

# The Future of Summarization

Eduard Hovy

USC Information Sciences Institute

# Crazy questions that bother me

- Why so little inter-human agreement?
- Especially on generic summaries: what is ‘the author’s point of view’? What does a point of view look like, when represented?

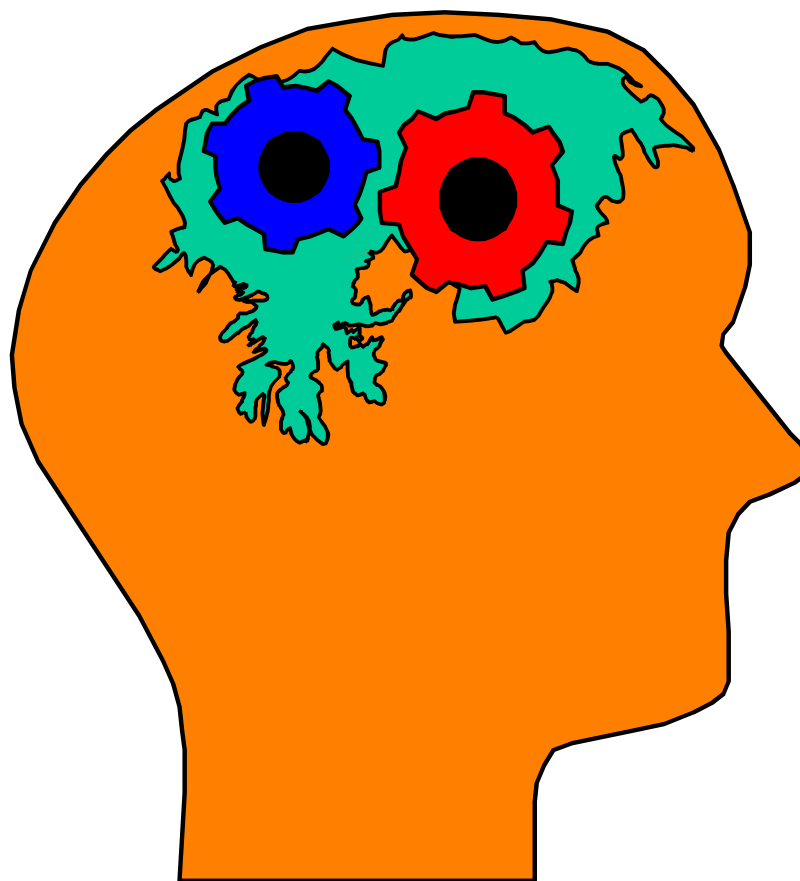
# Outrageous claims for the day

- Summarization is simply a step during a cycle of repeated drill-down QA.
- A summary is an intermediate teaching device to let you know what question you really wanted to ask.
- We should consider replacing summarization with QA.

# The (lunatic?) fringe

- We are info-gathering machines.
- Our state of info always incomplete...  
...there are the things you know, and then  
at the fringe are numerous **unanswered questions**.
- When you read, incoming info either:
  - matches existing knowledge—confirmation,
  - contradicts existing knowledge—problem,
  - connects nowhere—irrelevant,
  - **connects to unanswered Qs**—informative....and often opens up **new Qs**.

# Life on the fringe



Disaster: How do I avoid a:

- tax audit?
- speeding ticket?
- earthquake?
- plane delay?
- ...

My job:

- Research:
  - How can I build a big ontology?
  - How can I fix the EM learning of QA patterns?
- Admin:
  - How can I have fewer meetings?
- ...

Hobbies

- Should I join an orchestra?
- ...

Getting people to like me:

- be quieter?
- smile more?
- ...

The world:

- Politics:
  - How is balance between Bush and Congress working?
  - ...
- Religion
- The economy
- ...

People I know:

- Donna H:
  - What was her thesis about?
  - How many kids?
- Aunt Margrit:
  - How's her health?
- ...

# Every Reader has a different fringe

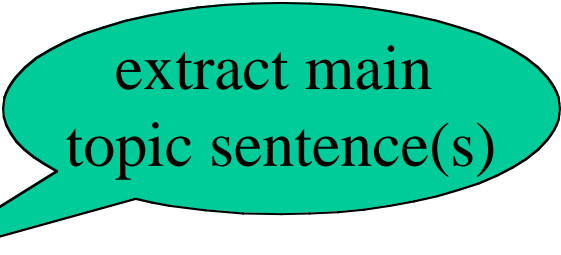
The operationally useful summary is the one that advances the fringe.

The summary *must address the Reader's fringe* Qs.

Every Reader has a different fringe.

The best the system can do:

1. set the context—activate relevant parts of the fringe,
2. determine the active Qs there,
3. make the summary provide answers to them.



extract main topic sentence(s)



how??



question answering?

