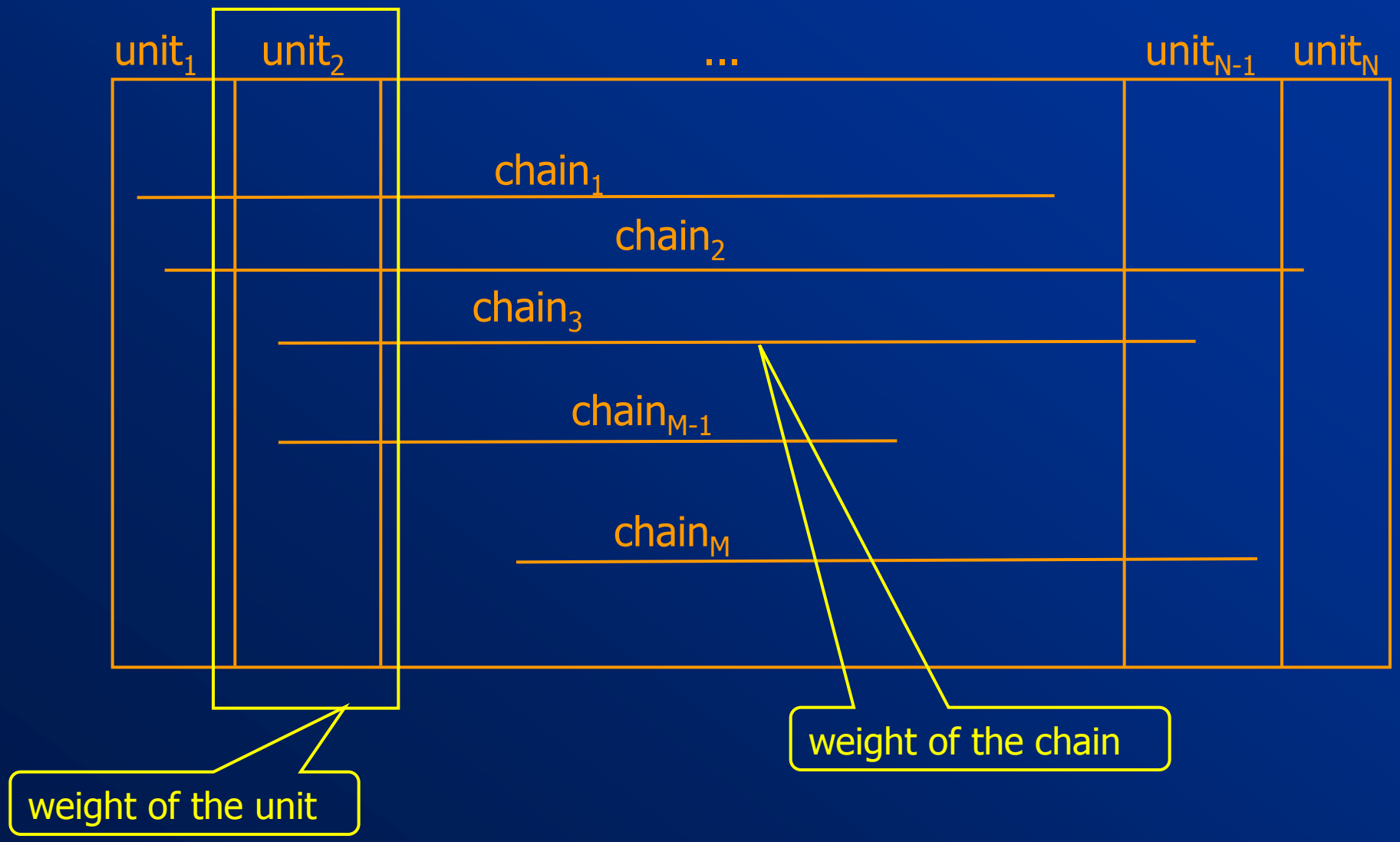
- Binwahlan et al. 2009
  - used particle swarm optimization to investigate the impact of the feature structure on the features selection process.
  - In each iteration, the system selects some features. Subsequently, the corresponding weights of those features are used to score the sentences, and then the top-ranking sentences are selected as a summary.
- Suanmali et al 2011
  - relied on Genetic Algorithm (GA) to feature selection
- Nandhini and Balasundaram 2013
  - GA to increase readability through sentence cohesion by extracting the optimal combination of sentences. In addition to the informative features, they considered several learner readability-related features such as percentage of trigger words, average sentence length, percentage of noun entity occurrences and percentage of polysyllabic words

Lexical Chains 1

- Regina Barzilay (1997)
- Factors for defining the chain
  - Word coocurrence, coreference, relevance of words, semantic relations, ...
- Fuentes, Rodríguez, 2002 use a relative classification
  - μ_S is the mean of scores of all the chains
  - σ_S standard deviation

$$
\begin{aligned}
\tau &= \mu_s + 2 \cdot \sigma_s \\
Strong &= \{c \mid score_c \geq \tau\} \\
Medium &= \{c \mid \tau > score_c \geq \tau/2\} \\
Light &= \{c \mid \tau/2 > score_c\}
\end{aligned}
$$

