Summarization
Machine Translation

Summarization

- *Text summarization is the process of distilling the most important information from a text to produce an abridged version for a particular task and user*
  - Definition adapted from Mani and Maybury 1999
- Types of summaries in current research:
  - Outlines or abstracts of any document, article, etc.
  - Snippets summarizing a Web page or a search engine results page
  - Action items or other summaries of a business meeting
  - Summaries of email threads
  - Simplifying text by compressing sentences

Single vs. Multiple Documents

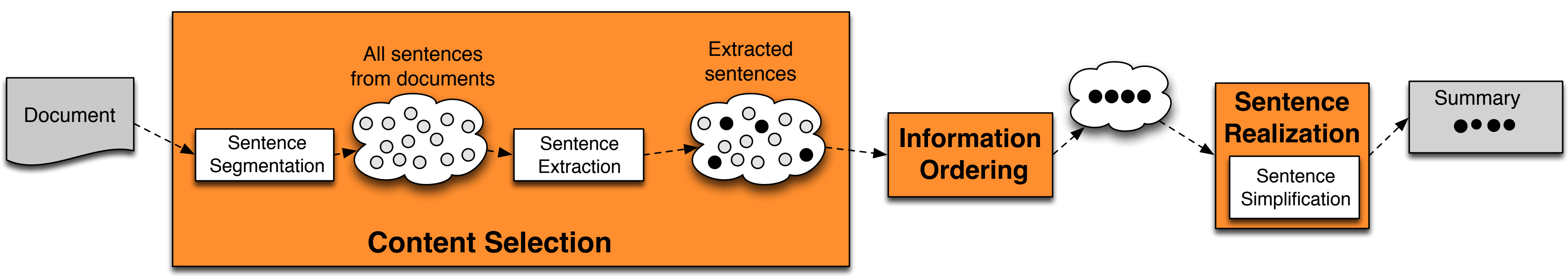
- **Single-document summarization**
  - Given a single document, produce
    - abstract
    - outline
    - headline
- **Multiple-document summarization**
  - Given a group of documents, produce a gist of the content, and create a cohesive answer that combines information from each document
    - a series of news stories on the same event
    - a set of web pages about some topic or question

Extractive vs. Abstractive

- Extractive summarization:
  - create the summary from phrases or sentences in the source document(s)
- Abstractive summarization:
  - express the ideas in the source documents using (at least in part) different words

Typical approaches to general problem

- Currently, achieve extraction instead of a true re-phrasing
  – **Content Selection**
    - Identify the sentences or clauses to extract
  – **Information Ordering**
    - How to order the selected units
  – **Sentence Realization**
    - Perform cleanup on the extracted units so that they are fluent in their new context.



Content Selection

- Simple approach is to select sentences that have more informative words according to saliency defined from a topic signature of the document
- Centroid-based summarization uses log-likelihood ratios for words, computing the probability of observing the word in the input more often than in the background corpus
- Other centrality methods try to rank the sentences according to a centrality score
- Methods based on rhetorical parsing use coherence relations to identify satellite and nuclues sentences
- Machine learning methods use features based on
  - Position, cue phrases, word informativeness, sentence length, cohesion (computing lexical chains of the document)

Information Ordering

- Simplest is to keep the document ordering
- Chronological ordering:
  - Order sentences by the date of the document (for summarizing news)..

  (Barzilay, Elhadad, and McKeown 2002)
- Coherence:
  - Choose orderings that make neighboring sentences similar (by cosine).
  - Choose orderings in which neighboring sentences discuss the same entity (Barzilay and Lapata 2007)
- Topical ordering
  - Learn the ordering of topics in the source documents

Simplifying Sentences

Zajic et al. (2007), Conroy et al. (2006), Vanderwende et al. (2007)

- Simplest method: parse sentences, use rules to decide which modifiers to prune
  - (more recently a wide variety of machine-learning methods)

