Segmentation & Wrap-up: Q-A

Ling571
Deep Processing Techniques for NLP
March 11, 2015

TextTiling (Hearst '97)

- Lexical cohesion-based segmentation
  - Boundaries at dips in cohesion score
  - Tokenization, Lexical cohesion score, Boundary ID
- Tokenization
  - White-space delimited words
  - Stopped
  - Stemmed
  - 20 words = 1 pseudo sentence

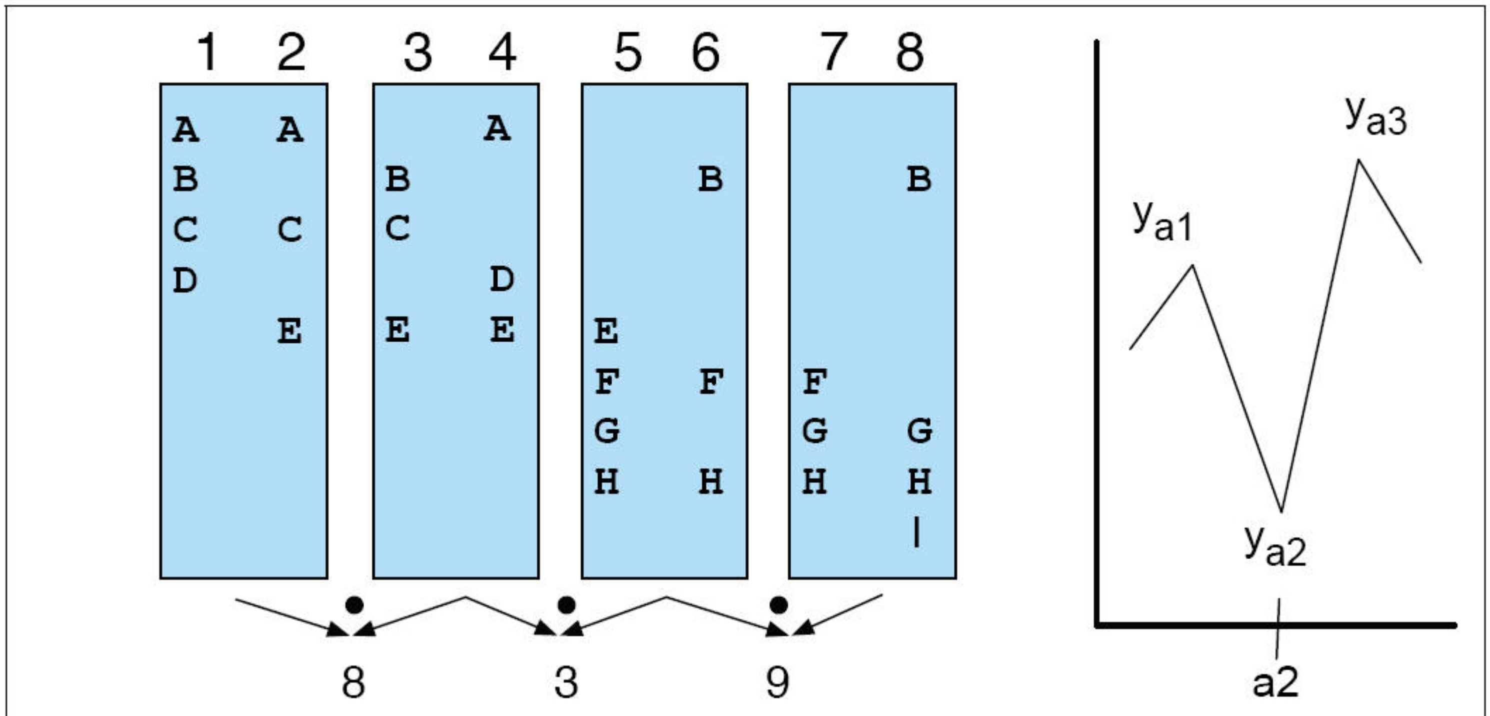
Lexical Cohesion Score

- Similarity between spans of text
  - b = 'Block' of 10 pseudo-sentences before gap
  - a = 'Block' of 10 pseudo-sentences after gap
  - How do we compute similarity?
    - Vectors and cosine similarity (again!)

$$sim_{cosine}(\vec{b},\vec{a}) = \frac{\vec{b} \bullet \vec{a}}{|\vec{b}||\vec{a}|} = \frac{\sum_{i=1}^{N} b_i \times a_i}{\sqrt{\sum_{i=1}^{N} b_i^2}\sqrt{\sum_{i=1}^{N} a_i^2}}$$

Segmentation

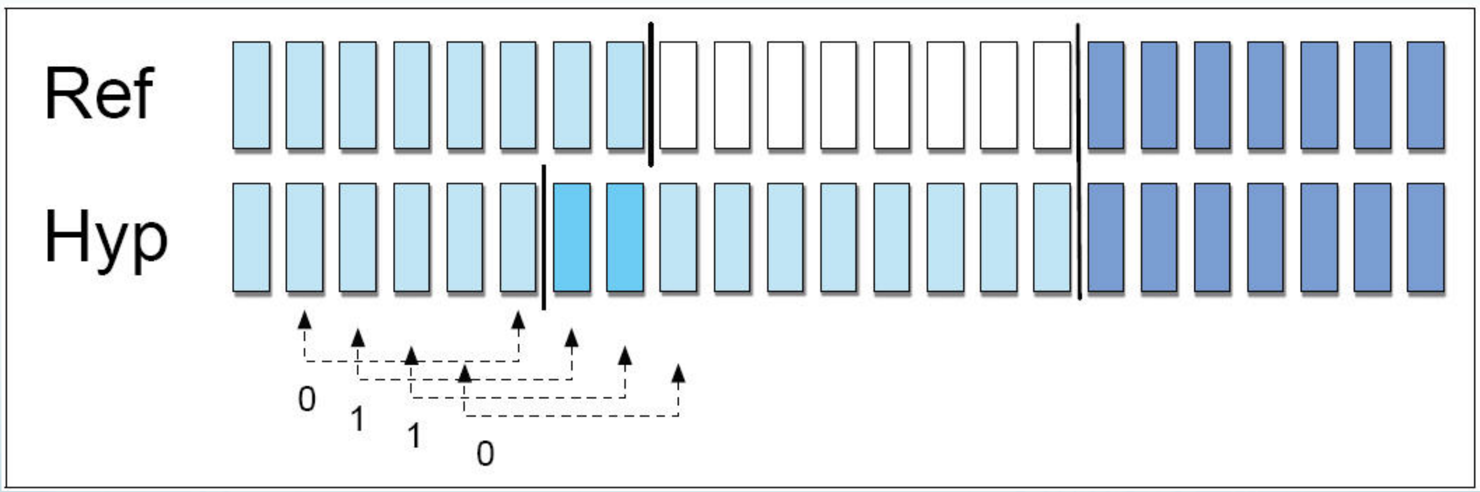
- Depth score:
  - Difference between position and adjacent peaks
  - E.g., $(y_{a1}-y_{a2})+(y_{a3}-y_{a2})$



(a) (b)

Evaluation

- How about precision/recall/F-measure?
  - Problem: No credit for near-misses
- Alternative model: WindowDiff



$$WindowDiff(ref,hyp) = \frac{1}{N-k}\sum_{i=1}^{N-k}\left(\left|b(ref_i, ref_{i+k}) - b(hyp_i, hyp_{i+k})\right| \neq 0\right)$$

Discussion

- Overall: Auto much better than random
- Issues

Discussion

- Overall: Auto much better than random
- Issues: Summary material
  - Often not similar to adjacent paras
- Similarity measures

Discussion

- Overall: Auto much better than random
- Issues: Summary material
  - Often not similar to adjacent paras
- Similarity measures
  - Is raw tf the best we can do?
  - Other cues??
- Other experiments with TextTiling perform less well – Why?

Summary

- Computational discourse:
  - Cohesion and Coherence in extended spans

  - Key tasks:
    - Reference resolution
      - Constraints and preferences
      - Heuristic, learning, and sieve models

    - Discourse structure modeling
      - Linear topic segmentation
    - Exploiting shallow and deep language processing

Question-Answering: Shallow & Deep Techniques for NLP

Deep Processing Techniques for NLP
Ling 571
March 11, 2015

(Examples from Dan Jurafsky)

Roadmap

- Question-Answering:
  - Definitions & Motivation
- Basic pipeline:
  - Question processing
  - Retrieval
  - Answering processing
- Shallow processing: Aranea (Lin, Brill)
- Deep processing: LCC (Moldovan, Harabagiu, et al)
- Wrap-up

Why QA?

- Grew out of information retrieval community
- Web search is great, but...
  - Sometimes you don't just want a ranked list of documents
  - Want an answer to a question!
    - Short answer, possibly with supporting context

Why QA?

- Grew out of information retrieval community
- Web search is great, but...
  - Sometimes you don't just want a ranked list of documents
  - Want an answer to a question!
    - Short answer, possibly with supporting context
- People ask questions on the web
  - Web logs:
    - Which English translation of the bible is used in official Catholic liturgies?
    - Who invented surf music?
    - What are the seven wonders of the world?

Why QA?

