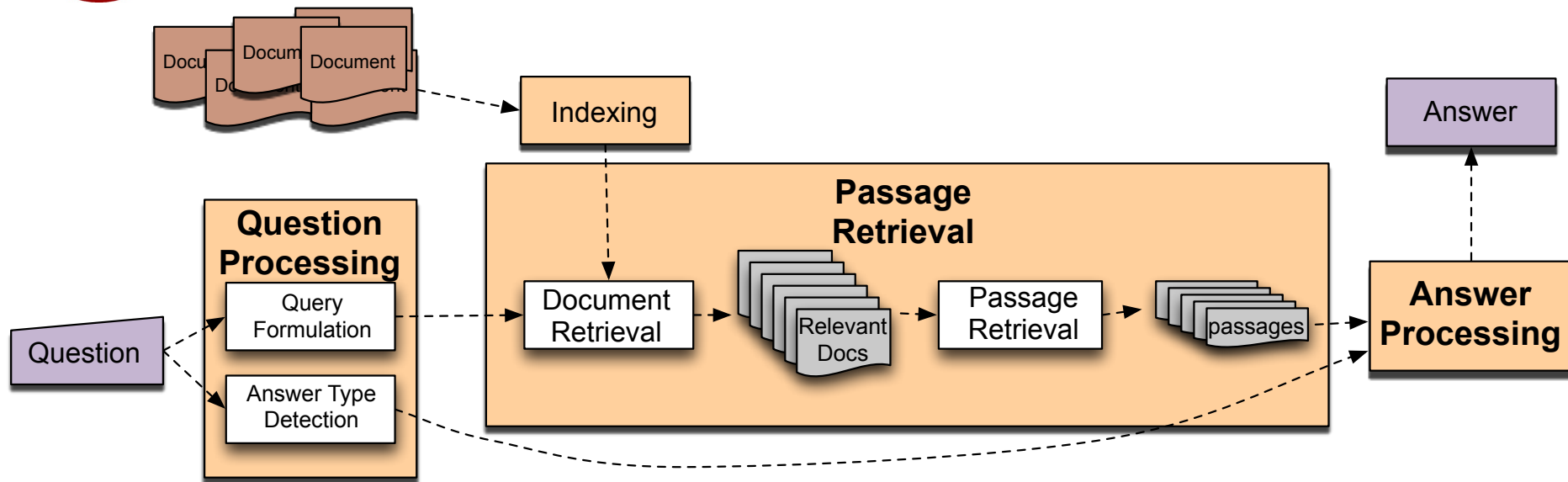




Factoid Q/A





Question Processing

Things to extract from the question

- Answer Type Detection
 - Decide the **named entity type** (person, place) of the answer
- Query Formulation
 - Choose **query keywords** for the IR system
- Question Type classification
 - Is this a definition question, a math question, a list question?
- Focus Detection
 - Find the question words that are replaced by the answer
- Relation Extraction
 - Find relations between entities in the question



Question Processing

They're the two states you could be reentering if you're crossing Florida's northern border

- Answer Type: US state
- Query: two states, border, Florida, north
- Focus: the two states
- Relations: borders(Florida, ?x, north)



Answer Type Detection: Named Entities

- *Who founded Virgin Airlines?*
 - PERSON
- *What Canadian city has the largest population?*
 - CITY.