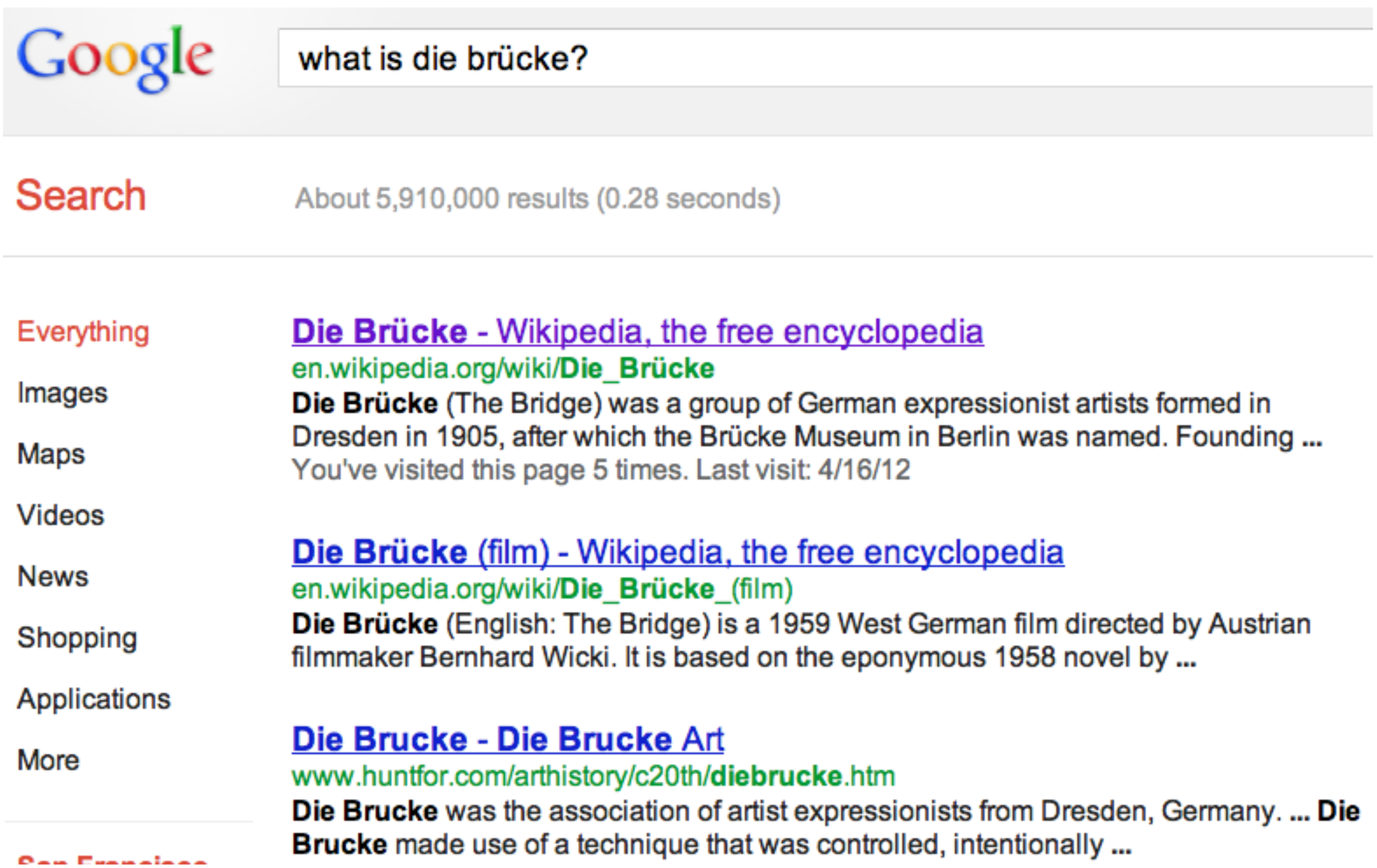
appositives	Rajam, ~~28, an artist who was living at the time in Philadelphia~~, found the inspiration in the back of city magazines.
attribution clauses	Rebels agreed to talks with government officials, ~~international observers said Tuesday.~~
PPs without named entities	The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [~~PP to a sustainable number~~]]
initial adverbials	"~~For example~~", "~~On the other hand~~", "~~As a matter of fact~~", "~~At this point~~"