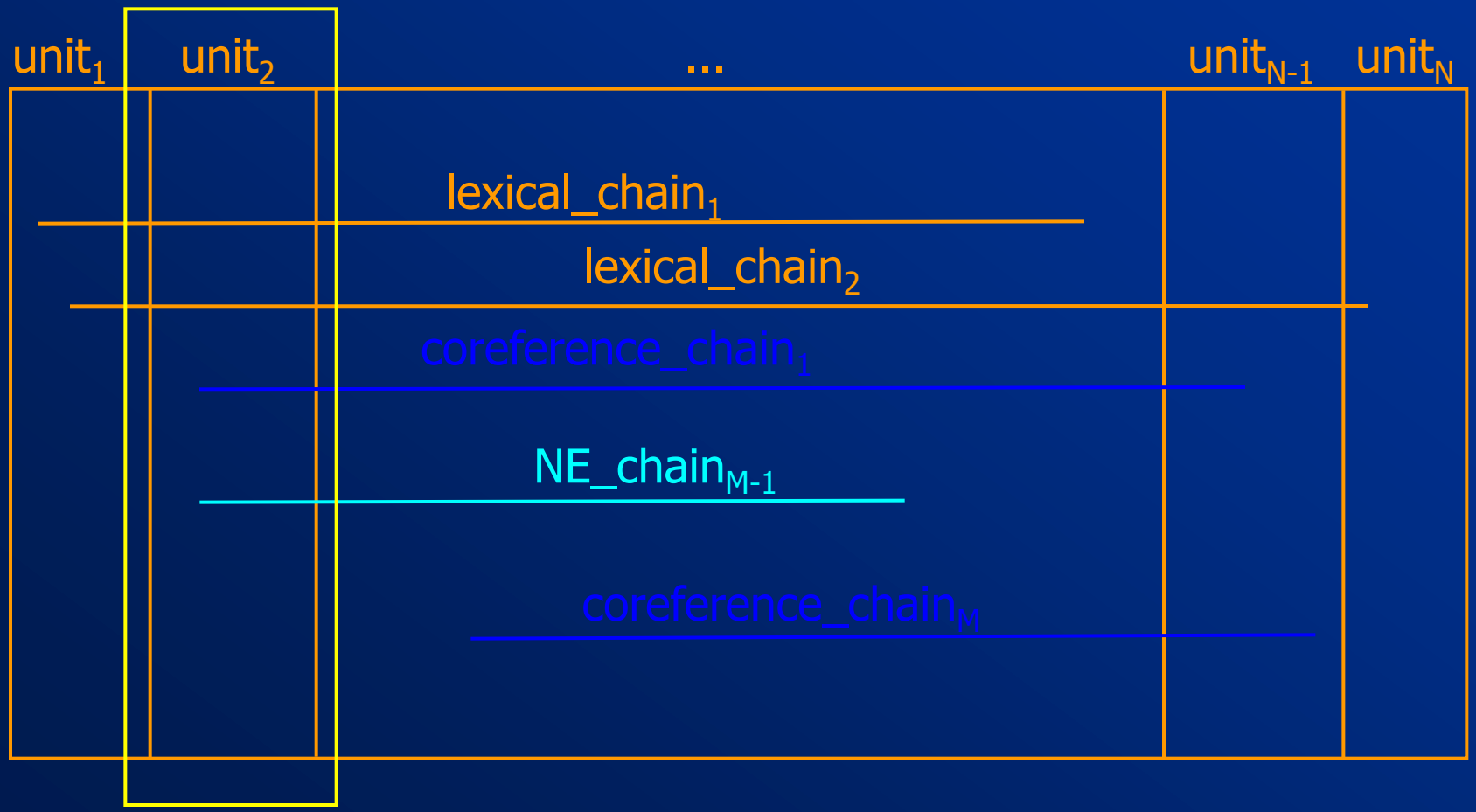
Lexical Chains 2



Lexical Chains 3

3 types of chains:

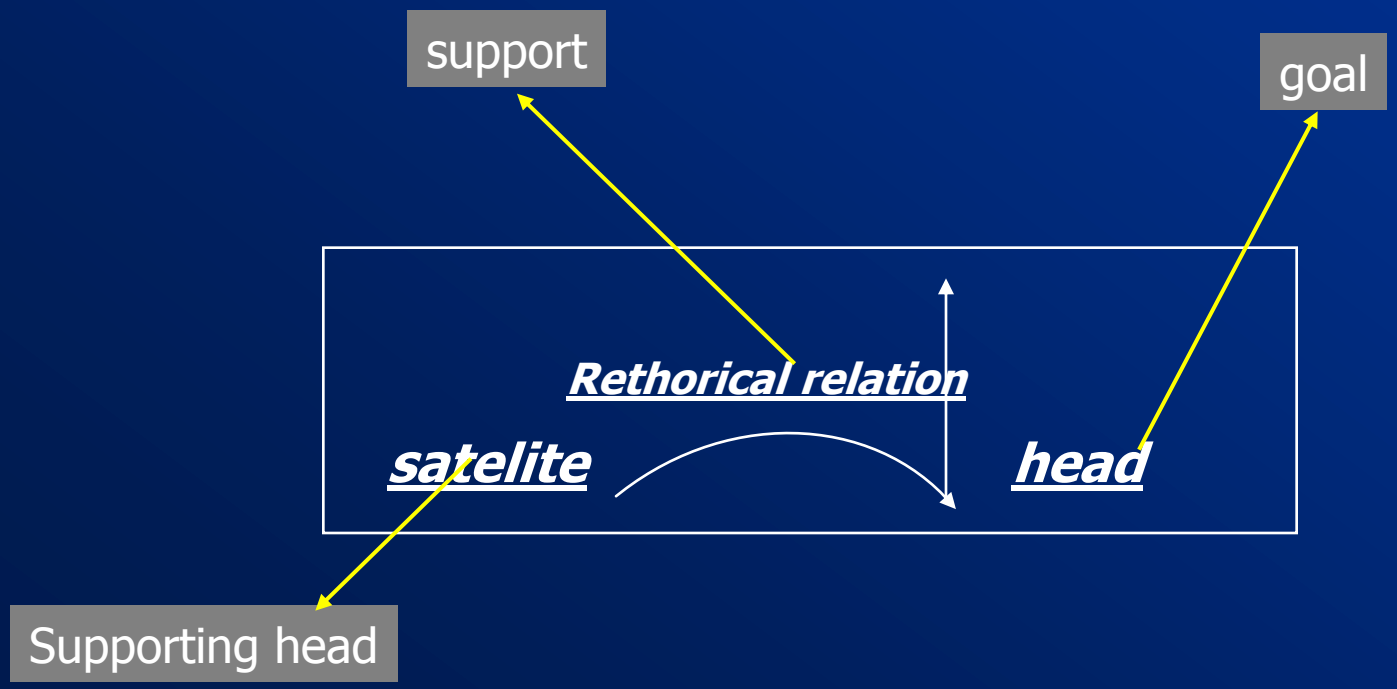
- Lexical chains
- coreference chains
- NE chains



RST 1

RST states that the elements in the text are related by rethorical relations supporting it. The basic idea is that the text can be organized into text spans, expressing the goals of talkers related by these rethorical relations

