CS474 Intro to NLP

Question Answering

– Last class
  • History
  • Open-domain QA
  • Basic system architecture

– Today
  • Finish basic system architecture
  • Predictive indexing methods
  • Pattern-matching methods

List Questions

• **List questions**

1915: List the names of chewing gums.

Stimorol	Orbit	Winterfresh	Double Bubble
Dirol	Trident	Spearmint	Bazooka
Doublemint	Dentyne	Freedent	Hubba Bubba
Juicy Fruit	Big Red	Chiclets	Nicorette

• **Can't just rely on a single document**
• **Performance**
  – TREC 2003: F .40
  – TREC 2004: F .62

Definition Questions

• **Who is Colin Powell?**
• **What is mold?**

• **Hard to evaluate**
  – Who is the audience?
  – Evaluation requires matching *concepts* in the desired response to *concepts* in a system response
    • TREC 2003:
      – Audience: questioner is an adult, a native speaker of English, and an "average" reader of US newspapers
      – Results: F .55

Context Task

• **Track a target discourse object through a series of questions**

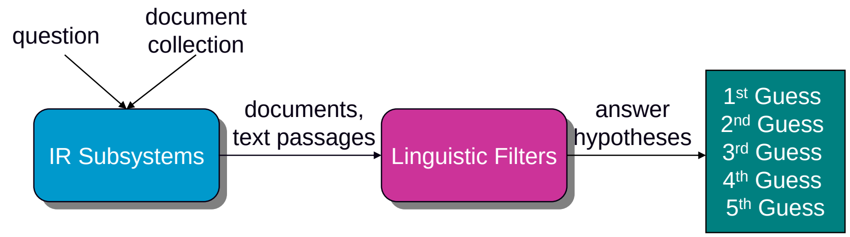
21	Club Med		
	21.1	FACTOID	How many Club Med vacation spots are there worldwide?
	21.2	LIST	List the spots in the United States.
	21.3	FACTOID	Where is an adults-only Club Med?
	21.4	OTHER	

• **Performance**
  – TREC 2004
    • Factoids: .84 initial; .74 non-initial
    • Lists: .62 F
    • Other: .46 F

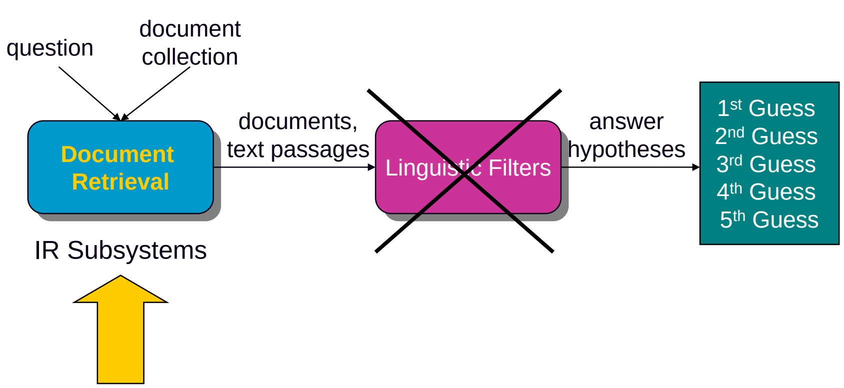
Question answering

- Overview and task definition
- History
- Open-domain question answering
- **Basic system architecture**
  [Cardie et al., ANLP 2000]
- Predictive indexing methods
- Pattern-matching methods

