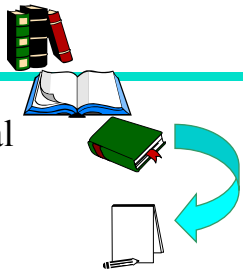
Automated Text Summarization

Extract from tutorial at COLING/ACL'98

Eduard Hovy and Daniel Marcu

Information Sciences Institute
University of Southern California
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292
http://www.isi.edu/natural-language/people/{hovy.html,marcu.html}

an exciting challenge...



...put a book on the scanner, turn the dial to '2 pages', and read the result...

...download 1000 documents from the web, send them to the summarizer, and select the best ones by reading the summaries of the clusters...

...forward the Japanese email to the summarizer, select '1 par', and skim the translated summary.

Headline news — informing

TIME.com

TIME Daily
> News Wire
> Editor's Letter
> Comments
> News Features
> Text Only

Magazine
Community
Special Reports

LIFE Picture of the Day

Address
Password
go

Get TIME Daily delivered to your desktop every day with Microsoft Internet Explorer

June 30, 1998

U.S. Plane Fires a Missile On Iraq
An Iraqi radar station targets an Allied plane, and a U.S. F-16 responds quickly -- with deadly force. Is another showdown with Saddam on the way?
Full Story


Responding with Force: A U.S. Air Force F-16 flies over Kuwait. U.S. AIR FORCE/AP

Starr Plays the Tripp Card
The former confidante's grand jury appearance puts the squeeze on Ms. Lewinsky.

Down to Business in Shanghai
President Clinton spends some time in the city he wants the rest of China to turn into.

·Poll: Does the U.S. have the right to impose its idea of human rights on China?

·Postcards From the Middle Kingdom: TIME's Jay Branegan says President Clinton is in full campaign mode in China. But the big question is, why isn't the pressing the flesh?

Boris Duels With the Duma
If Russian president Yeltsin wants to make other Russian pols look bad, he should stop making a fool of himself first.

TV-GUIDES — decision making

2:30am **VC2 – 76**
The Jackal
Movie: Bruce Willis excels as "The Jackal," a cunning assassin who uses many disguises in this 1997 thriller. Richard Gere and Sidney Poitier costar as players from different sides of the law who unite to stop him.

3:00am **KCOP – 13**
The Untouchables
Movie: Eliot Ness (Kevin Costner) and ''The Untouchables'' take on Robert De Niro's flamboyant Al Capone in the pulse–pounding 1987 adaptation of the popular TV series.\Sean Connery won an Oscar as the Irish beat cop who shows Ness ''the Chicago way.''\ Brian De Palma directed the feature;\David Mamet wrote the script.\And yes, film majors, the scene at Union Station was lifted directly from the

3:05am **STARZ – 25**
Grosse Pointe Blank
Movie: A razor–sharp script and a fine turn by John Cusack as a troubled hit man mark 1997's "Grosse Pointe Blank," a dark comedy in which the assassin encounters his old flame (Minnie Driver of "Good Will Hunting") at a high–school reunion. Cusack's sister Joan ("In and Out") is hilarious as the killer's devoted assistant, and Alan Arkin makes the most of his small role as Cusack's terrified the

Abstracts of papers — time saving

An Incremental Interpreter for High-Level Programs with Sensing

Giuseppe De Giacomo
Dipartimento di Informatica e Sistemistica
Universita di Roma "La Sapienza"
Via Salaria 113, 00198 Rome, Italy
degiacomo@dis.uniroma1.it

Hector Levesque
Department of Computer Science
University of Toronto
Toronto, Canada M5S 3H5
hector@cs.toronto.edu

Abstract

Like classical planning, the execution of high-level agent programs requires a reasoner to look all the way to a final goal state before even a single action can be taken in the world. This deferral is a serious problem in practice for large programs. Furthermore, the problem is compounded in the presence of sensing actions which provide necessary information, but only after they are executed in the world. To deal with this, we propose (characterize formally in the situation calculus, and implement in Prolog) a new incremental way of interpreting such high-level programs and a new high-level language construct, which together, and without loss of generality, allow much more control to be exercised over when actions can be executed. We argue that such a scheme is the only practical way to deal with large agent programs containing both nondeterminism and sensing.

