Multilingual Summarization

Leo Wanner
ICREA and DTIC, Pompeu Fabra University

leo.wanner@upf.edu

ESSIR 2015, Thessaloniki

Google

climate change causes

Web Imatges Vídeos Llibres Maps Més Eines de cerca

Aproximadament 91.400.000 resultats (0,52 segons)

Articles acadèmics per **climate change causes**

Stern Review: The economics of **climate change** - Stern - Citat per 5019

… variation confirms that **climate change causes** birds to … - Both - Citat per 288

Causes of **climate change** over the past 1000 years - Crowley - Citat per 1710

Climate Change: Vital Signs of the Planet: Causes - Global ...

climate.nasa.gov/**causes** ▾ Tradueix aquesta pàgina

Vital Signs of the Planet: Global **Climate Change** and Global Warming. What is the "greenhouse effect"? What is **causing** it? Are humans to blame? What does ...

Causes of Climate Change | Climate Change | US EPA

www.epa.gov/**climatechange**/science/**causes**.html ▾ Tradueix aquesta pàgina

Before humans, **changes** in **climate** resulted entirely from natural **causes** such as **changes** in Earth's orbit, **changes** in solar activity, or volcanic eruptions.

Imatges sobre climate change causes

Informeu de les imatges

[LLIBRE] Stern Review: The economics of **climate change**
NH Stern, HM Treasury - 2006 - hm-treasury.gov.uk
... sradley@eef-fed.org.uk www.eef.org.uk ref: SR/Stern/161205 Dear Sir Nicholas, Stern Review on the Economics of **Climate Change** I have pleasure in enclosing out initial submission to your review on the economics of **climate change**. **...**
Citat per 5019 Articles relacionats Cita Desa Més

Large–scale geographical variation confirms that **climate change causes** birds to lay earlier
C Both, AV Artemyev, B Blaauw... - ... of the Royal ..., 2004 - rspb.royalsocietypublishing.org
Abstract Advances in the phenology of organisms are often attributed to **climate change**, but alternatively, may reflect a publication bias towards advances and may be caused by environmental factors unrelated to **climate change**. Both factors are investigated using the **...**
Citat per 288 Articles relacionats Totes les 26 versions Cita Desa

Causes of **climate change** over the past 1000 years
TJ Crowley - Science, 2000 - sciencemag.org
Abstract Recent reconstructions of Northern Hemisphere temperatures and **climate** forcing over the past 1000 years allow the warming of the 20th century to be placed within a historical context and various mechanisms of **climate change** to be tested. Comparisons of **...**
Citat per 1710 Articles relacionats Totes les 40 versions Cita Desa

[LLIBRE] ... 1995: The science of **climate change**: contribution of working group I to the second assessment report of the Intergovernmental Panel on **Climate Change**
JT Houghton - 1996 - books.google.com
... of **Climate Change** 65 3 Observed **Climate** Variability and **Change** 133 4 **Climate** Processes 193 5 **Climate** Models-Evaluation 229 6 **Climate** Models-Projections of Future **Climate** 285 7 Changes in Sea Level 359 8 Detection of **Climate Change** and Attribution of **Causes** 407 9 **...**
Citat per 3302 Articles relacionats Totes les 3 versions Cita Desa Més

[LLIBRE] **Climate change**: **causes**, effects, and solutions
JT Hardy - 2003 - books.google.com
This book addresses civilization's most important environmental challenge: **climate change**.

Causes of Climate Change Over the Past 1000 Years

Thomas J. Crowley

Recent reconstructions of Northern Hemisphere temperatures and climate forcing over the past 1000 years allow the warming of the 20th century to be placed within a historical context and various mechanisms of climate change to be tested. Comparisons of observations with simulations from an energy balance climate model indicate that as much as 41 to 64% of preanthropogenic (pre-1850) decadal-scale temperature variations was due to changes in solar irradiance and volcanism. Removal of the forced response from reconstructed temperature time series yields residuals that show similar variability to those of control runs of coupled models, thereby lending support to the models' value as estimates of low-frequency variability in the climate system. Removal of all forcing except greenhouse gases from the ~1000-year time series results in a residual with a very large late-20th-century warming that closely agrees with the response predicted from greenhouse gas forcing. The combination of a unique level of temperature increase in the late 20th century and improved constraints on the role of natural variability provides further evidence that the greenhouse effect has already established itself above the level of natural variability in the climate system. A 21st-century global warming projection far exceeds the natural variability of the past 1000 years and is greater than the best estimate of global temperature change for the last interglacial.

The origin of the late-20th-century increase in global temperatures has prompted considerable discussion. Detailed comparisons of climate model results with observations (*1*) increase in solar irradiance or a reduction in volcanism might account for a substantial amount of the observed 20th-century warming (*1, 3–10*). Although many studies have ad-

Data

The data used in this study include physically based reconstructions of Northern Hemisphere temperatures and indices of volcanism, solar variability, and changes in GHGs and tropospheric aerosols.

Northern Hemisphere temperatures. Four indices of millennial Northern Hemisphere temperature have been produced over the past 3 years (*11–14*). The analysis here uses the mean annual temperature reconstructions of Mann *et al.* (*11*) and of Crowley and Lowery (CL) (*12*), because the energy balance model used in this study calculates only this term [the other records (*13, 14*) are estimates of warm-season temperature at mid-high latitudes]. The Mann *et al.* reconstruction was determined (*8*) by first regressing an empirical orthogonal function analysis of 20th-century mean annual temperatures against various proxy indices (such as tree rings, corals, and ice cores). Past changes in temperature are estimated from variations in the proxy data (*15*). The Mann *et al.* reconstruction has a varying number of records per unit of time (although the number in the earlier part of the record is still greater than in CL). The CL reconstruction is a more heterogeneous mix of data than the Mann *et al.* reconstruction, but the number of records is nearly constant in time. It is a simple composite of Northern

Google

cambio climático causas

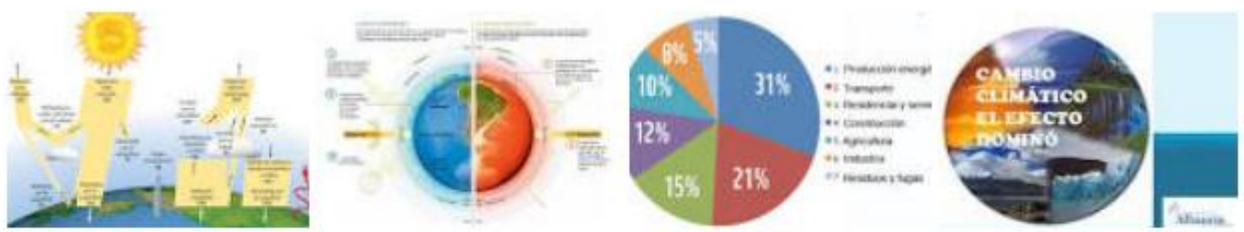
Web   Imatges   Vídeos   Maps   Més ▾   Eines de cerca

Aproximadament 736.000 resultats (0,24 segons)

Causas del Cambio Climático - Cambio Climático Global
cambioclimaticoglobal.com/**causas** ▾ Tradueix aquesta pàgina
Las **causas** del **cambio climatico** se dividen en dos categorías: **causas** naturales (por ej. volcanes) y **causas** antrópicas como la quema de combustibles fósiles.

WWF España - Causas
www.wwf.es › Qué hacemos › Cambio Climático ▾ Tradueix aquesta pàgina
El **cambio climático** es síntoma de un planeta enfermo. La huella que los seres humanos dejamos sobre la Tierra es cada vez más extensa y profunda.

Imatges sobre cambio climático causas   Informeu de les imatges



Més imatges per a cambio climático causas

Las causas | Greenpeace México
www.greenpeace.org/...**cambio-climatico**/Las-**ca**... ▾ Tradueix aquesta pàgina
El **cambio climático** que hoy enfrentamos está vinculado a la quema de combustibles fósiles, especialmente carbón, gas y petróleo, y a la deforestación, es ...

Cambio climático - Wikipedia, la enciclopedia libre

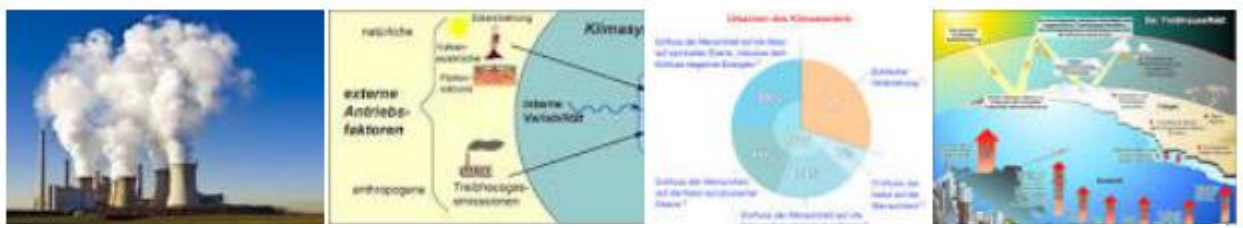
Klimawandel Ursachen

Web Imatges Vídeos Maps Més Eines de cerca

Aproximadament 467.000 resultats (0,34 segons)

Die Ursachen des Klimawandels - WWF Deutschland
www.wwf.de › ... › Klimawandel ▾ Tradueix aquesta pàgina
Die **Ursachen** des **Klimawandels**. Hauptquelle für Treibhausgase, insbesondere CO2, ist die Erzeugung von Energie. Dazu werden auch heute noch in ...

Imatges sobre Klimawandel Ursachen Informeu de les imatges



Ursachen des Klimawandels | Klima sucht Schutz
www.klima-sucht-schutz.de › ... › Klimawandel ▾ Tradueix aquesta pàgina
Ursachen des **Klimawandels**. Seit vielen Jahren ist sich die Wissenschaft einig: Eine erhöhte Konzentration von Treibhausgasen in der Atmosphäre führt zu ...

Klimawandel: Ursachen & Auswirkungen | Klima sucht Schutz
www.klima-sucht-schutz.de › Klimaschutz ▾ Tradueix aquesta pàgina
Ursachen und Folgen des **Klimawandels**, häufige Irrtümer und Diskussion. Jetzt informieren zu **Klimawandel** & globaler Erwärmung.

[PDF] Globaler Klimawandel: Ursachen, Folgen ... - Germanwatch
germanwatch.org/klima/gkw11.pdf ▾ Tradueix aquesta pàgina
Abb. 14: Sicherheitsrisiken durch **Klimawandel**: ausgewählte Brennpunkte. 36. Abb. 15: ..... Realität des **Klimawandels** oder dessen **Ursachen**. Vielmehr ...

Welche Ursachen hat der Klimawandel? - Helles Köpfchen
www.helles-koepfchen.de/artikel/2439.html ▾ Tradueix aquesta pàgina
★★★★★ Puntuació: 4,6 - 69 vots
Der Lebensraum von Mensch und Tier ist zunehmend bedroht. Was sind die **Ursachen** für den **Klimawandel**? Was ist der Treibhaus-Effekt? Was muss getan ...

Do we need to read all this?

Or can we obtain a gist of it?

…in the language of our preference…

… Summarization

… Multilingual summarization

… Cross-language summarization

What is a summary?

Wikipedia

A **summar**y means to write something in short like shortening a passage or a write up without changing its meaning but by using different words and sentences.

Does not need to be by using different words and sentences, though…

We can just pick the important statements

An (automatic) **summar**y is a condensation of the original material to certain (possibly predefined) length, while preserving its main message (content).

What can summarization be good for?

newspaper material

social media

Improved Compressed Indexes for Full-Text Document Retrieval*

Djamal Belazzougui[1] and Gonzalo Navarro[2]

[1] LIAFA, Univ. Paris Diderot - Paris 7, France
dbelaz@liafa.jussieu.fr
[2] Department of Computer Science, University of Chile
gnavarro@dcc.uchile.cl

Abstract. We give new space/time tradeoffs for compressed indexes that answer document retrieval queries on general sequences. On a collection of D documents of total length n, current approaches require at least $|\mathsf{CSA}| + O(n\frac{\lg D}{\lg\lg D})$ or $2|\mathsf{CSA}| + o(n)$ bits of space, where CSA is a full-text index. Using monotone minimum perfect hash functions, we give new algorithms for document listing with frequencies and top-k document retrieval using just $|\mathsf{CSA}| + O(n \lg\lg\lg D)$ bits. We also improve current solutions that use $2|\mathsf{CSA}| + o(n)$ bits, and consider other problems such as colored range listing, top-k most important documents, and computing arbitrary frequencies.

1 Introduction and Related Work

Full-text document retrieval is the problem of, given a collection of D *documents* (i.e., general sequences over alphabet $[1,\sigma]$), concatenated into a text $T[1,n]$, preprocess T so as to later answer various queries of significance in IR. The problem has received much attention recently [16,22,24,11,8,7,4,12] as a natural evolution over plain full-text indexing (which just counts and locates all the individual occurrences of a pattern $P[1,m]$ in T) and for its applications in IR on Oriental languages, software repositories, and bioinformatic databases. As space is a serious problem in classical solutions [16,11], most of the focus has been on extending compressed full-text indexes to answer various document retrieval queries. The most studied queries, among several others, are the following.