Mann, Thomson propose 25 rethorical relations



RST 2

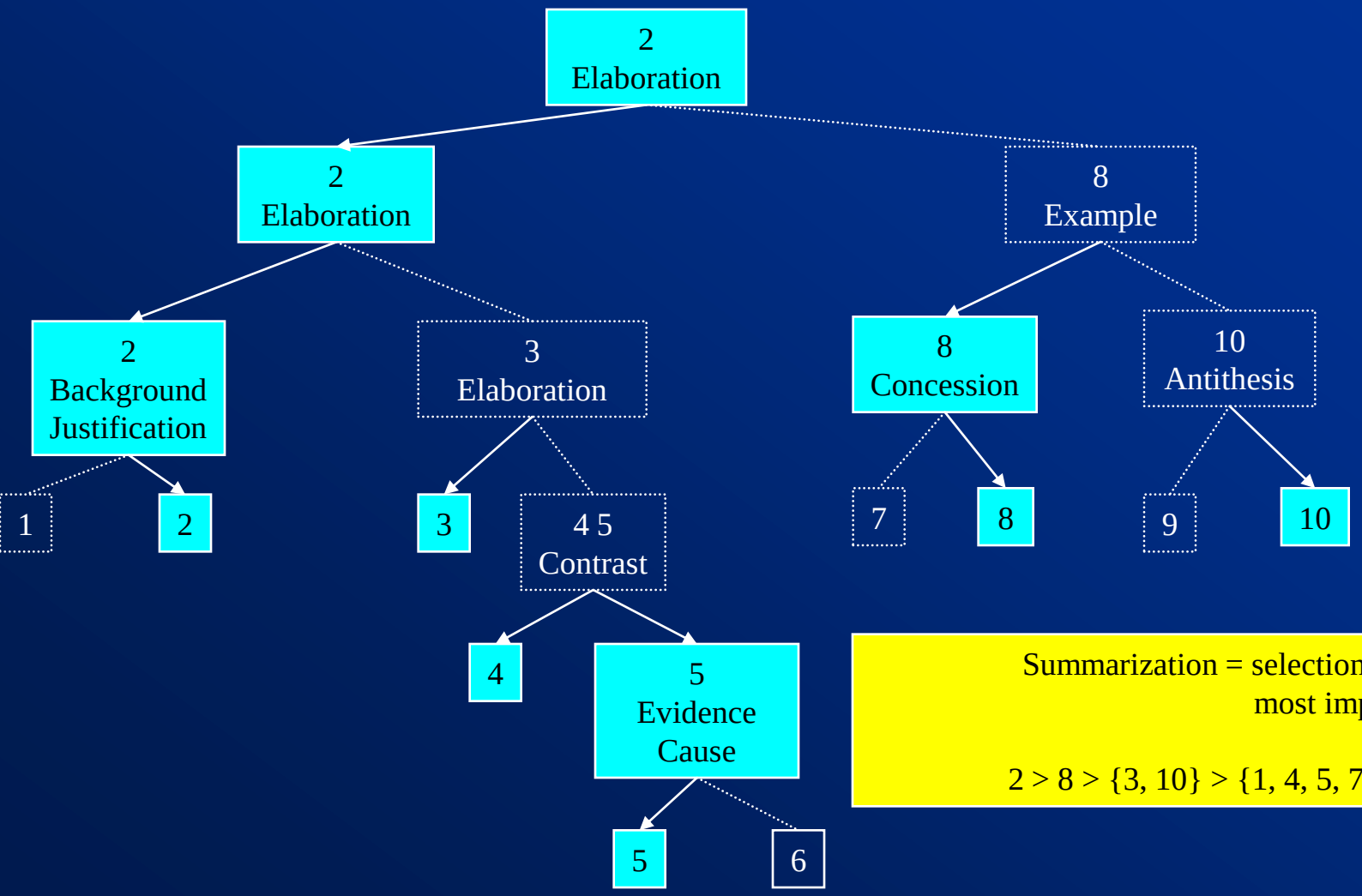
- Construction of RTS tree
- RST parsing (usually tied to recognition of Rethorical markers)
- Aplication of measures of cohesion (e.g. vocabulary overlap) for connecting the trees
- Alonso, 2005
  - dm neutral, nuc relevante, sat irrelevante
  - r right directed, b both sides

discourse marker	relevance	syntactic
because	dm	b
although	sat	r
first_of_all	dm	r
however	nuc	r
in_case	sat	b
in_order_to	nuc	b

Rhetorical Parsing (Marcu) 1

[With its distant orbit {– 50 percent farther from the sun than Earth –} and slim atmospheric blanket,[1]] [Mars experiences frigid weather conditions.[2]] [Surface temperatures typically average about –60 degrees Celsius (–76 degrees Fahrenheit) at the equator and can dip to –123 degrees C near the poles.[3]] [Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion,[4]] [but any liquid water formed that way would evaporate almost instantly[5]] [because of the low atmospheric pressure.[6]] [Although the atmosphere holds a small amount of water, and water-ice clouds sometimes develop,[7]] [most Martian weather involves blowing dust or carbon dioxide.[8]] [Each winter, for example, a blizzard of frozen carbon dioxide rages over one pole, and a few meters of this dry-ice snow accumulate as previously frozen carbon dioxide evaporates from the opposite polar cap.[9]] [Yet even on the summer pole, {where the sun remains in the sky all day long,} temperatures never warm enough to melt frozen water.[10]]

Rhetorical Parsing (Marcu) 2



Summarization = selection of the most important units

2 > 8 > {3, 10} > {1, 4, 5, 7, 9} > 6

Statistical methods 1

- Bayesian models
  - Schlesinger, Baker, 2001, Daumé III, Marcu, 2005
- HMM
  - Conroy, Schlesinger, ... 2001, ... , 2006
- Logistic Regression
  - Conroy, Schlesinger, ... 2001,
- Language Models
  - Kraaij et al, 2001, 2002
- Noisy Channel Model
  - Berger 2001, Berger Mittal, 2000
- Combination, mixture models
  - Lloret et al, 2009, Chandra et al, 2011, Bhatia et al, 2012

Statistical methods 2

Conroy et al, 2001

- Naive Bayes
- HMM
- Logistic regression
- features:
  - Position in the sentence, # tokens, # query terms in the sentence, distance to the query terms

Kraaij et al, 2001, 2002

- Combination of a mixture LM
  - content-based (smoothed unigram)
- and Naive Bayes
  - non content features (sentence length, 1st sentence, ...) combined using odds of salience

Statistical methods 3

Berger, 2001, Berger, Mittal, 2000

- Noisy Channel Model Approach
  - OCELOT
    - summarizing Web pages
    - Learning from Open Directory Project
    - Content selection (models of gisting), word ordering, searching
  - Learning from FAQ files
    - query-relevant summarization
    - relevance vs fidelity
    - shrinkage for smoothing

Statistical methods 4

Lloret et al, 2009

- Compute the relevance of a sentence by using the frequency of words in combination with the length of noun phrases

Chandra et al, 2011

- Use a K-mixture probabilistic model. They proposed a K-mixture semantic relationship significance (KSRS) approach. It employs the K-mixture probabilistic model to establish term weights. It also determines the term relationships to derive the semantic relationship significance of nouns. Depending on sentences semantic relationship significance values, sentences are ranked and extracted.

Bhatia al, 2013

- Also a K-mixture model. Sentences are ranked and extracted depending on their semantic relationship significance values

CLASSY 1

- Conroy, Schlesinger, O'leary, 2001, …, 2006
  - IDA, Univ. Maryland
- Clustering Linguistic and Statistics for Summarization Yield
- MDS
  - Sentence splitting & trimming
  - Sentence scoring
  - Sentence selection for redundancy removal and improving flow
- 50% terms in human summaries do not occur in document (s)
- Agreement in human summarizers vocabulary 40%

