QA

Hal Daumé III

School of Computing
University of Utah

me@hal3.name

THE UNIVERSITY OF UTAH

Types of Information Needs

- *Ad hoc* retrieval: find me documents "like this"

  Identify positive accomplishments of the Hubble telescope since it was launched in 1991.

  Compile a list of mammals that are considered to be endangered, identify their habitat and, if possible, specify what threatens them.

- Question answering

"Factoid"	Who discovered Oxygen? When did Hawaii become a state? Where is Ayer's Rock located? What team won the World Series in 1992?
"List"	What countries export oil? Name U.S. cities that have a "Shubert" theater.
"Definition"	Who is Aaron Copland? What is a quasar?

Central Idea of Factoid QA

- Determine the semantic type of the expected answer

  "Who won the Nobel Peace Prize in 1991?" is looking for a PERSON

- Retrieve documents that have keywords from the question

  Retrieve documents that have the keywords "won", "Nobel Peace Prize", and "1991"

- Look for named-entities of the proper type near keywords

  Look for a PERSON near the keywords "won", "Nobel Peace Prize", and "1991"

An Example

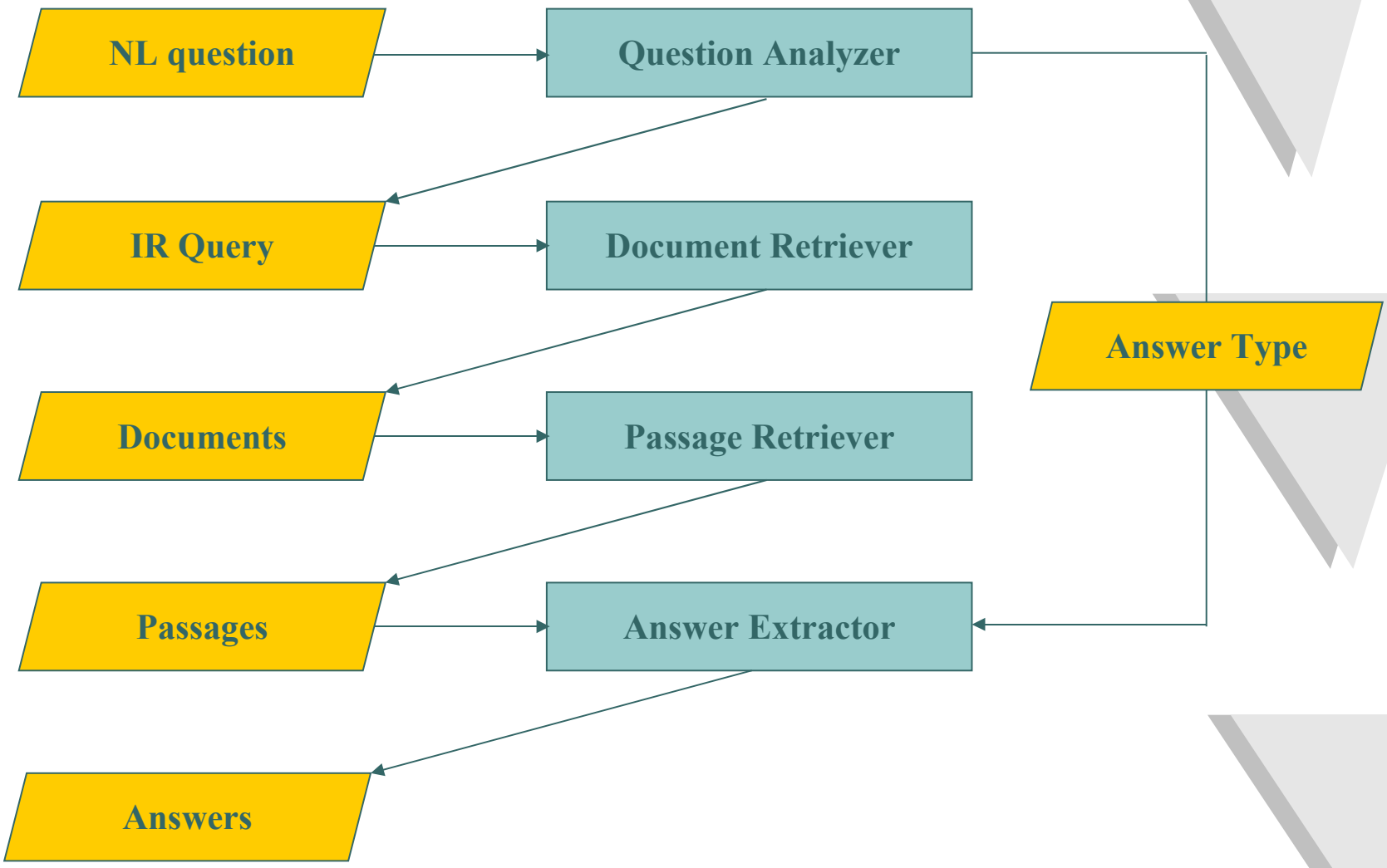
Who won the Nobel Peace Prize in 1991?

But many foreign investors remain sceptical, and western governments are withholding aid because of the Slorc's dismal human rights record and the continued detention of Ms Aung San Suu Kyi, the opposition leader who won the Nobel Peace Prize in 1991.

The military junta took power in 1988 as pro-democracy demonstrations were sweeping the country. It held elections in 1990, but has ignored their result. It has kept the 1991 Nobel peace prize winner, Aung San Suu Kyi - leader of the opposition party which won a landslide victory in the poll - under house arrest since July 1989.

The regime, which is also engaged in a battle with insurgents near its eastern border with Thailand, ignored a 1990 election victory by an opposition party and is detaining its leader, Ms Aung San Suu Kyi, who was awarded the 1991 Nobel Peace Prize. According to the British Red Cross, 5,000 or more refugees, mainly the elderly and women and children, are crossing into Bangladesh each day.