# Overview

1. **Introduction:** effective summarization is relative to the Reader's knowledge state
2. **Topic-based summarization:** when the Reader can describe his/her knowledge state
3. **Generic summarization:** when not
4. **Conclusion**



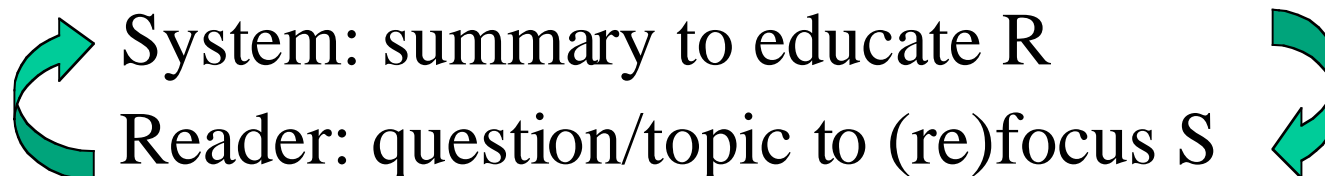
# Navigating the fringe

## Case 1: When the Reader can help (topic-based):

- **Problems:**

1. underspecificity: Reader seldom gives fringe Qs exactly or fully; just gives topic(s) as shorthand
  - ➡ system must infer Qs—may need several tries.
2. topic drift: on reading a summary, the Reader learns, and forms new Qs
  - ➡ system must make new summary.

- **Result:** repeated cycle of drilling down:



# Donna's MDS challenge

Topic-driven summarization of CL texts:

*tell me about Wordsense disambiguation!*

(download papers from comp-lg (now called Computing Research Repository (CoRR) <http://arXiv.org/> ; follow CS)

- **Problem:** This is not a news story: how to summarize *technical papers* together?
- **Small experiment:**
  1. categorized papers into topic buckets (one on WSD),
  2. found five recent papers on WSD; extracted abstracts,
  3. drill-down: extracted sentences containing specific key words,
  4. presented summary; then step 3 again with new key words.

# Window (“method”/“algorithm”/“model”)

- 1.1 In this paper Schapire and Singer’s AdaBoost.MH boosting algorithm is applied to the Word Sense Disambiguation (WSD) problem.
- 2.1 This paper describes an experimental comparison between two standard supervised learning methods, namely Naive Bayes and Exemplar-based classification, on the Word Sense Disambiguation (WSD) problem.
- 3.1 This paper describes a set of comparative experiments, including cross-corpus evaluation, between five alternative algorithms for supervised Word Sense Disambiguation (WSD), namely Naive Bayes, Exemplar-based learning, SNoW, Decision Lists, and Boosting.
- 4.1 This dissertation analyses the computational properties of current performance-models of natural language, in particular Data Oriented Parsing (DOP), points out some of their major shortcomings and suggests suitable solutions.
- 5.1 This paper presents the use of probabilistic class-based lexica for disambiguation in target-word selection.

# Results

## **WSD cluster. Papers about *methods/algorithms/models*:**

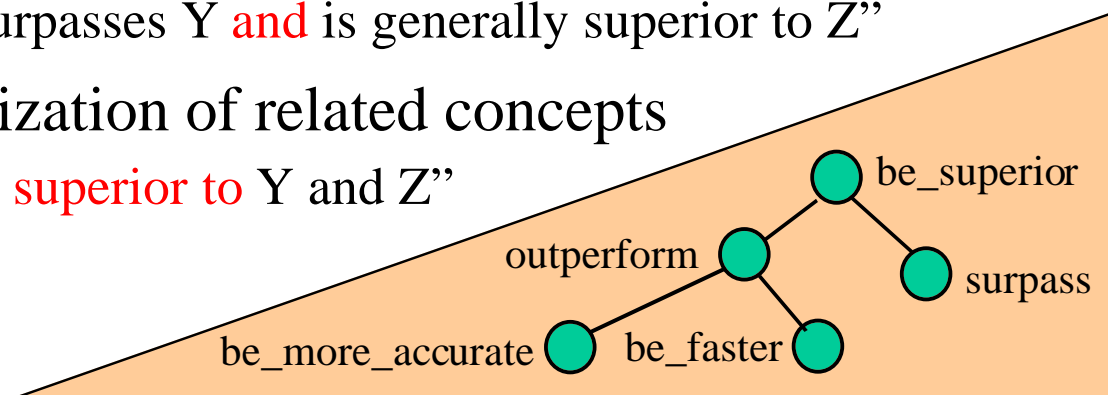
- Schapire and Singer's AdaBoost.MH boosting (paper 1.1)
- Naive Bayes and Exemplar-based classification (paper 2.1)
- Naive Bayes, Exemplar-based learning, SNoW, Decision Lists, and Boosting (paper 3.1)
- Data Oriented Parsing (paper 4.1)
- *one additional paper* (paper 5)

## **WSD cluster. Papers about *result/show*:**

- show that the boosting approach surpasses Naive Bayes and Exemplar-based approaches (paper 1.2)
- Results show that the Exemplar-based approach to WSD is generally superior to the Bayesian approach (paper 2.8)
- resulting Specialized DOP (SDOP) models to the original DOP models with encouraging results. (paper 4.9)
- shows promising results in an evaluation on real-world translations. (paper 5.5)
- *one additional paper* (paper 3)