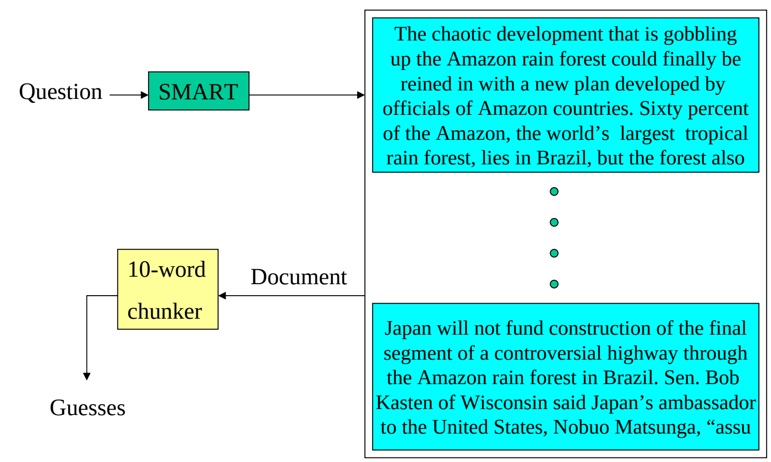
Basic system architecture



System architecture: document retrieval



QA as document retrieval



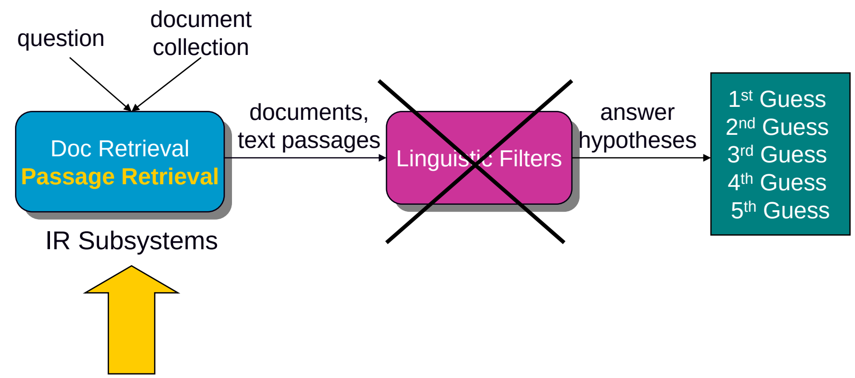
Baseline evaluation

- **Document retrieval only**
- **Corpus**
  – TREC-8 development corpus (38 questions)
  – TREC-8 test corpus (200 questions)

	Development (38)		Test (200)	
	Correct	**MAR**	**Correct**	**MAR**
Smart	3	3.33	29	3.07

MAR = Mean Answer Rank

System architecture: passage retrieval



Passage retrieval

[Salton *et al.*]

Query-dependent text summarization

Which country has the largest part of the Amazon rain forest?

[The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists from around the world.] [“That’s some of the most encouraging news about the Amazon rain forest in recent years,” said Thomas Lovejoy, an Amazon specialist.] [“It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon.”]

[Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil.]

Extract passages that best summarize each document w.r.t. the query

Query-dependent text summarization

- **Basic algorithm**

1. Decide on a summary length (10% of document length).
2. Use standard ad-hoc retrieval algorithm to retrieve top documents.
3. ***Treat each sentence/paragraph in top N documents as a document itself.***

   Use standard document similarity equations to assign a similarity score to the sentence/paragraph.
4. Return highest-scoring sentences/paragraphs as the summary, subject to the length constraint.

Passage retrieval

[Salton *et al.*]

Query-dependent text summarization

Which country has the largest part of the Amazon rain forest?

[The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists from around the world.] [“That’s some of the most encouraging news about the Amazon rain forest in recent years,” said Thomas Lovejoy, an Amazon specialist.] [“It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon.”]