CLASSY 2

- Human variation: unigram bow
- Topic τ
- Probability a human includes t in a summary $P(t|\tau)$
- For a sentence x

$$\omega(x) = \frac{1}{|x|} \sum_{t \in T} x(t) P(t|\tau)$$

- Where x(t) = 1 if t occurs in x and zero otherwise

CLASSY 3

- Learning P(t|τ)
  - MLE from h sample summaries where
    - $C_{it}(\tau) = 1$ if i-th summary contains t
  - Oracle score $\bar{\omega}(x)$ computed from the estimate of P
  - Another approximation
    - τ topic description → query terms
    - collection of relevant documents → signature terms
  - Improvements
    - Including NEs as third set of terms
      - NO IMPROVEMENT
    - Pseudo-relevance feedback

$$\hat{P}(t|\tau) = \frac{1}{h}\sum_{i=1}^{h} c_{it}(\tau).$$

$$P_{qs}(t|\tau) = \frac{1}{2}q_t(\tau) + \frac{1}{2}s_t(\tau)$$

$$\omega_{qs}(x) = \frac{1}{|x|}\sum_{t\in T} x(t)P_{qs}(t|\tau)$$

CLASSY 4

- Pseudo-relevance feedback
  - 1) compute ω_{qs} for each sentence
  - 2) select k tops
- A term-sentence matrix
  - k tops correspond to columns $j_1, \ldots, j_k$
- $\rho(t)$ expectation that a term is used in the extract of these k sentences

$$\rho(t) = \frac{1}{k}\sum_{i=1}^{k} A(t, j_i).$$

$$P_{qs\rho}(t|\tau) = \frac{1}{2}P(t|\tau) + \frac{1}{2}\rho(t).$$

is given by

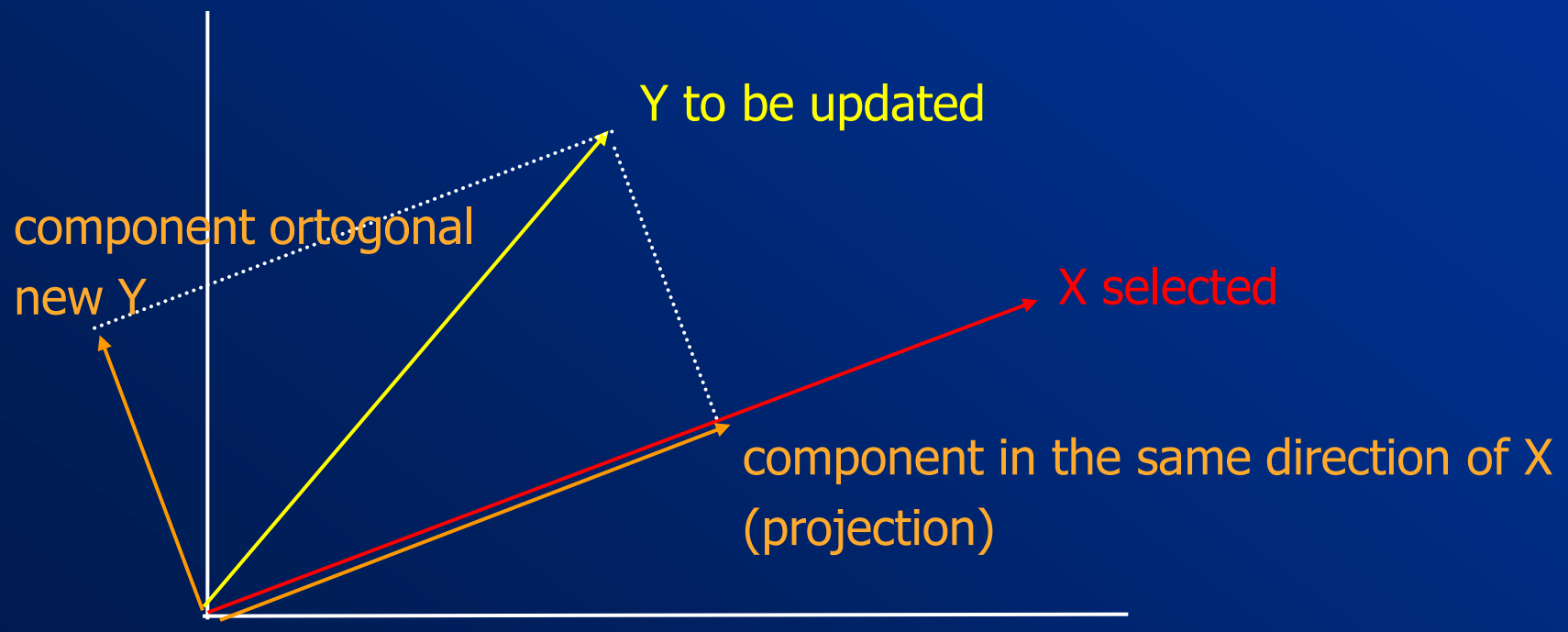
$$\omega_{qs\rho}(x) = \frac{1}{|x|}\sum_{t\in T} x(t)P_{qs\rho}(t|\tau)$$

CLASSY 5

- Reducing the redundancy of the selected sentences
- Most popular method: MMR
- They use pivoted QR until DUC-2005, in DUC-2006 non-negative-QR decomposition
  - Conroy et al, 2001
- Gram-Schmidt ortogonalization method
- B term-sentence matrix
  - T columns, m rows
  - B_{ij} = 1 if j contains i
- A_w normalized
  - Each column is normalized dividing each element by its 2-norm (euclidean distance)
- for i= 1, ... , min(m,T)
  - choose column l that has the highest norm (a_l)
  - set $Index_i = l$
  - set $q_i = a_l / || a_l ||$
  - update the other columns to make them ortogonal to a_l
    - for each unchosen column a_j
      - $r_{ij} = a_j^T q_i$
      - $a_j = a_j - r_{ij} q_i$
- The summary of length k will contain $Index_i$, ... , $Index_k$

CLASSY 6

- The idea of pivoted QR is in every iteration, after the selection is made, substracting off from all the remaining columns the components thar are in the direction of the selected column



Graph-based measures of importance 1

- Mihalcea, Radev tutorial on RANLP 2005
- Centrality, Relevance, Saliency of nodes
  - paragraphs, sentences, words, ...
- Several forms of represention
  - graphs (undirected)
  - digraphs
  - weighted or binary
  - bigraphs

Graph-based measures of importance 2

- Random Walk Algorithms
  - Decide the importance of a vertex within a graph
  - A link between two vertices = a vote
    - Iterative voting => Ranking over all vertices
  - Model a random walk on the graph
    - A walker takes random steps
    - Converges to a stationary distribution of probabilities
    - Guaranteed to converge if the graph is
      - Aperiodic – holds for non-bipartite graphs
      - Irreducible – holds for strongly connected graphs
  - Examples
    - PageRank (Brin, Page, 1998)
    - HITS (Kleinberg, 1999)

Graph-based measures of importance 3

- PageRank
  - Random walks over the Web following hyperlinks
  - some probability of jumping randomly out of the page
  - The fraction of time spent in each page is the PageRank
- HITS (Hyperlinked Induced Topic Search)
  - Defines 2 types of importance:
    - Authority
      - Can be thought of as being an expert on the subject
      - "You are important if important hubs point to you"
    - Hub
      - Can be thought of as being knowledgeable about the subject
      - "You are important if important authorities point to you"
  - Hubs and Authorities mutually reinforce each other
  - random walk where:
    - Odd steps you walk from hubs to authorities (follow outlink)
    - Even steps you walk from authorities to hubs (follow inlink)

Graph-based measures of importance 4

- PageRank
  - simple

$$PR(V_i) = (1-d) + d \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} PR(V_j)$$

$$d \in [0,1]$$

  - weighted graphs

$$PR^w(V_i) = (1-d) + d \sum_{j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} PR^w(V_j)$$