Generic QA Architecture

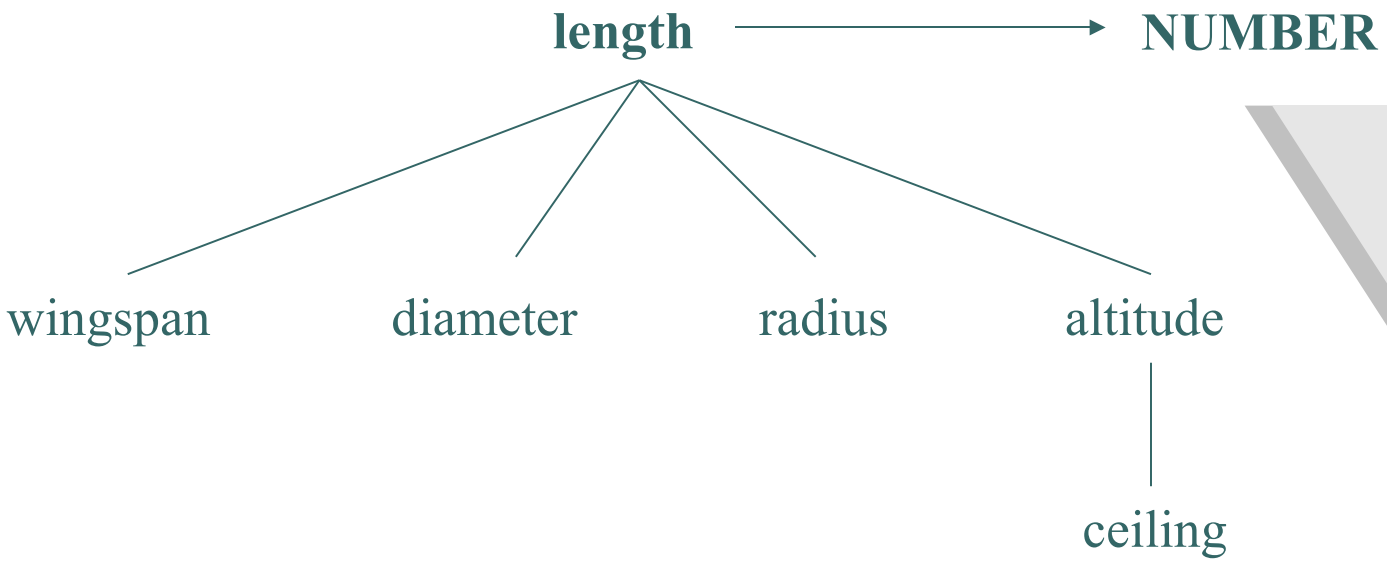


Question analysis

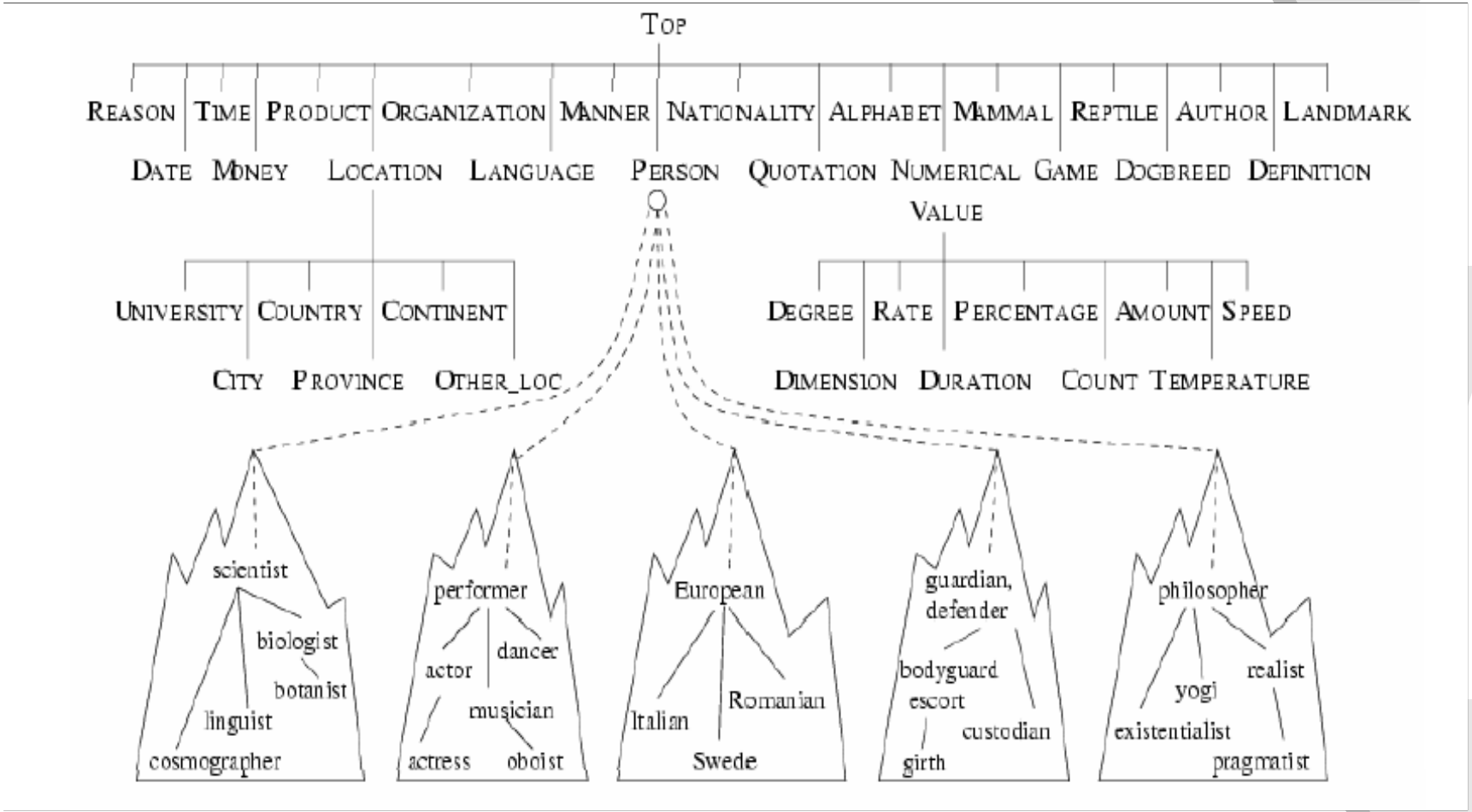
- Question word cues
  - Who → person, organization, location (e.g., city)
  - When → date
  - Where → location
  - What/Why/How → ??
- Head noun cues
  - What city, which country, what year…
  - Which astronaut, what blues band, …
- Scalar adjective cues
  - How long, how fast, how far, how old, …

Using WordNet

What is the service ceiling of an U-2?