# Doing this is not impossible...

- Simple approach, little magic:
  1. find “algorithm/method/model” or “result/show” as the key words
  2. determine extract windows around these keys
  3. synthesize the extracts ‘coherently’
  
- Next steps:
  - Syntactic aggregation of overlapping phrases  
 “show that X surpasses Y” / “show that X is generally superior to Z”  
 ➡ “show that X surpasses Y **and** is generally superior to Z”
  - Semantic generalization of related concepts  
 ➡ “show that X is **superior to** Y and Z”



# Searching blindly

## Case 2: When the Reader cannot help (generic):

(even if the main topic is right there, probably in one sentence)

- **Problems:**

- you don't (can't) know the author's fringe (and don't care),
- you haven't been given the Reader's fringe
- ... so you have to guess the Reader's Qs, or use your own.

(basic fringe Qs: 5W1H)

- **Solutions:**

- Top-down: predefined Qs — templates and IE
- Bottom-up: evidence for Qs — extraction heuristics

# Summarization as template extraction

**Easy case:** often, the story is stereotypical enough:

*“There was another instance of X,  
and as you know, Xs have the important features A, B, C,  
and here are the values (parameters) for A, B, and C: ...”*

Earthquake: location, magnitude, number of casualties, after-effects, assistance

Robbery: valuables, perpetrators, owners, police action, arrest or escape

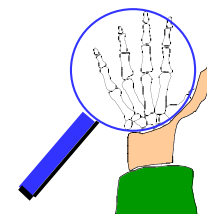
New medicine: disease, cure, inventor, owner, could-I-have-it-too



- Much summarization is template-driven IE.
- Represent templates as lists of Qs → summary skeleton.  
Then summarization = QA over the skeleton's Qs.
- **Challenge:** how to learn the Qs from Reader feedback.

# Summarization as heuristic search

- **Hard case:** no predefined template applies.
- Extraction summarizers use heuristics that exploit
  - nature of genre: presence of titles, abstracts, etc.
  - rules of text structure: position policy (lead sentence).
  - rules of language: cue/stigma words, word counts, etc.
- Heuristics are 2nd-order approximation to Qs—**they model Qs' effects on summary content and structure.**
- **Problem:** how to know when they apply.
- **Problem:** what to do when they don't.
- **Challenge:** we know the heuristics already... now we must understand relationship of Qs to effects.





# Qs and effects

## Extraction heuristics:

1. segment text into units,
2. each heuristic specialist assigns a score to each unit:
  - frequency: from *tf* to language models
  - position: title words, conventionalized structure, rhetorical/discourse structure
  - indicators: cue words (“*note that...*”), format (bold font)...
3. integrate each unit’s scores,
4. return top  $N$  units.

Which (kinds of) fringe Qs are best answered by which heuristics?

Can one automatically construct a suitable heuristic for each Q (type)?

# Human summaries

- For generic summaries, each human summarizer uses own personal fringe.
- Thus: **low inter-human agreement**, after introducing topic  
(SUMMAC-98:  $\kappa_{\text{ad hoc}} = 0.38$ ,  $\kappa_{\text{cat}} = 0.29$ ).
- **Prediction:** higher agreement for shorter summaries.
- **Actually:** who knows?
  - SUMMAC: no data available
  - (Jing et al. 98):
  - DUC01: the opposite (all numbers averaged; 2 humans)

# The fringe as topic keywords

- **Fringe Qs:**
  - approximate them by topic keywords
  - learn/infer them from Reader feedback
- **Summaries:**
  - learn/create Q ‘packets’ as summary skeletons
  - apply QA-style matching/extraction to texts
  - compose passages into summaries
- **Magic:**
  - Qs: extraction of Qs from Reader feedback
  - fringe: representation and organization
  - synthesis: semantic generalization

## Earthquake:

location,  
magnitude,  
number of  
casualties,  
after-effects,  
assistance

## Robbery:

valuables,  
perpetrators,  
owners, police  
action, arrest or  
escape

## New medicine:

disease, cure,  
inventor,  
owner, could-I-  
have-it-too

# The future of summarization

- Systems **integrate summarization and QA**
  - fringe represented as list of Qs (sorted in topic hierarchy?).
- Systems **perform drill-down** with their users
  - and while focusing, systems can record Qs.
- Systems **maintain fringe(s) as user profile(s)**
  - user can edit fringe, adding or deleting topics and questions. System can be user-customized.
- Systems can **follow their own interests (?)**
  - self-motivated learning agents, trying to answer their Qs.

Thank you

# There's nothing new under the sun

(so most stories are not really that interesting)

Provide a single central topic, for context. Hope that it's known.

1. Most stories are an instance or small tweak of something already known

Another earthquake? Another war? Another Mafia Don? Another politician elected? Another disease? Another suspicious character?

2. Some stories include a significant variation or extension  
The Bush-Gore election

3. Some stories provide novel answers to procedural Qs  
A new way to avoid meetings! Avoid losing your luggage!