Introduction

In [4] it was argued that when it comes to providing high level control to autonomous agents or robots, the notion of *high-level program execution* offers an alternative to classical planning that may be more practical in many applications. Briefly, instead of looking for a sequence of actions $\vec{a}$ such that

$$Axioms \models Legal(do(\vec{a}, S_0)) \wedge \phi(do(\vec{a}, S_0))$$

where ϕ is the goal being planned for, we look for a sequence $\vec{a}$ such that

$$Axioms \models Do(\delta, S_0, do(\vec{a}, S_0))$$

to find a sequence with the right properties. This can involve considerable search when δ is very nondeterministic, but much less search when δ is more deterministic. The feasibility of this approach for AI purposes clearly depends on the expressive power of the programming language in question. In [4], a language called ConGolog is presented, which in addition to nondeterminism, contains facilities for sequence, iteration, conditionals, concurrency, and prioritized interrupts. In this paper, we extend the expressive power of this language by providing much finer control over the nondeterminism, and by making provisions for sensing actions. To do so in a way that will be practical even for very large programs requires introducing a different style of on-line program execution.

In the rest of this section, we discuss on-line and off-line execution informally, and show why sensing actions and nondeterminism together can be problematic. In the following section, we formally characterize program execution in the language of the situation calculus. Next, we describe an incremental interpreter in Prolog that is correct with respect to this specification. The final section contains discussion and conclusions.

Off-line and On-line execution

To be compatible with planning, the ConGolog interpreter presented in [4] executes in an *off-line* manner, in the sense that it must find a sequence of actions constituting an entire legal execution of a program *before* actually executing any of them in the world.[1] Consider, for example, the following program:

Questions

- **What kinds** of summaries do people want?
  - What are *summarizing, abstracting, gisting*,...?
- **How sophisticated** must summ. systems be?
  - Are statistical techniques sufficient?
  - Or do we need symbolic techniques and deep understanding as well?
- **What milestones** would mark quantum leaps in summarization theory and practice?
  - How do we measure summarization quality?

Table of contents

'Genres' of Summary?

- Indicative *vs*. informative
  ...used for quick categorization vs. content processing.
- Extract *vs*. abstract
  ...lists fragments of text vs. re-phrases content coherently.
- Generic *vs*. query-oriented
  ...provides author's view vs. reflects user's interest.
- Background *vs*. just-the-news
  ...assumes reader's prior knowledge is poor vs. up-to-date.
- Single-document *vs*. multi-document source
  ...based on one text vs. fuses together many texts.

Aspects that Describe Summaries

- **Input** (Sparck Jones 97)
  - *subject type*: domain
  - *genre*: newspaper articles, editorials, letters, reports...
  - *form*: regular text structure; free-form
  - *source size*: single doc; multiple docs (few; many)
- **Purpose**
  - *situation*: embedded in larger system (MT, IR) or not?
  - *audience*: focused or general
  - *usage*: IR, sorting, skimming...
- **Output**
  - *completeness*: include all aspects, or focus on some?
  - *format*: paragraph, table, etc.
  - *style*: informative, indicative, aggregative, critical...

Table of contents

Making Sense of it All...

To understand summarization, it helps to consider several perspectives simultaneously:

1. **Approaches**: basic starting point, angle of attack, core focus question(s): *psycholinguistics, text linguistics, computation...*
2. **Paradigms**: theoretical stance; methodological preferences: *rules, statistics, NLP, Info Retrieval, AI...*
3. **Methods**: the nuts and bolts: modules, algorithms, processing: *word frequency, sentence position, concept generalization...*

Psycholinguistic Approach: 2 Studies

- Coarse-grained summarization protocols from professional summarizers (Kintsch and van Dijk, 78):
  - Delete material that is trivial or redundant.
  - Use superordinate concepts and actions.
  - Select or invent topic sentence.

- 552 finely-grained summarization strategies from professional summarizers (Endres-Niggemeyer, 98):
  - **Self control**: make yourself feel comfortable.
  - **Processing**: produce a unit as soon as you have enough data.
  - **Info organization**: use "Discussion" section to check results.
  - **Content selection**: the table of contents is relevant.