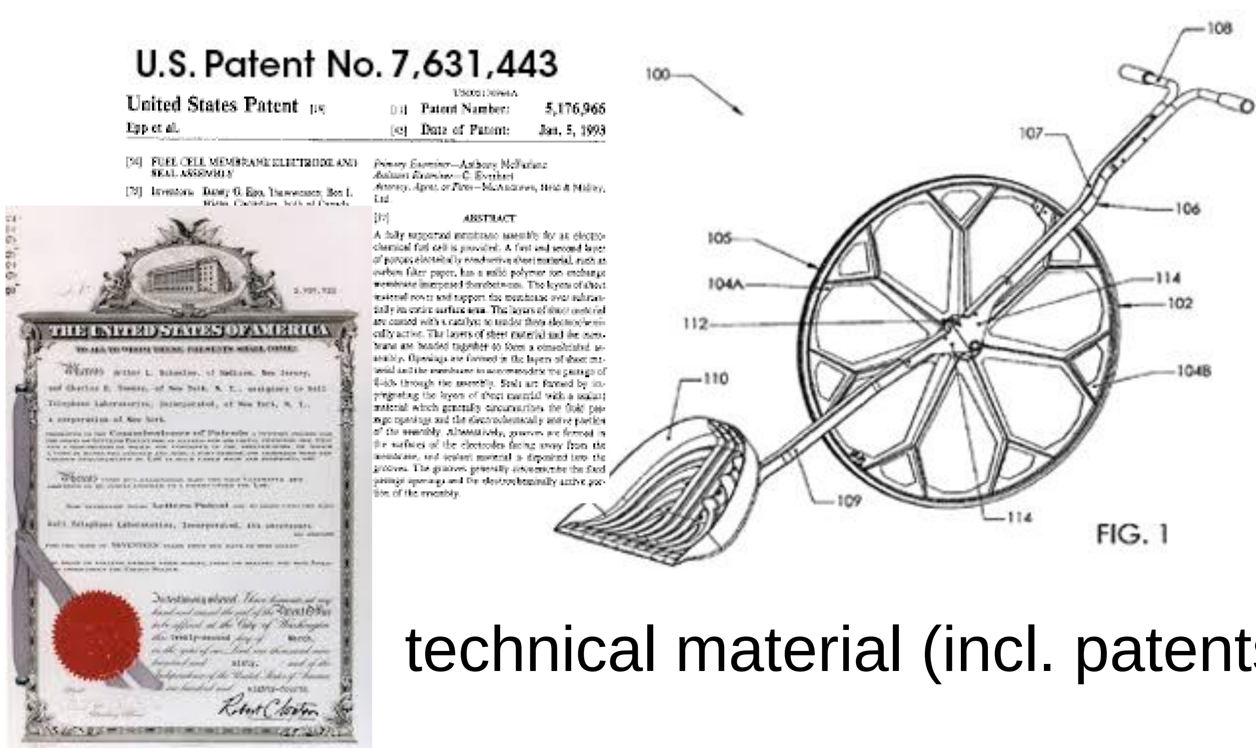
Document listing: List the distinct documents where P appears.
Document listing with frequencies: List the distinct documents where P appears, and the frequency (number of occurrences) of P in each.
Top-k retrieval: List the k documents where P appears most times.

A compressed full-text index [17] is used as the base data structure. This is usually a compressed suffix array of T (we call this structure CSA and its bit space $|\mathsf{CSA}|$). The CSA simulates the suffix array $A[1,n]$ [13], where $A[i]$ points

* Partially funded by Fondecyt Grant 1-110066, Chile. First author also partially supported by the French ANR-2010-COSI-004 MAPPI Project.

R. Grossi et al. (Eds.): SPIRE 2011, LNCS 7024, pp. 386–397, 2011.

scientific articles



technical material (incl. patents)

books

Depending on the type of material we might want a different type of summary…

Parameters of a summary

- Purpose of the summary
  - Indicative
    - resumes the main message of the original

      Thai police arrested a suspect in connection with the Bangkok bomb.

  - Informative
    - includes qualitative and/or quantitative information

      Thai police have arrested a foreign man in connection with the bomb attack on a Hindu temple in central Bangkok that killed 20 people..

  - Evaluative
    - Interprets / assesses the presented information

      Thai police arrested a suspect in connection with the Bangkok bomb, the first potential breakthrough in a case that appeared to have stalled.

Parameters of a summary

- Methodology followed to obtain the summary
  - Extractive vs. abstractive
    - select among the original wordings vs. summarize in “own words”
  - Language-specific vs. language-neutral (multilingual)
    - exploit language-specific characteristics vs. use only language-universal parameters
  - Language preserving vs. cross-language
    - provide the summary in the language of the original vs. in the language of the preference of the user
- Dimensions
  - Single document/text vs. document / text collection
- Function
  - As a pull service (query-driven) or as a push service (generic)

Criteria to assess a summary

- Informativeness
  - Ability to reconstruct the content of the original material from the summary
- Redundancy
  - Amount / Lack of information redundancy in the summary
- Cohesion / Coherence
  - Availability of a cohesive / coherent discourse structure
- Compression ratio
  - How long (compared to the original) is the summary
- …

(Very) coarse-grained time line of automatic summarization

- 50ies – 70ies
  - Experimental statistical techniques

    Exploration of individual distributional features or of a combination thereof to create sentence or paragraph extracts
- 80ies
  - AI techniques

    Exploration of partial abstracting, analysis and template generation strategies
- 90ies
  - Revival of statistical techniques; AI techniques continued
- 00ies – 10ies
  - Exploration of a great variety of ML techniques
  - Multilingual (language-independent) and multi-document summarization
  - Scaling up (Web!)

Outline of the Lecture

- Some term clarification
  - extractive summarization vs. abstractive summarization
  - single document vs. multiple document summarization
  - multilingual summarization vs. cross-language summarization
- Extractive summarization
  - Getting the right metrics: from term distribution to discourse structure
- Abstractive summarization
  - Doing it half-way: using machine translation
  - “Genuine” abstractive summarization
- Cross-language summarization
- Evaluation of summarization
- Wrap up

Clarification of some terms (1)

Extractive summarization

- Extraction of linguistic units (sentences, phrases, words) from the original material, in accordance with a combination of a number of linguistic unit relevance metrics, for inclusion into the summary
- The summary often does not claim cohesion and coherence (although some surface-oriented cohesion techniques may be used)

Abstractive summarization

- (Partial or complete) analysis of the original material down to an intermediate or a content representation
- Selection of the content from the content representation, in accordance with given content relevance metrics, for inclusion into the summary
- Use of text generation techniques to obtain a cohesive and cohesive summary

Clarification of some terms (2)

Single-document vs. multi-document summarization

- The summary reflects the material of a single document
  vs.
- The summary reflects the material of several documents (can be “reshuffled” or document-wise)

Monolingual vs. multilingual summarization

- Language-specific vs. (mainly) language-neutral techniques

Original-language vs. cross-lingual summarization

- Summary is in the same language as the original
  vs.
- Material written in one language is summarized in another language

Tasks in Summarization

- Text interpretation
  - Unit (content el./sentence/ phrase/word) identification
  - Unit analysis
- Unit selection
  - Selection of content el./sentences/… for inclusion into the summary
- Compression
  - Redundancy elimination
  - Fusion
  - Generalization
- Realization
  - Generation (abstractive summaries)
  - Ordering
  - Paraphrasing

Extractive summarization

- Text interpretation
  - Sentence splitting / chunking
- Unit selection Main challenge
  - Selection of content/**sentences**/phrases/keywords for inclusion into the summary
- Compression
  - Redundancy elimination
  - Fusion
- Realization
  - Ordering

Compute informativeness of the units in the text

Computing informativeness

Different perspectives, the same techniques…

Textual perspective	Topological perspective	Methodology perspective
❑ Looking at features that describe ▪ text surface ▪ text cohesion ▪ text coherence	❑ Looking at features that are ▪ single value-oriented ▪ vector-oriented ▪ graph-oriented	❑ Looking at the approach ▪ supervised ▪ unsupervised ▪ rule-based

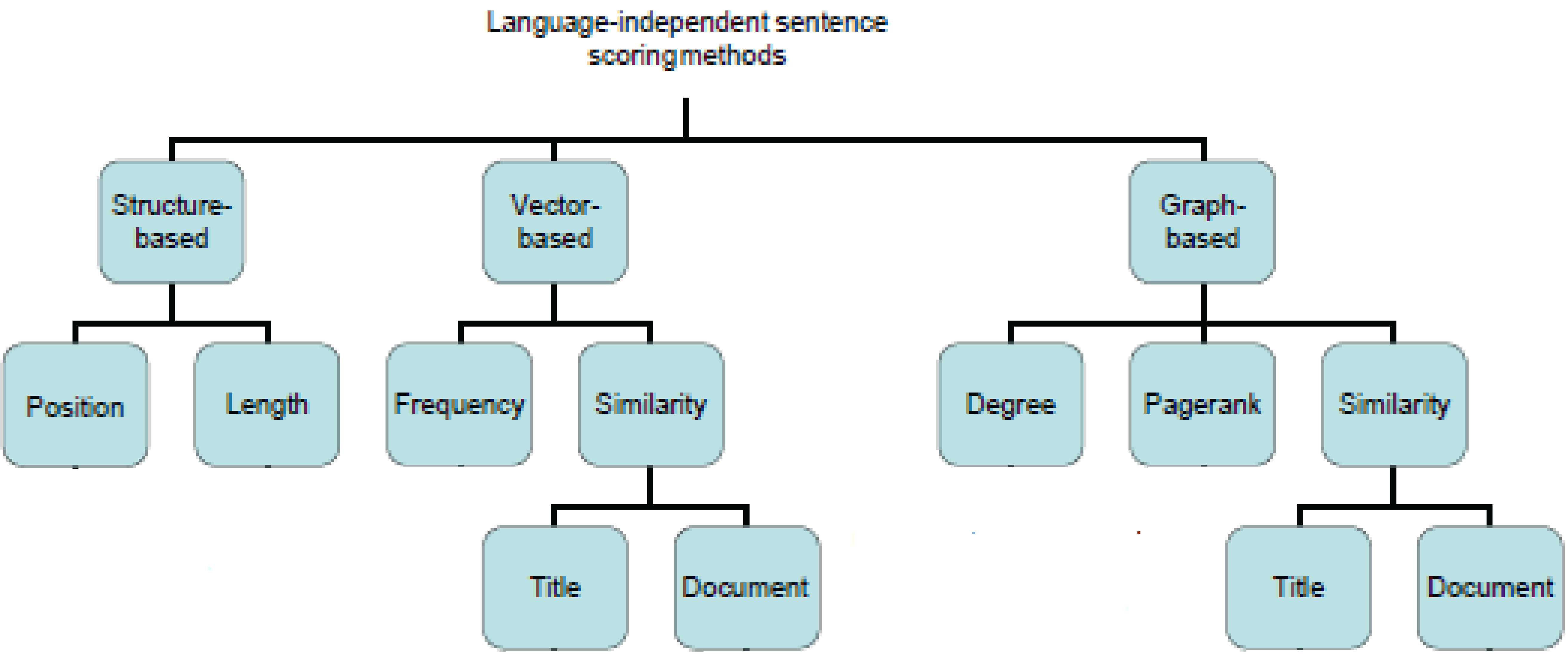
Surface-based / topic-oriented features

- Word distribution / Term frequency
- tf*idf
- Topic words / phrases
- Keywords
- Title words
- Prominent words in a reference corpus

- Proper nouns
- Acronyms
- Pronouns
- Upper case words
- Highlighted words
- Sentence position
- Sentence length
- …

Computing informativeness

Topological perspective



Source: (Litvak et al., 2010)

Surface-based / topic-oriented features

- Word distribution / Term frequency based metric
  - Words that appear significantly often in a text are likely to be indicative of the topic of the text
  - Luhn (1958): Use of term frequency to determine summary relevant sentences
    - for each word, its probability is calculated
    - a sentence is assigned a score according to the probabilities of the words it contains

Straightforward but it seems reasonable…

Human created summaries buttress this view (see, e.g., Nenkova et al., 2006)

Surface-based / topic-oriented features

- *tf*idf* based metric (used in most summarizers)
  - Accounts for the “locality” of pure word distribution
    - *tf* (term frequency): # of times a term *t* appears in the document

$$\mathrm{tf}(t,d) = 0.5 + \frac{0.5 \times \mathrm{f}(t,d)}{\max\{\mathrm{f}(t,d) : t \in d\}}$$

    - *idf* (inverse term frequency): # of documents in the reference corpus that contain *t*

$$\mathrm{idf}(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

… Can we go beyond computing word probabilities?

… and try to obtain an insight of the content?

… Topic words / Topic signatures

Surface-based / topic-oriented features

- Metric based on cue/topic words
  - Not all words are equally relevant for inclusion into the summary – depends on the topic!

How to know?