- Grew out of information retrieval community
- Document retrieval is great, but…
  - Sometimes you don't just want a ranked list of documents
  - Want an answer to a question!
    - Short answer, possibly with supporting context
- People ask questions on the web
  - Web logs:
    - Which English translation of the bible is used in official Catholic liturgies?
    - Who invented surf music?
    - What are the seven wonders of the world?
  - Account for 12-15% of web log queries

Search Engines and Questions

- What do search engines do with questions?

Search Engines and Questions

- What do search engines do with questions?
  - Increasingly try to answer questions
    - Especially for wikipedia infobox types of info
  - Backs off to keyword search
- How well does this work?

Search Engines and Questions

- What do search engines do with questions?
  - Increasingly try to answer questions
    - Especially for wikipedia infobox types of info
  - Backs off to keyword search
- How well does this work?
  - *Who invented surf music?*

Search Engines and Questions

- What do search engines do with questions?
  - Increasingly try to answer questions
    - Especially for wikipedia infobox types of info
  - Backs off to keyword search
- How well does this work?
  - *Who invented surf music?*
    - Rank #2 snippet:
      - Dick Dale *invented surf music*
    - Pretty good, but…

Search Engines & QA

- *Who was the prime minister of Australia during the Great Depression?*

Search Engines & QA

- *Who was the prime minister of Australia during the Great Depression?*
  - Rank 1 snippet:
    - The conservative *Prime Minister of Australia*, Stanley Bruce

Search Engines & QA

- *Who was the prime minister of Australia during the Great Depression?*
  - Rank 1 snippet:
    - The conservative *Prime Minister of Australia*, Stanley Bruce
  - Wrong!
    - Voted out just before the Depression
- *What is the total population of the ten largest capitals in the US?*

Search Engines & QA

- ***Who was the prime minister of Australia during the Great Depression?***
  - Rank 1 snippet:
    - The conservative *Prime Minister of Australia*, Stanley Bruce
  - Wrong!
    - Voted out just before the Depression
- ***What is the total population of the ten largest capitals in the US?***
  - Rank 1 snippet:
    - The table below lists the *largest* 50 *cities in the United States* .....

Search Engines & QA

- *Who was the prime minister of Australia during the Great Depression?*
  - Rank 1 snippet:
    - The conservative *Prime Minister of Australia*, Stanley Bruce
  - Wrong!
    - Voted out just before the Depression

- *What is the total population of the ten largest capitals in the US?*
  - Rank 1 snippet:
    - The table below lists the *largest* 50 *cities in the United States* .....
  - The answer is in the document – with a calculator..

Search Engines and QA

Search Engines and QA

- Search for exact question string
  - "Do I need a visa to go to Japan?"
    - Result: Exact match on Yahoo! Answers

    - Find 'Best Answer' and return following chunk

Search Engines and QA

- Search for exact question string
  - “Do I need a visa to go to Japan?”
    - Result: Exact match on Yahoo! Answers

    - Find ‘Best Answer’ and return following chunk

- Works great if the question matches exactly
    - Many websites are building archives
  - What if it doesn’t match?

Search Engines and QA

- Search for exact question string
  - "Do I need a visa to go to Japan?"
    - Result: Exact match on Yahoo! Answers
    - Find 'Best Answer' and return following chunk
- Works great if the question matches exactly
    - Many websites are building archives
  - What if it doesn't match?
    - 'Question mining' tries to learn paraphrases of questions to get answer

Perspectives on QA

- TREC QA track (~2000---)
  - Initially pure factoid questions, with fixed length answers
    - Based on large collection of fixed documents (news)
    - Increasing complexity: definitions, biographical info, etc
      - Single response

Perspectives on QA

- TREC QA track (~2000---)
  - Initially pure factoid questions, with fixed length answers
    - Based on large collection of fixed documents (news)
    - Increasing complexity: definitions, biographical info, etc
      - Single response
- Reading comprehension (Hirschman et al, 2000---)
  - Think SAT/GRE
    - Short text or article (usually middle school level)
    - Answer questions based on text
  - Also, 'machine reading'

Perspectives on QA

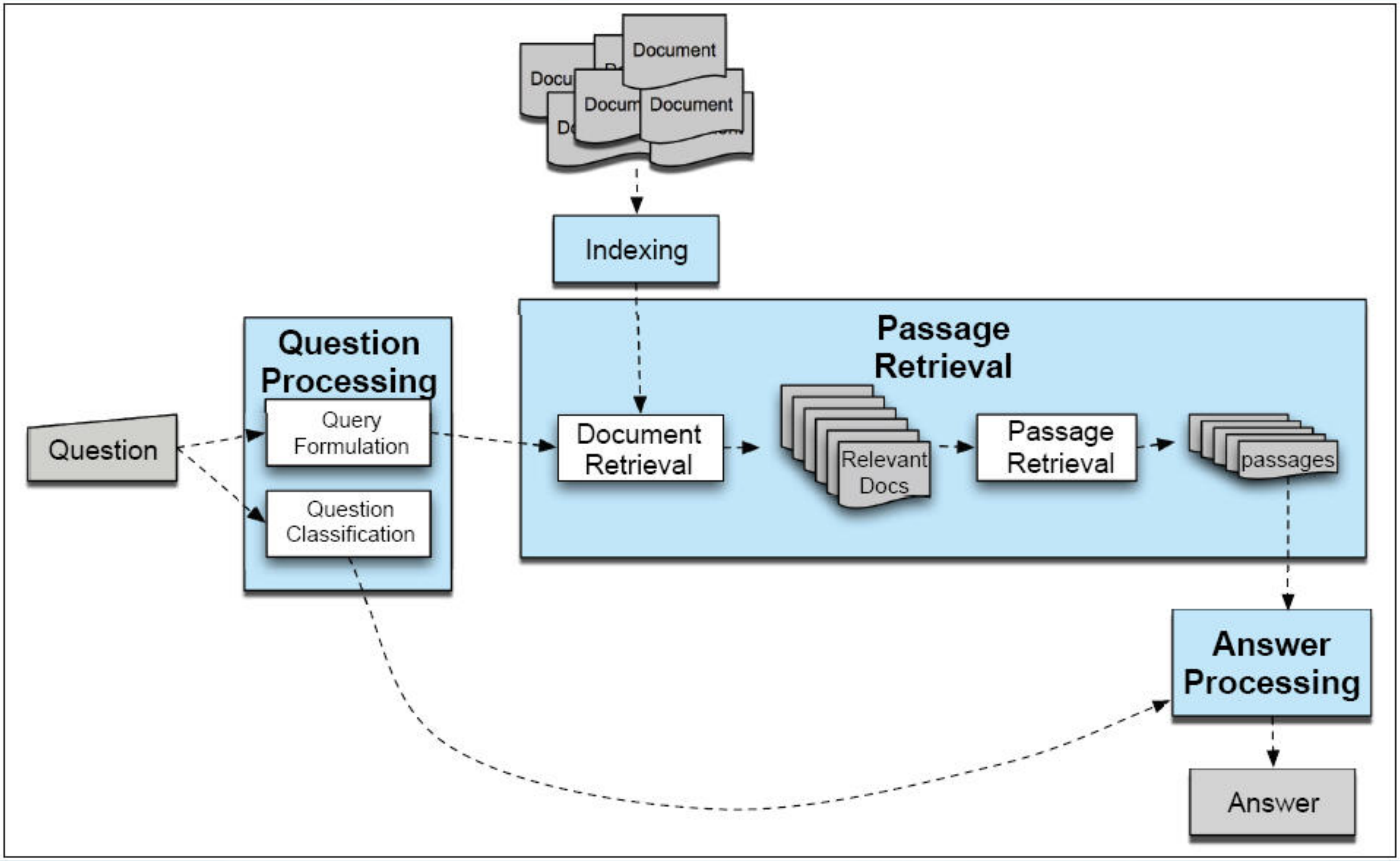
- TREC QA track (~2000---)
  - Initially pure factoid questions, with fixed length answers
    - Based on large collection of fixed documents (news)
    - Increasing complexity: definitions, biographical info, etc
      - Single response
- Reading comprehension (Hirschman et al, 2000---)
  - Think SAT/GRE
    - Short text or article (usually middle school level)
    - Answer questions based on text
  - Also, 'machine reading'
- And, of course, Jeopardy! and Watson

Question Answering (a la TREC)

Question	Answer
Where is the Louvre Museum located?	in Paris, France
What's the abbreviation for limited partnership?	L.P.
What are the names of Odin's ravens?	Huginn and Muninn
What currency is used in China?	the yuan
What kind of nuts are used in marzipan?	almonds
What instrument does Max Roach play?	drums
What's the official language of Algeria?	Arabic
What is the telephone number for the University of Colorado, Boulder?	(303)492-1411
How many pounds are there in a stone?	14

Basic Strategy

- Given an indexed document collection, and
- A question:
- Execute the following steps:
  - Query formulation
  - Question classification
  - Passage retrieval
  - Answer processing
  - Evaluation



Document
Document
Document
Indexing
Question
Question Processing
Query Formulation
Question Classification
Passage Retrieval
Document Retrieval
Relevant Docs
Passage Retrieval
passages
Answer Processing
Answer