Computational Approach: Basics

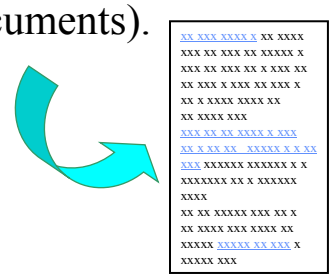
Top-Down:	*Bottom-Up*:
• *I know what I want! — don't confuse me with drivel!* 	• *I'm dead curious: what's in the text?* 
• User needs: only certain types of info	• User needs: anything that's important
• System needs: *particular criteria of interest*, used to focus search	• System needs: *generic importance metrics*, used to rate content

Query-Driven vs. Text-DRIVEN Focus

- **Top-down: Query-driven focus**
  - *Criteria of interest* encoded as search specs.
  - System uses specs to filter or analyze text portions.
  - Examples: *templates* with slots with semantic characteristics; *termlists* of important terms.
- **Bottom-up: Text-driven focus**
  - *Generic importance metrics* encoded as strategies.
  - System applies strategies over rep of whole text.
  - Examples: degree of *connectedness* in semantic graphs; *frequency* of occurrence of tokens.

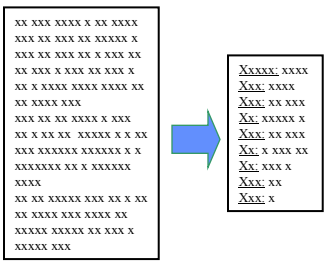
Bottom-Up, using Info. Retrieval

- **IR task**: Given a query, find the relevant document(s) from a large set of documents.
- **Summ-IR task**: Given a query, find the relevant passage(s) from a set of passages (i.e., from one or more documents).
- **Questions**:
  1. IR techniques work on large volumes of data; can they scale down accurately enough?
  2. IR works on words; do abstracts require abstract representations?



Top-Down, using Info. Extraction

- **IE task**: Given a template and a text, find all the information relevant to each slot of the template and fill it in.
- **Summ-IE task**: Given a query, select the best template, fill it in, and generate the contents.
- **Questions**:
  1. IE works only for very particular templates; can it scale up?
  2. What about information that doesn't fit into any template—is this a generic limitation of IE?

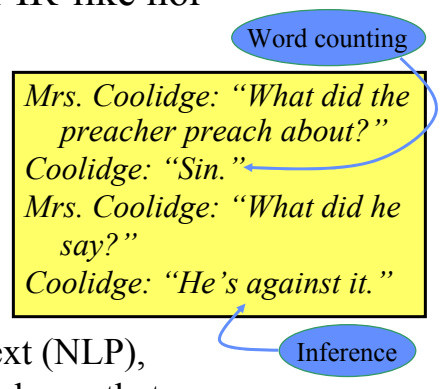


Paradigms: NLP/IE vs. IR/Statistics

NLP/IE:	IR/Statistics:
• **Approach**: try to 'understand' text—re-represent content using 'deeper' notation; then manipulate that.	• **Approach**: operate at lexical level—use word frequency, collocation counts, etc.
• **Need**: rules for text analysis and manipulation, at all levels.	• **Need**: large amounts of text.
• **Strengths**: higher quality; supports abstracting.	• **Strengths**: robust; good for query-oriented summaries.
• **Weaknesses**: speed; still needs to scale up to robust open-domain summarization.	• **Weaknesses**: lower quality; inability to manipulate information at abstract levels.

Toward the Final Answer...

- **Problem**: What if neither IR-like nor IE-like methods work?
  - sometimes counting and templates are insufficient,
  - and then you need to do inference to *understand*.
- **Solution**:
  - semantic analysis of the text (NLP),
  - using adequate knowledge bases that support inference (AI).



The Optimal Solution...