- Precompile a list of “cue” words, assigning a weight to each of them

Renewable energy

Earth, power, electricity, renewable, energy, solar, wind, water, hydrogen, biomass, field, ocean, summer, storm, mill, temperature, atmosphere, movement, heating, alternative, …

- Use a corpus from a different domain to distinguish between topic and non-topic words

Portable (and multilingual)

Surface-based / topic-oriented features

- Generic topic word determination (Lin & Hovy, 2000)

Given a word *w*,

a collection C_T of articles on a given domain/topic and

a corpus C_B on a different domain (background corpus)

If *w has a similar occurrence probability in* C_T *and in* C_B*, it is not a topic word*

H_1: $p = P(w \mid C_T) \cong P(w \mid C_B)$ *w* is not a topic word

H_2: $p_1 = P(w \mid C_T)$, $p_2 = P(w \mid C_B)$, and $p_1 >> p_2$ *w* is a topic word

Surface-based / topic-oriented features

… assuming a binomial distribution

$$f(k;n,p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

(for $k = 1,2,\ldots$; n as the length of the text and p as the probability)

… and

- F_{11}: frequency of w_i in C_T
- F_{12}: frequency of w_i in C_B
- F_{21}: frequency of w_j ($w_i \neq w_j$) in C_T
- F_{22}: frequency of w_j ($w_i \neq w_j$) in C_B

Likelihood of $L(H_1)$ respectively $L(H_2)$ is:

$$L(H_1) = f(F_{11};F_{11}+F_{12},p)\ \ f(F_{21};F_{21}+F_{22},p)$$

$$L(H_2) = f(F_{11};F_{11}+F_{12},p_1)\ \ f(F_{21};F_{21}+F_{22},p_2)$$

Confidence that w_i is a topic term can be estimated via the log-likelyhood ratio of H_1 and H_2

$$-2log = L(H_1) \,/\, L(H_2)$$

Surface-based / topic-oriented features

- Metric based on topic models
  - Learn topic models drawing on a reference corpus (RC)
    - Topic: a theme often discussed in the RC
    - Hypothesis: Each topic has its own “word profile”
    - Topic representation = word probability table
  - Approaches in the literature
    - E.g., LSA for generic topic models
    - E.g., HMMs, Bayesian networks, ... for domain-specific topic models

Surface-based / topic-oriented features

Using LSA for deriving topic models (Gong and Liu, 2001; Hachey et al., 2006, Steinberger et al., 2007)

$$A = UPV^T$$

with

- A_{ij} as the word x sentence matrix

  A_i: weighted word frequency vector of sentence i in the text

  A_j: occurrence count of word j in each of the sentences

- V^T as the projection of the column vectors j of A to the columns of V^T

- U as the projection of the row vectors i of A to the rows of U

- P as a nxn diagonal matrix with non-negative singular values as diagonal elements (sorted in descending order)

Surface-based / topic-oriented features

Using LSA for deriving topic models

(Source: Gong and Liu, 2001)

1. Decompose the document D into the set of individual sentences S, set $k = 1$.
2. Construct the words by sentences matrix A for D.
3. Perform the SVD on A to obtain the singular value matrix P and the right singular vector matrix V^T. In the singular vector space, each sentence i is represented by the column vector $\psi_{\mathrm{i}} = [v_{i1}, v_{i2}, \ldots, v_{ir}]^T$ of V^{T}
4. Select the k-th right singular vector from V^T
5. Select the sentence which has the largest index value with the k-th right singular vector and include it into the summary
6. If k reaches the predetermined number, terminate the operation; otherwise, increment k by 1 and go to step 4

Surface-based / topic-oriented features

Deriving domain-specific topic models

Learning phase

- Cluster sentences from a domain-specific corpus in accordance with their similarity (using, e.g., cosine as measure)
- Obtain a word probability table for each cluster topic

Summarization phase

- Select sentences from the text to be summarized with the highest probability for the individual topics

"Topic" may mean here "aspect" of the information (event, cause of the event, location of the event, …)

Surface-based / topic-oriented features

Using HMM for deriving domain-specific topic models (Barzilay and Lee, 2004)

- Model
  - States: Topic clusters
  - Transition probability (topic shift): sentence precedence in the corpus

$$p(s_j|s_i) = \frac{D(c_i, c_j) + \delta_2}{D(c_i) + \delta_2 m}.$$

  - Emission probability (inclusion of a sentence into the summary) bigram language model

$$p_{s_i}(w'|w) \stackrel{def}{=} \frac{f_{c_i}(ww') + \delta_1}{f_{c_i}(w) + \delta_1 |V|},$$

- To summarize
  1. Assign to each of the sentences a topic using the content model
  2. Select the sentences with the highest probability

… One feature might be not enough to assess the relevance of a sentence to a summary

… What about a combination thereof?

…first a linear combination…

Surface-based / topic-oriented features

- Linear Combination of features
  - It is often a combination of features, rather than a single feature that leads to the best sentence selection metric
  - Edmundson (1969): Experimental Combination of 4 features
    - Features: (1) title word, (2) keyword, (3) cue word, (4) position of the sentence
    - Given a sentence *S*

$$\text{Score}(S) = \alpha Score_{Title}(S) + \beta Score_{kw}(S) + \gamma Score_{cw}(S) + \delta Score_{Pos}(S)$$

    - Adjust α, β, γ, δ using gold summaries

Surface-based/topic oriented summarization

- Combination of features

Given a text $T = \{S_1, S_2, \dots S_n\}$ (with S_i being a sentence of T) and a set of features $F = \{f_1, \dots, f_m\}$ that characterize S_i

for $i = 1$ to n
    do for each $j = 1$ to m
    do v_{ij} := get_feature_value(S_i, f_j)
    enddo
        s_i := combine_features(v_{ij}), $j = 1, \dots m$
    enddo
$T' \leftarrow$ Sort(S_i, s_i) with respect to s_i in descending order
SELECT L top ranked sentences from $T' \rightarrow SUM$
DISPLAY $S_k \in SUM$ in their order in T

Surface-based / topic-oriented features

- Linear Combination of features: MEAD (Radev et al., 2004 http://www.summarization.com/mead/)

- Most important features

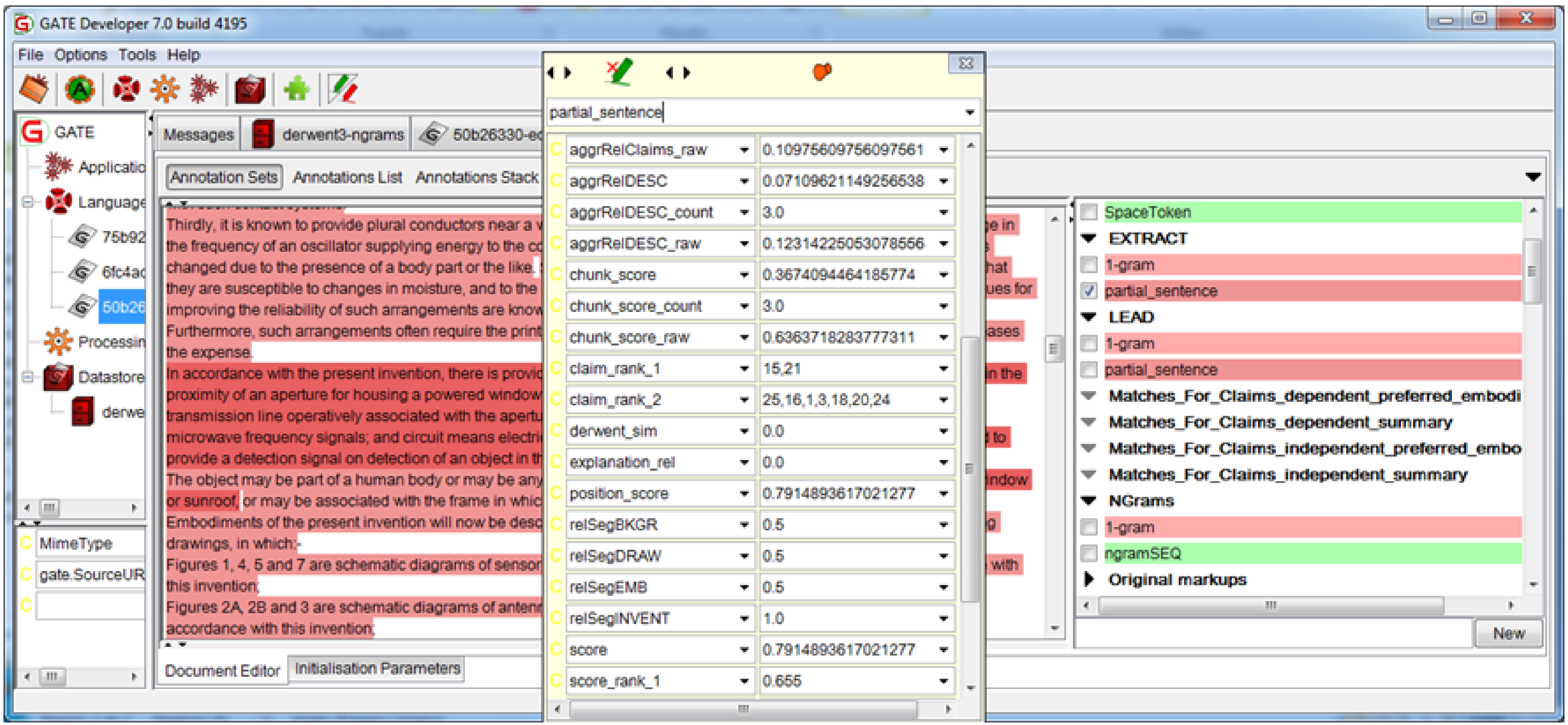
  - **Centroid**: cosine overlap with the centroid vector of the cluster (Radev et al., 2004),
  - **SimWithFirst**: cosine overlap with the first sentence in the document (or with the title, if it exists),
  - **Length**: 1 if the length of the sentence is above a given threshold and 0 otherwise,
  - **RealLength**: the length of the sentence in words,
  - **Position**: the position of the sentence in the document,
  - **QueryOverlap**: cosine overlap with a query sentence or phrase,
  - **KeyWordMatch**: full match from a list of keywords.

Surface-based / topic-oriented features

- Linear Combination of features: SUMMA (Saggion, 2008)
  - GATE-based summarization platform
    - GATE (http://gate.ac.uk)
      - Framework for development /execution of NLP applications
      - Graphical user interface supports the access to, editing of and visualisation of resources and experimentation
      - Java library (gate.jar) for programmers to implement and pack applications
    - Computes a linear score for each sentence; top ranked sentences are selected for inclusion into the summary
    - Allows for flexible introduction of new features

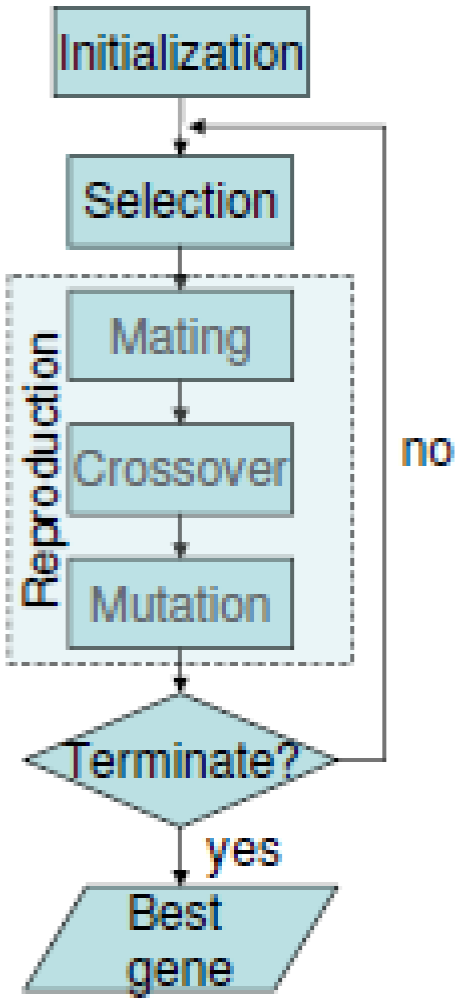
Surface-based / topic-oriented features

- Feature selection in SUMMA for patent summarization (Brügmann et al., 2015)



Surface-based / topic-oriented features

- Linear Combination of features: MUSE (Litvak et al., 2010)
  - Combination of 31 features
  - Use of a genetic algorithm



- Initialization: Random selection of N=500 genes
  gene: weighting vector $v_i = (w_1, w_2, \ldots w_L)$, with $31 \geq L$
- Selection: Use ROUGE to select the best genes
- Mating: Select randomly two parent genes v_m and v_f
- Crossover: Create a new gene from v_m and v_f with

  $$v_c = \lambda * v_m + (1 - \lambda) * v_f \qquad (\lambda = 0.5)$$

- Mutation: Mutation operator that changes (with a probability of 3%) a randomly chosen w_i by a factor within the range [–0.3, 0.3]

… One feature might be not enough to assess the relevance of a sentence to a summary

… What about a combination thereof?

…now let’s have a look at a statistical combination of features…

Surface-based / topic-oriented features

- Statistical combination of features
  - Kupiec et al., (1995): Combination of 5 features

    (1) Sentence length feature (S with |S| < N receive a score ‘0’)

    (2) Fixed phrase feature (S with a specific phase in it (e.g., “to conclude”) receive a score ‘1’)

    (3) Paragraph feature (S receives a position score if it is paragraph-initial / paragraph-final, or paragraph-medial in the first ten or last five paragraphs)

    (4) Thematic word feature (S receives a score ‘1’ if it contains a certain number of most frequent content words)

    (5) Upper case word feature (S receives a score ‘1’ if it contains a certain number of proper nouns)

Applying Bayes

$$p(s \in E \mid f_1, \ldots, f_n) = \frac{p(f_1, \ldots, f_n \mid s \in E) . p(s \in E)}{p(f_1, \ldots, f_n)}$$

By now we looked at surface-oriented (isolated or distributional features)

… What is the problem with them?

Problems with purely surface-oriented metrics

- Lack of cohesion

Joe Atkins has a contacts book that includes George Lucas, the Sultan of Brunei and the senior management teams at Apple and Maserati. For 30 years he has built Bowers & Wilkins (B&W) from a niche British speaker manufacturer into a world-beating consumer audio brand.
It is almost 50 years since John Bowers, a communications engineer in the second world war and classical music nut, opened a modest shop selling audio kit in Worthing, West Sussex. Now the company that bears his name is a global electronics brand employing 1,100 people with an annual turnover of £120m.

For 30 years he has built Bowers & Wilkins (B&W) from a niche British speaker manufacturer into a world-beating consumer audio brand.
Now the company that bears his name is a global electronics brand employing 1,100 people with an annual turnover of £120m.

Problems with purely surface-oriented metrics

- Lack of coherence

Joe Atkins has a contacts book that includes George Lucas, the Sultan of Brunei and the senior management teams at Apple and Maserati. For 30 years he has built Bowers & Wilkins (B&W) from a niche British speaker manufacturer into a world-beating consumer audio brand.
It is almost 50 years since John Bowers, a communications engineer in the second world war and classical music nut, opened a modest shop selling audio kit in Worthing, West Sussex. Now the company that bears his name is a global electronics brand employing 1,100 people with an annual turnover of £120m.

[Joe Atkins has a contacts book that includes George Lucas, the Sultan of Brunei and the senior management teams at Apple and Maserati.] For 30 years he has built Bowers & Wilkins (B&W) from a niche British speaker manufacturer into a world-beating consumer audio brand.
Now the company that bears his name is a global electronics brand employing 1,100 people with an annual turnover of £120m.

… Use features that ensure cohesion and coherence to improve …

… For instance, lexical chains or discourse structure …

* …

Cohesion-oriented features

Lexical chain

A word sequence in a text where the words are related by a semantic similarity relation

Julian Assange has said he advised the NSA whistleblower Edward Snowden against seeking asylum in Latin America because he could have been kidnapped and possibly killed there. The WikiLeaks editor-in-chief said he told Snowden to ignore concerns about the “negative PR consequences” of sheltering in Russia because it was one of the few places in the world where the CIA’s influence did not reach. In a wide-ranging interview with the Times, Assange also said he feared he would be assassinated if he was ever able to leave the Ecuadorian embassy in London, where he sought asylum in 2012 to avoid extradition. He accused US officials of breaking the law in their pursuit of him and his whistle blowing organisation, and in subjecting his connections to a campaign of harassment.