- HITS

$$HITS_H(V_i) = \sum_{j \in Out(V_i)} HITS_A(V_j)$$

$$HITS_A(V_i) = \sum_{j \in In(V_i)} HITS_H(V_j)$$

Salton et al, 1999

- Salton, Singhal, Mitra, Buckley, 1999
- Representing documents and paragraphs using VSM
- Similarity: escalar product
- Representing document as undirected graphs
  - Nodes = paragraphs
  - Edges weigthed with sim
- Salient nodes are those connected with many others with sim over a threshold

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

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

TextRank 1

- Mihalcea & Tarau, 2004
- Graph-based ranking algorithms for extractive summarization
- Weighted Graph representation
  - nodes: sentences
  - edges: similarity between sentences
- Run a Graph-based ranking algorithm
  - PR^w, $HITS_A{}^w$, $HITS_H{}^w$
- Get the N best ranked sentences
- MDS
  - cascaded summarization
  - using SDS for each document
  - using SDS for performing summary of summaries

LexRank 1

- Erkan, Radev, 2004
- Idea: Sentences more similar to others are more central (salient)
- Centroid-based
  - 3 measures
    - Degree
    - LexRank with threshold
    - Continuous LexRank
- Sentence similarity

$$\text{idf-modified-cosine}(x, y) = \frac{\sum_{w \in x,y} \text{tf}_{w,x}\text{tf}_{w,y}(\text{idf}_w)^2}{\sqrt{\sum_{x_i \in x}(\text{tf}_{x_i,x}\text{idf}_{x_i})^2} \times \sqrt{\sum_{y_i \in y}(\text{tf}_{y_i,y}\text{idf}_{y_i})^2}}$$

LexRank 2

- Similarity graph (SG)
  - Nodes: sentences
  - Edges: similarity scores
- Degree centrality = degree of a sentence in SG
- SG is represented by a adjacency matrix A
- A can be normalized as B

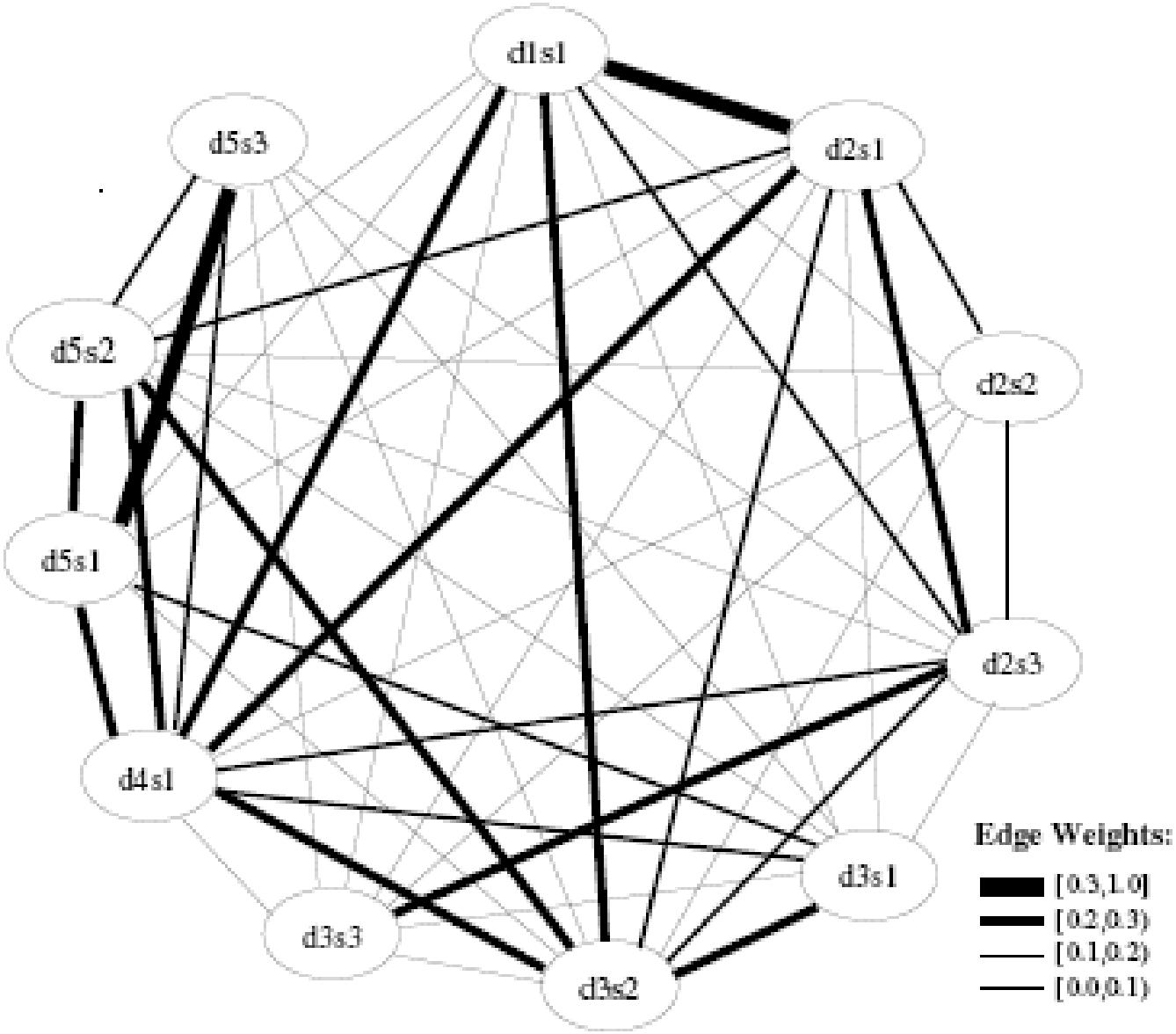
$$\mathbf{B}(i,j) = \frac{\mathbf{A}(i,j)}{\sum_k \mathbf{A}(i,k)}$$