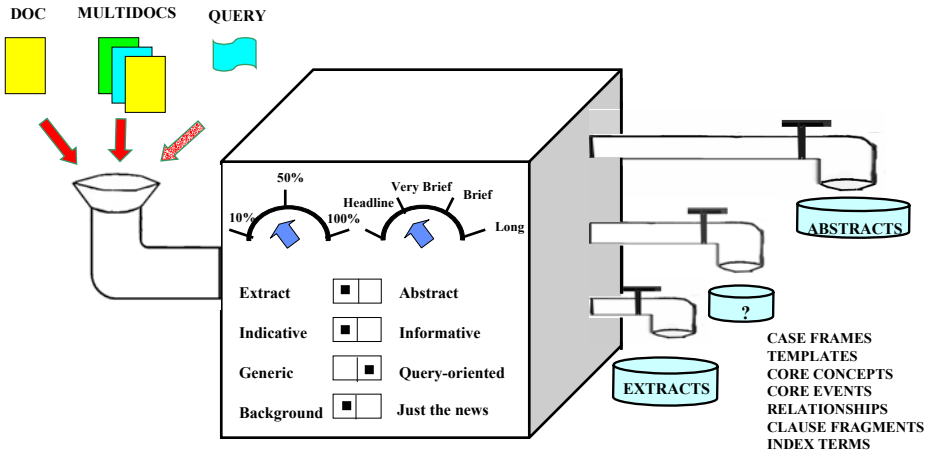
Combine strengths of both paradigms…

...use IE/NLP when you have suitable template(s),

...use IR when you don't...

...but how exactly to do it?

A Summarization Machine



The Modules of the Summarization Machine

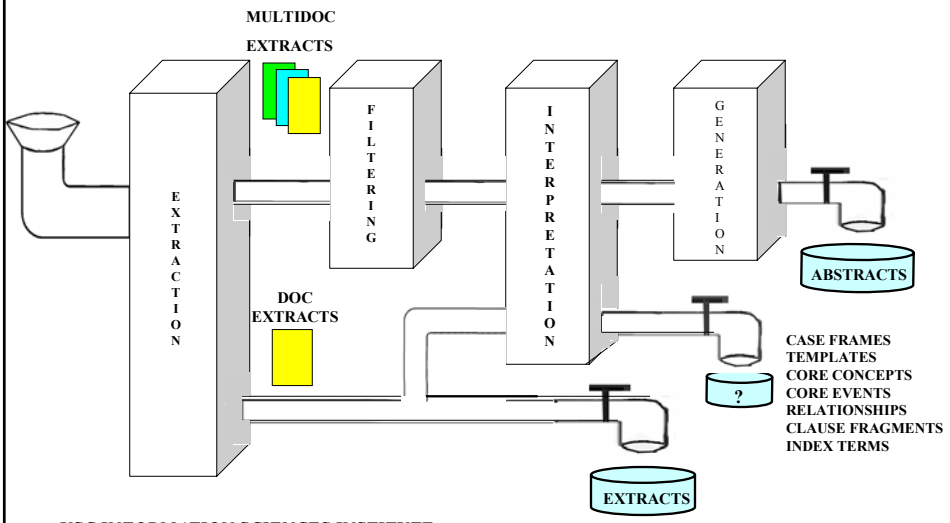


Table of contents

Overview of Extraction Methods

- Position in the text
  - lead method; optimal position policy
  - title/heading method
- Cue phrases in sentences
- Word frequencies throughout the text
- Cohesion: links among words
  - word co-occurrence
  - co-reference
  - lexical chains
- Discourse structure of the text
- Information Extraction: parsing and analysis

POSition-based method (1)

- **Claim**: Important sentences occur at the beginning (and/or end) of texts.
- **Lead method**: just take first sentence(s)!
- Experiments:
  - In 85% of 200 individual paragraphs the topic sentences occurred in initial position and in 7% in final position (Baxendale, 58).
  - Only 13% of the paragraphs of contemporary writers start with topic sentences (Donlan, 80).

Optimum Position Policy (OPP)

- **Claim**: Important sentences are located at positions that are genre-dependent; these positions can be determined automatically through training (Lin and Hovy, 97).
  - **Corpus**: 13000 newspaper articles (ZIFF corpus).
  - **Step 1**: For each article, determine overlap between sentences and the index terms for the article.
  - **Step 2**: Determine a partial ordering over the locations where sentences containing important words occur: Optimal Position Policy (OPP)

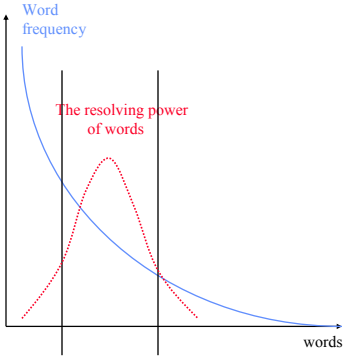
Title-Based Method (1)

- **Claim**: Words in titles and headings are positively relevant to summarization.
- Shown to be statistically valid at 99% level of significance (Edmundson, 68).
- Empirically shown to be useful in summarization systems.