Cohesion-oriented summarization

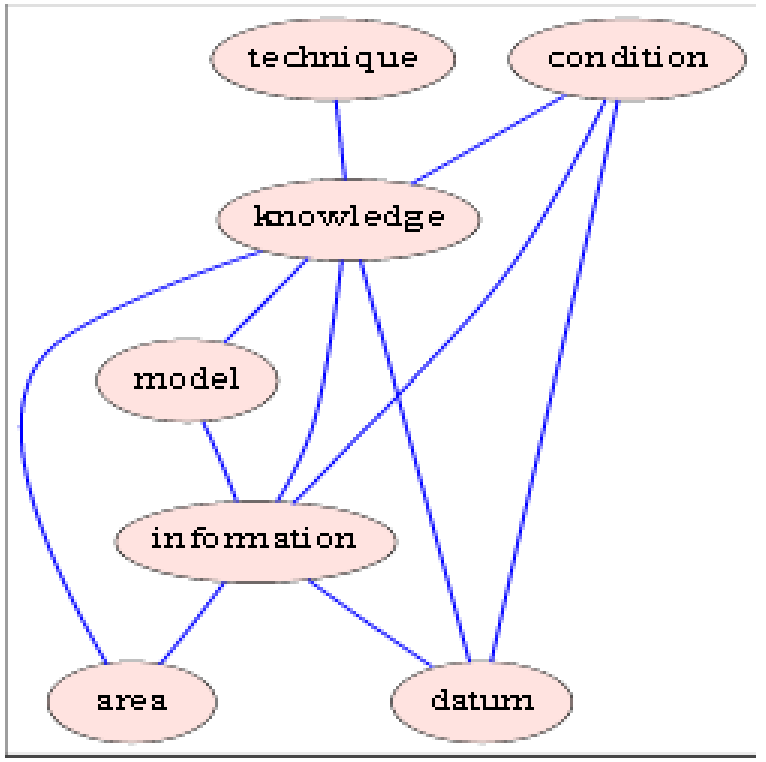
Lexical chain-based summarization (Barzilay and Elhadad, 1997; Silber and McCoy, 2002; Doran et al., 2004; Ye et al., 2007; Berker and Güngör, 2012)

- A chain captures the distribution of a concept through the text
  - Often, WordNet relations are used to determine this distribution
    - synonymy, hyponymy and hypernymy

synonymy: *concern* vs. *company*

hyponymy: *SME vs. company*

hypernymy: *institution vs. institute*



Cohesion-oriented summarization

- Lexical chain-based summarization
  - Barzilay and Elhadad (1997)
    - Chain Scoring function

$$\text{Score(Chain)} = \text{Length} * \text{HomogeneityIndex}$$

    - Alternative sentence selection heuristics (applied to chains with a score above the threshold)

A. For each chain, choose the sentence that contains the first appearance of a chain member in the text.

B. For each chain, choose the sentence that contains the first appearance of a representative chain member in the text.

C. For each chain, find the text unit where the chain is highly concentrated. Extract the sentence with the first chain appearance in this central unit.

$$\text{Centrality: } \frac{\text{\# chain members occurrences in a segment}}{\text{\# nouns in the segment}}$$

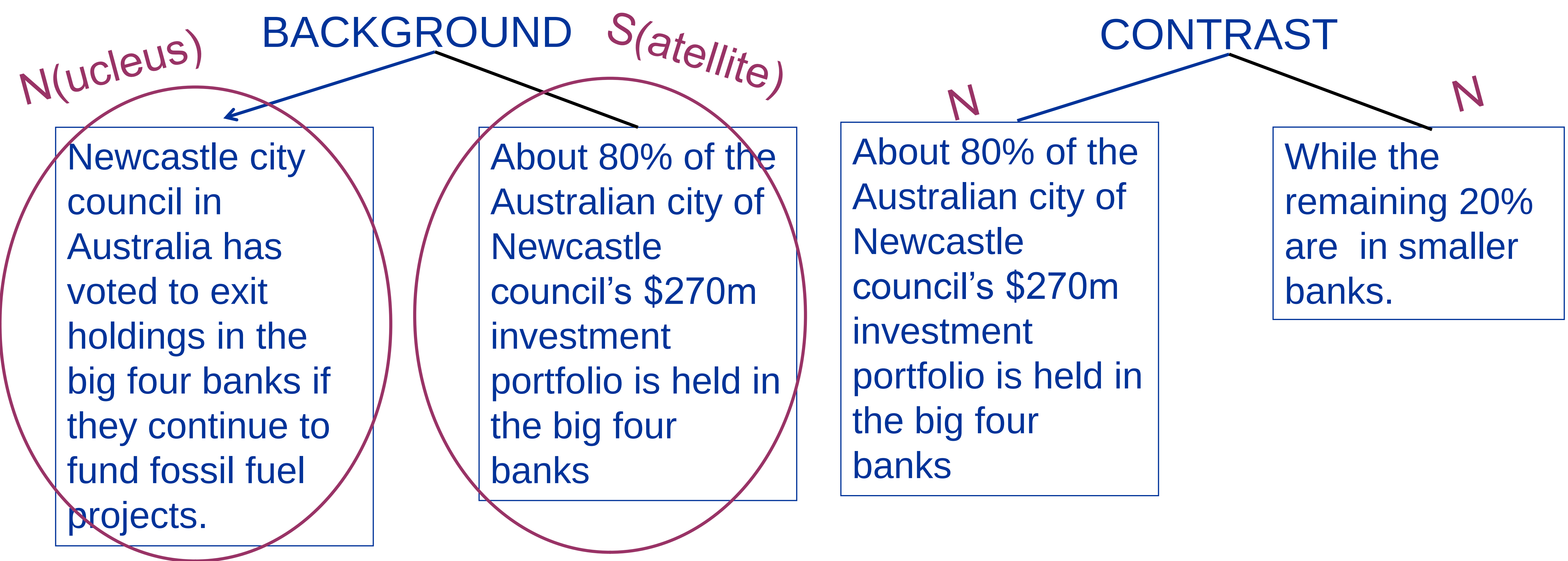
Text coherence / structure-oriented summarization

- Discourse structure-based summarization (Marcu, 1997, 2001; Bouayad-Agha et al., 2009; Louis et al., 2010; Wang et al., 2015)

[Newcastle city council in Australia has voted to exit holdings in the big four banks if they continue to fund fossil fuel projects.]$_1$ [About 80% of the Australian city of Newcastle council's $270m investment portfolio is held in the big four banks,]$_2$ [mostly through term deposits]$_3$. [Those investments are spread evenly across the big four.]$_4$ [But after the council passed a motion on Tuesday, six votes to five, it will dump holdings in the banks for more "environmentally and socially responsible" institutions when deposits come up for renewal.]$_5$ [This will be done]$_6$ [only if the rate of return is comparable with the council's holdings in the big four and the council's credit rating criteria is met.]$_7$ [Minerals Council of Australia coal chief Greg Evans said the "stock in trade of the divestment activists is to get public attention through empty symbolism.]$_8$

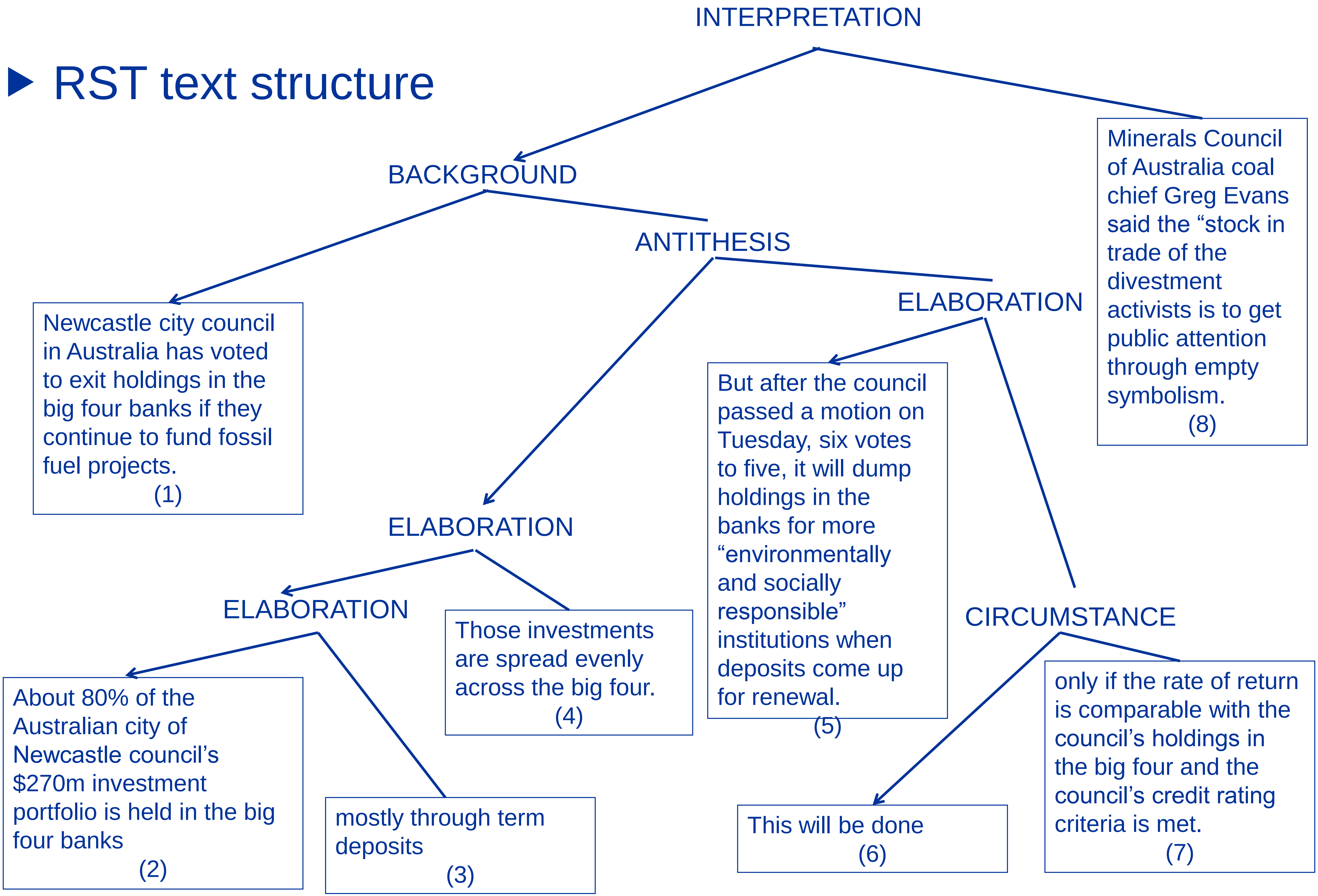
Text coherence / structure-oriented summarization

- Discourse structure
  - Clauses in a text are related by “rhetorical” or “discourse” relations
  - Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) is often used to model discourse in NLP
    - Restricted set of asymmetric and symmetric relations



Text coherence / structure-oriented summarization

- RST text structure

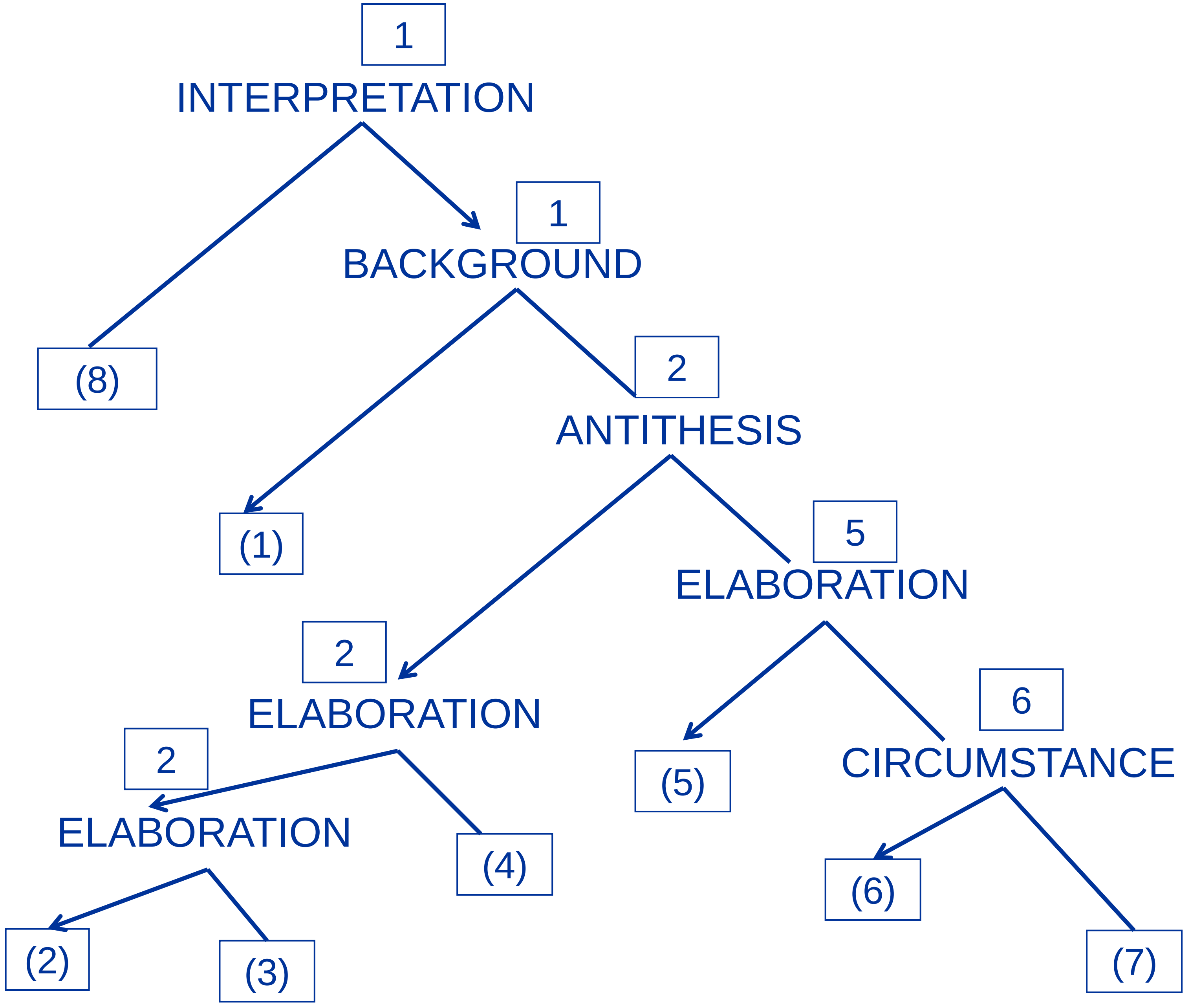


Text coherence / structure-oriented summarization

- Discourse-structure based summarization
  - (Marcu, 1997)
    - Propagate a discourse element upwards, summing up its weight along the nucleus links of an RST tree
    - Order the discourse elements with respect to their obtained weights
    - Select for the summary the highest ranked elements