LexRank 3

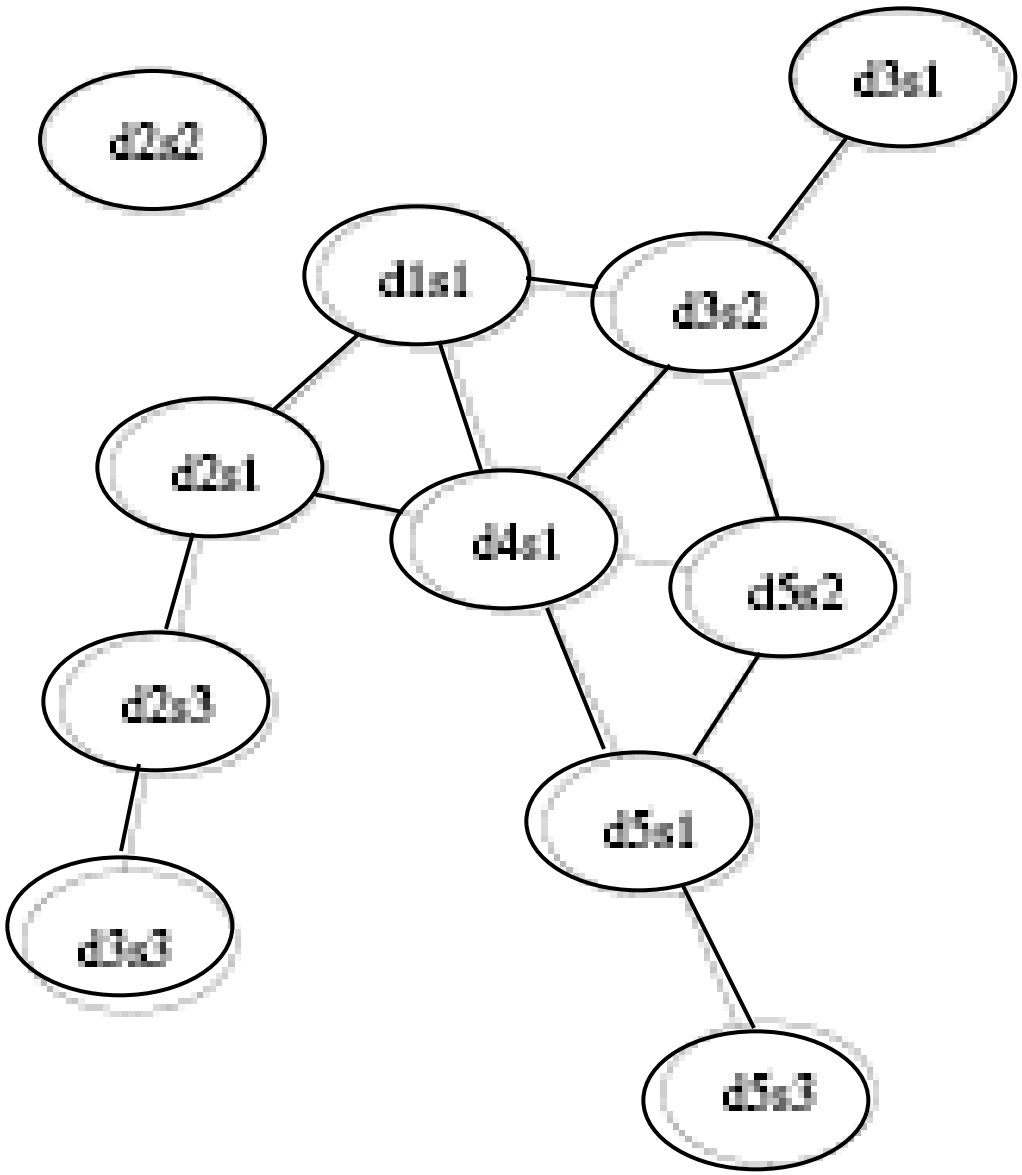
	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.45	0.02	0.17	0.03	0.22	0.03	0.28	0.06	0.06	0.00
2	0.45	1.00	0.16	0.27	0.03	0.19	0.03	0.21	0.03	0.15	0.00
3	0.02	0.16	1.00	0.03	0.00	0.01	0.03	0.04	0.00	0.01	0.00
4	0.17	0.27	0.03	1.00	0.01	0.16	0.28	0.17	0.00	0.09	0.01
5	0.03	0.03	0.00	0.01	1.00	0.29	0.05	0.15	0.20	0.04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

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

LexRank 4


d1s1
d2s1
d2s2
d2s3
d3s1
d3s2
d3s3
d4s1
d5s1
d5s2
d5s3
Edge Weights:
[0.3,1.0]
[0.2,0.3)
[0.1,0.2)
[0.0,0.1)

LexRank 5



LexRank 6

- Distribute centrality between neighbours
  - $p = B^T p$
  - $p^T B = p^T$

$$p(u) = \sum_{v \in adj[u]} \frac{p(v)}{deg(v)}$$

- p^T is the left eigenvector of B corresponding to the eigenvalue 1
- B has to be irreducible and aperiodic

LexRank 7

- Following PageRank (Page et al, 1998)
  - Lexical PadeRank
  - Reserving some low probability of jumping to any node in SG

$$p(u) = \frac{d}{N} + (1-d) \sum_{v \in adj[u]} \frac{p(v)}{deg(v)}$$

$$\mathbf{p} = [d\mathbf{U} + (1-d)\mathbf{B}]^{\mathrm{T}}\mathbf{p}$$

LexRank 8

- Using a weighted graph
- Continuous LexRank

$$p(u) = \frac{d}{N} + (1-d) \sum_{v \in adj[u]} \frac{\text{idf-modified-cosine}(u,v)}{\sum_{z \in adj[v]} \text{idf-modified-cosine}(z,v)} p(v)$$

LexNet

Radev et al, 2006

LexRank 9

- Topic sensitive LexRank
- Ottebacher et al, 2006
- Relevance to the query
- Computing idf_w for each word (after stemming) in the set

$$\mathrm{idf}_w = \log\left(\frac{N+1}{0.5 + sf_w}\right)$$

- Relevance of a sentence

$$\mathrm{rel}(s|q) = \sum_{w \in q} \log(tf_{w,s} + 1) \times \log(tf_{w,q} + 1) \times \mathrm{idf}_w$$

LexRank 10

- Mixture model
  - Similarity

$$\mathrm{sim}(x,y) = \frac{\sum_{w \in x,y} \mathrm{tf}_{w,x}\,\mathrm{tf}_{w,y}(\mathrm{idf}_w)^2}{\sqrt{\sum_{x_i \in x}(\mathrm{tf}_{x_i,x}\,\mathrm{idf}_{x_i})^2} \times \sqrt{\sum_{y_i \in y}(\mathrm{tf}_{y_i,y}\,\mathrm{idf}_{y_i})^2}}$$

  - Relevance of s

$$p(s|q) = d\,\frac{\mathrm{rel}(s|q)}{\sum_{z \in C}\mathrm{rel}(z|q)} + (1-d)\sum_{v \in C}\frac{\mathrm{sim}(s,v)}{\sum_{z \in C} sim(z,v)}\,p(v|q)$$

$$\mathbf{p} = [d\mathbf{A} + (1-d)\mathbf{B}]^{\mathrm{T}}\mathbf{p}$$

  - When A is uniform and B normalized binary the algorithm reduces to LexRank

Zha, 2002 1

- Generic Text Summarization + Keyphrase extraction
- SDS or MDS
- Mutual Reinforcement Principle
- Unsupervised approach
- Approach
  - 1) Clustering sentences into topical groups
  - 2) Within each topical group select keyphrases and sentences by saliency score

Zha, 2002 2

- Bigraph representation of a document
  - Set of terms $T = \{t_1, \ldots, t_n\}$
  - Set of sentences $S = \{s_1, \ldots, s_m\}$
  - Weighted bipartite graph
    - Nodes: $T \cup S$
    - Edges
      - (t_i, s_j) is weighted by the number of times t_i occur in s_j
    - G(T,S,W)
- Saliency
  - A term should have a high salience if it appears in many sentences with high salience while a a sentence has a hig salience if it contains many terms with high salience

Zha, 2002 3

- Salience of a term

$$u(t_i) \propto \sum_{v(s_j) \sim u(t_i)} w_{ij} . v(s_j)$$

- Salience of a sentence

$$v(s_j) \propto \sum_{v(s_j) \sim u(t_i)} w_{ij} . u(t_i)$$

- Salience of terms

$$u = \frac{1}{\sigma} W v$$

- Salience if sentences

$$v = \frac{1}{\sigma} W^T u$$

Zha, 2002 4

- For σ = Max eigenvalue of W components of u and v (eigen vectors) are non negative
- Max {u, σ ,v} over W
- Iteration till convergence of equations u and v
- On convergence σ is computed

$$u = Wv \qquad v = W^T u$$

$$u = \frac{u}{\|u\|} \qquad v = \frac{v}{\|v\|}$$

$$\sigma = u^T W v$$

Graph-based

Lin et al, 2009

- Unsupervised approach. Optimizing sub-modular functions defined on the semantic graph. The enhancement is theoretically ensured to be near-optimal under the framework of submodularity.