Cue-Phrase method (1)

- **Claim 1**: Important sentences contain 'bonus phrases', such as *significantly, In this paper we show,* and *In conclusion*, while non-important sentences contain 'stigma phrases' such as *hardly* and *impossible*.
- **Claim 2**: These phrases can be detected automatically (Kupiec et al. 95; Teufel and Moens 97).
- **Method**: Add to sentence score if it contains a bonus phrase, penalize if it contains a stigma phrase.

Word-frequency-based method (1)

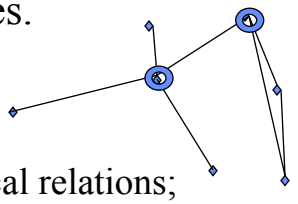


(Luhn, 59)

- **Claim**: Important sentences contain words that occur "somewhat" frequently.
- **Method**: Increase sentence score for each frequent word.
- **Evaluation**: Straightforward approach empirically shown to be mostly detrimental in summarization systems.

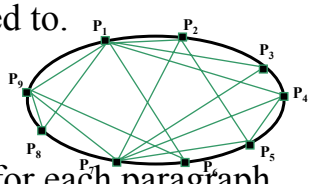
Cohesion-based methods

- **Claim**: Important sentences/paragraphs are the highest connected entities in more or less elaborate semantic structures.
- Classes of approaches
  - word co-occurrences;
  - local salience and grammatical relations;
  - co-reference;
  - lexical similarity (WordNet, lexical chains);
  - combinations of the above.



Cohesion: WORD co-occurrence (1)

- Apply IR methods at the document level: texts are collections of paragraphs (Salton et al., 94; Mitra et al., 97; Buckley and Cardie, 97):
  - Use a traditional, IR-based, word similarity measure to determine for each paragraph P_i the set S_i of paragraphs that P_i is related to.
- **Method**:
  - determine relatedness score S_i for each paragraph,
  - extract paragraphs with largest S_i scores.



Discourse-based method

- **Claim**: The multi-sentence coherence structure of a text can be constructed, and the 'centrality' of the textual units in this structure reflects their importance.
- Tree-like representation of texts in the style of *Rhetorical Structure Theory* (Mann and Thompson,88).
- Use the discourse representation in order to determine the most important textual units. Attempts:
  - (Ono et al., 94) for Japanese.
  - (Marcu, 97) for English.

Information Extraction method (1)

- **Idea**: content selection using templates
  - Predefine a template, whose slots specify what is of interest.
  - Use a canonical IE system to extract from a (set of) document(s) the relevant information; fill the template.
  - Generate the content of the template as the summary.
- **Previous IE work**:
  - FRUMP (DeJong, 78): '*sketchy scripts*' of terrorism, natural disasters, political visits...
  - (Mauldin, 91): templates for conceptual IR.
  - (Rau and Jacobs, 91): templates for business.
  - (McKeown and Radev, 95): templates for news.

Information Extraction method (2)

- Example template:

MESSAGE:ID	TSL-COL-0001
SECSOURCE:SOURCE	Reuters
SECSOURCE:DATE	26 Feb 93 Early afternoon
INCIDENT:DATE	26 Feb 93
INCIDENT:LOCATION	World Trade Center
INCIDENT:TYPE	Bombing
HUM TGT:NUMBER	AT LEAST 5

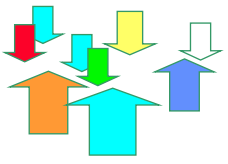
Review of Methods

Bottom-up methods

- Text location: title, position
- Cue phrases
- Word frequencies
- Internal text cohesion:
  - word co-occurrences
  - local salience
  - co-reference of names, objects
  - lexical similarity
  - semantic rep/graph centrality
- Discourse structure centrality

Top-down methods

- Information extraction templates
- Query-driven extraction:
  - query expansion lists
  - co-reference with query names
  - lexical similarity to query



Finally: Combining the Evidence

- **Problem**: which extraction methods to believe?
- **Answer**: assume they are independent, and combine their evidence: merge individual sentence scores.
- **Studies**:
  - (Kupiec et al., 95; Aone et al., 97, Teufel and Moens, 97): Bayes' Rule.
  - (Mani and Bloedorn,98): SCDF, C4.5, inductive learning.
  - (Lin and Hovy, 98b): C4.5.
  - (Marcu, 98): rhetorical parsing tuning.