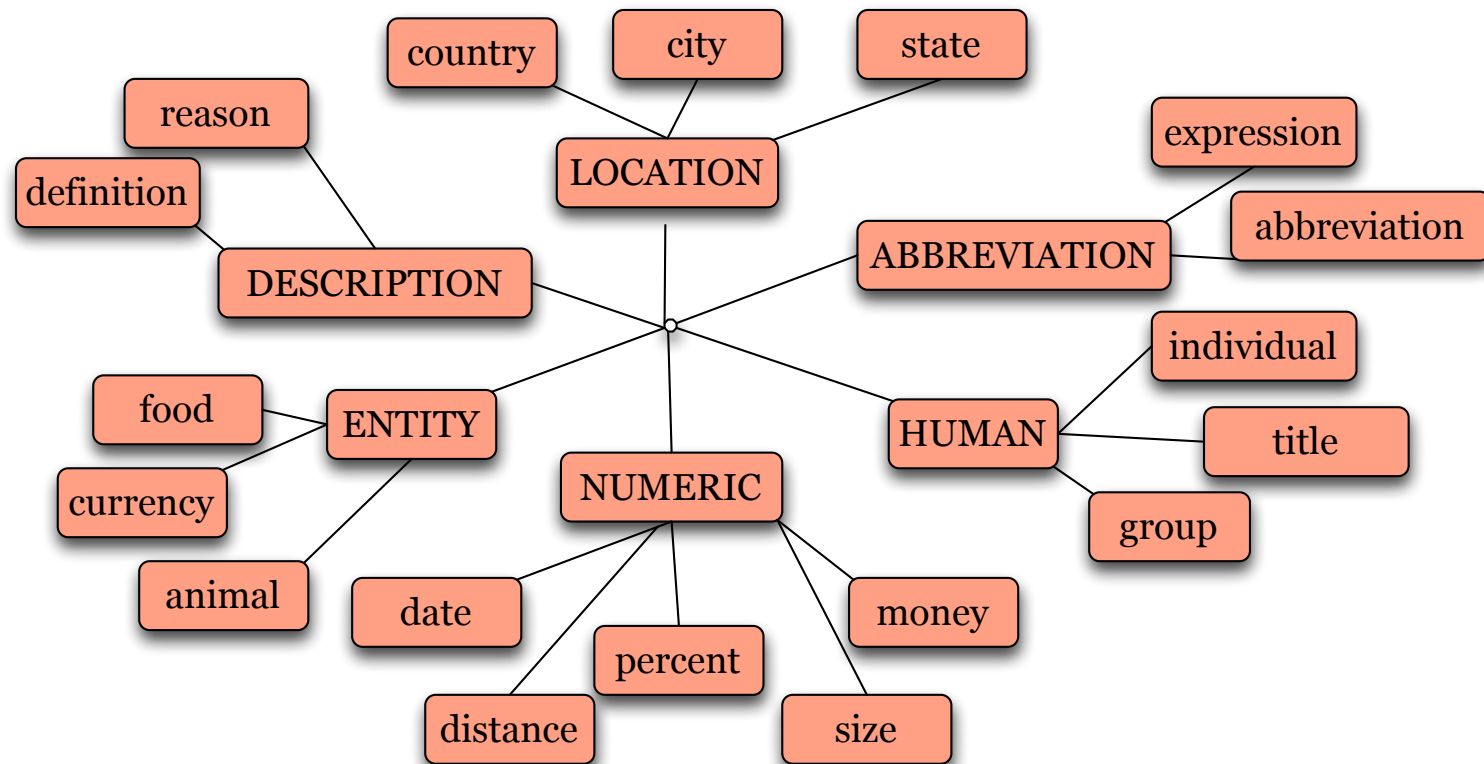
Answer Type Taxonomy

Xin Li, Dan Roth. 2002. Learning Question Classifiers.
COLING'02

- 6 coarse classes
 - ABBEVIATION, ENTITY, DESCRIPTION, HUMAN, LOCATION, NUMERIC
- 50 finer classes
 - LOCATION: city, country, mountain...
 - HUMAN: group, individual, title, description
 - ENTITY: animal, body, color, currency...



Part of Li & Roth's Answer Type Taxonomy





Answer Types

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? |



More Answer Types

| | |
|-------------|--|
| 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? |



Answer types in Jeopardy

Ferrucci et al. 2010. Building Watson: An Overview of the DeepQA Project. AI Magazine. Fall 2010. 59-79.

- 2500 answer types in 20,000 Jeopardy question sample
- The most frequent 200 answer types cover < 50% of data
- The 40 most frequent Jeopardy answer types

he, country, city, man, film, state, she, author, group, here,
company, president, capital, star, novel, character, woman, river,
island, king, song, part, series, sport, singer, actor, play, team,
show, actress, animal, presidential, composer, musical,
nation, book, title, leader, game



Answer Type Detection

- Hand-written rules
- Machine Learning
- Hybrids



Answer Type Detection

- Regular expression-based rules can get some cases:
 - Who {is|was|are|were} PERSON
 - PERSON (YEAR – YEAR)
- Other rules use the **question headword**:
(the headword of the first noun phrase after the wh-word)
 - Which **city** in China has the largest number of foreign financial companies?
 - What is the state **flower** of California?



Answer Type Detection

- Most often, we treat the problem as machine learning classification
 - **Define** a taxonomy of question types
 - **Annotate** training data for each question type
 - **Train** classifiers for each question class using a rich set of features.
 - features include those hand-written rules!