Query Processing

- Query reformulation
  - Convert question to suitable form for IR

Query Processing

- Query reformulation
  - Convert question to suitable form for IR
    - E.g. ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs

Query Processing

- Query reformulation
  - Convert question to suitable form for IR
    - E.g. ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs
- Question classification
  - Answer type recognition
    - Who

Query Processing

- Query reformulation
  - Convert question to suitable form for IR
    - E.g. ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs
- Question classification
  - Answer type recognition
    - Who → Person; What Canadian city →

Query Processing

- Query reformulation
  - Convert question to suitable form for IR
    - E.g. ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs
- Question classification
  - Answer type recognition
    - Who → Person; What Canadian city → City
    - What is surf music →

Query Processing

- Query reformulation
  - Convert question to suitable form for IR
    - E.g. 'stop structure' removal:
      - Delete function words, q-words, even low content verbs
- Question classification
  - Answer type recognition
    - Who → Person; What Canadian city → City
    - What is surf music →Definition
  - Train classifiers to recognize expected answer type
    - Using POS, NE, words, synsets, hyper/hypo-nyms

Tag	Example
ABBREVIATION	
abb	What's the abbreviation for limited partnership?
exp	What does the "c" stand for in the equation E=mc2?
DESCRIPTION	
definition	What are tannins?
description	What are the words to the Canadian National anthem?
manner	How can you get rust stains out of clothing?
reason	What caused the Titanic to sink ?
ENTITY	
animal	What are the names of Odin's ravens?
body	What part of your body contains the corpus callosum?
color	What colors make up a rainbow ?
creative	In what book can I find the story of Aladdin?
currency	What currency is used in China?
disease/medicine	What does Salk vaccine prevent?
event	What war involved the battle of Chapultepec?
food	What kind of nuts are used in marzipan?
instrument	What instrument does Max Roach play?
lang	What's the official language of Algeria?
letter	What letter appears on the cold-water tap in Spain?
other	What is the name of King Arthur's sword?
plant	What are some fragrant white climbing roses?
product	What is the fastest computer?
religion	What religion has the most members?
sport	What was the name of the ball game played by the Mayans?
substance	What fuel do airplanes use?
symbol	What is the chemical symbol for nitrogen?
technique	What is the best way to remove wallpaper?
term	How do you say " Grandma " in Irish?
vehicle	What was the name of Captain Bligh's ship?
word	What's the singular of dice?

HUMAN	
description	Who was Confucius?
group	What are the major companies that are part of Dow Jones?
ind	Who was the first Russian astronaut to do a spacewalk?
title	What was Queen Victoria's title regarding India?
LOCATION	
city	What's the oldest capital city in the Americas?
country	What country borders the most others?
mountain	What is the highest peak in Africa?
other	What river runs through Liverpool?
state	What states do not have state income tax?
NUMERIC	
code	What is the telephone number for the University of Colorado?
count	About how many soldiers died in World War II?
date	What is the date of Boxing Day?
distance	How long was Mao's 1930s Long March?
money	How much did a McDonald's hamburger cost in 1963?
order	Where does Shanghai rank among world cities in population?
other	What is the population of Mexico?
period	What was the average life expectancy during the Stone Age?
percent	What fraction of a beaver's life is spent swimming?
speed	What is the speed of the Mississippi River?
temp	How fast must a spacecraft travel to escape Earth's gravity?
size	What is the size of Argentina?
weight	How many pounds are there in a stone?

Passage Retrieval

- Why not just perform general information retrieval?

Passage Retrieval

- Why not just perform general information retrieval?
  - Documents too big, non-specific for answers
- Identify shorter, focused spans (e.g., sentences)

Passage Retrieval

- Why not just perform general information retrieval?
  - Documents too big, non-specific for answers
- Identify shorter, focused spans (e.g., sentences)
  - Filter for correct type: answer type classification
  - Rank passages based on a trained classifier
  - Or, for web search, use result snippets

Answer Processing

- Find the specific answer in the passage

Answer Processing

- Find the specific answer in the passage
- Pattern extraction-based:
  - Include answer types, regular expressions
    - Can use syntactic/dependency/semantic patterns

Answer Processing

- Find the specific answer in the passage
- Pattern extraction-based:
  - Include answer types, regular expressions
    - Can use syntactic/dependency/semantic patterns
    - Leverage large knowledge bases

Pattern	**Question**	**Answer**
<AP> such as <QP>	What is autism?	“, developmental disorders such as autism”
<QP>, a <AP>	What is a caldera?	“the Long Valley caldera, a volcanic crater 19 miles long”

Evaluation

- Classical:
  - Return ranked list of answer candidates

Evaluation

- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score

  - Measure: Mean Reciprocal Rank (MRR)

Evaluation

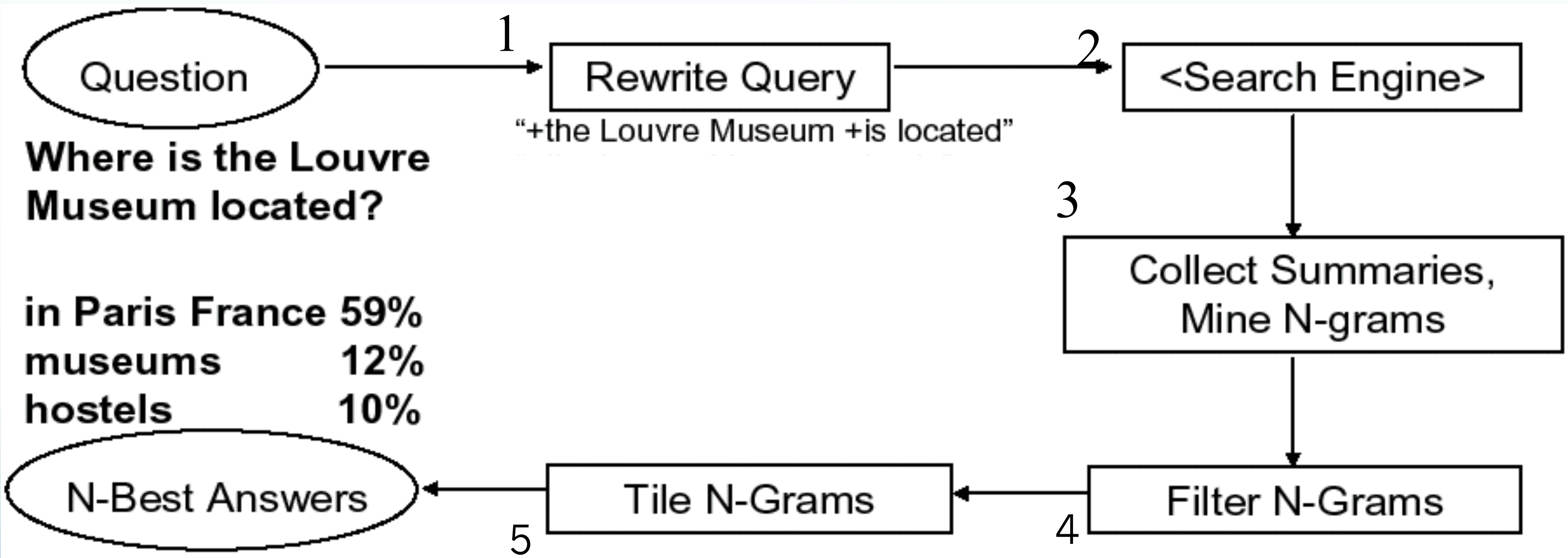
- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score

  - Measure: Mean Reciprocal Rank (MRR)
    - For each question,
      - Get reciprocal of rank of first correct answer
        - E.g. correct answer is 4 => ¼
        - None correct => 0
    - Average over all questions

$$MRR = \frac{\sum_{i=1}^{N} \frac{1}{rank_i}}{N}$$

AskMSR/Aranea (Lin, Brill)

- Shallow Processing for QA



Intuition

- Redundancy is useful!
  - If similar strings appear in many candidate answers, likely to be solution
    - Even if can't find obvious answer strings

Intuition

- Redundancy is useful!
  - If similar strings appear in many candidate answers, likely to be solution
    - Even if can't find obvious answer strings
- Q: How many times did Bjorn Borg win Wimbledon?
    - Bjorn Borg blah blah blah Wimbledon blah 5 blah
    - Wimbledon blah blah blah Bjorn Borg blah 37 blah.
    - blah Bjorn Borg blah blah 5 blah blah Wimbledon
    - 5 blah blah Wimbledon blah blah Bjorn Borg.

Intuition

- Redundancy is useful!
  - If similar strings appear in many candidate answers, likely to be solution
    - Even if can't find obvious answer strings
- Q: How many times did Bjorn Borg win Wimbledon?
    - Bjorn Borg blah blah blah Wimbledon blah 5 blah
    - Wimbledon blah blah blah Bjorn Borg blah 37 blah.
    - blah Bjorn Borg blah blah 5 blah blah Wimbledon
    - 5 blah blah Wimbledon blah blah Bjorn Borg.
  - Probably 5

Query Reformulation

- Identify question type:
  - E.g. Who, When, Where,...
- Create question-type specific rewrite rules:

Query Reformulation

- Identify question type:
  - E.g. Who, When, Where,…
- Create question-type specific rewrite rules:
  - Hypothesis: Wording of question similar to answer
    - For ‘where’ queries, move ‘is’ to all possible positions
      - Where is the Louvre Museum located? =>
        - Is the Louvre Museum located
        - The is Louvre Museum located
        - The Louvre Museum is located, .etc.