Table of contents

Some Types of Interpretation

- Concept generalization:
  Sue ate apples, pears, and bananas ⇒ *Sue ate fruit*
- Meronymy replacement:
  Both wheels, the pedals, saddle, chain… ⇒ *the bike*
- Script identification: (Schank and Abelson, 77)
  He sat down, read the menu, ordered, ate, paid, and left ⇒ *He ate at the restaurant*
- Metonymy:
  A spokesperson for the US Government announced that… ⇒ *Washington announced that...*

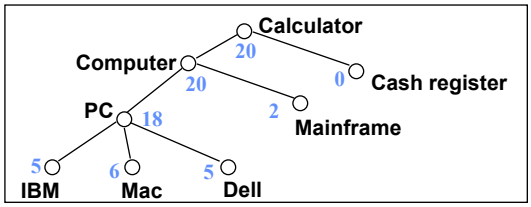
General Aspects of Interpretation

- Interpretation occurs at the conceptual level...
  …words alone are polysemous (*bat* = *animal* and *sports instrument*) and combine for meaning (*alleged murderer* ≠ *murderer*).
- For interpretation, you need world knowledge...
  …the fusion inferences are not in the text!
- Little work so far: (Lin, 95; McKeown and Radev, 95; Reimer and Hahn, 97; Hovy and Lin, 98).

Concept Generalization: Wavefront

- **Claim**: Can perform concept generalization, using WordNet (Lin, 95).

- Find most appropriate summarizing concept:



1. Count word occurrences in text; score WN concs
2. Propagate scores upward
3. $R = Max\{scores\} / \Sigma scores$
4. Move downward until no obvious child: $R<R_t$
5. Output that concept

Table of contents

How can You Evaluate a Summary?

- **When you already have a summary…**
  ...then you can compare a new one to it:
  1. choose a granularity (clause; sentence; paragraph),
  2. create a similarity measure for that granularity (word overlap; multi-word overlap, perfect match),
  3. measure the similarity of each unit in the new to the most similar unit(s) in the gold standard,
  4. measure Recall and Precision.

  e.g., (Kupiec et al., 95).

…….. but when you don't?

Toward a Theory of Evaluation

- **Two Measures:**

  Compression Ratio: *CR = (length S) / (length T)*
  Retention Ratio: *RR = (info in S) / (info in T)*

- **Measuring length:**
  - Number of letters? words?
- **Measuring information:**
  - *Shannon Game*: quantify information content.
  - *Question Game*: test reader's understanding.
  - *Classification Game*: compare classifiability.

Table of contents

1. Motivation.
2. Genres and types of summaries.
3. Approaches and paradigms.
4. Summarization methods.
5. Evaluating summaries.
6. The future.

The Future — **There's much to do!**

- Data preparation:
  - Collect large sets of texts with abstracts, all genres.
  - Build large corpora of <*Text, Abstract, Extract*> tuples.
  - Investigate relationships between extracts and abstracts (using <*Extract, Abstract*> tuples).
- Types of summary:
  - Determine characteristics of each type.
- Topic Identification:
  - Develop new identification methods (discourse, etc.).
  - Develop heuristics for method combination (train heuristics on <*Text, Extract*> tuples).

The Future (2)

- Concept Interpretation (Fusion):
  - Investigate types of fusion (semantic, evaluative…).
  - Create large collections of fusion knowledge/rules (e.g., signature libraries, generalization and partonymic hierarchies, metonymy rules…).
  - Study incorporation of User's knowledge in interpretation.
- Generation:
  - Develop Sentence Planner rules for dense packing of content into sentences (using <*Extract, Abstract*> pairs).
- Evaluation:
  - Develop better evaluation metrics, for types of summaries.