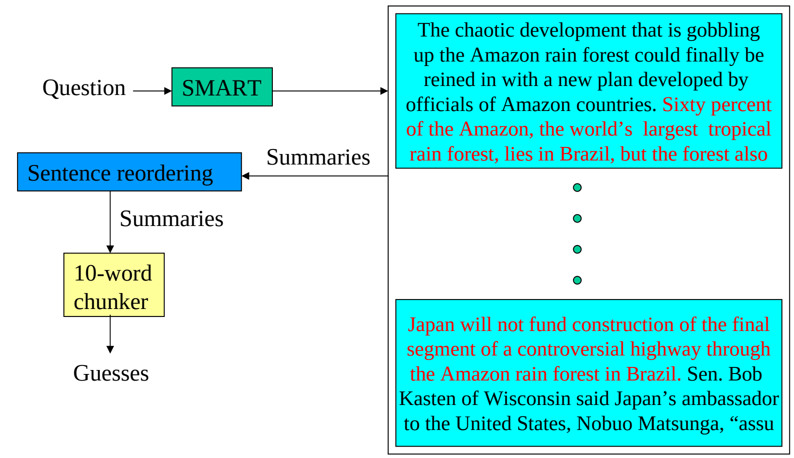
[Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil.]

Sort summary extracts across top k documents

ordered list of summary extracts

answer hypotheses

QA as query-dependent text summarization



Evaluation: text summarization

	Development (38)		Test (200)	
	Correct	MAR	Correct	MAR
Smart	3	3.33	29	3.07
Text Summarization	4	2.25	45	2.67

MAR = Mean Answer Rank

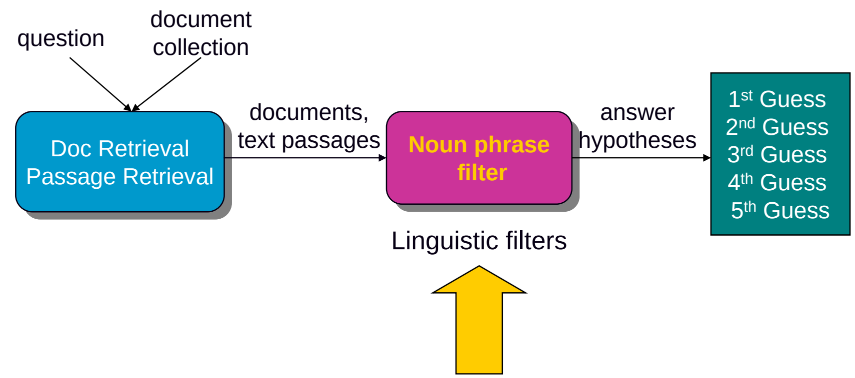
Evaluation: text summarization

- **Summarization method can limit performance**
  - Development corpus
    - In only 23 of the 38 developments questions (61%) does the correct answer appear in the summary for one of the top *k*=7 documents
  - Test corpus
    - In only 135 of the 200 developments questions (67.5%) does the correct answer appear in the summary for one of the top (*k*=6) documents

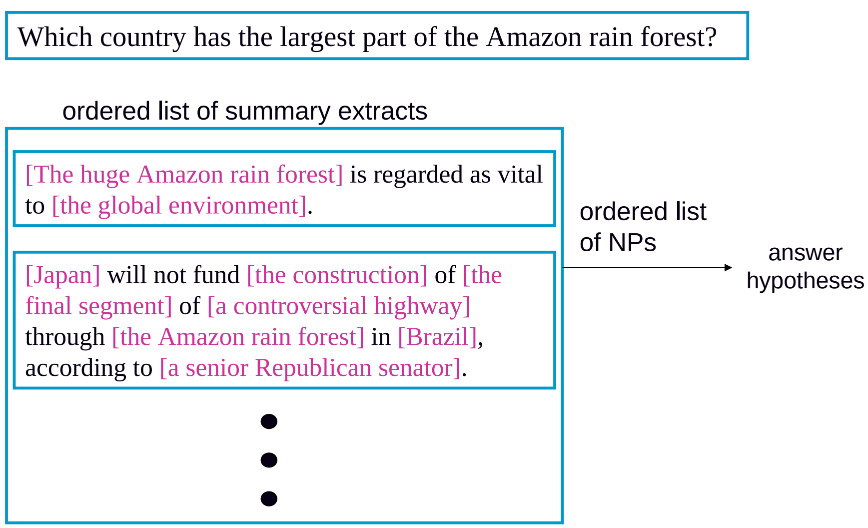
Linguistic filters

- **50 byte answer length effectively eliminates *how* or *why* questions**
- **almost all of the remaining question types are likely to have noun phrases as answers**
  - development corpus: 36 of 38 questions have noun phrase answers
- **consider adding at least a simple linguistic filter that considers only noun phrases as answer hypotheses**