Query Reformulation

- Identify question type:
  - E.g. Who, When, Where,...
- Create question-type specific rewrite rules:
  - Hypothesis: Wording of question similar to answer
    - For 'where' queries, move 'is' to all possible positions
      - Where is the Louvre Museum located? =>
        - Is the Louvre Museum located
        - The is Louvre Museum located
        - The Louvre Museum is located, .etc.
- Create type-specific answer type (Person, Date, Loc)

Retrieval, N-gram Mining & Filtering

- Run reformulated queries through search engine
  - Collect (lots of) result snippets

Retrieval, N-gram Mining & Filtering

- Run reformulated queries through search engine
  - Collect (lots of) result snippets
  - Collect n-grams from snippets

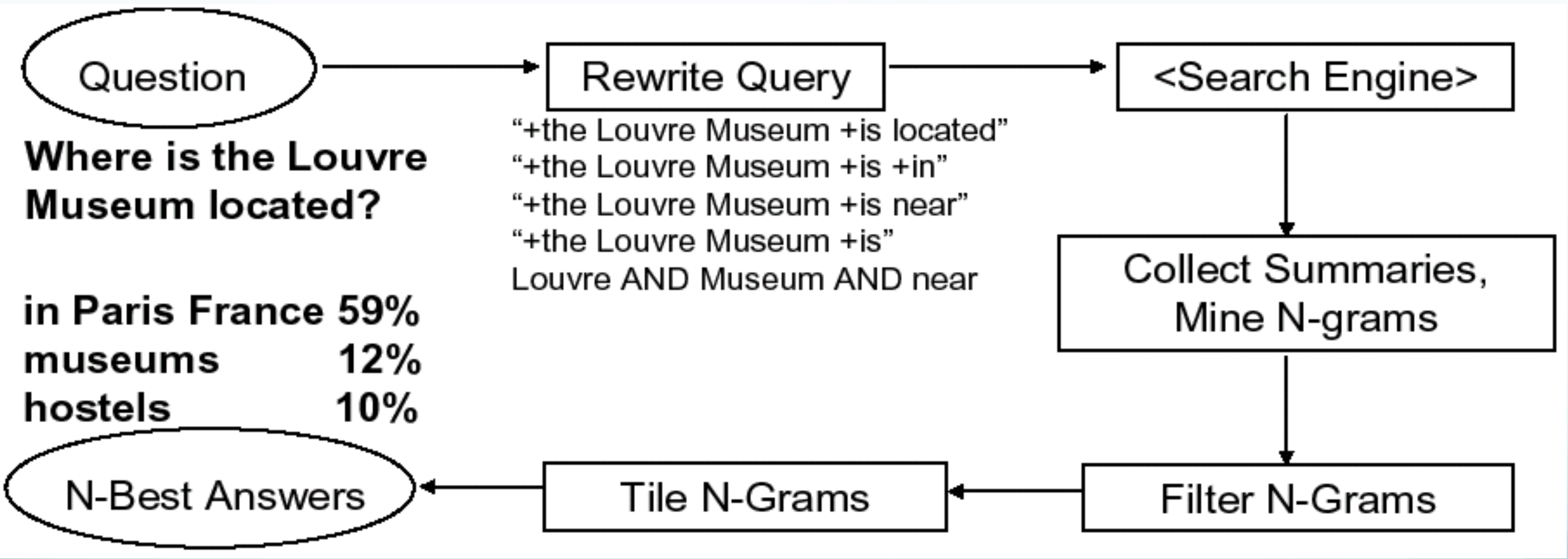
Retrieval, N-gram Mining & Filtering

- Run reformulated queries through search engine
  - Collect (lots of) result snippets
  - Collect n-grams from snippets
  - Weight each n-gram summing over occurrences

Retrieval, N-gram Mining & Filtering

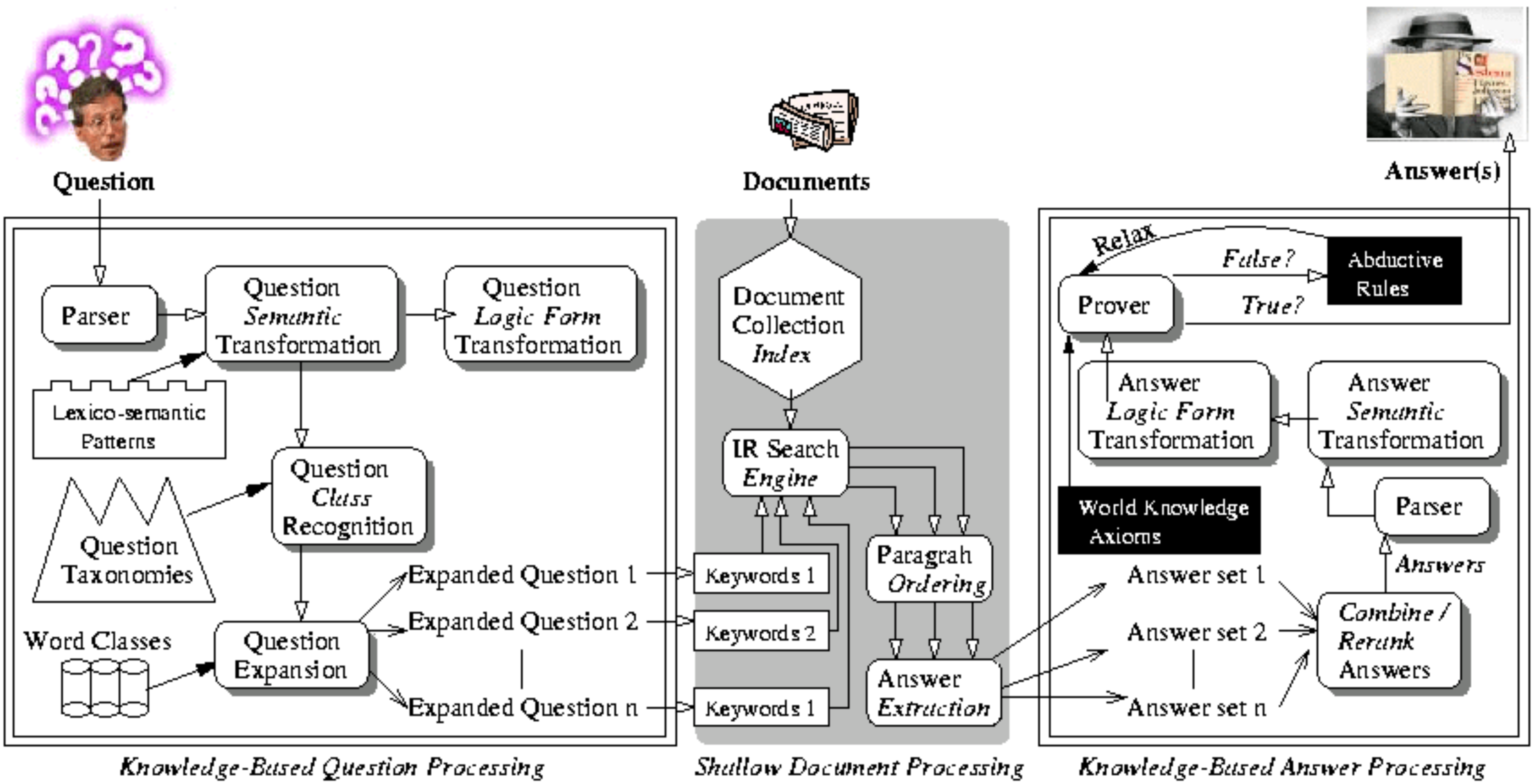
- Run reformulated queries through search engine
  - Collect (lots of) result snippets
  - Collect n-grams from snippets
  - Weight each n-gram summing over occurrences
  - Concatenate n-grams into longer answers
    - E.g. Dickens, Charles Dickens, Mr. Charles ➔
      - Mr. Charles Dickens

Example Redux



Deep Processing Technique for QA

- LCC, PowerAnswer, Qanda (Moldovan, Harabagiu, et al)



Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess

Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess
- Parsing creates syntactic representation:
  - Focused on nouns, verbs, and particles
    - Attachment