Answer Type Hierarchy



Question-Answering from the Web

- The notion of getting computers to give reasonable answers to questions has been around for quite awhile
- Three kinds of systems
  1) Finding answers in text collections
  2) Interfaces to relational databases
  3) Mixed initiative dialog systems

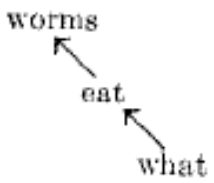
Finding Answers in Text

- Not a new idea... (Simmons et al 1963)
  - Take an encyclopedia and load it onto a computer.
  - Take a question and parse it into a logical form
  - Perform simple information retrieval to get relevant texts
  - Parse those into a logical form
  - Match and rank

Simmons, Klein, McConlogue 1963: Parse Q+A using dependency parser (Hays 1962)

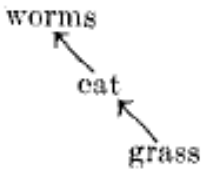
Question:

(a) What do worms eat?

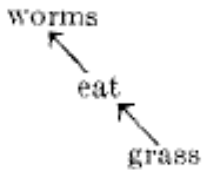


Answers:

(b) Worms eat grass

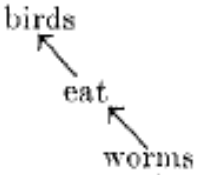


(c) Grass is eaten by worms → worms eat grass



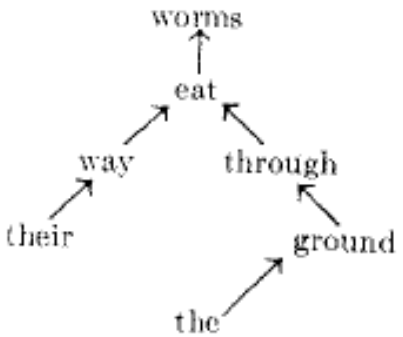
(complete agreement of dependencies)

(d) Birds eat worms



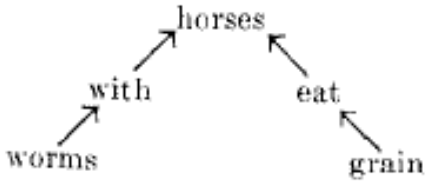
(no agreement)

(e) Worms eat their way through the ground



(partial agreement)

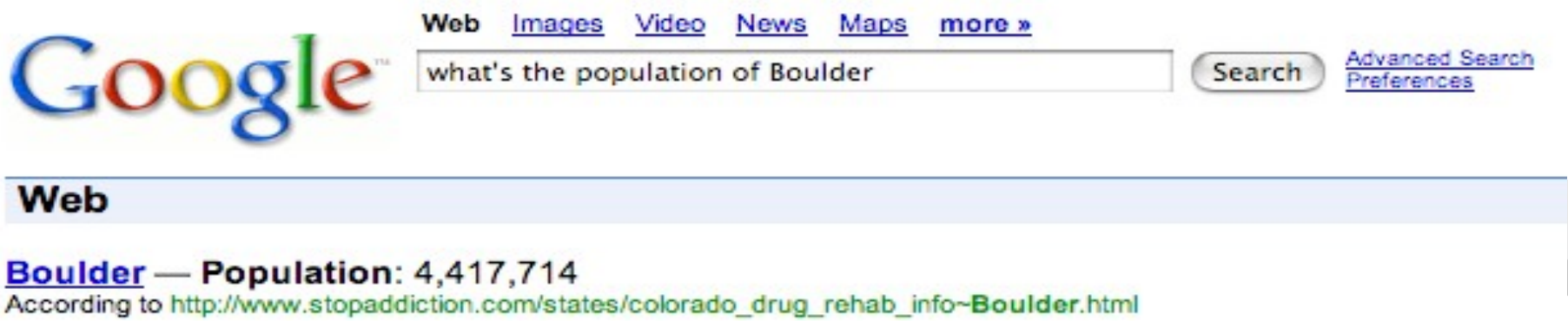
(f) Horses with worms eat grain



(partial agreement)

Web QA


Live Search
what's the population of boulder
Web
Images
News
Maps
QnA Beta
More
what's the population of boulder Page 1 of 112,364 results · Options
Boulder, Colorado Population, total: 92,196
Is this useful?
2004 estimate · US Census Bureau


Google
Web
Images
Video
News
Maps
more »
what's the population of Boulder
Search
Advanced Search
Preferences
Web
Boulder — Population: 4,417,714
According to http://www.stopaddiction.com/states/colorado_drug_rehab_info~Boulder.html

Finding Answers in Text

- Fundamentally, this is about modifying, processing, enriching or marking up both the question and potential answer texts to allow a simple match.
- All current systems do pretty much that.

People *do* ask questions…

Examples from search engine query logs

Which english translation of the bible is used in official Catholic liturgies?
How tall is the sears tower
How can i find someone in texas
Where can i find information on puritan religion?
What are the 7 wonders of the world
How can i eliminate stress
What vacuum cleaner does Consumers Guide recommend