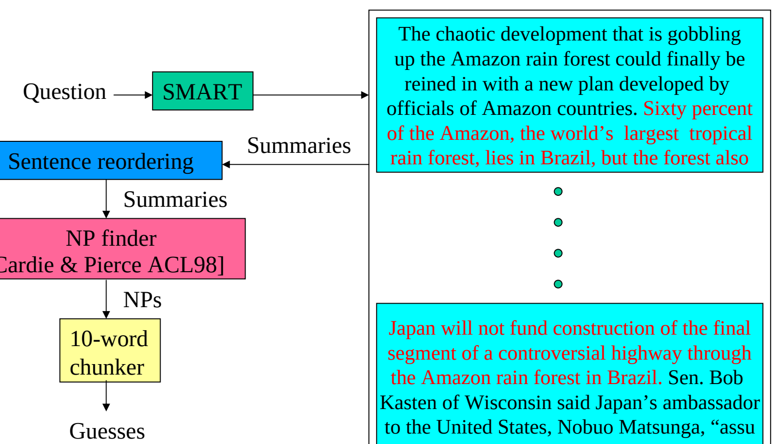
System architecture: linguistic filters



The noun phrase filter



QA using the NP filter



Chunking answer hypotheses: BAD

Which country has the largest part of the Amazon rain forest?

Question-Answering System

"Japan Brazil a new plan Amazon countries A section"

"Northwestern Brazil A plan the Amazon region"

"eight surrounding countries A Brazilian company"

"union leader his modest wooden house The people"

"defense Brazilian wealth the international market"

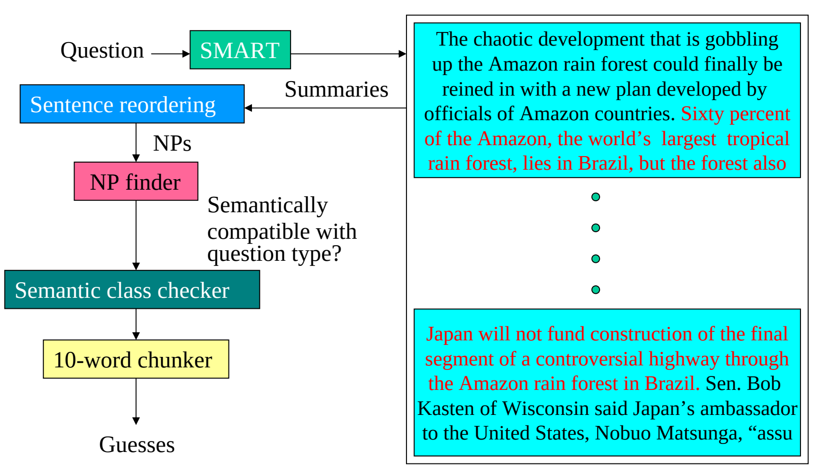
Evaluation: NP filter

	Development (38)		Test (200)	
	Correct	MAR	Correct	MAR
Smart	3	3.33	29	3.07
Text Summarization	4	2.25	45	2.67
TS + NPs	7	2.29	50	2.66

MAR = Mean Answer Rank

- **Using NP finder of Cardie & Pierce (1998)**
  - ~94% precision and recall on Wall Street Journal text
- **How much does the (unnatural) NP "chunking" help?**
  - Without it, only 1 and 20 questions answered for each corpus, respectively
  - NP filter is extracting good guesses, but better linguistic processing is needed to promote the best guesses to the top of the ranked guess list

Semantic class checking



Semantic class checking

- **Approximate question type using question word**

  Who is the president of the U.S.?

  person

  Which country has the largest part of the Amazon rain forest?

  country

  Where is the Connecticut River?

  state? county? country? location?