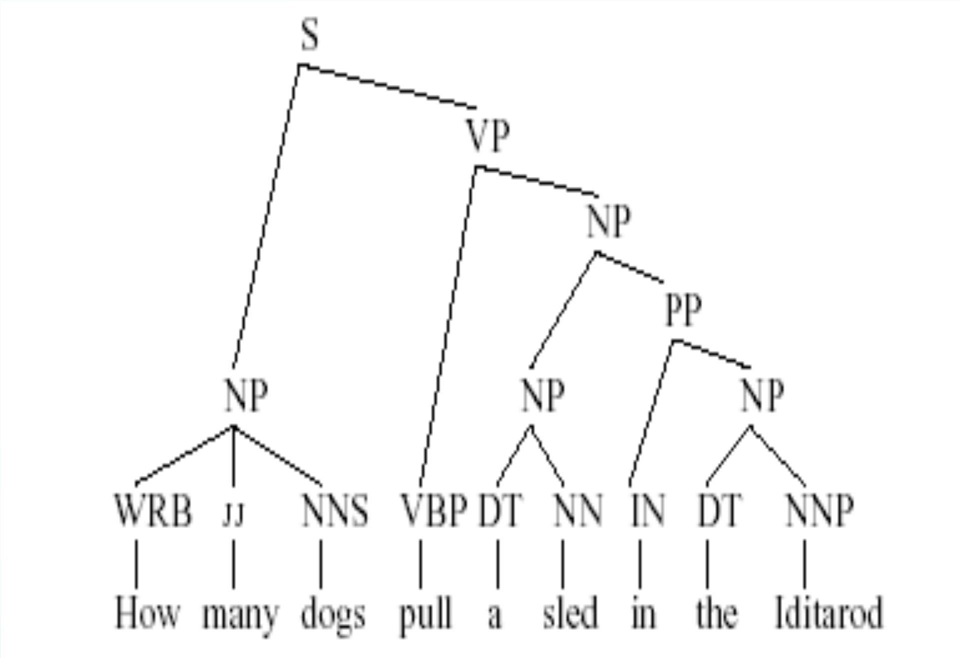
Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess
- Parsing creates syntactic representation:
  - Focused on nouns, verbs, and particles
    - Attachment
- Coreference resolution links entity references

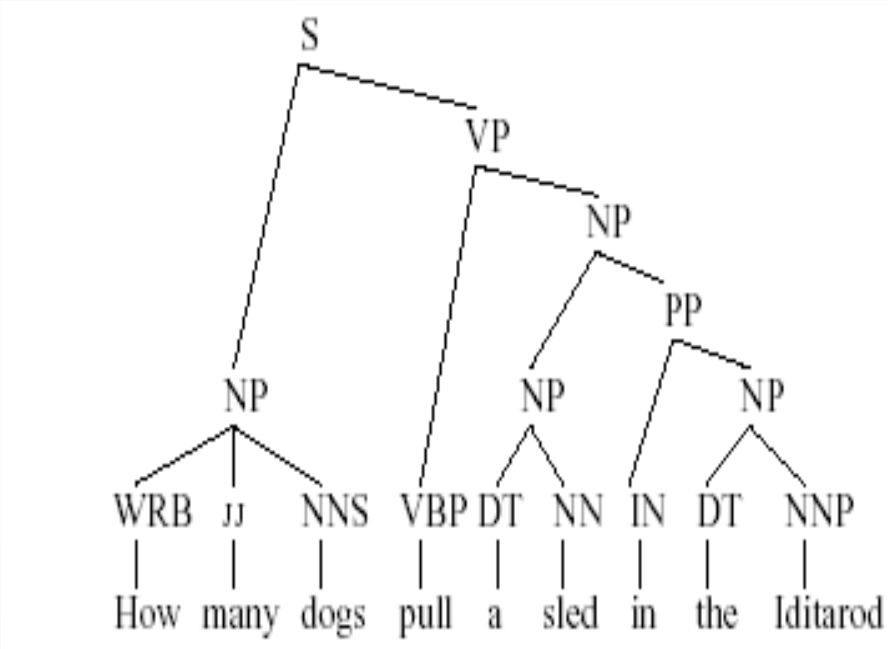
Deep Processing: Query/Answer Formulation

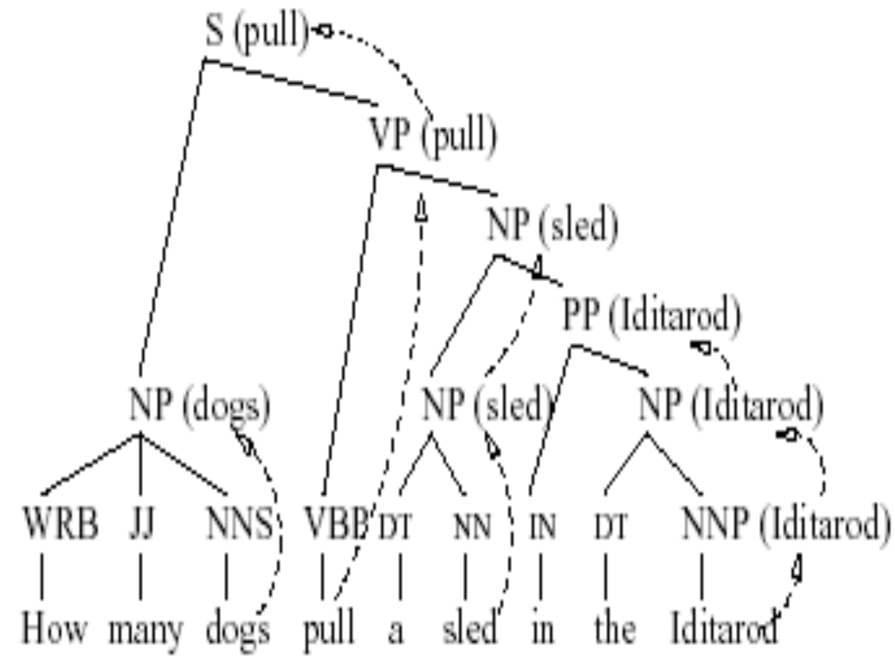
- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess
- Parsing creates syntactic representation:
  - Focused on nouns, verbs, and particles
    - Attachment
- Coreference resolution links entity references
- Translate to full logical form
  - As close as possible to syntax

Syntax to Logical Form

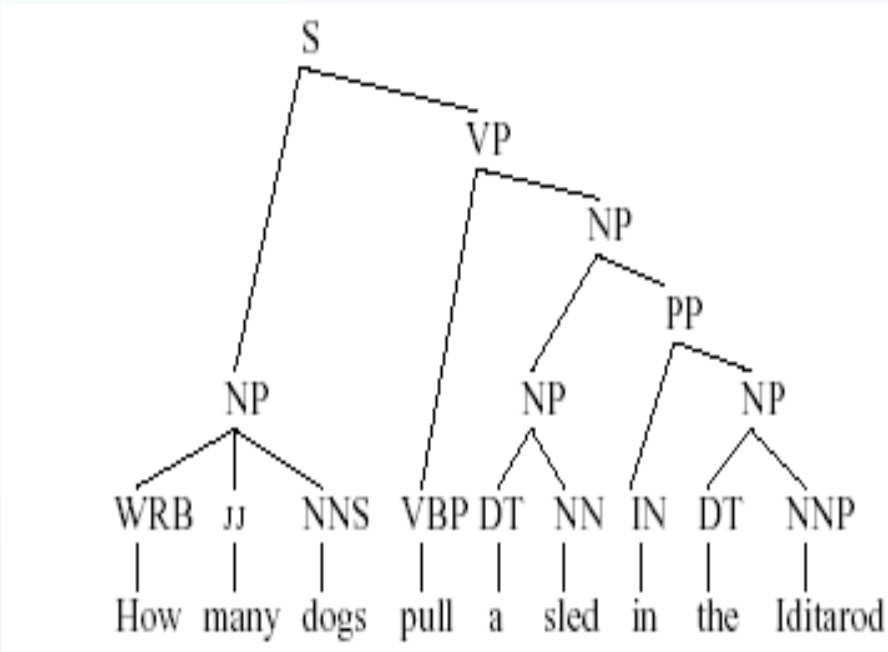


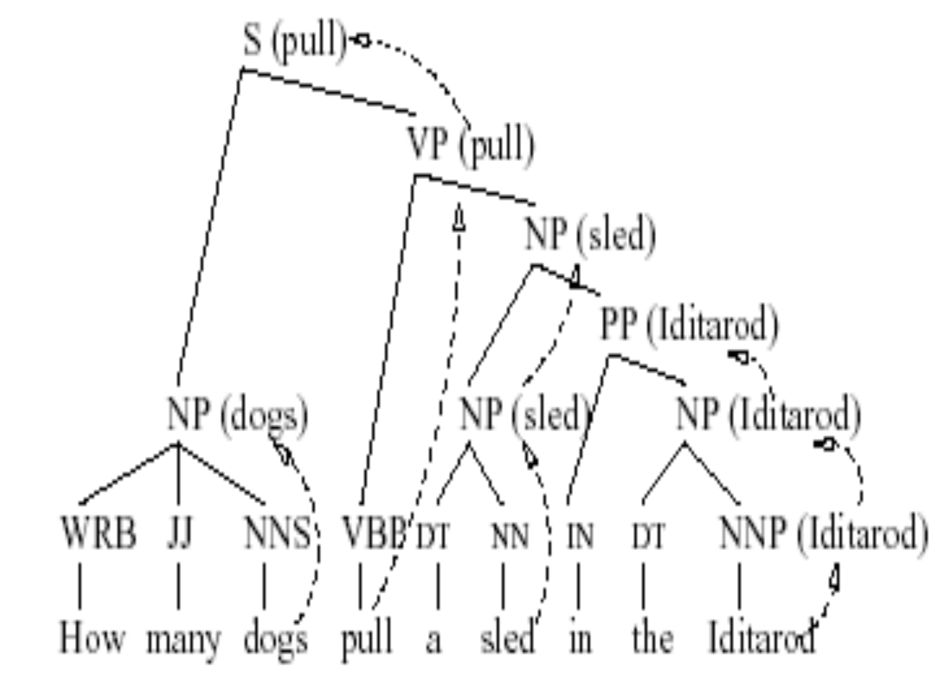
Syntax to Logical Form

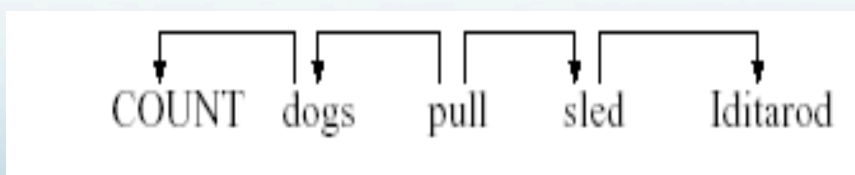




Syntax to Logical Form







Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection

Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
  - Create connections between synsets through topicality
    - *Q: When was the internal combustion engine invented?*
    - *A: The first internal-combustion engine was built in 1867.*
    - invent → create_mentally → create → build

Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
  - Create connections via WordNet synsets
    - *Q: When was the internal combustion engine invented?*
    - *A: The first internal-combustion engine was built in 1867.*
    - invent → create_mentally → create → build

- Perform abductive reasoning
  - Tries to justify answer given question

Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
  - Create connections via WordNet synsets
    - *Q: When was the internal combustion engine invented?*
    - *A: The first internal-combustion engine was built in 1867.*
    - invent → create_mentally → create → build
- Perform abductive reasoning
  - Tries to justify answer given question
  - Yields 30% improvement in accuracy!

Question Answering Example

- How hot does the inside of an active volcano get?
  - get(TEMPERATURE, inside(volcano(active)))

Question Answering Example

- How hot does the inside of an active volcano get?
  - get(TEMPERATURE, inside(volcano(active)))
- "lava fragments belched out of the mountain were as hot as 300 degrees Fahrenheit"
  - fragments(lava, TEMPERATURE(degrees(300)),
    belched(out, mountain))
  - volcano ISA mountain
  - lava ISPARTOF volcano
    - lava inside volcano
  - fragments of lava HAVEPROPERTIESOF lava

Question Answering Example

- How hot does the inside of an active volcano get?
  - get(TEMPERATURE, inside(volcano(active)))
- "lava fragments belched out of the mountain were as hot as 300 degrees Fahrenheit"
  - fragments(lava, TEMPERATURE(degrees(300)),
    belched(out, mountain))
  - volcano ISA mountain
  - lava ISPARTOF volcano
    - lava inside volcano
  - fragments of lava HAVEPROPERTIESOF lava
- The needed semantic information is in WordNet definitions, and was successfully translated into a form that was used for rough 'proofs'

A Victory for Deep Processing

Run Tag	Score	#	%	Inexact	Prec	Recall
LCCmain2002	0.856	415	83.0	8	0.578	0.804
exactanswer	0.691	271	54.2	12	0.222	0.848
pris2002	0.610	290	58.0	17	0.241	0.891
IRST02D1	0.589	192	38.4	17	0.167	0.217
IBMPQSQACYC	0.588	179	35.8	9	0.196	0.630
uwmtB3	0.512	184	36.8	20	0.000	0.000
BBN2002C	0.499	142	28.4	18	0.182	0.087
isi02	0.498	149	29.8	15	0.385	0.109
limsiQalir2	0.497	133	26.6	11	0.188	0.196
ali2002b	0.496	181	36.2	15	0.156	0.848
ibmsqa02c	0.455	145	29.0	44	0.224	0.239
FDUT11QA1	0.434	124	24.8	6	0.139	0.957
aranea02a	0.433	152	30.4	36	0.235	0.174
nuslamp2002	0.396	105	21.0	17	0.000	0.000

Aranea: 0.30 on TREC data; 0.42 on TREC queries w/full web

Conclusions

- Deep processing for QA
  - Exploits parsing, semantics, anaphora, reasoning
  - Computationally expensive
    - But tractable because applied only to
    - Questions and Passages

Conclusions

- Deep processing for QA
  - Exploits parsing, semantics, anaphora, reasoning
  - Computationally expensive
    - But tractable because applied only to
    - Questions and Passages
- Trends:
  - Systems continue to make greater use of
    - Web resources: Wikipedia, answer repositories
    - Machine learning!!!!

Summary

- Deep processing techniques for NLP
  - Parsing, semantic analysis, logical forms, reference,etc
  - Create richer computational models of natural language
    - Closer to language understanding

Summary

- Deep processing techniques for NLP
  - Parsing, semantic analysis, logical forms, reference,etc
  - Create richer computational models of natural language
    - Closer to language understanding
- Shallow processing techniques have dominated many areas
  - IR, QA, MT, WSD, etc
    - More computationally tractable, fewer required resources

Summary

- Deep processing techniques for NLP
  - Parsing, semantic analysis, logical forms, reference,etc
  - Create richer computational models of natural language
    - Closer to language understanding
- Shallow processing techniques have dominated many areas
  - IR, QA, MT, WSD, etc
    - More computationally tractable, fewer required resources
- Deep processing techniques experiencing resurgence
  - Some big wins – e.g. QA

Summary

- Deep processing techniques for NLP
  - Parsing, semantic analysis, logical forms, reference,etc
  - Create richer computational models of natural language
    - Closer to language understanding
- Shallow processing techniques have dominated many areas
  - IR, QA, MT, WSD, etc
    - More computationally tractable, fewer required resources
- Deep processing techniques experiencing resurgence
  - Some big wins – e.g. QA
  - Improved resources: treebanks (syn/disc, Framenet, Propbank)
  - Improved learning algorithms: structured learners,…
  - Increased computation: cloud resources, Grid, etc

Notes

- Last assignment posted – Due March 18

- Course evaluation web page posted:
  - Please respond!

- THANK YOU!

Text Coherence

- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse

Text Coherence

- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse
  - Examples:
    - **Result:** Infer state of S_0 cause state in S_1
      - The Tin Woodman was caught in the rain. His joints rusted.

Text Coherence

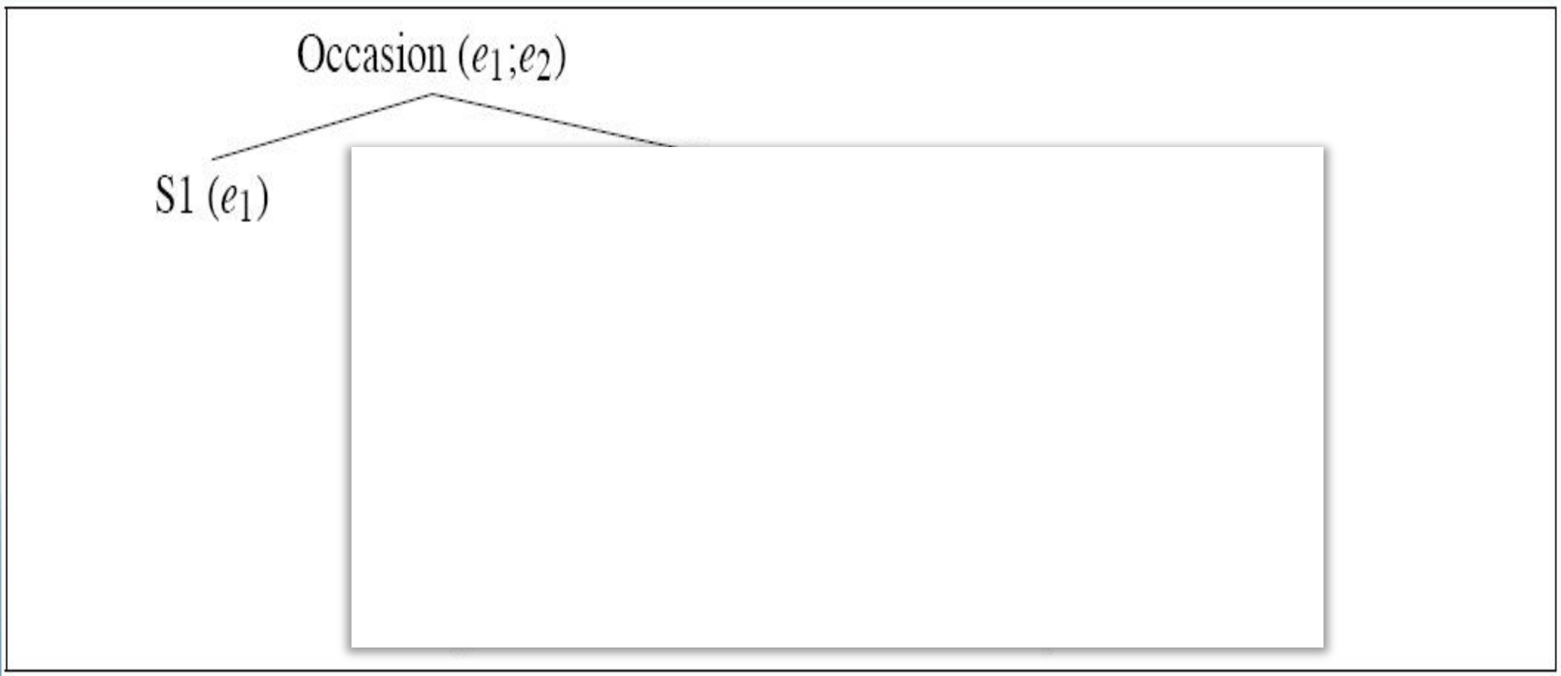
- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse
  - Examples:
    - **Result:** Infer state of S_0 cause state in S_1
      - The Tin Woodman was caught in the rain. His joints rusted.
    - **Explanation**: Infer state in S_1 causes state in S_0
      - John hid Bill's car keys. He was drunk.