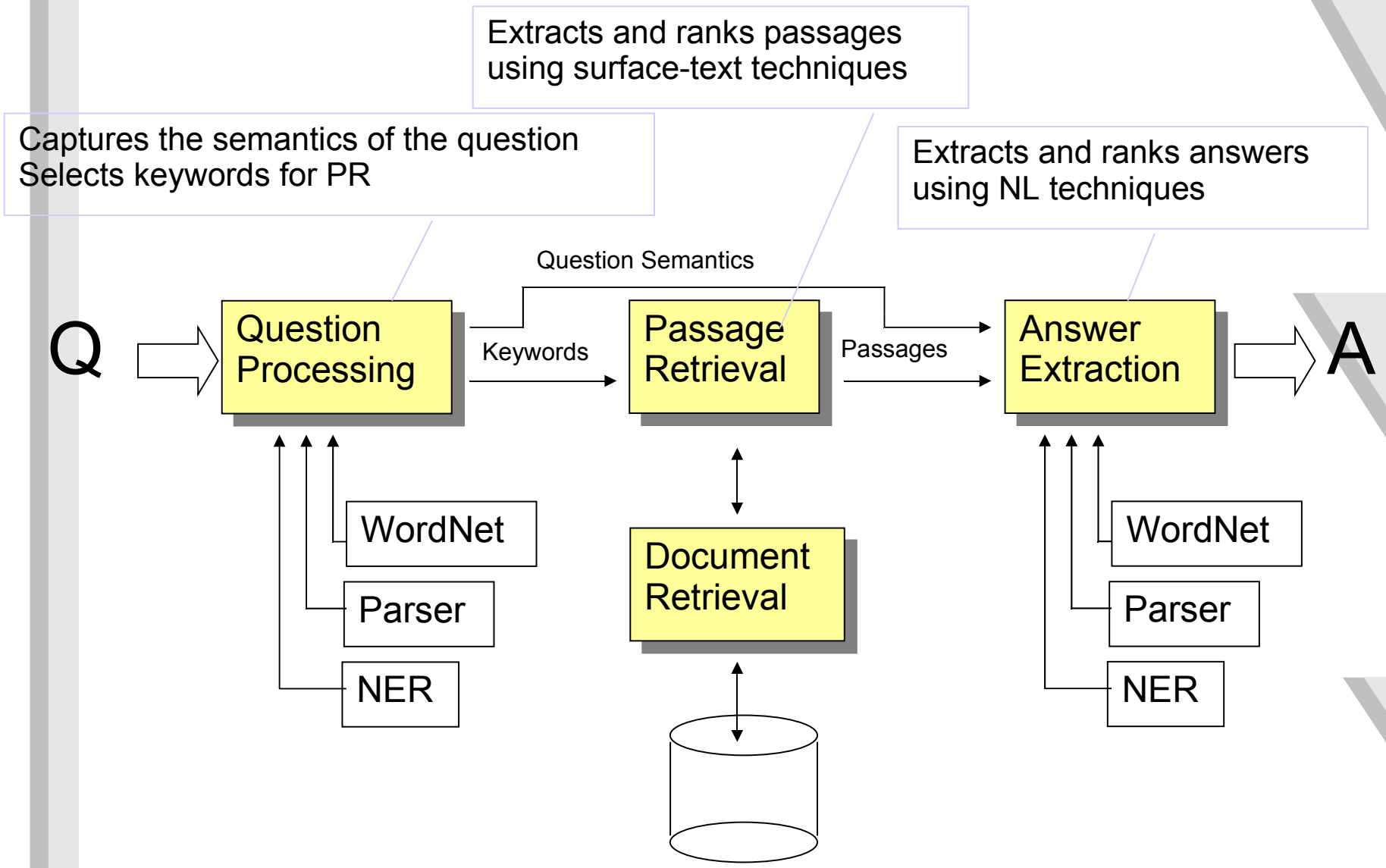
Full-Blown Heavy-Weight System

- Parse and analyze the question
- Formulate queries suitable for use with an IR system (search engine)
- Retrieve ranked results
- Break into suitable units
- Perform NLP on those rank units
- Re-Rank snippets based on NLP processing
- Done

UT Dallas Q/A Systems

- This system contains many components used by other systems, but more complex in some ways
- Next slides based mainly on:
  - Paşca and Harabagiu, *High-Performance Question Answering from Large Text Collections*, SIGIR'01.
  - Paşca and Harabagiu, *Answer Mining from Online Documents*, ACL'01.
    - Harabagiu, Paşca, Maiorano: *Experiments with Open-Domain Textual Question Answering*. COLING'00

QA Block Architecture



Question Processing

- Two main tasks
  - Determining the type of the answer
    - →If you know the type of the answer you can focus your processing only on docs that have things of the right type
  - Extract keywords from the question and formulate a query
    - →Assume that a generic IR search engine can find docs with an answer (and lots that don't). Ie. The NLP/QA system is dealing with precision not recall

Answer Types

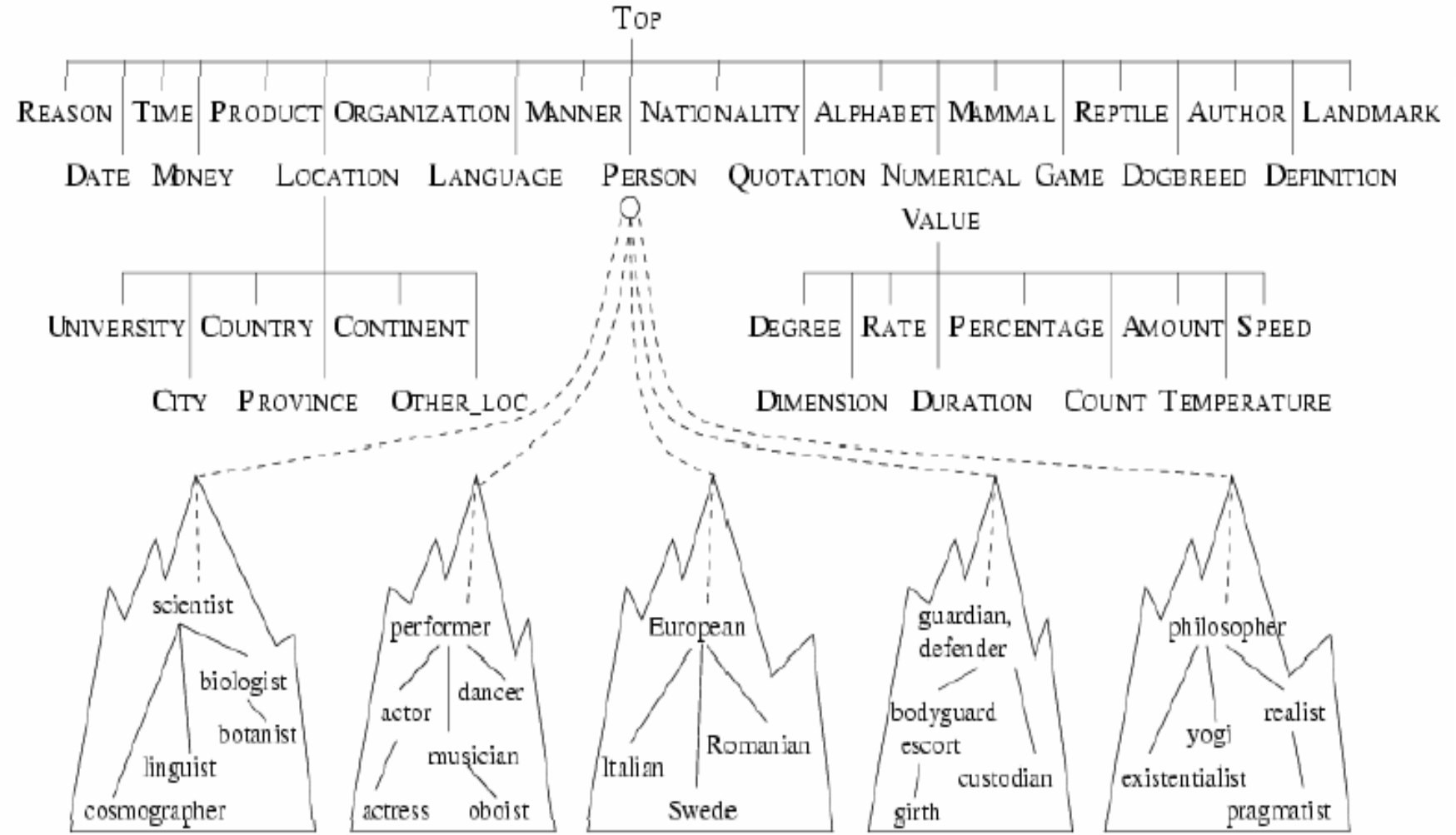
- Factoid questions…
  - Who, where, when, how many…
  - The answers fall into a limited and somewhat predictable set of categories
    - Who questions are going to be answered by…
    - Where questions…
  - Generally, systems select answer types from a set of Named Entities, augmented with other types that are relatively easy to extract