Kumaresh and Ramakrishnan, 2012

- Graph traversal technique with constraint for cohesion guarantee.
- A better summary can be generated by recognizing both the unstructured and the structured features

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- **Multi-document Systems**
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Multidocument Summarization (MDS) 1

Goals

- Content of a collection of documents
- Briefing
  - concise summary of the factual matter of a set of news articles on the same or related events (SUMMONS, Radev, 1999)
- Information updating
- Looking for user information needs in a collection of documents

MDS 2

Differences SDS MDS

- Lower compression level
- anti-redundancy
- temporal dimension
- Coreference more challenging

MDS 3

Requirements

- Clustering of documents and pasages
  - Similarity meadures between units (documents, pasages, paragraphs, sentences)
- recall
- anti-redundancy
- Summary cohesion
- quality
  - readable
  - relevant
  - context
- Sources inconsistencies
- updating

MDS 4

Types

- Common sections of documents of the collection
- Common sections + unique sections
- Centroids
- Centroids + outliers
- Last document + outliers
- Common sections + unique sections + temporal weighting score

MDS 5

- Information Fusion (Barzilay et al, 1999)
  - articles presenting different descriptions of the same news
  - Repetition is a good indicador of relevance
  - automatically generate a concise summary by identifying similarities and differences across a set of related documents.
    - Identifying themes
    - Information Fusion
    - Generation (Reformulation)

MDS 6

Techniques

- Most are the same used in SDS
- Clustering
- IE
  - Context modeling
    - (Lal, Rueger, 2002), (Kan, McKeown, 1999)
  - Templates adapted to domains or events
    - (Harabagiu, Lacatusu, 2002), (Daumé et al, 2002)
- Centroid-based
  - (Radev et al, 2004), (Saggion, Gaizauskas, 2004)
- Graph-based
  - (Radev et al, 2004), (Erkan, Radev, 2004), (Mihalcea, Tarau, 2004), (Zha, 2002)
- SVM
  - (Hirao et al, 2003)

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- **Headline extraction**
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Headline Extraction 1

- A headline is a highly concise representation of the most relevant points contained in a document. It can consist of a sentence, either extracted from the document or automatically generated, or, sometimes, of a list of relevant terms.
- The main characteristic of a headline is its extremely small length (usually between 5 and 10 words).
- 3 main tasks
  - Identification of relevance signaling words.
  - Combination of the results proposed by simpler individual methods.
  - Some form of compression or simplication of the extracted sentences.

Headline Extraction 2

- Identification of relevance signaling words
  - (Schlessinger et al 2001)
    - lead property, specificity of verbs, use of Concept Sets derived from WN
  - (Kraaij et al, 2002)
    - identify the most informative topical NP:
      - Sentences are ranked by an hybrid model that merges i) a unigram LM, mixing cluster and document models, for scoring sentences and ii) a Bayesian model on content features like cue phrases, position, length, etc. A Trigger Word Pool is built from the ranked sentences and the most salient trigger word is selected.
      - Then, all maximal NPs containing the trigger word are found, and the one in the highest ranked sentence is taken as a headline.
  - (Daumé et al, 2002)
    - go beyond words or NP and perform a full parsing of the document for getting the main entities and their relations. From them the headline is built.
  - (Zajik et al, 2002)
    - Use a Noisy Channel Model (NCM), implemented by means of a HMM, for selecting a stream of headline words from a stream of story words. The first is modelled by a bigram LM and the latter by unigram one. The headline is generated by a simple Viterbi decoder constrained by the model. A set of penalty constraints (length, position, gap and string) is added.

Headline Extraction 3

- Combination of the results proposed by simpler individual methods
  - (Lin, 1999)
    - linear combination
  - (Aone et al, 1997)
    - DT
  - MEAD (Erkan, Radev, 2004)
    - combining centroid, position and length
- Some form of compression or simplication of the extracted sentences
  - (Knight, Marcu, 2000
    - translation from full text to summary using NCM or DT
  - (Lal, Rueger, 2002)
    - lexical simplification
  - TOPIARY (Zajik et al, 2004), (Dorr et al, 2003)
    - sentence compression

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- **Query-based Systems**
- Multimedia and Multilingual Systems
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Query_based Summarization 1

- User information need expressed with a query
  - NL question
  - IR query
  - DUC 2005 and 2006
- Two tasks
  - Identification of relevant sentences
  - Similarity between sentence and query