Text coherence / structure-oriented summarization

- (Marcu, 1997)



2 > 1 > 5, 6 > 3, 4, 7, 8

Text coherence / structure-oriented summarization

- (Bouayad-Agha et al., 2009)

  Discourse structure-based summarization of patent claims

  - Pruning of the RST tree using three types of criteria
    - Depth of the tree: Prune the tree branches of depth d > n
    - Specific discourse markers: Prune the tree branches introduced by specific discourse markers (*when*, *by*, *for*, …)
    - Discourse relation relevance hierarchy: Prune branches of relations that are less relevant to the summary
      - PURPOSE > CAUSE > MEANS > ELABORATION-OBJ > ELABORATION-LOC
  - Complementary pruning of the syntactic dependency trees of discourse elements kept for the summary
  - Syntactic generation to ensure cohesion

Text coherence / structure-oriented summarization

- (Bouayad-Agha et al., 2009), Example

An optical disk drive comprising: a laser light source for emitting a laser beam; an optical system for conversing the laser beam from the laser light source on a signal plane of optical disk on which signal marks are formed and for transmitting the light reflected from the signal plane; one or more optical components, arranged in the optical path between the laser light source and the optical disk, for making the distribution of the laser beam converged by the conversing means located on a ring belt just after the passage of an aperture plane of the optical system; a detection means for detecting the light reflected from the optical disk; and a signal processing circuit for generating a secondary differential signal by differentiating the signals detected by the detection means and for detecting the edge positions of the signal marks by comparing the secondary differential signal with a detection level.

An optical disk drive comprises a laser light source, an optical system, a detection means, and a signal processing circuit. The laser light source emits a laser beam. The optical system converses the laser beam from the laser light source on a signal plane of optical disk. On the latter, signal marks are formed. The detection means detects the light reflected from the optical disk. The signal processing circuit generates a secondary differential signal. To do so, it differentiates the signals detected by the detection means.

With discourse and syntactic structures, we moved from vector representations to graph representations…

… But discourse and syntactic structure parsing are language-specific…

… Let’s have a look at a multilingual graph model …

* …

Graph-oriented summarization

Graph model-based summarization (Erkan and Radev, 2004; Mihalcea and Tarau, 2004; Leskovec et al., 2005; Wei et al., 2010; Ge, 2011; …)

- Nodes
  - sentences
  - discourse units
- Edges
  - links between similar sentences
  - links between syntactically related units
- Sentence similarity metrics
  - E.g., cosine similarity between TF*IDF weighted vector representations

Graph-oriented summarization

❑ LexRank (Erkan and Radev, 2004)

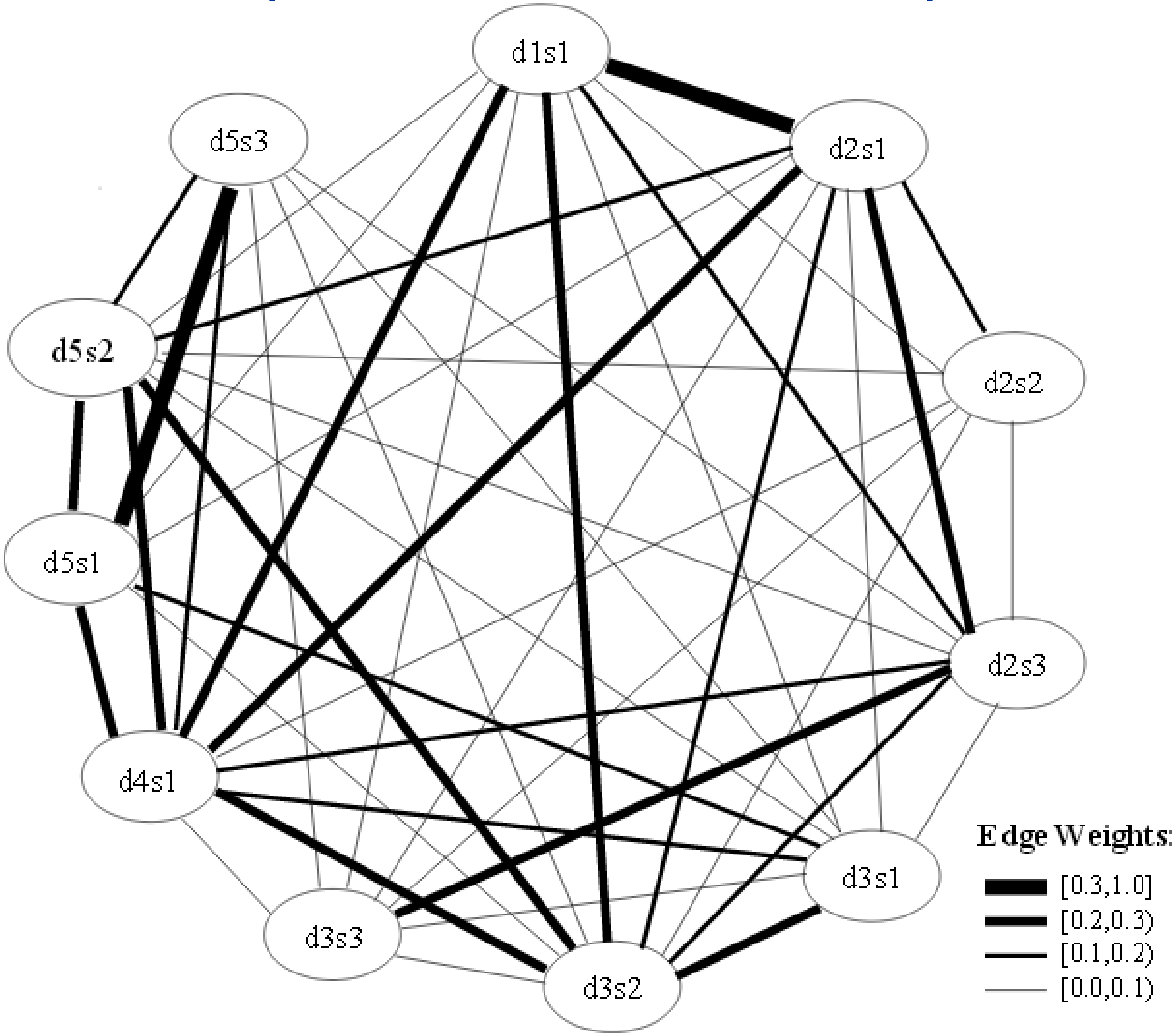
SNo	ID	Text
1	d1s1	Iraqi Vice President Taha Yassin Ramadan announced today, Sunday, that Iraq refuses to back down from its decision to stop cooperating with disarmament inspectors before its demands are met.
2	d2s1	Iraqi Vice president Taha Yassin Ramadan announced today, Thursday, that Iraq rejects 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 resume its cooperation with the to a military operation.
5	d3s1	The Russian Foreign Minister, against using force against Iraq him, seven years of difficult dip the regional situation in the are
6	d3s2	Ivanov contended that carrying to cooperate with the United N tremendous work achieved by t seven years and will complicate
7	d3s3	Nevertheless, Ivanov stressed t

Sentence similarity

	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

Graph-oriented summarization

- LexRank (Erkan and Radev, 2004)



That much on extractive single document summarization…

Let’s move now to extractive multi-document summarization

* …

Multi-document summarization

- Starting from a collection of related documents / texts, produce a coherent, redundancy-free summary

Thousands of Malaysians took to the streets of Kuala Lumpur calling for the resignation of the prime minister, Najib Razak, who is battling the fallout from a financial scandal. The government condemned the weekend rallies as illegal and blocked the website of the organisers, a coalition of non-governmental organisations.
Thousands of Malaysians have made their way back to the centre of the capital, assembling again in an illegal demonstration to call for the resignation of the prime minister, Najib Razak, who is battling the fallout from a financial scandal. Some people in the 34-hour protest had slept in the streets overnight in an unusually calm demonstration of public outrage by the group Bersih, a coalition of non-governmental organisations, which means “clean” in Malay.
Thousands of demonstrators have continued their protests in Kuala Lumpur for a second day to demand the resignation of Malaysian Prime Minister Najib Razak over a financial scandal. The crowd of yellow-clad protesters, who slept on the streets near the city's Independence Square, woke on Sunday to mass exercises and a resumption of the previous day's peaceful demonstration.

Multi-document summarization

- Centroid-based multi-document summarization (Radev et al., 2004; Saggion and Gaizauskas, 2004; Nedunchelian, 2008)
  - Create a centroid of the collection of texts to be summarized
    - Centroid: vector of statistically significant words across the texts
  - Calculate the centroid score for each sentence in the texts
    - Centroid score: similarity of the sentence with the centroid
  - Combine the centroid score with the scores of structural and vectorial features of each text
  - Detect and eliminate redundancy between highly scored sentences applying a similarity metric

Multi-document summarization

- Graph-based multi-document summarization (Christensen et al., 2013)
  - Construct a multi-document discourse graph
    - Nodes: sentences
    - Edges: discourse relations between sentences forming a coherent segment
      - $s_i \rightarrow s_j$: s_j can appear after s_i in the summary
    - Use of textual cues (discourse markers, derivational morphology, …) to identify relations; no relation labeling
      - discourse markers: *but, when, nonetheless, …*
      - derivational morphology: *… attack… the attackers*
      - coreferences: *… an attack… the attack*
      - lexical chains: *… murderer… the criminal*
      - …

Multi-document summarization

- Graph-based multi-document summarization (Christensen et al., 2013), contd.
  - Construct a multi-document discourse graph
    - …
    - Weight the edges of the constructed graph
      - Weight: number of indicators leading to the relation
  - Navigate in the graph
    - Return the summary of length |X| that maximizes an objective function over coherence, salience and redundancy.

$$
\begin{array}{rc}
\textbf{maximize:} & F(x) \triangleq Sal(X) + \alpha Coh(X) - \beta |X| \\
s.t. & \sum_{i=1..|X|} len(x_i) < B \\
 & \forall x_i, x_j \in X : \text{redundant}(x_i, x_j) = 0
\end{array}
$$

Multi-document summarization

- Graph-based multi-document summarization (Christensen et al., 2013), contd.
  - Coherence, salience and redundancy to obtain the score of a summary
    - coherence $$Coh(X) = \sum_{i=1..|X|-1} w_{G+}(x_i, x_{i+1}) + \lambda w_{G-}(x_i, x_{i+1})$$
      with '*wG+*' as positive edge weight and '*wG-*' as negative edge weight
    - salience $Sal(X) = \sum_i Sal(x_i)$
      - salience of x_i obtained using a linear regression classifier over the DUC '03 data set and a number of surface features
    - redundancy
      - convert sentences into predicate-argument tuples
      - mark as redundant tuples whose predicates or at least one of the arguments are synonymous and have the same 24h time stamp

weight	feature
-0.037	position in document
0.033	from first three sentences
-0.035	number of people mentions
0.111	contains money
0.038	sentence length $>$ 20
0.137	length of sentence
0.109	#sentences verbs appear in (any form)
0.349	#sentences common nouns appear in
0.355	#sentences proper nouns appear in

Summary
extractive single and multiple document summarization

- Extractive summarization
  - exploits structural, distributional, semantic co-occurrence, syntactic, and discourse characteristics of a text or text collection
    - different features may be combined to form a unique relevance score of a sentence
  - captures these characteristics in terms of single value functions, vectors, and graphs
  - implements a large variety of symbolic and ML-based techniques

Summary extractive single and multiple document summarization

… what we could not look at

- Summarization as a binary classification task (include a sentence into the summary vs. do not)
  - SVMs (Chali et al., 2009; Xie and Liu, 2010)
  - Conditional Random Fields (Galley, 2006; Shen et al., 2007)
  - Neural networks (Kaikhah, 2004)
  - Regression (Ulrich, 2009)
  - …
- Summarization as an optimization problem
  - Integer Linear Programming (Galanis et al., 2012)
  - Dynamic Programming (McDonald, 2007)
  - …
- …

Let’s move now to abstractive summarization

* …

A short glance at abstractive summarization

- Genuine abstractive summarization
  - “Understand” a text and summarize its content
- Template-based abstractive summarization
  - Look for chunks of the text that are understood and are relevant and summarize its content
- Syntax-oriented abstractive summarization
  - Parse sentences selected using a relevance metric, fuse similar syntactic trees and generate a summary out of the fused trees
- Hybrid abstractive summarization

A short glance at abstractive summarization

- Genuine abstractive summarization (Saggion and Lapalme, 2002; Moawad and Aref, 2012)
  - Map the sentences of the text(s) to be summarized onto a semantic representation
    - Deep parsing
    - Direct projection
  - Select the most relevant content from the obtained semantic representation
    - Graph navigation
    - Graph reduction
  - Generate from the selected content a cohesive and coherent summary using automatic text generation techniques

Some words on cross-language summarization

What about the language of the summary?