  What fabric should one use to make curtains?

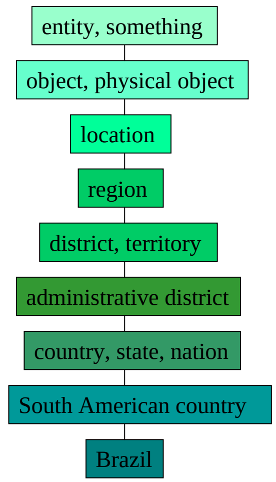
  fabric???
- **Check that head noun (i.e. the last noun) of answer NP is of the same type**

  a <u>man</u> = person

  <u>Massachusetts</u> = state, location

Semantic type checking

- **Use lexical resource to determine semantic compatibility**
  - WordNet!
- **Proper names handled separately since they are unlikely to appear in WordNet**
  - Small set (~20) rules



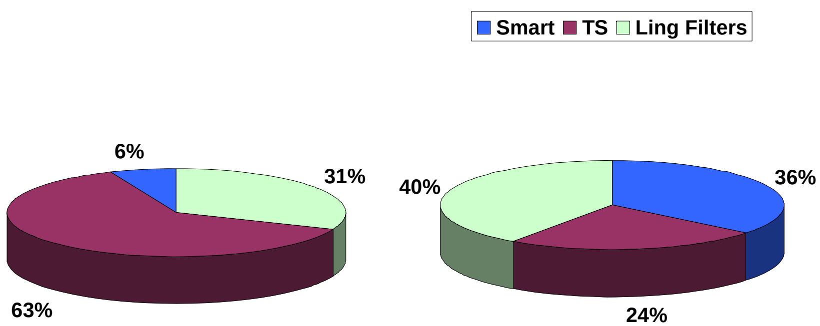
Evaluation: semantic class filter

	Development (38)		Test (200)	
	Correct	**MAR**	**Correct**	**MAR**
Smart	3	3.33	29	3.07
Text Summarization	4	2.25	45	2.67
TS + NPs	7	2.29	50	2.66
TS + NPs + Semantic Type	21	1.38	86	1.90

MAR = Mean Answer Rank

- **Weak syntactic and semantic information allows large improvements**
- **Problems?**

Sources of error



Development questions (38)

Test questions (200)

Question answering

- Overview and task definition
- History
- Open-domain question answering
- Basic system architecture
- **Predictive indexing methods**
  - Slides based on those of Jamie Callan, CMU
- Pattern-matching methods

Indexing with predictive annotation

- **Some answers belong to well-defined semantic classes**
  – People, places, monetary amounts, telephone numbers, addresses, organizations
- **Predictive annotation: index a document with “concepts” or “features” that are expected to be useful in (many) queries**
  – E.g. people names, location names, addresses, etc.
- **Add additional operators for use in queries**
  – E.g. Where does Ellen Vorhees work? “Ellen Vorhees” NEAR/10 *organization

Predictive annotation

In the early part of this century, the only means of transportation for travelers and mail between **<LOCATION> Europe </LOCATION>** and **<LOCATION> North America </LOCATION>** was by passenger steamship. By <DATE> 1907 </DATE>, the **<COMPANY> Cunard Steamship Company </COMPANY>** introduced the largest and fastest steamers in the **<LOCATION> North Atlantic </LOCATION>** service: the **<NAME> Lusitania </NAME>** and the **<NAME> Mauritania </NAME>**. Each had a gross tonnage of **<WEIGHT> 31,000 tons </WEIGHT>** and a maximum speed of **<SPEED> 26 knots </SPEED>**.

– From K. Felkins, H.P. Leighly, Jr., and A. Jankovic. “The Royal Mail Ship Titanic: Did a Metallurgical Failure Cause a Night to Remember?” *JOM*, 50 (1), 1998, pp. 12-18.