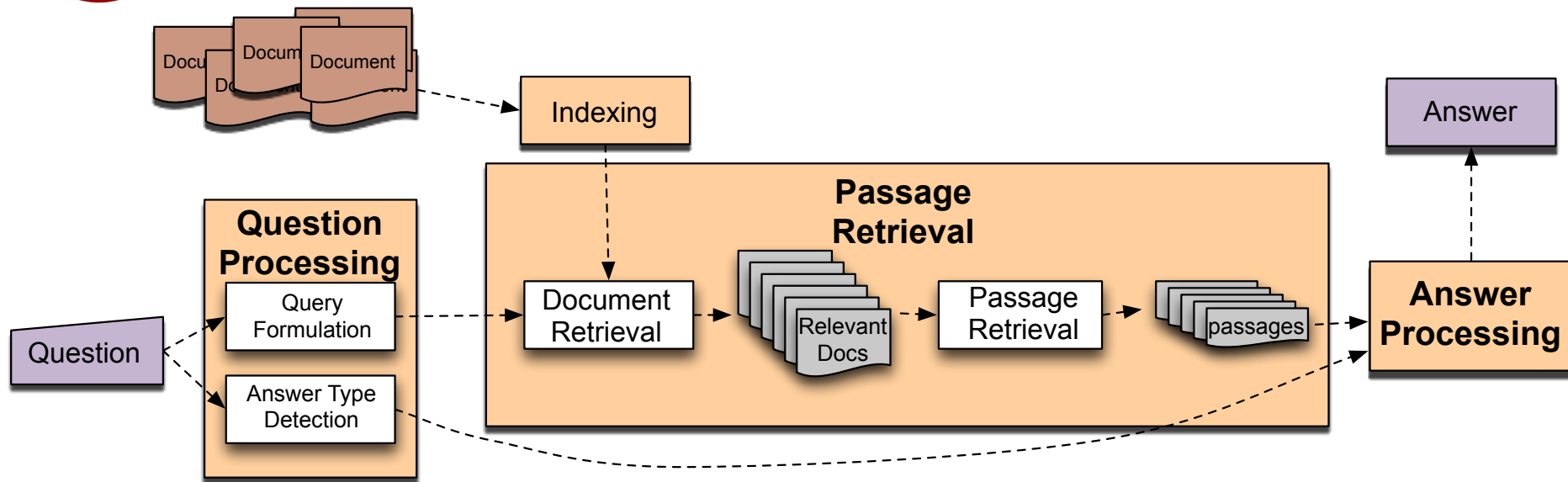


Features for Answer Type Detection

- Question words and phrases
- Part-of-speech tags
- Parse features (headwords)
- Named Entities
- Semantically related words



Factoid Q/A





Keyword Selection Algorithm

Dan Moldovan, Sanda Harabagiu, Marius Păcă, Rada Mihalcea, Richard Goodrum, Roxana Girju and Vasile Rus. 1999. Proceedings of TREC-8.

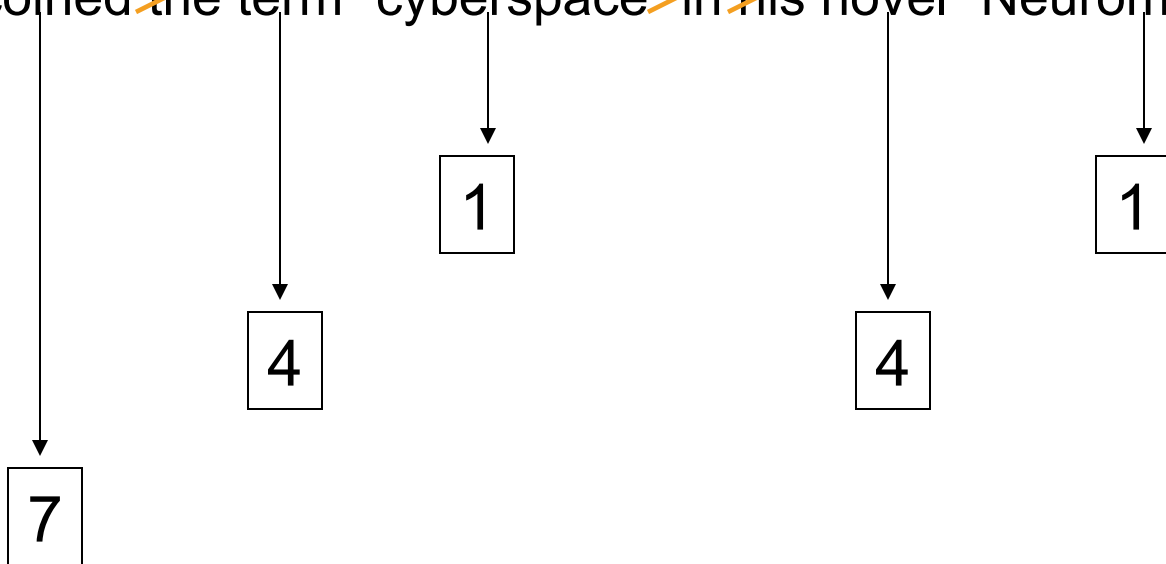
1. Select all non-stop words in quotations
2. Select all NNP words in recognized named entities
3. Select all complex nominals with their adjectival modifiers
4. Select all other complex nominals
5. Select all nouns with their adjectival modifiers
6. Select all other nouns
7. Select all verbs
8. Select all adverbs
9. Select the QFW word (skipped in all previous steps)
10. Select all other words



Choosing keywords from the query

Slide from Mihai Surdeanu

Who coined the term “cyberspace” in his novel “Neuromancer”?



cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7