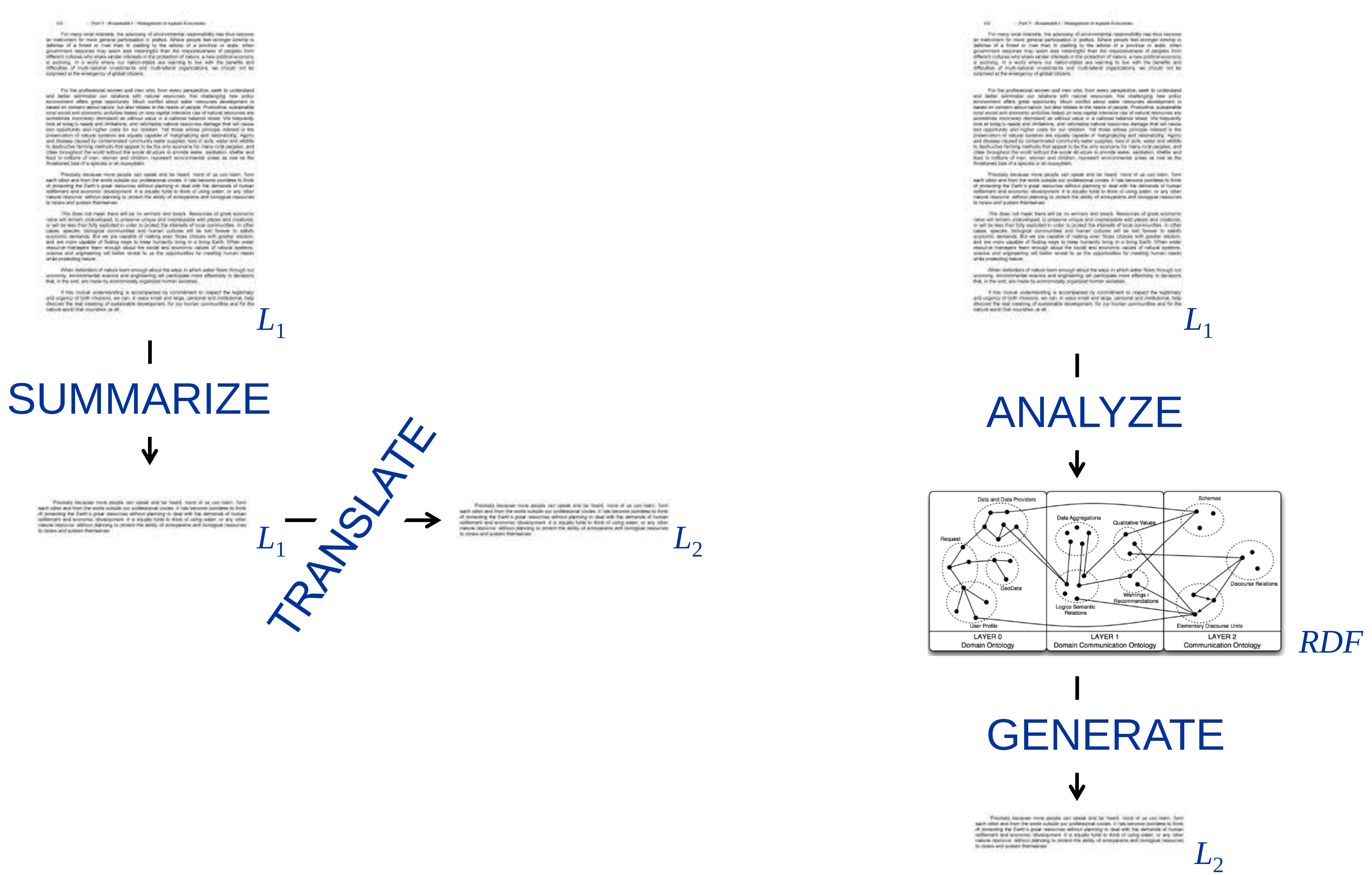
- Relevant material on a given topic often comes in different languages
- The user often wishes to read a summary in the language of their preference

While most of the discussed extractive techniques are per se multilingual, i.e., language-independent, they do not solve the issue

… what we need is cross-language summarization

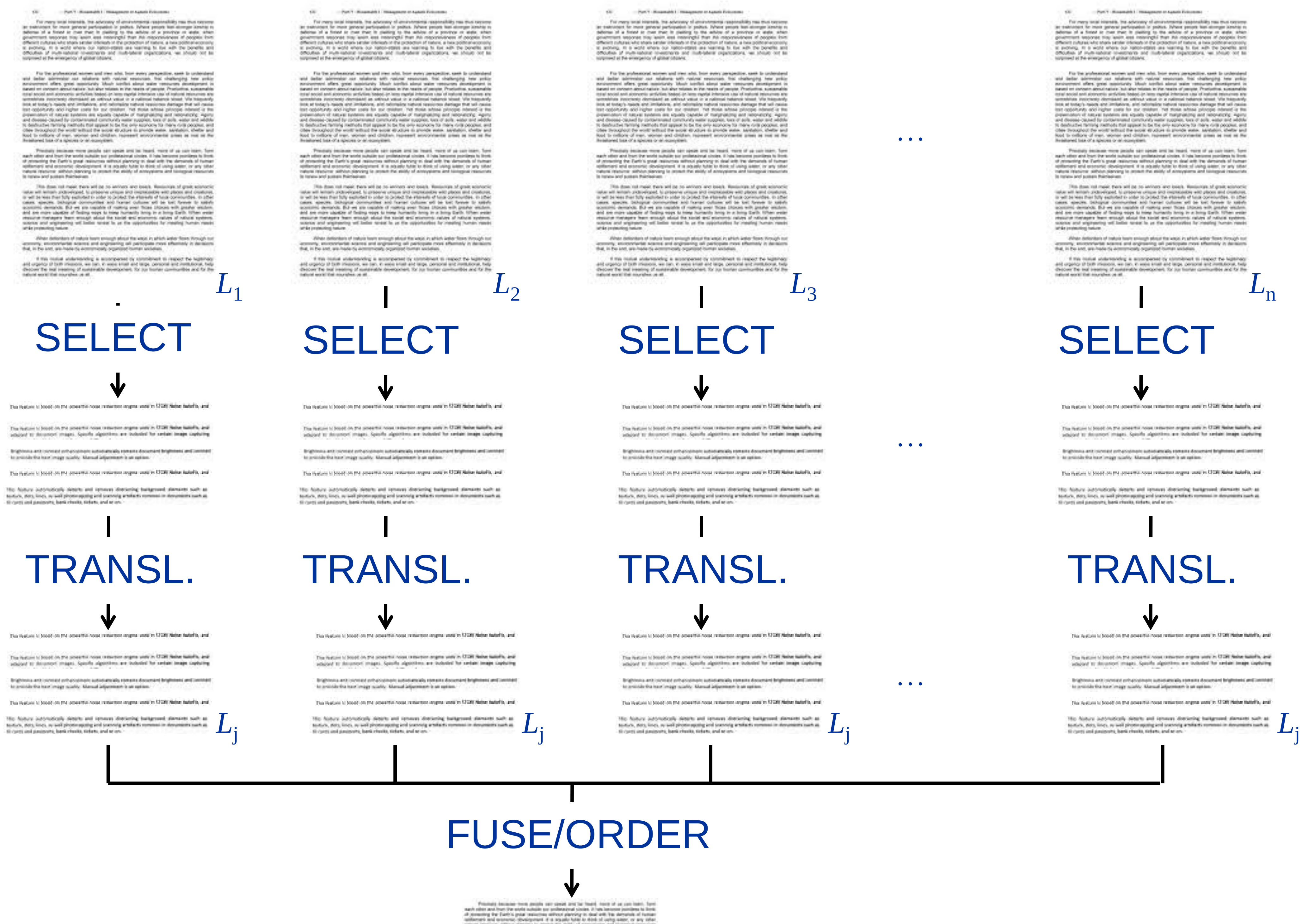
Cross-language summarization

- Given a text in the language L_1, produce a summary in L_2



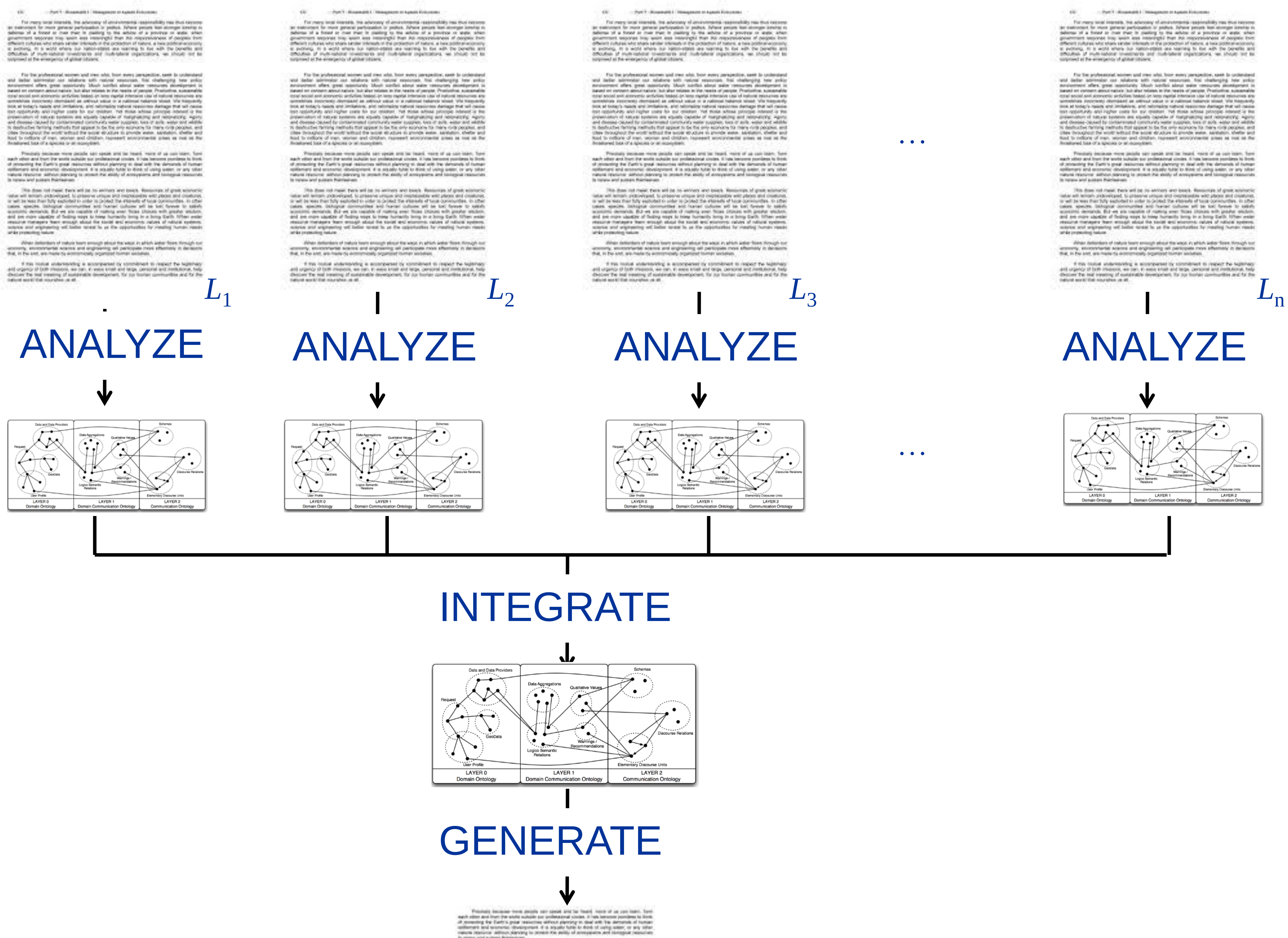
Cross-language summarization

- Given texts in the languages L_1, L_2, L_3, … produce a summary in L_j



Cross-language summarization

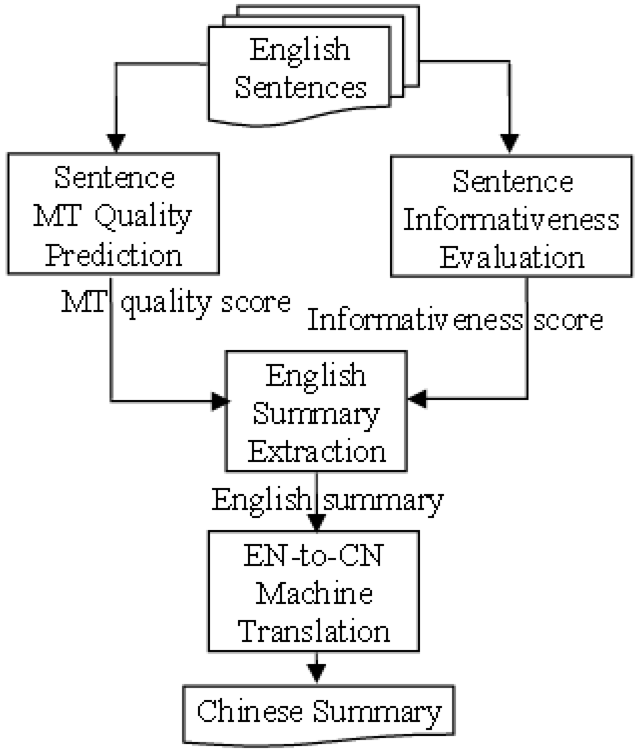
- Given texts in the languages L_1, L_2, L_3, … produce a summary in L_j



Cross-language oriented summarization

- Machine translation based CLS (Wan et al., 2010; Boudin et al., 2011)
  - ❑ Predict MT Quality of source-to-target sentences translation
    - Classifier-oriented
  - ❑ Apply MT
    - Google Translate
    - MOSES
    - …
  - ❑ Score original sentences with respect to their informativeness (taking MT-quality into account)
  - ❑ Generate the target language summary

(Wan et al, 2010)



Cross-language oriented summarization

- Deep parsing and generation-based CLS
  - Parse the source language sentences with a deep dependency parser
    - Semantic Role Labeler
    - Deep-syntactic parser
    - …
  - Map the parse trees/graphs onto a language-neutral semantic (e.g, RDF-based) representation
    - Usually: language-specific aligned ontologies or multiple language labeled ontologies
  - Run content selection on the semantic representation
  - Generate the summary in the target language

How can the quality of a summary be evaluated?