Answer Types

- Of course, it isn't that easy…
  - Who questions can have organizations as answers
    - Who sells the most hybrid cars?
  - Which questions can have people as answers
    - Which president went to war with Mexico?

Answer Type Taxonomy

- Contains ~9000 concepts reflecting expected answer types
- Merges named entities with the WordNet hierarchy



Answer Type Detection

- Most systems use a combination of hand-crafted rules and supervised machine learning to determine the right answer type for a question.
- But remember our notion of matching. It doesn't do any good to do something complex here if it can't also be done in potential answer texts.

Keyword Selection

- Answer Type indicates *what* the question is looking for, but that doesn't really help in finding relevant texts (i.e. Ok, let's look for texts with people in them)
  - Lexical terms (keywords) from the question, possibly expanded with lexical/semantic variations provide the required context.

Lexical Terms Extraction

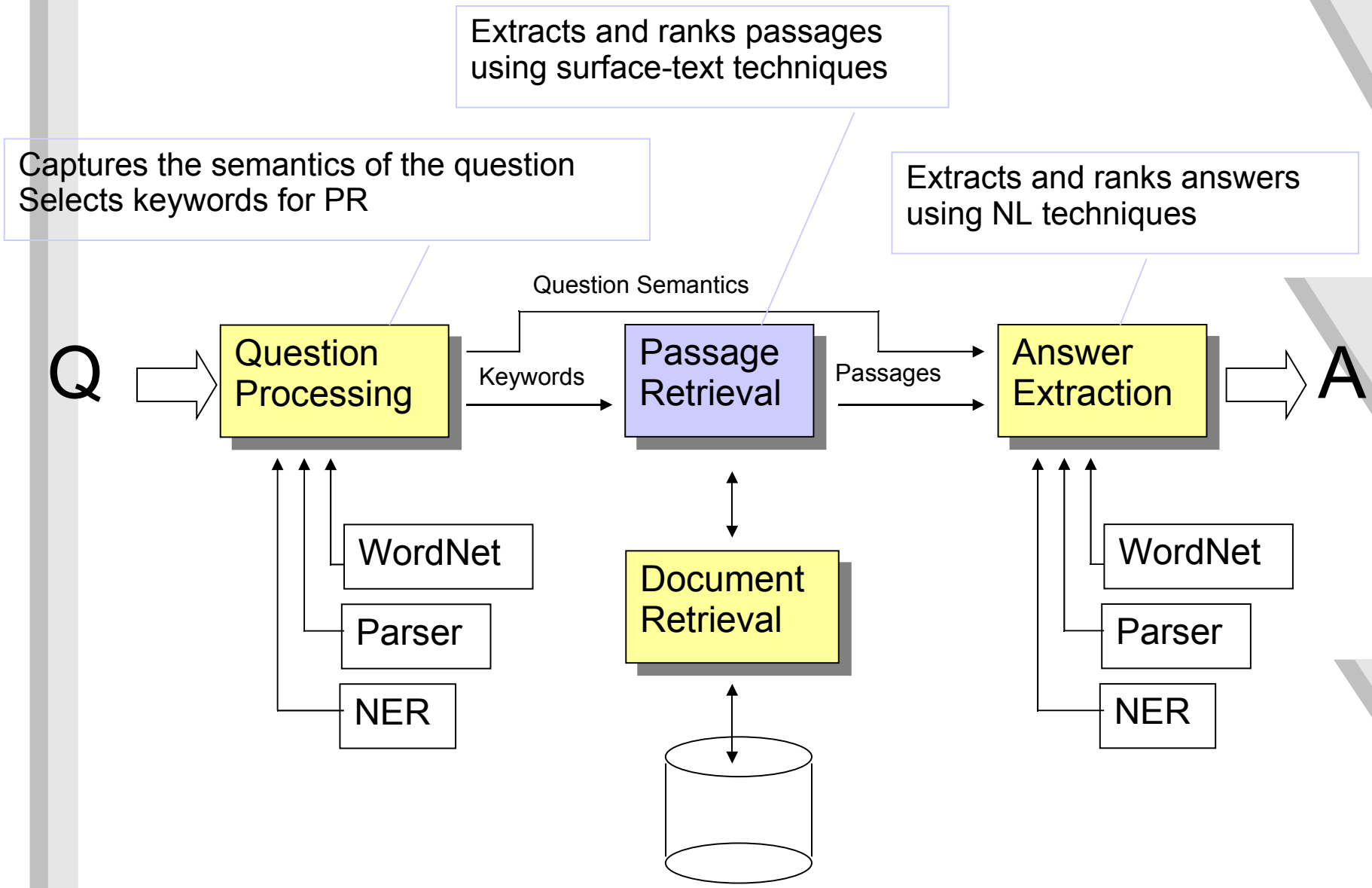
- Questions approximated by sets of unrelated words (lexical terms)
- Similar to bag-of-word IR models