Query_based Summarization 2

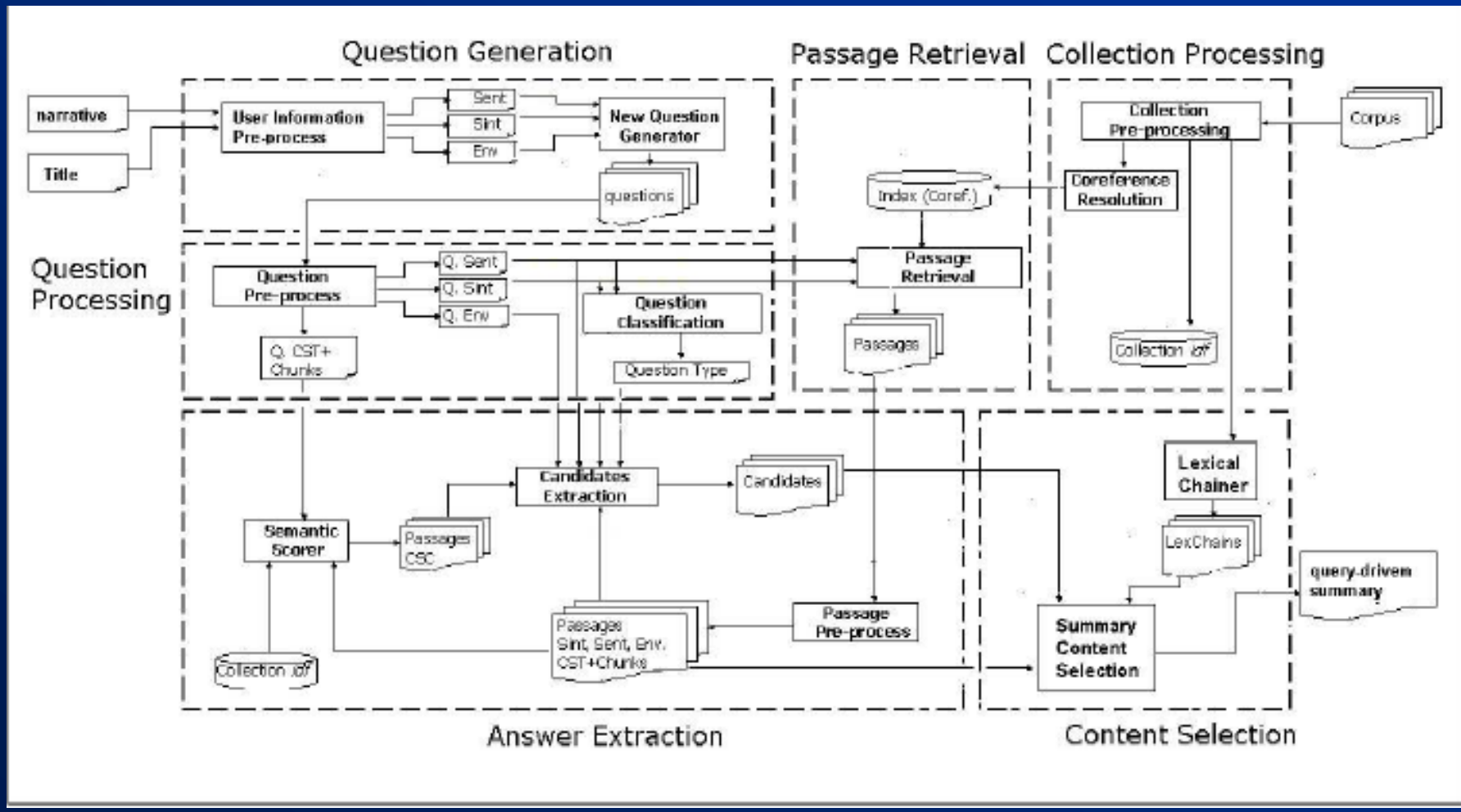
- Identification of relevant sentences
  - frequency
    - (Nenkova, Vanderwende, 2005)
  - oracle score
    - (Conroy et al, 2005)
  - Bayes
    - (Daumé, Marcu, 2005), (Daumé, 2006)
  - Perceptron (sentence ranking)
    - (Fisher, Roark, 2006)
  - Centroid
  - Graphs
- Similarity between sentence and query

Query_based Summarization 3

- Similarity between sentence and query
  - Tree similarity
    - (Schilder, McInnes, 2006)
  - Q&A
    - (Fuentes et al, 2005), (Lacatusu et al, 2006), (McKeown et al, 2006), (Mollá, 2006)
  - Enriched VSM
    - Query expansion
      - (Alfonseca et al, 2006)
    - Syntactic and semantic relations
      - (Lacatusu et al, 2006)
    - LSA
      - (Miller, 2003), (Hachey et al, 2005)

Query_based Summarization 4

QASUM-TALP



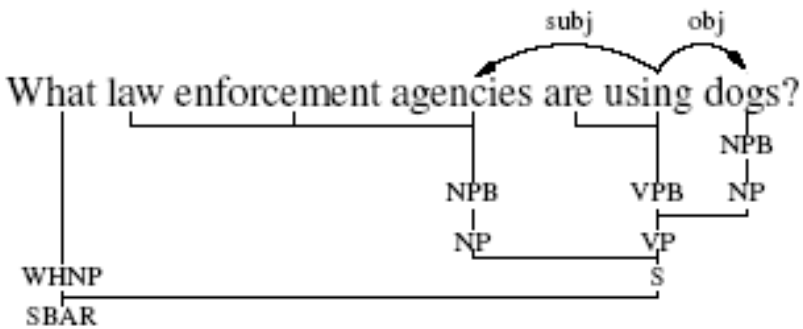
M. Fuentes (2008)

Query_based Summarization 5

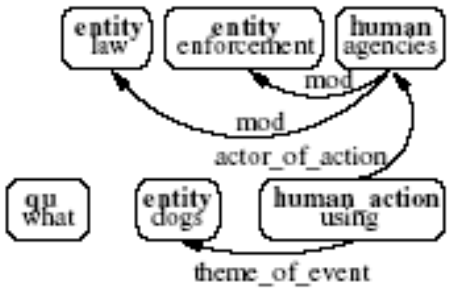
D426d

What law enforcement agencies are using dogs?

```
sent(3, 16, [
        (word('What'), lemma("what"), pos("WP"), nil, nil, []),
        (word('law'), lemma("law"), pos("NN"), nil, nil, [n00361262,... ]),
        ...
        (word('?'), lemma("?"), pos("."), nil, nil, [])
      ]).
```




```
sint(3,[ch(1,7,sbar,1,[ch(1,1,whnp,1,[tk(1)]),ch(2,7,s,6,[ch(2,4,np,4,
        [ch(2,4,npb,4,[tk(2),tk(3),tk(4)])]),ch(5,7,vp,6,[ch(5,6,vpb,6,[tk(5),tk(6)])
        ch(7,7,np,7,[ch(7,7,npb,7,[tk(7)])])])])])],[subj(6, 4), obj(6, 7)]).
```




```
env(3, [human_action(6), qu(1), human(4), mod(4,2), entity(2),
    mod(4,3),entity(3),entity(7),actor_of_action(6,4),theme_of_event(6,7)]).
```

Generated Questions

When agencies are using dogs?
Where agencies are using dogs?
Who **agencies are using dogs?**

Error due to missclassification of agencies as human

M. Fuentes (2008)

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- **Multimedia and Multilingual Systems**
- Multitask Systems
- Related tasks
  - Text simplification
  - Paraphrasing

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- **Multitask Systems**
- Related tasks
  - Text simplification
  - Paraphrasing

Outline

- Introduction
- Problems of AS
- Basic Architectures
- Evaluations
- Single-document Systems
- Multi-document Systems
- Headline extraction
- Query-based Systems
- Multimedia and Multilingual Systems
- Multitask Systems
- **Related tasks**
  - Text simplification
  - Paraphrasing

Text simplification 1

- Saggion 2013, 2015
  - The process of transforming a text into an equivalent which is more readable and/or understandable by a target audience
  - During simplification, complex sentences are split into simple ones and uncommon vocabulary is replaced by more common expressions
  - Identify and measure sources of complexity / difficulty
    - Create a "paraphrase" which is easy to read
    - Several NLP expertises involved: summarization, natural, language generation, sentence compression, word sense, disambiguation, machine translation, etc.

Text simplification 2

- Saggion 2013, 2015
  - Induced or hand-crafted rules for reducing syntactic complexity
  - Lexical substitution procedures
  - Combined approach treating the problem as MT
  - Extracting Lexical Simplifications from Wikipedia
  - Biran et al. (2011) use English Wikipedia(EW) and Simple English Wikipedia (ESW)