Summarization quality evaluation

- Human judges (assessors) evaluation
  - Ad hoc assessment of each generated summary
  - Contrastive assessment of generated summaries against corresponding gold standard
    - Pyramid
- Automatic evaluation
  - Use of an evaluation metric to calculate a similarity score between a generated summary and its corresponding gold standard
    - precision
    - recall
    - F-score
    - ROUGE
    - MeMoG
    - NPowER

Expert evaluation of summaries

- Ad hoc evaluation of each summary

  Human judges rate the quality of a summary with respect to several dimensions, e.g., on the 5-value Likert scale

  - Relevance of the presented information
  - Omission of relevant information
  - Cohesion and coherence of the summary (if applicable)
  - …

✓ Allows for the judgment of the meaning accuracy of the summary

✓ Allows for relative assessment across several summaries

✗ Risk of judgments of different rigor across summaries

✗ Hard to revise a taken decision

✗ Very subjective

✗ Expensive and slow

Expert evaluation of summaries

- Contrastive assessment of generated summaries against corresponding gold standard

  Human experts create ground truth summaries

  - Pyramid evaluation strategy (Nenkova et al., 2007)
    - Draws upon a set of documents that are to be summarized
    - Attempts to generalize over different surface renderings of the same or similar meaning
    - Based on Semantic Content Units (SCUs), which are weighted with respect to their relevance for inclusion into the summary
      - SCU: A semantically motivated, subsentential (max. clausal) unit

Expert evaluation of summaries

Pyramid-based evaluation

- Summary Content Units

S1 Officials had initially said that as many as 50 corpses, thought to be migrants, were in the abandoned vehicle, but they revised the death toll after working through the night.

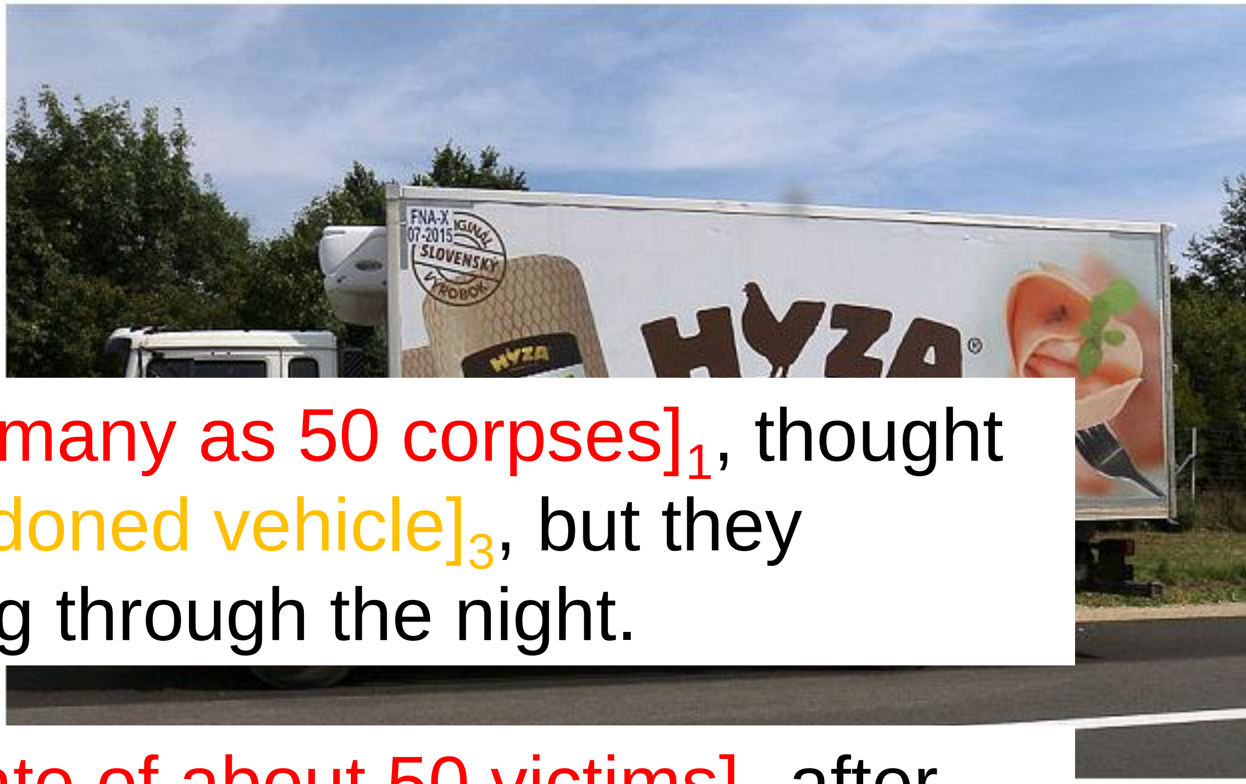
S2 Death toll rises from initial estimate of about 50 victims after badly decomposed remains are found on Austrian motorway

S3 Austria freezer truck death toll rises as more than 70 refugees found dead after becoming trapped. The grim discovery was made by workers and police have said identifying the number of victims has been difficult as they were so badly decomposed

S4 The Austrian government confirms a higher than expected death toll after the discovery of the truck on a parking strip off the highway in Burgenland state.

Expert evaluation of summaries

Pyramid-based evaluation



- Summary Content Units

S1 [Officials]$_0$ had [initially said that as many as 50 corpses]$_1$, thought to [[be migrants]$_2$, were in the abandoned vehicle]$_3$, but they [revised the death toll]$_4$ after working through the night.

S2 [Death toll rises]$_4$ from [initial estimate of about 50 victims]$_1$ after [badly decomposed remains [are found]$_5$ [on Austrian motorway]$_6$

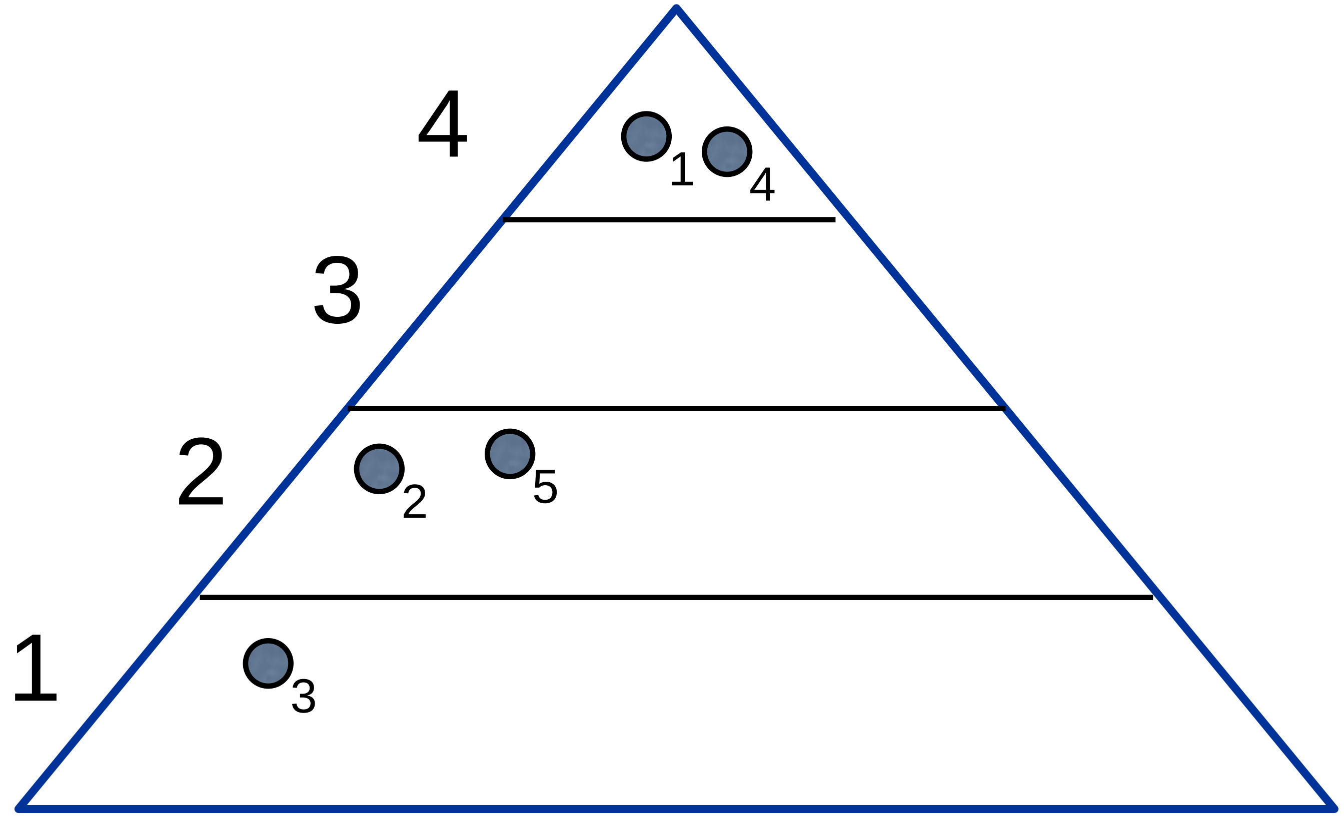
S3 [Austria freezer truck]$_7$ [death toll rises]$_4$ as [more than 70 [refugees]$_2$ found dead]$_1$ after becoming trapped. The grim discovery was made by workers and police have said [identifying the number of victims]$_1$ has been difficult as they [were so badly decomposed]$_5$

S4 [The Austrian government]$_0$ confirms a [higher than expected death toll]$_4$ after the discovery of the [truck]$_7$ on a parking strip [off the highway in Burgenland state]$_6$.

Expert evaluation of summaries

Pyramid-based evaluation

- SCUs with the same relevance weight form a “tier”
  - Relevance of an SCU is reflected by the number of its occurrences across human generated gold summaries



Expert evaluation of summaries

Pyramid-based evaluation

- An optimal summary contains all SCUs from tier *n-i* before any SCU from the tier *n*-(*i*-*1*) is expressed

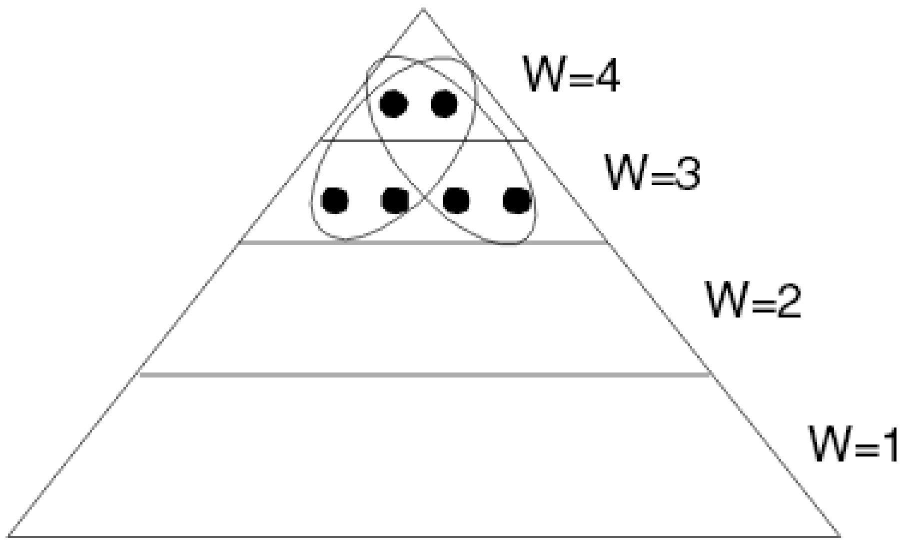


Fig. 1. Two of six optimal summaries with 4 SCUs.

(from Nenkova et al., 2007)

- The informartiveness of a summary is the ratio of the sum of the weights of its SCUs to the weight of an optimal summary with the same number of SCUs

Expert evaluation of summaries, Pyramid-based evaluation

- Total weight of a summary $$D = \sum_{i=j}^{n} i \times D_i$$

  with i as the number of tier, j as the "lowest tier (usually 1) and D_i as the score at tier i

- Weight of an optimal summary $$Max = \sum_{i=j+1}^{n} |T_i| + j \times D_j$$

- Informativeness of a summary $$D \,/\, Max$$

Precision and Recall

for evaluation of the quality of summaries

.

- Selection of the most important sentences of a text by a human expert (S_H)

    - Precision:

$$\frac{S_A \cap S_H}{S_A}$$

    - Recall:

$$\frac{S_A \cap S_H}{S_H}$$

S_A: automatic summary

ROUGE (Lin, 2004)

"Recall-Oriented Understudy for Gisting Evaluation"

.

N-gram co-occurrence statistics between generated and gold standard summaries.

- A collection of different evaluation metrics
  - ROUGE-N
  - ROUGE-L
  - ROUGE-N_{Multi}
  - ROUGE-W
  - ROUGE-S
  - …
- The metrics assess the quality of a generated summary by measuring its similarity with an ideal summary
- De facto standard in text summarization
- Has been shown to correlate with human judgments

ROUGE-N

- N-gram recall based

$$\mathrm{ROUGE\text{-}n} = \frac{\sum_{S \in \{Refs\}} \sum_{n\text{-}gram \in S} count_{match}(n\text{-}gram)}{\sum_{S \in \{Refs\}} \sum_{n\text{-}gram \in S} count(n\text{-}gram)}$$

S1. police killed the gunman
S2. police kill <u>the gunman</u>
S3. <u>the gunman</u> kill police

with Rouge-2 S2 and S3 equal

ROUGE-L

- Longest common subsequence

$$R_{lcs} = \frac{LCS(X,Y)}{m}$$

$$P_{lcs} = \frac{LCS(X,Y)}{n}$$

$$F_{lcs} = \frac{(1+\beta^2)R_{lcs}P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}}$$

(with summaries *X* of length *m* and *Y* of length *n*)

What about *R* and *P* of S_2 and S_3 from the previous slide?