Predictive annotation

- **How is annotated text stored in the index?**

In the early part of this century, the only means of transportation for travelers and mail between **<$LOCATION, Europe>** and **<$LOCATION North> <$LOCATION America>** was by passenger steamship. By <$DATE 1907>, the **<$COMPANY, Cunard> <$COMPANY, Steamship> <$COMPANY, Company>** introduced the largest and fastest steamers in the **<$LOCATION, North> <$LOCATION, Atlantic>** service: the **<$NAME, Lusitania>** and the **<$NAME, Mauritania>**. Each had a gross tonnage of **<$WEIGHT, 31,000> <$WEIGHT, tons>** and a maximum speed of **<$SPEED, 26> <$SPEED, knots>**.

- **Treat <$QA-token, term> as meaning that $QA-token and term occur at the same location in the text**
  – Or use phrase indexing approach to index as a single item

Issues for predictive annotation

- **What makes a good QA-token?**
  – Question that would use the token
    - Can be recognized with high reliability (high precision)
    - Occurs frequently enough to be worth the effort
- **How do you want the system to make use of the QA-tokens?**
  – Filtering step?
  – Transform original question into an ad-hoc retrieval question that incorporates QA-tokens and proximity operators?
- **Common approaches to recognizing QA-tokens**
  – Tables, lists, dictionaries
  – Heuristics
  – Hidden Markov models

Advantages and disadvantages

+ **Most of the computational cost occurs during indexing**
  - Allows use of more sophisticated methods
+ **Annotator has access to complete text of document**
  - Important for recognizing some types of features
– **Must know ahead of time which types of features/concepts are likely to be important**
– **Increases size of index considerably**
  - E.g. by an order of magnitude if many features

- **Used (in varying amounts) by almost all open-domain Q/A systems**

Question answering

- **Overview and task definition**
- **History**
- **Open-domain question answering**
- **Basic system architecture**
- **Predictive indexing methods**
- **Pattern-matching methods**
  - Slides based on those of Jamie Callan, CMU

Simple pattern-based QA

- **Observation: there are many questions…but fewer <u>types</u> of questions**
- **Each type of question can be associated with**
  - Expectations about answer string characteristics
  - Strategies for retrieving documents that might have answers
  - Rules for identifying answer strings in documents

Example

- **Who is the President of Cornell?**
  - Expectation: answer string contains a person name
    - Named entity identification
  - Search query: “president Cornell *PersonName”
  - Rule: “*PersonName, President of Cornell”
    - Matches “…David Skorten, President of Cornell”
    - Answer = “David Skorten”

Question analysis

- **Input: the question**
- **Output**
  - Search query
  - Answer expectations
  - Extraction strategy
- **Requires**
  - Identifying named entities
  - Categorizing the question
  - Matching question parts to templates
- **Method: pattern-matching**
  - Analysis patterns still created manually…

Question analysis example

- **"Who is Elvis?"**
  - Question type: "who"
  - Named-entity tagging: "Who is <person-name>Elvis</person-name>"
  - Analysis pattern: if question type = "who" and question contains <person-name> then
    - Search query doesn't need to contain a *PersonName operator
    - Desired answer probably is a description
    - Likely answer extraction patterns
      - "Elvis, the X"
        - » "…Elvis, the <u>king of rock and roll</u>…"
      - "the X Elvis"
        - » "the <u>legendary entertainer</u> Elvis"

Question analysis

Frequency of question types on an Internet search engine

- 42% what
- 21% where
- 20% who
- 8% when
- 8% why
- 2% which
- 0% how

Relative difficulty of question types

- What is difficult
  - What time…
  - What country…
- Where is easy
- Who is easy
- When is easy
- Why is hard
- Which is hard
- How is hard

Example: What is Jupiter?