Question (from TREC QA track)	Lexical terms
Q002: What was the monetary value of the Nobel Peace Prize in 1989?	monetary, value, Nobel, Peace, Prize
Q003: What does the Peugeot company manufacture?	Peugeot, company, manufacture
Q004: How much did Mercury spend on advertising in 1993?	Mercury, spend, advertising, 1993
Q005: What is the name of the managing director of Apricot Computer?	name, managing, director, Apricot, Computer

Keyword Selection Algorithm

- Select all non-stopwords in quotations
- Select all NNP words in recognized named entities
- Select all complex nominals with their adjectival modifiers
- Select all other complex nominals
- Select all nouns with adjectival modifiers
- Select all other nouns
- Select all verbs
- Select the answer type word

Passage Retrieval



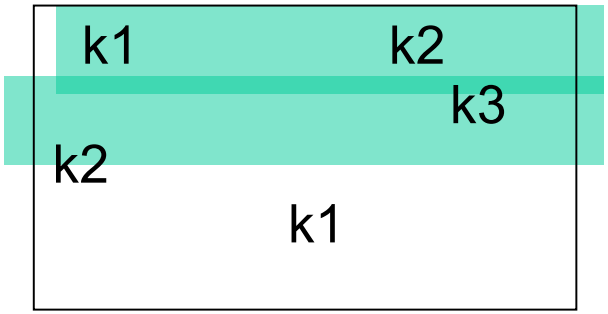
Passage Extraction Loop

- Passage Extraction Component
  - Extracts passages that contain all selected keywords
  - Passage size dynamic
  - Start position dynamic
- Passage quality and keyword adjustment
  - In the first iteration use the first 6 keyword selection heuristics
  - If the number of passages is lower than a threshold ⇒ query is too strict ⇒ drop a keyword
  - If the number of passages is higher than a threshold ⇒ query is too relaxed ⇒ add a keyword

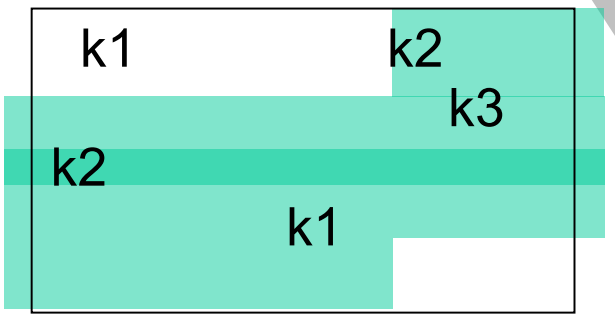
Passage Scoring

- Passages are scored based on keyword windows
  - For example, if a question has a set of keywords: {k1, k2, k3, k4}, and in a passage k1 and k2 are matched twice, k3 is matched once, and k4 is not matched, the following windows are built:

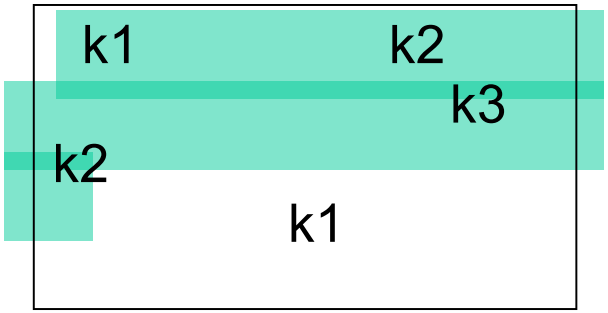
Window 1



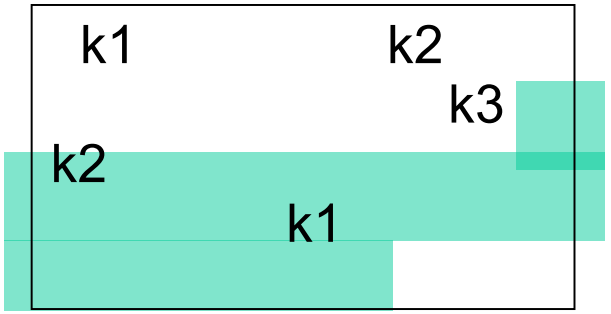
Window 2



Window 3



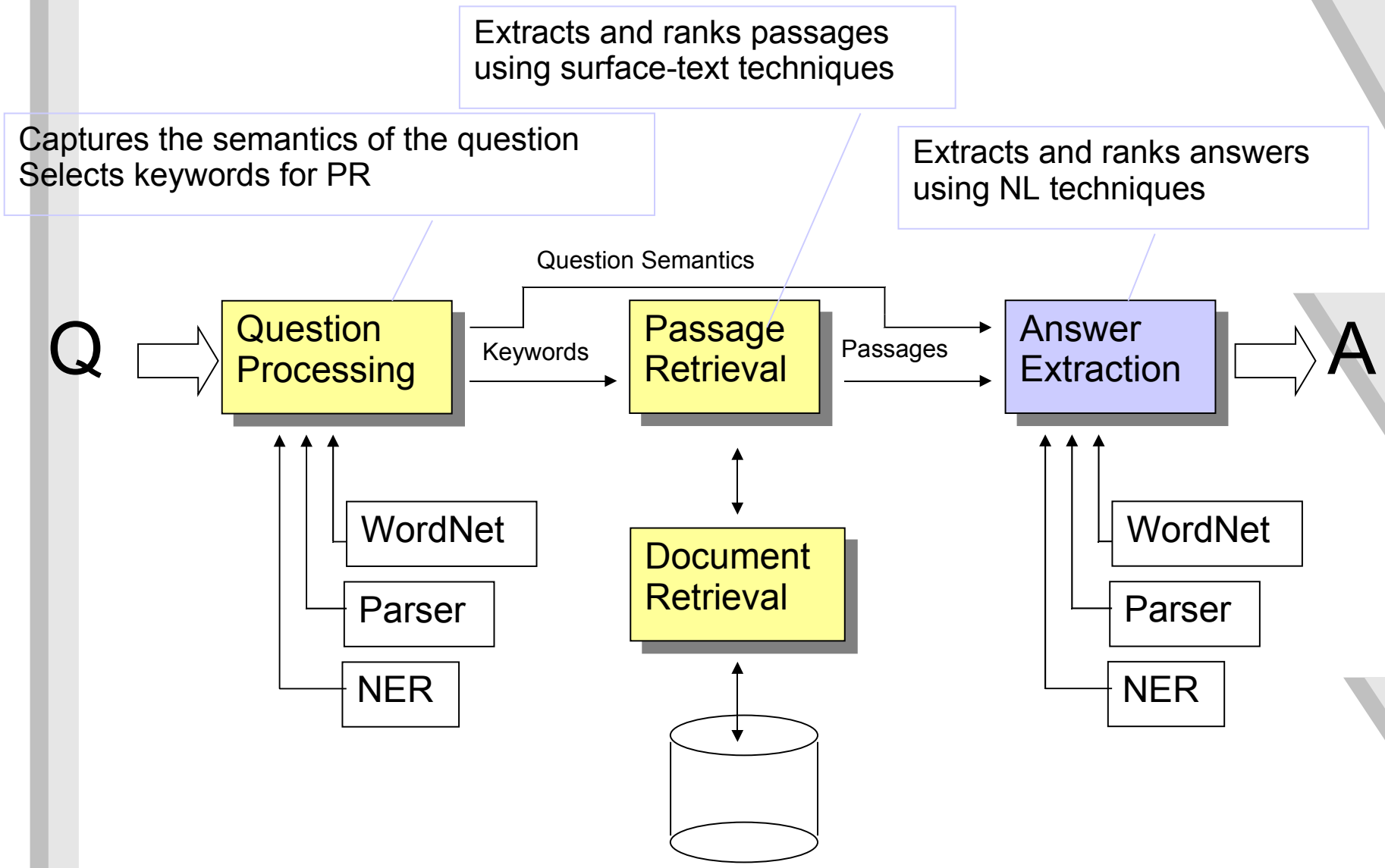
Window 4



Passage Scoring

- Passage ordering is performed using a trained re-ranking algorithm that involves three scores:
  - The number of words from the question that are recognized in the same sequence in the window
  - The number of words that separate the most distant keywords in the window
  - The number of unmatched keywords in the window

Answer Extraction



Ranking Candidate Answers

Q066: Name the first private citizen to fly in space.

- Answer type: Person
- Text passage:

  "Among them was Christa McAuliffe, the first private citizen to fly in space. Karen Allen, best known for her starring role in "Raiders of the Lost Ark", plays McAuliffe. Brian Kerwin is featured as shuttle pilot Mike_Smith..."

Ranking Candidate Answers

Q066: Name the first private citizen to fly in space.

- Answer type: Person
- Text passage:

  "Among them was Christa McAuliffe, the first private citizen to fly in space. Karen Allen, best known for her starring role in "Raiders of the Lost Ark", plays McAuliffe. Brian Kerwin is featured as shuttle pilot Mike_Smith..."

- Best candidate answer: Christa McAuliffe

Features for Answer Ranking

- Number of question terms matched in the answer passage
- Number of question terms matched in the same phrase as the candidate answer
- Number of question terms matched in the same sentence as the candidate answer
- Flag set to 1 if the candidate answer is followed by a punctuation sign
- Number of question terms matched, separated from the candidate answer by at most three words and one comma
- Number of terms occurring in the same order in the answer passage as in the question
- Average distance from candidate answer to question term matches

Evaluation

- Evaluation of this kind of system is usually based on some kind of TREC-like metric.
- In Q/A the most frequent metric is
  - Mean Reciprocal Rank

    You're allowed to return N answers. Your score is based on 1/Rank of the first right answer.

    Averaged over all the questions you answer.

Is the Web Different?

- In TREC (and most commercial applications), retrieval is performed against a closed relatively homogeneous collection of texts.
- The diversity/creativity in how people express themselves necessitates all that work to bring the question and the answer texts together.
- But…

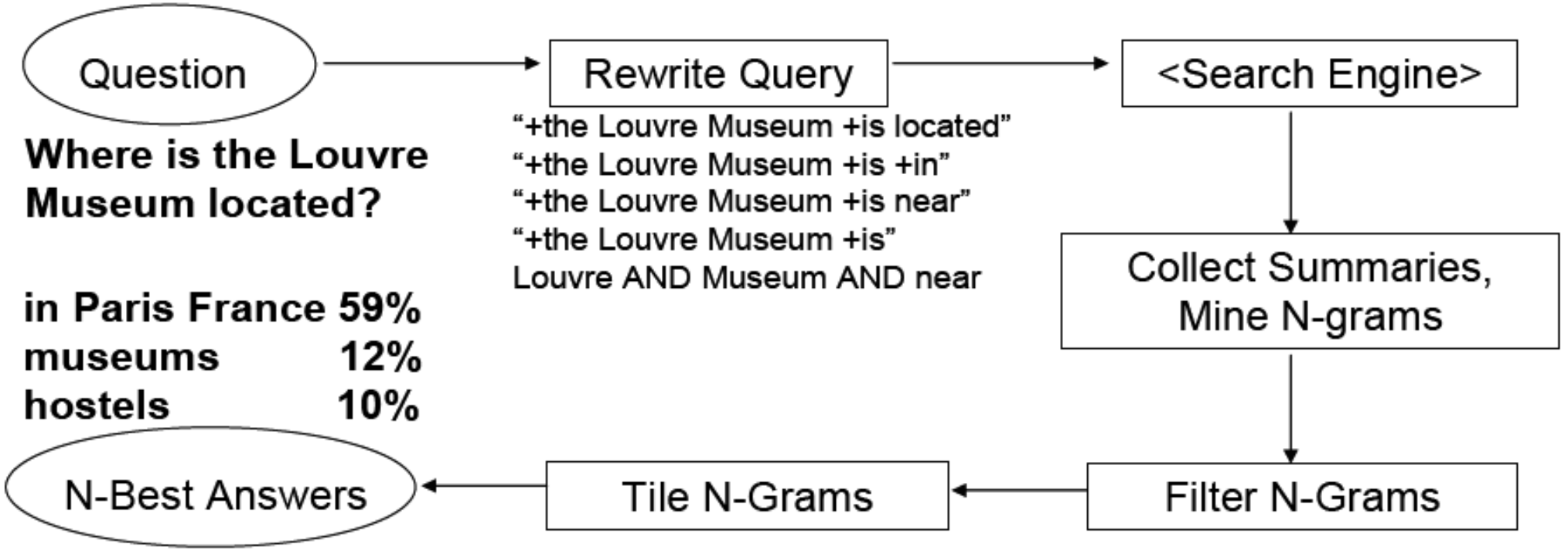
The Web is Different

- On the Web popular factoids are likely to be expressed in a gazzilion different ways.
- At least a few of which will likely match the way the question was asked.
- So why not just grep (or agrep) the Web using all or pieces of the original question.