ROUGE-i

- ROUGE-N_{Multi}

  extends ROUGE-N to multiple references

- ROUGE-W

  weighted longest common subsequence, favours consecutive matches

- ROUGE-S

  Skip-bigram recall metric

- ROUGE-SU

  adds unigrams to ROUGE-S

- ….

In total, 17 metrics have been presented at DUC 2004

Summarization in IR

To conclude…

- Summarization is an important add-on to any IR system
- Especially multilingual and cross-lingual multi-document summarization is of relevance
- With the progress in
  - stochastic deep parsing and text generation
  - ontological representations and their processing

“genuine” abstractive summarization is about to gain a more important say

Bibliography

- Selected Tutorials on Text Summarization
  - Nenkova A., S. Maskey, and Y. Liu. “Automatic Summarization”. Tutorial at the Annual Conference of the Association of Computational Linguistics (ACL), 2011.
  - Saggion, H. “Introduction to Text Summarization and Other Information Access Technologies”. Tutorial at Biannual Language Resources and Evaluation Conference (LREC) 2008.
  - Radev, D. “Text Summarization”. Tutorial at ACM SIGIR, 2004
- Selected major shared tasks related to (Multilingual) Text Summarization
  - Summarization Tracks at Text Analysis Conferences (TAC), organized by NIST (2008 – 2011, 2014). http://www.nist.gov/tac/tracks
  - Multilingual Summarization Challenges
    - MultiLing 2011 http://users.iit.demokritos.gr/~ggianna/TAC2011/MultiLing2011.html
    - MultiLing 2013 http://multiling.iit.demokritos.gr/pages/view/662/multiling-2013
    - MultiLing 2015 http://multiling.iit.demokritos.gr/pages/view/1516/multiling-2015

Bibliography

– partially based on (Nenkova et al., 2011) –

- Barzilay, R. and L. Lee Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization. In Proceedings of HLT-NAACL 2004, 2004.
- Barzilay, R. and M. Elhadad. 2009. Text summarizations with lexical chains. In: I. Mani and M. Maybury (eds.): Advances in Automatic Text Summarization.
- Barzilay, R. and M. Elhadad. 1997. Using Lexical Chains for Text Summarization. In Proc. of ACL Workshop on Intelligent Scalable Text Summarization.
- Berker, M. and T. Güngör. 2012. Using Genetic Algorithms with Lexical Chains for Automatic Text Summarization. ICAART (1) 2012: 595-600.
- Bouayad-Agha, N., G. Casamayor, G. Ferraro, S. Mille, V. Vidal, and L. Wanner. 2009. Improving the Comprehension of Legal Documentation: The Case of Patent Claims. In Proc. of the International Conference on Artificial Intelligence in Law. Barcelona.
- Boudin, F. S. Huet, and J.-M. Torres-Moreno. 2011. A Graph-based Approach to Cross-language Multi-document Summarization. Polibits 43:113-118.
- Brügmann, S., N. Bouayad-Agha, A. Burga, S. Carrascosa, A. Ciaramella, M. Ciaramella, J. Codina-Filba, E. Escorsa, A. Judea, S. Mille, A. Müller, H. Saggion, P. Ziering, H. Schütze, and L. Wanner. 2015. Towards content-oriented patent document processing: Intelligent patent analysis and summarization. World Patent Information
- Carbonell, J. . http://dx.doi.org/10.1016/j.wpi.2014.10.003 and J. Goldstein. 1998. The Use of MMR, Diversity-Based reranking for Reordering Documents and Producing Summaries. Proc. of the 21st Annual International ACM SIGIR.
- Chali, Y., S.A. Hasan, and SR. Joty. 2009. A SVM-Based Ensemble Approach to Multi-Document Summarization. Proc. Of. Canadian AI Conference.
- Christensen, J., Mausam, S. Soderland, and O. Etzionin. 2013. Towards coherend multi-document summarization. Proc.of NAACL-HLT.
- Conroy, J. and D. O'Leary. 2001. Text Summarization via Hidden Markov Models. Proc. of SIGIR.

Bibliography

- Conroy, J.M., J. D. Schlesinger, and D. P. OLeary. 2006. Topic-Focused Multi-Document Summarization Using an Approximate Oracle Score. Proc. COLING/ACL 2006. pp. 152-159.
- Doran, W.P., N. Stokes, E. Newman, J. Dunnion, J. Carthy, F. Toolan. 2004. News Story Gisting at University College Dublin. In the Proc. of DUC.
- Edmundson, H.P. 1969. New Methods in Automatic Extracting. Journal of the Association for Computing Machinery, 16(2):264–285.
- Erkan G. and D. R. Radev.2004. LexRank: Graph-based Centrality as Salience in Text Summarization. Journal of Artificial Intelligence Research (JAIR).
- Fung, P. G. Ngai, and P. Cheung. 2003. Combining optimal clustering and hidden Markov models for extractive summarization. Proceedings of ACL Workshop on Multilingual Summarization.
- Galanis, D., G. Lampouras and I. Androutsopoulos. 2012. Extractive Multi-Document Summarization with Integer Linear Programming and Support Vector Regression. Proc. of COLING
- Galley. M. 2006. A Skip-Chain Conditional Random Field for Ranking Meeting Utterances by Importance Proc. of EMNLP
- Ge, S.S., Z. Zhang, and H. He. 2011. Weighted graph model based sentence clustering and ranking for document summarization. Interaction Sciences.
- Gillick, D. and B. Favre. 2009. A scalable global model for summarization. Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing.
- Gillick, D., K. Riedhammer, B. Favre, and D. Hakkani-Tur. 2009. A global optimization framework for meeting summarization. Proc. of ICASSP.
- Goldstein, J., V. Mittal, J. Carbonell, and M. Kantrowitz. 2000. Multi-document summarization by sentence extraction. Proc. of the 2000 NAACL-ANLP Workshop on Automatic Summarization.
- Gong, Y. and X. Liu. 2001. Generic text summarization using relevance measure and latent semantic analysis. Proc. Of ACM SIGIR.
- Gurevych, I. and T. Nahnsen. 2005. Adapting Lexical Chaining to Summarize Conversational Dialogues. Proc. Of RANLP.
- Hachey, B., G. Murray, and D. Reitter.2006. Dimensionality reduction aids term co-occurrence based multidocument summarization. In: SumQA 06: Proc. of the Workshop on Task-Focused Summarization and Question Answering.

Bibliography

- Haghighi, A. and L. Vanderwende. 2009. Exploring content models for multi-document summarization. Proc. of NAACL-HLT.
- He, L., E. Sanocki, A. Gupta, and J. Grudin. 2000. Comparing presentation summaries: Slides vs. reading vs. listening. Proc. of SIGCHI on Human factors in computing systems.
- Kaikhah, K. 2004. Automatic Text Summarization with Neural Networks. Proc. of IEEE Conference on Intelligent Systems.
- Kupiec, J., J. Pedersen, and F. Chen. 1995. A Trainable Document Summarizer. Proc. of SIGIR.
- Leskovec, J., N. Milic-frayling, and M. Grobelnik. 2005. Impact of Linguistic Analysis on the Semantic Graph Coverage and Learning of Document Extracts. Proc. of AAAI.
- Lin, C-Y. 2004. ROUGE: A Package for Automatic Evaluation of Summaries, Workshop on Text Summarization Branches Out.
- Lin, C.-Y. and E. Hovy. 2000. The automated acquisition of topic signatures for text summarization. Proc. of COLING.
- Lin, H. and J. Bilmes. 2010. Multi-document summarization via budgeted maximization of submodular functions. Proc. of NAACL.
- Lin, H., J. Bilmes and S. Xie. 2009. Graph-based Submodular Selection for Extractive Summarization. Proceedings of ASRU.
- Litvak, M. N. Vanetik. 2013. Mining the Gaps: Towards Polynomial Summarization, Prof. of IJCNLP.
- Litvak, M.,, M. Last, and M. Friedman. 2010. A new Approach to Improving Multilingual Summarization using a Genetic Algorithm. Proc. of ACL.
- Louis, A., A. Joshi, A. Nenkova. 2010. Discourse indicators for content selection in summarization. Proc. of SIGDIAL.
- Louis, A. and A. Nenkova. 2009. Automatically evaluating content selection in summarization without human models. Proceedings of EMNLP
- Luhn, H.P. 1958. The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development 2(2).
- Mana-Lopez, M.J., M De Buenaga, and J. M. Gomez-Hidalgo. 2004. Multidocument summarization: An added value to clustering in interactive retrieval. ACM Trans. Inf. Systems.

Bibliography

- Mani, I., G. Klein, D. House, L. Hirschman, T. Firmin, and B. Sundheim. 2002. SUMMAC: a text summarization evaluation. Natual Language Engineering. 8(1): 43-68.
- Mann, W.C. and S.A. Thompson. Rhetorical Structure Theory: towards a functional theory of text organization. Text, 8(3):243–281, 1988.
- Marcu, D. The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts. PhD thesis, Department of Computer Science, University of Toronto, 1997.
- McDonald, R. 2007. A Study of Global Inference Algorithms in Multi-document Summarization. Lecture Notes in Computer Science. Advances in Information Retrieval.
- McKeown, K..R., R. J. Passonneau, D. K. Elson, A. Nenkova, and J. Hirschberg. 2005. Do summaries help? Proc. of SIGIR.
- McKeown, K.R., J. L. Klavans, V. Hatzivassiloglou, R. Barzilay, and E. Eskin.1999. Towards multidocument summarization by reformulation: progress and prospects. Proc. AAAI 1999.
- Mihalcea, R. and P. Tarau .2004. Textrank: Bringing order into texts. Proc. of EMNLP 2004.
- Moawad, I.F. and M. Aref. 2012. Semantic graph reduction approach for abstractive Text Summarization. Proc. of Computer Engineering & Systems (ICCES).
- Nedunchelian, R. 2008. Centroid Based Summarization of Multiple Documents Implemented Using Timestamps. ICETET 2008:480-485.
- Nenkova, A. and R. Passonneau. 2004. Evaluating Content Selection in Summarization: The Pyramid Method. Proc. Of HLT-NAACL.
- Nenkova, A., L. Vanderwende, and K. McKeown. 2006. A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization. Proc. ACM SIGIR.
- Nenkova, A., R. Passonneau, and K. McKeown. 2007. The Pyramid Method: Incorporating human content selection variation in summarization evaluation. ACM Trans. Speech Lang. Processing.
- Osborne, M. 2002. Using maximum entropy for sentence extraction. Proc. of ACL Workshop on Automatic Summarization.
- Radev, D., T. Allison, M. Craig, S. Dimitrov, O. Kareem, M. Topper, and A. Winkel. 2004. A scaleable multi-document centroid-based summarizer. In Proc. of HLT-NAACL Demo Session.

Bibliography

- Radev, D.R. and K. R. McKeown. 1998. Generating natural language summaries from multiple on-line sources. Computational Linguistics, 24(3):469–500.
- Roussinov, D.G. and H. Chen. 2001. Information navigation on the web by clustering and summarizing query results. Inf. Process. Manag. 37, 6 (October 2001), 789-816
- Saggion, H. SUMMA. A Robust and Adaptable Summarization Tool. 2008. TAL 49(2):103-125.
- Saggion, H. and R. Gaizauskas. 2004. Multi-document summarization by cluster/profile relevance and redundancy removal. In Proc. of the Document Understanding Conference 2004. NIST.
- Saggion, H. and G. Lapalme. 2002. Generating indicative-informative summaries with SumUM. Computational Linguistics, 28 (4):497-526.
- Saggion, H., J.M. Torres-Moreno, I. da Cunha, and E. SanJuan. 2010. Multilingual Summarization Evaluation without Human Models. Proc. of COLING.
- Schiffman, B., A. Nenkova, and K. McKeown. 2002. Experiments in Multidocument Summarization. Proc. of HLT.
- Shen, D., J.-T. Sun, H. Li, Q. Yang, Z. Chen. 2007. Document Summarization using Conditional Random Fields. Proc. of IJCAI.
- Siddharthan, A., A. Nenkova, and K. Mckeown.2004. Syntactic Simplification for Improving Content Selection in Multi-Document Summarization. Proc. of COLING.
- Silber, H.G., and K. F. McCoy. 2002. Efficiently computed lexical chains as an intermediate representation for automatic text summarization. Computational. Linguist. 28(4):487-496.
- Steinberger, J., M. Poesio, M. A. Kabadjov, and K. Jeek. 2007. Two uses of anaphora resolution in summarization. Inf. Process. Manag. 43(6).
- Ulrich, J., G. Carenini, G. Murray and R. Ng. 2009. Regression-Based Summarization of Email Conversations. Proc. of ICWSM 2009.
- Vanderwende, L., H. Suzuki, C. Brockett, and A. Nenkova. 2007. Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion. Information Processing and Management 43.

Bibliography

- Wan, X., H.Li ,and J. Xiao. 2010. Cross-Language Document Summarization Based on Machine Translation Quality Prediction. In Proc. of ACL.
- Wang, X. et al. 2015. Summarization Based on Task-Oriented Discourse Parsing. Audio, Speech and Language Proc. 23(8).
- Wei, F, W. Li, Q. Li, Y. He. 2010. A document-sensitive graph model for multi-document summarization. Knowledge and Information Systems, 22(2):245-259.
- Wong, K.F., M. Wu, and W. Li. 2008. Extractive Summarization using Supervised and Semi-supervised learning. Proc. of ACL.
- Xie, S. and Y. Liu. 2010. Improving Supervised Learning for Meeting Summarization using Sampling and Regression. Computer Speech and Language. 24:495-514.
- Ye, S., T.-S. Chua, M.-Y. Kan, and L. Qiu. 2007. Document concept lattice for text understanding and summarization. Information Processing and Management 43(6).
- Yih, W., J. Goodman, L. Vanderwende, and H. Suzuki. 2007. Multi-Document Summarization by Maximizing Informative Content-Words. Proc. of IJCAI 2007.