Text Coherence

- Cohesion – repetition, etc – does not imply coherence
- Coherence relations:
  - Possible meaning relations between utts in discourse
  - Examples:
    - **Result:** Infer state of S_0 cause state in S_1
      - The Tin Woodman was caught in the rain. His joints rusted.
    - **Explanation**: Infer state in S_1 causes state in S_0
      - John hid Bill's car keys. He was drunk.
    - **Elaboration**: Infer same prop. from S_0 and S_1.
      - Dorothy was from Kansas. She lived in the great Kansas prairie.
  - Pair of locally coherent clauses: discourse segment

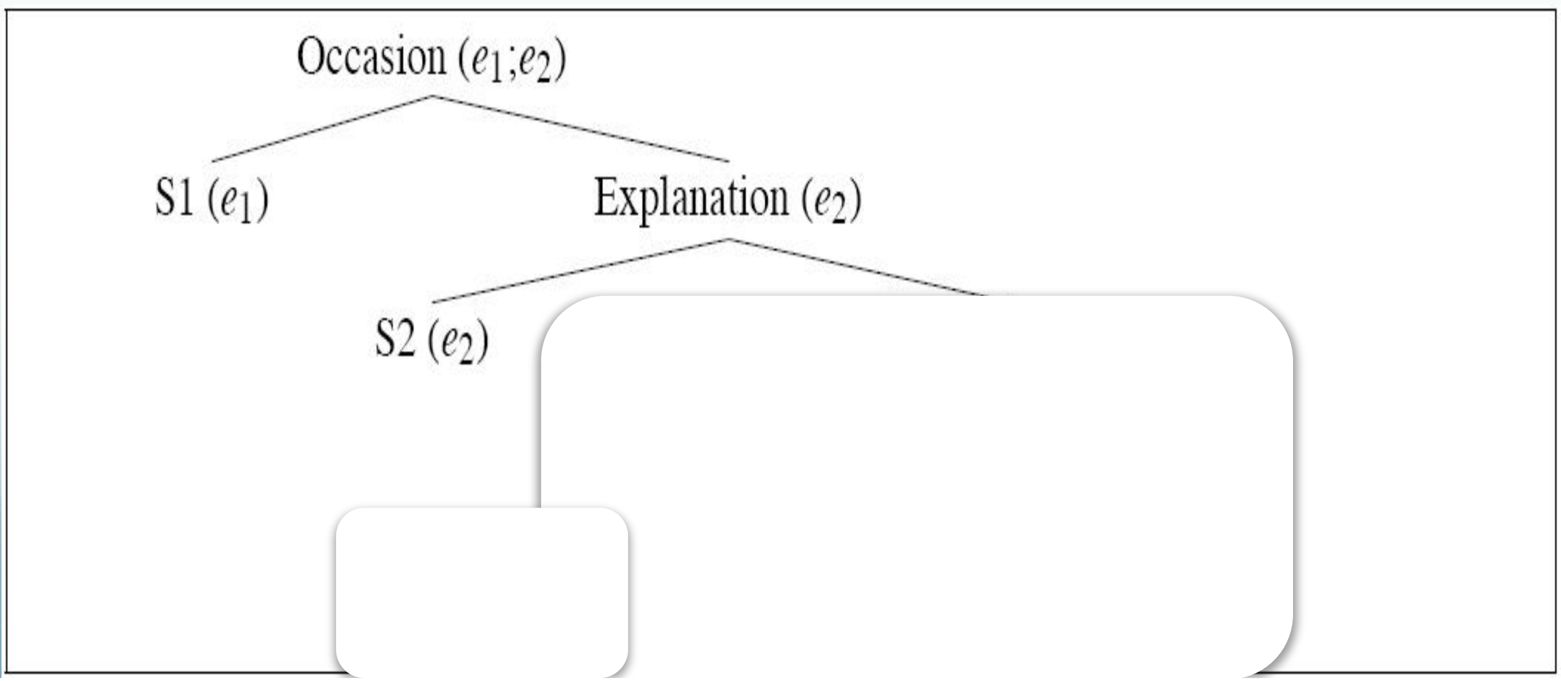
Coherence Analysis

S1: John went to the bank to deposit his paycheck.
S2: He then took a train to Bill’s car dealership.
S3: He needed to buy a car.
S4: The company he works now isn’t near any public transportation.
S5: He also wanted to talk to Bill about their softball league.



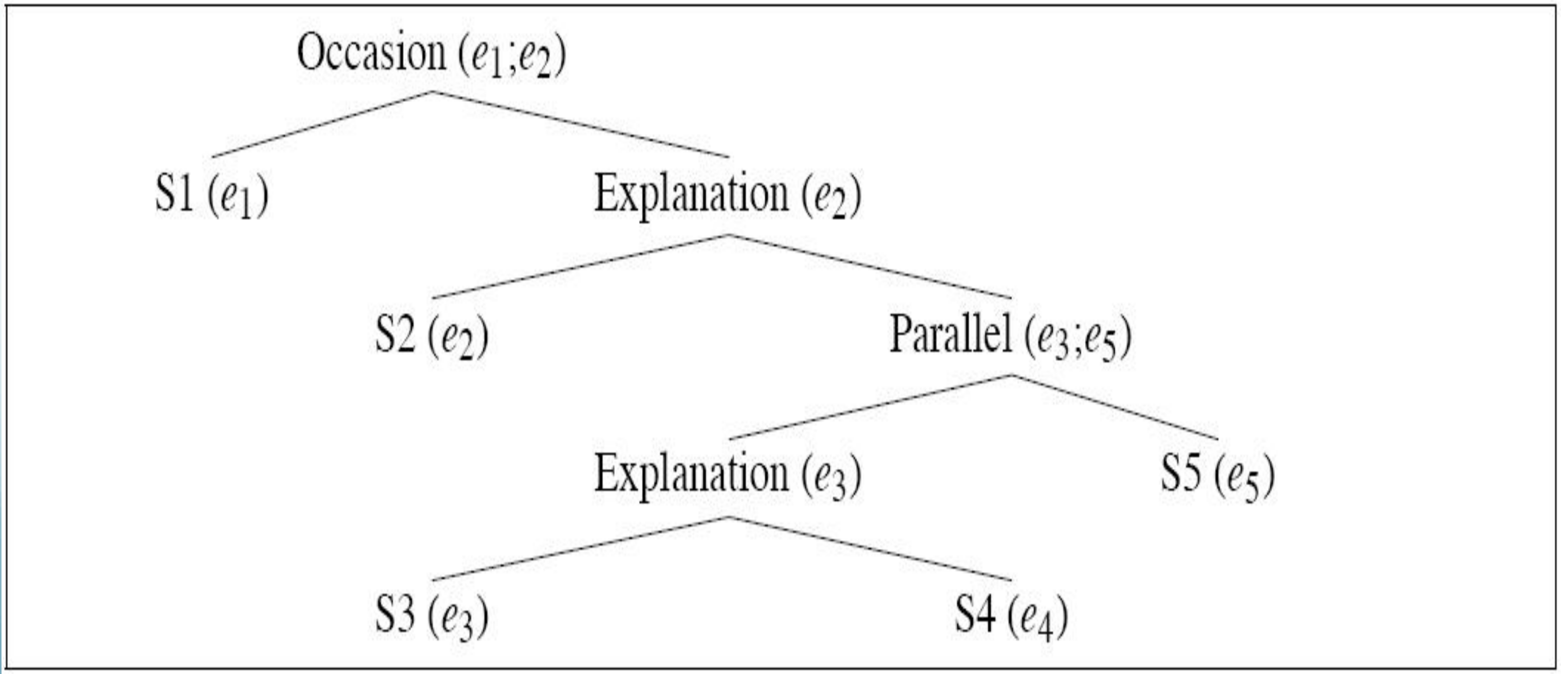
Coherence Analysis

S1: John went to the bank to deposit his paycheck.
S2: He then took a train to Bill's car dealership.
S3: He needed to buy a car.
S4: The company he works now isn't near any public transportation.
S5: He also wanted to talk to Bill about their softball league.



Coherence Analysis

S1: John went to the bank to deposit his paycheck.
S2: He then took a train to Bill's car dealership.
S3: He needed to buy a car.
S4: The company he works now isn't near any public transportation.
S5: He also wanted to talk to Bill about their softball league.