AskMSR

- Process the question by…
  - Simple rewrite rules to rewriting the original question into a statement
    - Involves detecting the answer type
- Get some results
- Extract answers of the right type based on
  - How often they occur

AskMSR


Question
Where is the Louvre Museum located?
in Paris France 59%
museums 12%
hostels 10%
Rewrite Query
"+the Louvre Museum +is located"
"+the Louvre Museum +is +in"
"+the Louvre Museum +is near"
"+the Louvre Museum +is"
Louvre AND Museum AND near
<Search Engine>
Collect Summaries, Mine N-grams
Filter N-Grams
Tile N-Grams
N-Best Answers

Step 1: Rewrite the questions

- Intuition: Users' questions are often syntactically quite close to sentences that contain the answer
  - Where is the Louvre Museum located?
    - The Louvre Museum is located in ***Paris***
  - Who created the character of Scrooge?
    - ***Charles Dickens*** created the character of Scrooge.

Query rewriting

Classify question into seven categories

- **Who** is/was/are/were...?
- **When** is/did/will/are/were ...?
- **Where** is/are/were ...?

a. Hand-crafted category-specific transformation rules
e.g.: For *where* questions, move 'is' to all possible locations
Look to the right of the query terms for the answer.

"Where is the Louvre Museum located?"
→ "is the Louvre Museum located"
→ "the is Louvre Museum located"
→ "the Louvre is Museum located"
→ "the Louvre Museum is located"
→ "the Louvre Museum located is"

Step 2: Query search engine

- Send all rewrites to a Web search engine
- Retrieve top N answers (100-200)
- For speed, rely just on search engine's "snippets", not the full text of the actual document

Step 3: Gathering N-Grams

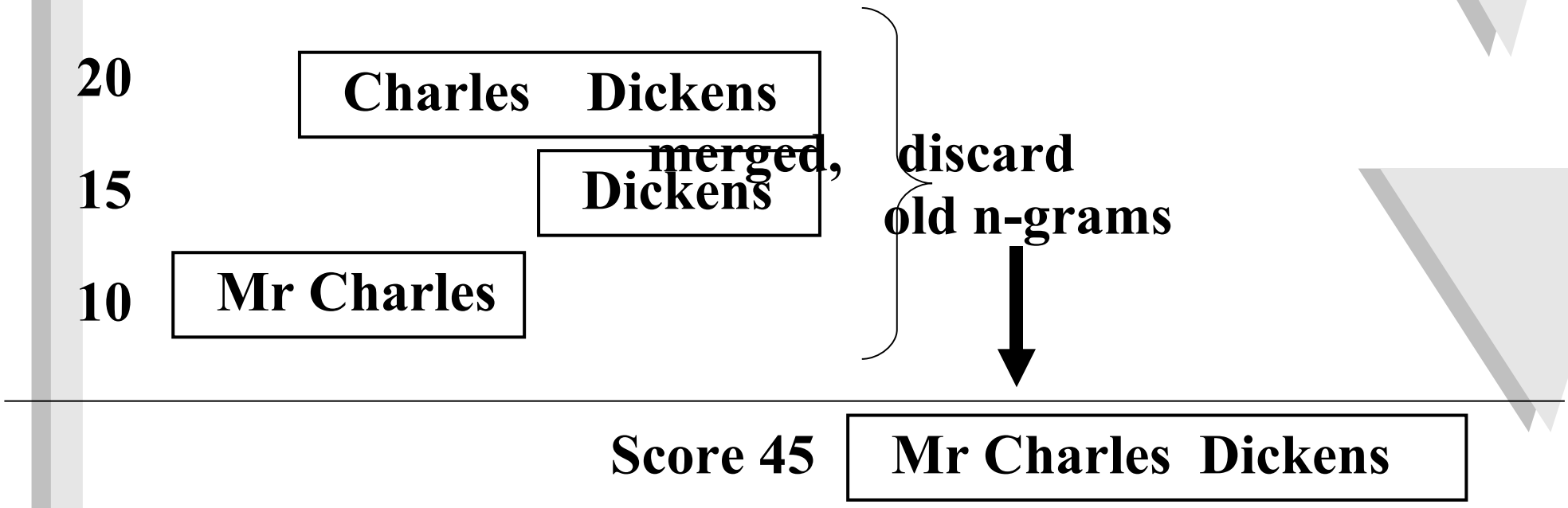
- Enumerate all N-grams (N=1,2,3) in all retrieved snippets
- Weight of an n-gram: occurrence count, each weighted by "reliability" (weight) of rewrite rule that fetched the document (can be trained).
  - Example: "Who created the character of Scrooge?"

Dickens		117
Christmas Carol	78	
Charles Dickens	75	
Disney		72
Carl Banks		54
A Christmas		41
Christmas Carol	45	
Uncle		31

Step 4: Filtering N-Grams

- Each question type is associated with one or more "data-type filters" = regular expressions for answer types
- Boost score of n-grams that match the expected answer type.
- Lower score of n-grams that don't match.
- For example
  - The filter for
    - How many dogs pull a sled in the Iditarod?
  - prefers a number
  - So disprefer candidate n-grams like
    - Dog race, run, Alaskan, dog racing
  - Prefer canddiate n-grams like
    - Pool of 16 dogs

Step 5: Tiling the Answers



Results

- Standard TREC contest test-bed (TREC 2001): 1M documents; 900 questions
  - Technique does ok, not great (would have placed in top 9 of ~30 participants)
    - MRR = 0.507
  - But with access to the Web… They do much better, would have come in second on TREC 2001
    - Be suspicious of any *after the bake-off is over* metrics