Summarization Evaluation

- Extrinsic (task-based) evaluation: humans are asked to rate the summaries according to how well they are enabled to perform a specific task
- Intrinsic (task-independent) evaluation
  - Human judgments to rate the summaries
  - ROUGE (Recall Oriented Understudy Gisting Evaluation)
    - Humans generate summaries for a document collection
    - System-generated summaries are rated according to how close they come to the human-generated summary
    - Measures have included unigram overlap, bigram overlap, and longest common subsequence
  - Pyramid method
    - Humans identify "units of meaning" and then an overlap measure is computed

Summarization for Question-Answering: Snippets

- **Create snippets** summarizing a web page for a query
  - Google: 156 characters (about 26 words) plus title and link



Machine Translation

- Translating text from one language to another is a task challenging even for humans to try to fully capture the style and nuanced meaning of the original
- While research focuses on trying to produce the fully-automatic, high-quality translation, there are many tasks for which a rough translation is sufficient
- The differences between languages include systematic differences that can be modeled in some way and idiosyncratic and lexical differences that must be dealt with one by one.

Why MT is hard

- Given the Japanese phrase
  fukaku hansei shite orimasu
- If this is translated to English as
  we apologize
  it is not faithful to the original meaning
- But if we translate it as
  we are deeply reflecting (on our past behavior, and what we did wrong, and how to avoid the problem next time)

  the translation is not fluent.

Example from Jurafsky and Martin text.

Differences between languages

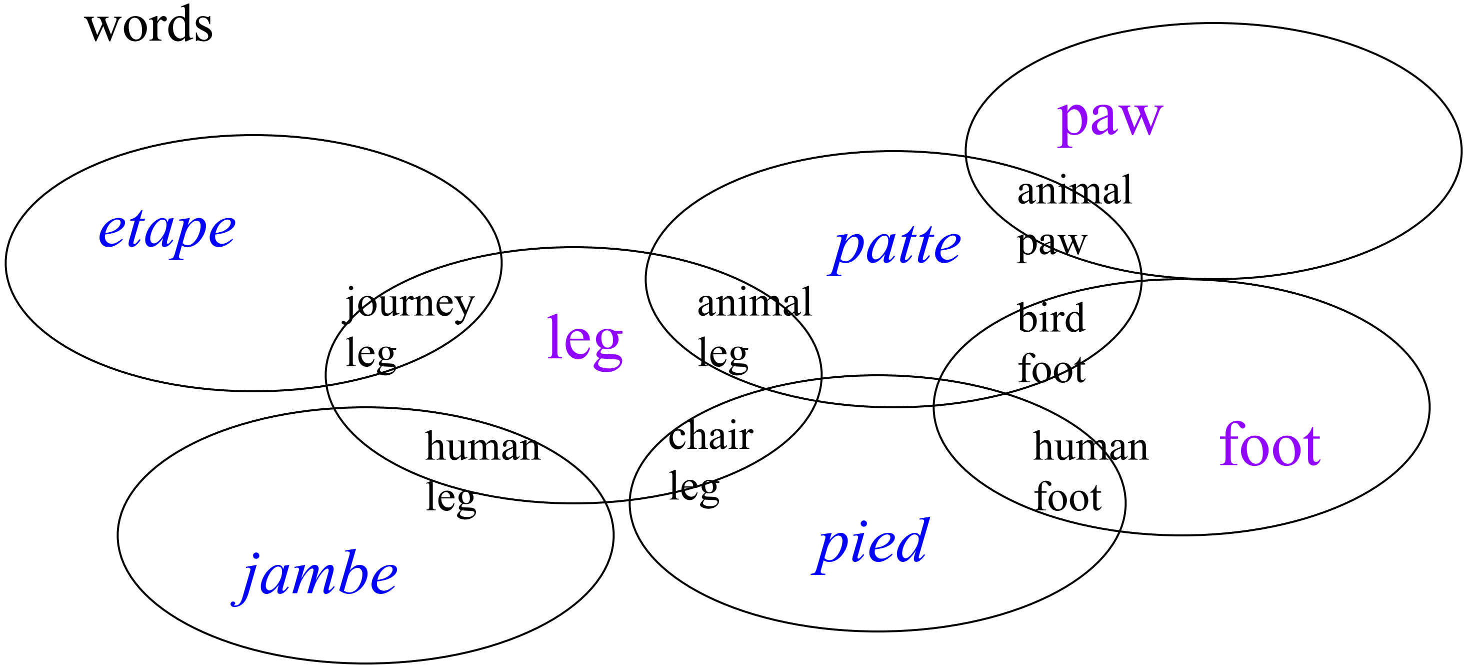
- Morphological differences:
  - Number of morphemes per word
    - Isolating languages: Vietnamese and Cantonese, each word has one morpheme
    - Polysynthetic languages: "Eskimo", a single word has many morphemes corresponding to a complete sentence.
  - Degree to which morphemes are segmentable
    - Agglutinative, morphemes have clean boundaries (Turkish)
    - Fusion languages, single affix may have multiple morphemes (Russian)