Rhetorical Structure Theory

- Mann & Thompson (1987)
- Goal: Identify hierarchical structure of text
  - Cover wide range of TEXT types
    - Language contrasts
  - Relational propositions (intentions)
- Derives from functional relations b/t clauses

Components of RST

- Relations:
  - Hold b/t two text spans, nucleus and satellite
    - Nucleus core element, satellite peripheral
    - Constraints on each, between
    - Effect: why the author wrote this

Components of RST

- Relations:
  - Hold b/t two text spans, nucleus and satellite
    - Nucleus core element, satellite peripheral
    - Constraints on each, between
    - Effect: why the author wrote this
- Schemas:
  - Grammar of legal relations between text spans
  - Define possible RST text structures
    - Most common: N + S, others involve two or more nuclei

Components of RST

- Relations:
  - Hold b/t two text spans, nucleus and satellite
    - Nucleus core element, satellite peripheral
    - Constraints on each, between
    - Effect: why the author wrote this
- Schemas:
  - Grammar of legal relations between text spans
  - Define possible RST text structures
    - Most common: N + S, others involve two or more nuclei
- Structures:
  - Using clause units, complete, connected, unique, adjacent

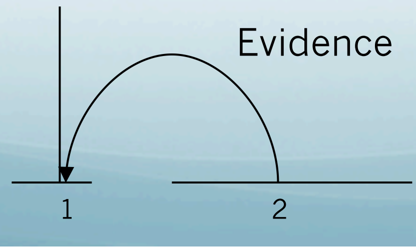
RST Relations

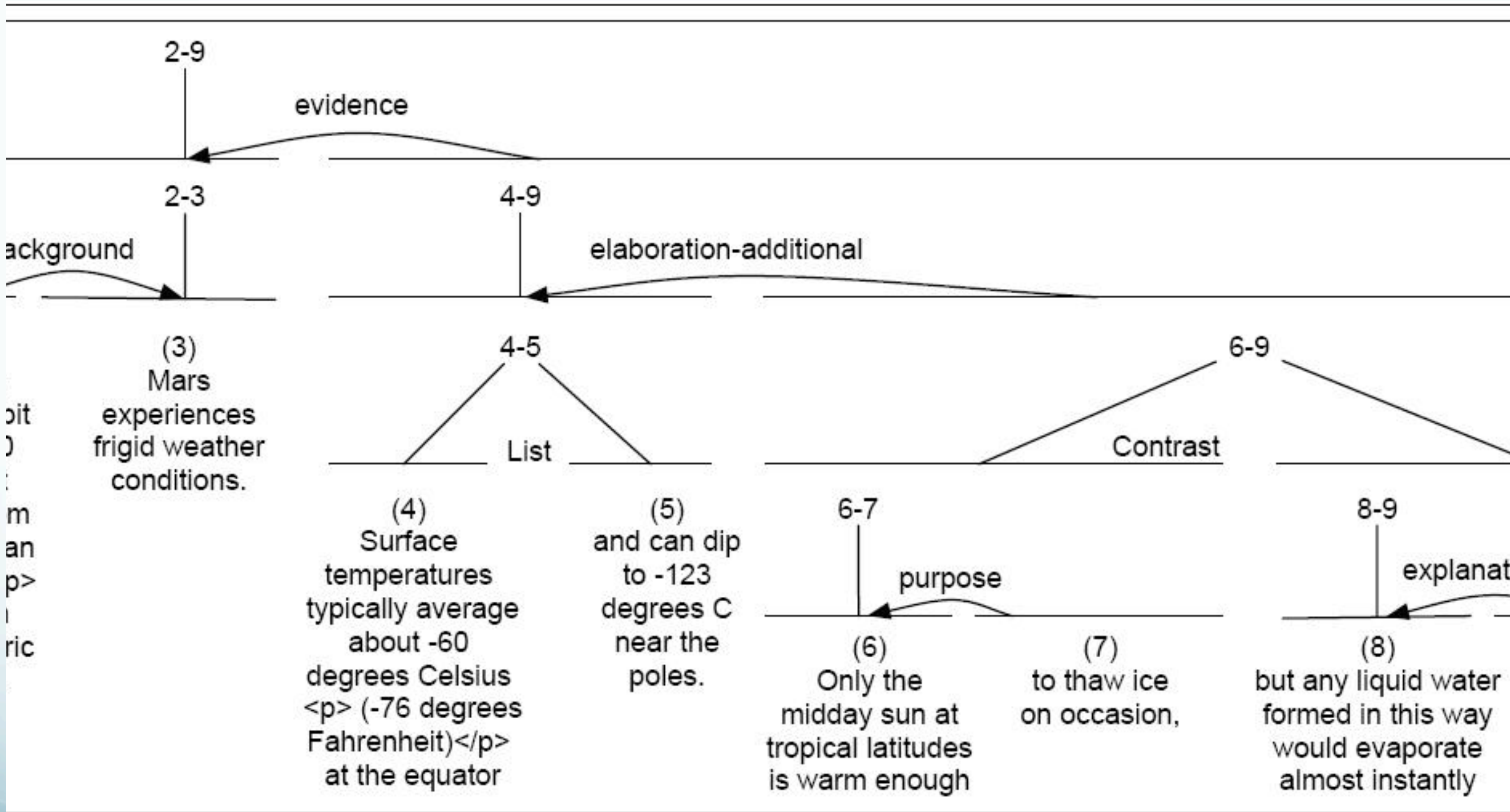
- Core of RST
  - RST analysis requires building tree of relations
  - Circumstance, Solutionhood, Elaboration. Background, Enablement, Motivation, Evidence, Justify, Vol. Cause, Non-Vol. Cause, Vol. Result, Non-Vol. Result, Purpose, Antithesis, Concession, Condition, Otherwise, Interpretation, Evaluation, Restatement, Summary, Sequence, Contrast

RST Relations

- Evidence
  - Effect: Evidence (Satellite) increases R's belief in Nucleus
    - The program really works. (N)
    - I entered all my info and it matched my results. (S)

Relation Name:	Evidence
Constraints on N:	R might not believe N to a degree satisfactory to W
Constraints on S:	R believes S or will find it credible
Constraints on N+S:	R's comprehending S increases R's belief of N
Effects:	R's belief of N is increased





2-9
evidence
2-3
4-9
ackground
elaboration-additional
(3)
Mars
experiences
frigid weather
conditions.
4-5
6-9
List
Contrast
(4)
Surface
temperatures
typically average
about -60
degrees Celsius
<p> (-76 degrees
Fahrenheit)</p>
at the equator
(5)
and can dip
to -123
degrees C
near the
poles.
6-7
purpose
(6)
Only the
midday sun at
tropical latitudes
is warm enough
(7)
to thaw ice
on occasion,
8-9
(8)
but any liquid water
formed in this way
would evaporate
almost instantly

RST Parsing

- Learn and apply classifiers for
  - Segmentation and parsing of discourse

RST Parsing

- Learn and apply classifiers for
  - Segmentation and parsing of discourse
- Assign coherence relations between spans

RST Parsing

- Learn and apply classifiers for
  - Segmentation and parsing of discourse
- Assign coherence relations between spans
- Create a representation over whole text => parse
- Discourse structure
  - RST trees
    - Fine-grained, hierarchical structure
      - Clause-based units

Identifying Segments & Relations

- Key source of information:

Identifying Segments & Relations

- Key source of information:
  - Cue phrases
    - Aka discourse markers, cue words, clue words
    - Although, but, for example, however, yet, with, and....
      - John hid Bill's keys **because** he was drunk.
- Issues:

Identifying Segments & Relations

- Key source of information:
  - Cue phrases
    - Aka discourse markers, cue words, clue words
    - Although, but, for example, however, yet, with, and....
      - John hid Bill's keys **because** he was drunk.
- Issues:
  - Ambiguity: discourse vs sentential use
      - **With** its distant orbit, Mars exhibits frigid weather.
      - We can see Mars **with** a telescope.
  - Ambiguity: cue multiple discourse relations
    - Because: CAUSE/EVIDENCE; But: CONTRAST/ CONCESSION

Cue Phrases

- Last issue:
  - Insufficient:

Cue Phrases

- Last issue:
  - Insufficient:
    - Not all relations marked by cue phrases
    - Only 15-25% of relations marked by cues

Learning Discourse Parsing

- Train classifiers for:
  - Segmentation
  - Coherence relation assignment
  - Discourse structure assignment
    - Shift-reduce parser transitions

- Use range of features:
  - Cue phrases
  - Lexical/punctuation in context
  - Syntactic parses

Evaluation

- Segmentation:
  - Good: 96%
    - Better than frequency or punctuation baseline
- Discourse structure:
  - Okay: 61% span, relation structure
- Relation identification: poor

Issues

- Goal: Single tree-shaped analysis of all text
  - Difficult to achieve

Issues

- Goal: Single tree-shaped analysis of all text
  - Difficult to achieve
    - Significant ambiguity
    - Significant disagreement among labelers

Issues

- Goal: Single tree-shaped analysis of all text
  - Difficult to achieve
    - Significant ambiguity
    - Significant disagreement among labelers
    - Relation recognition is difficult
      - Some clear “signals”, I.e. although
      - Not mandatory, only 25%