LAB

a simple extractive algorithm

- · reduce the document size of e.g., 10%, 20%, 30% (gride alle fire & 1 90,00,10%)
- (individuate the topic of the text being summarised; the topic can be referred to as a (set of) NASARI vector(s):

```
v_{t1} = \{\text{term}_{1}\_\text{score}, \text{term}_{2}\_\text{score}, ..., \text{term}_{10}\_\text{score} \}
v_{t2} = \{\text{term}_{1}\_\text{score}, \text{term}_{2}\_\text{score}, ..., \text{term}_{10}\_\text{score} \}
...
```

- 2 create the context, by collecting the vectors of terms herein (this step can be repeated, by dumping the contribution of the associated terms at each round);
- 3. retain paragraphs whose sentences contain the most salient terms, based on the Weighted Overlap, $WO(v_1,v_2)$
 - rerank paragraphs weight by applying at least one of the mentioned approaches (<u>title</u>, <u>cue</u>, <u>phrase</u>, <u>cohesion</u>).

51

NASARI (lexical) subset

- two distribution files are provided for NASARI, that require different resources allocation.
- <u>dd-nasari.txt.</u> a <u>subset of NASARI (obtained by truncating vectors at I0 features)</u>. 3,587,754 vectors, ~600MB;
 <u>https://goo.gl/85BubW</u>
- dd-small-nasari-15.txt. a subset of NASARI. same filtering as above, with 15 features + intersection with 60K lemmas in the Corpus of Contemporary American English: 13,084 vectors, 2MB storage (many entities removed here...).
- the second one has been extracted for starting our experimentation;
 the second one is intended to explore the resource in a richer (though reduced) flavour.



documents for summarisation

- text documents are provided for summarisation purposes:
 - Andy-Warhol.txt
 - Ebola-virus-disease.txt
 - Life-indoors.txt
 - Napoleon-wiki.txt
 - Trump-wall.txt
- do experiment with different compression rates: 10%, 20% and 30%.



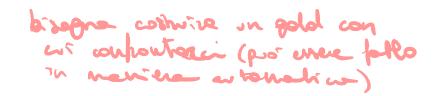
Daniele Radicioni - TLN 53

evaluation

- evaluation can be performed based on two complimentary metrics
- 1 BLEU (bilingual evaluation understudy) regarding precision; and
- ROUGE (Recall-Oriented Understudy for Gisting Evaluation) as regards as recall.



Daniele Radicioni - TLN 54



BLUE (bilingual evaluation understudy)

- scoring function that has been worked out to assess systems for automatic translation
- build a reference summary, as a list of relevant terms that should be present.
- compare the set of terms in the automatic summary (which we call candidate summary,) to those in the candidate summary.
- the BLEU score is computed as $P = m/w_t$ that is the fraction of terms from the candidate that are found in the reference, where m is the number of terms in the candidate that are in the reference, and w_t is the size of the candidate



©ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

- This metrics estimates in how far the words (and/or n-grams) in the human reference summaries appeared in the summaries built by the system
- ROUGE-N: Overlap of N-grams between candidate and reference summary.
- ROUGE-I refers to the overlap of unigram (each word) between the system and reference summaries.







Daniele Radicioni - TLN 57