Differences between languages

- Syntactic differences
  - Basic word order of verbs, subjects and objects
    - SVO: English, Mandarin, French, German, …
    - SOV: Hindi, Japanese
    - VSO: Classical Arabic and Biblical Hebrew
  - Head marking and dependent marking languages
    - Mark relation between dependent and head on the head
      - English marks possessive on dependent: *the man's house*
      - Hungarian marks possessive on the head noun: (Hungarian equivalent of:) *the man house-his*
  - Direction of motion with respect to verb
    - English direction on particle: *the bottle floated out*
    - Spanish direction on verb: *la botella salio' flotando*
  - Grammatical constraints on matching gender-marked words
  - Many others . . .

Differences between languages

- Semantic differences
  - Lexical gap
    - One language doesn’t have a word for concept in another
  - Differences in way that conceptual space is divided up for different words



The complex overlap between English leg, foot, etc. and various French translations. (Jurafsky & Martin, Figure 21.2)

Classical MT/Machine Translation

- In this line of MT research, approaches can be classified according to the level of unit of translation
  - Direct translation uses a word translation approach
  - Syntactic and semantic transfer approaches use syntactic phrase and semantic units, respectively, as the unit of translation

Statistical Approaches

- Build probabilistic models of faithfulness and fluency and combine the models to get the most probable translation.
- Modeled as a noisy channel "pretend that the foreign input F is a corrupted version of the target language output E and the task is to discover the hidden sentence E that generated the observed sentence F."
  - Informally, we refer to translating from French to English
- Requires two models
  - Language model to compute P(E), probability that any sequence E of English words is a sentence
  - Translation model to compute P(F|E), conditional probability that French sentence F was a translation of an English sentence E
- Given French sentence f, its translation e is

  arg max (all e in E) P(e) * P(f | e)

  - Note that this appears backwards to translate from English to French, but we invoke Bayes theorem to define the decoder.

Statistical Language Models

- Language model to compute P(E)
  - In practice, learn probabilities of bigrams in the language to be translated from instead of entire sentences
  - Translation has improved greatly due to large corpora
    - See Google Translate
- Translation model to compute P(F|E)
  - Learn probabilities from parallel corpora
  - Model the translation as word translation combined with alignment prob.
    - E: *And the program has been implemented.*
    - F: *Le programme a ete mis en application.*
    - Alignment variables: (2, 3, 4, 5, 6, 6, 6) gives

Le	*-> the*	*mis*	*-> implemented*
Programme	*-> program*	*en*	*-> implemented*
a	*-> has*	*application*	*-> implemented*
ete	*-> been*		

Alignment and Parallel Corpora

- The translation model uses probabilities of word alignment
- Word alignment models are automatically trained from parallel corpora
  - Hansard Corpus
    - Canadian parliament documents for French, English and a variety of native American languages
  - United Nations proceedings documents
  - LDC has corpora in several language pairs
- Literary parallel corpora are not as suitable because of the stronger presence of literary devices, such as metaphor

MT Evaluation

- Human raters can evaluate along the two dimensions of fluency and fidelity (and there are several individual metrics for each of these dimensions)
- BLEU automatic evaluation system
  - Evaluation corpus contains human generated translations
  - Metrics evaluate how closely the system-generated translations correspond to the human ones