1. **What We Will Learn from Galileo**
2. **The Nature of Things: Jupiter's shockwaves—How a comet's bombardment has sparked activity on Earth**
3. **Jupiter-Bound Spacecraft Visits Earth on 6-Year Journey**
4. **STAR OF THE MAGI THEORIES ECLIPSED?**
5. **Marketing & Media: Hearst, Burda to Scrap New Astrology Magazine**
6. **Greece, Italy Conflict On Cause Of Ship Crash That Kills 2, Injures 54**
7. **Interplanetary Spacecraft To `Visit` Earth With LaserGraphic**
8. **A List of Events During NASA's Galileo Mission to Jupiter**
9. **SHUTTLE ALOFT, SENDS GALILEO ON 6-YEAR VOYAGE TO JUPITER**
10. **Rebuilt Galileo Probe readied For Long Voyage To Jupiter**

Answer extraction

- **Select highly ranked sentences from highly ranked documents**
- **Perform named-entity tagging (or extract from index) and perform part of speech tagging**
  - "The/DT planet/NN <location>Jupiter/NNP</location> and/CC its/PRP moons/NNS are/VBP in/IN effect/NN a/DT mini-solar/JJ system/NN ,/, and/CC <location>Jupiter/NNP</location> itself/PRP is/VBZ often/RB called/VBN a/DT star/NN that/IN never/RB caught/VBN fire/NN ./."
- **Apply extraction patterns**
  - the/DT X Y, Y=Jupiter -> the <u>planet</u> Jupiter -> "planet"

Simple pattern-based Q/A: assessment

- **Extremely effective when**
  - Question patterns are predictable
    - Fairly "few" patterns cover the most likely questions
      - Could be several hundred
    - Not much variation in vocabulary
      - Simple word matching works
    - The corpus is huge (e.g., Web)
      - Odds of finding an answer document that matches the vocabulary and answer extraction rule improves
- **Somewhat labor intensive**
  - Patterns are created and tested manually

Common problem: matching questions to answers

- **Document word order isn't exactly what was expected**
- **Solution: "soft matching" of answer patterns to document text**
  - Approach: use distance-based answer selection when no rule matches
    - E.g. for "What is Hunter Rawlings' address?"
      - Use the address <u>nearest to</u> the words "Hunter Rawlings"
      - User the address in the <u>same sentence</u> as "Hunter Rawlings"

Common problem: matching questions to answers

- **Answer vocabulary doesn't exactly match question vocabulary**
- **Solution: bridge the vocabulary mismatch**
  - Approach: use WordNet to identify simple relationships
    - "astronaut" is a type of "person"
    - "astronaut" and "cosmonaut" are synonyms

Common problem: improving the set of retrieved documents

- **Sometimes the IR system can't find any documents that have answers (even though the right documents are in the corpus)**
- **Solution: get a broader set of documents**
  - Approach: if answer extractor fails to find an answer, send the question back to the search engine with instructions to widen the search
    - Assumes answer extractors can tell when they fail
  - Approach: use a variety of retrieval strategies to retrieve documents
    - E.g., all words within one sentence, then all words within one paragraph, then within same document, …
    - E.g. add synonyms to query or do query expansion
    - Simple, but much higher computational expense

Common problem: improving answer extraction patterns

- **Word sequence patterns have limited power**
- **Solution: create patterns that use syntactic information**
  - Partial syntactic parsing of documents
    - Is this noun the subject or the object of the sentence?
  - Allows more complex patterns
    - Question: "Who shot Kennedy?"
    - "Who" implies a person that should be subject of answer sentence/clause
    - "Kennedy" should be direct object of answer
    - Pattern: <subject> shot Kennedy
    - Matching text: "Oswald shot Kennedy"

Common problem: selecting/ranking the answer

- **Multiple answer candidates**
- **Solutions**
  - Features used to represent answer candidates
    - Frequency
    - Distance to question words
    - Location in answer passage(s)
    - …
  - Selection functions
    - Created manually
    - Learned from training data