In the name of God
the Compassionate, the Merciful

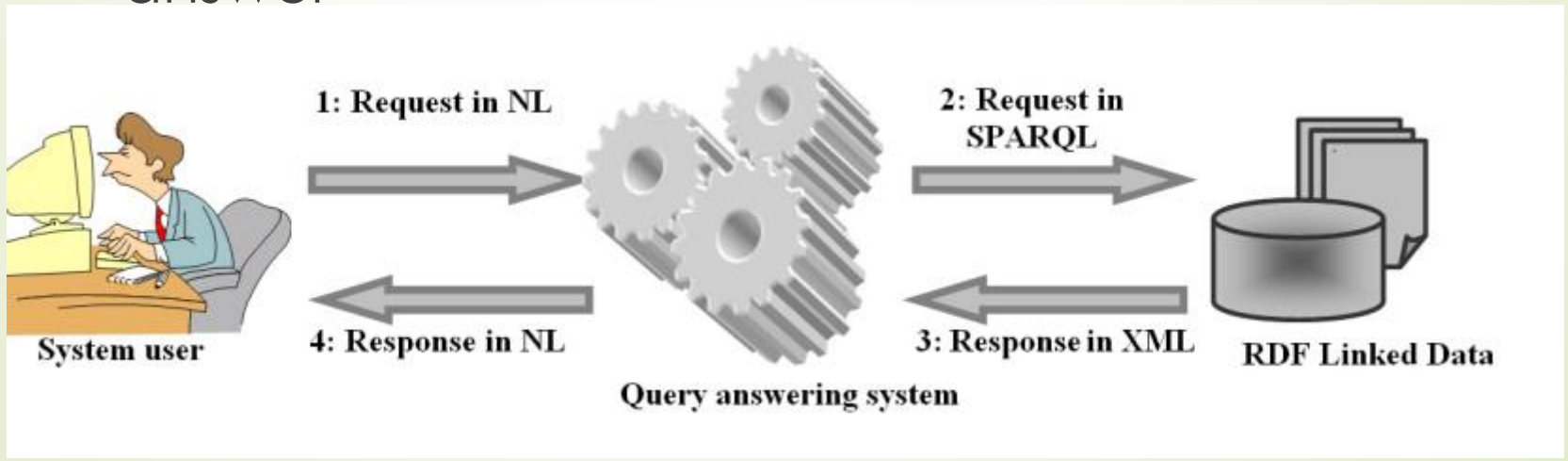
Question Answering and Information Retrieval

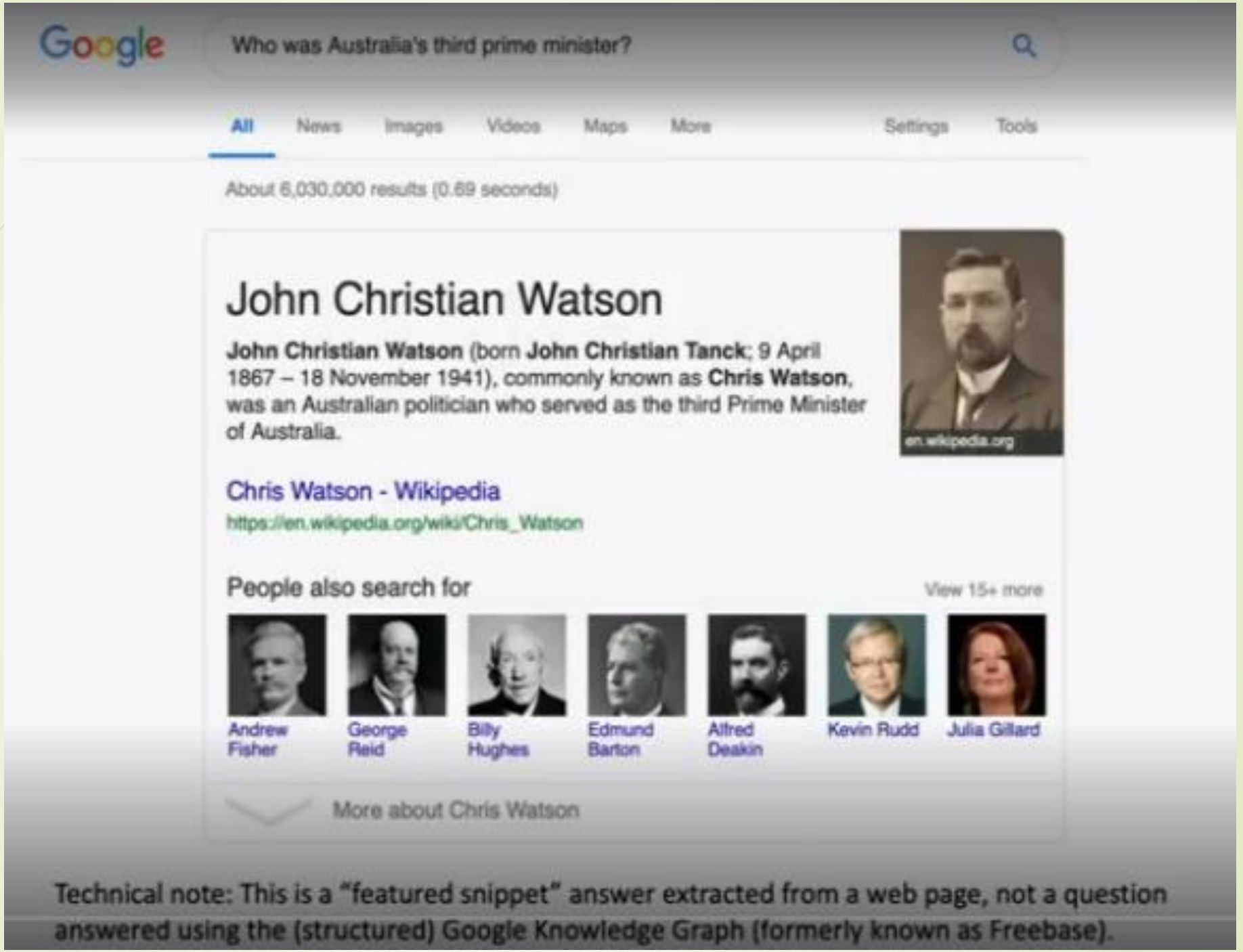
Content

- Problem Definition
- Architectures
- Benchmarks

Definition

- HCI
- Input: NL (not Keywords)
- Output: NL (not document)
- Difference with Search Engine: no doc, we want exact answer





Technical note: This is a "featured snippet" answer extracted from a web page, not a question answered using the (structured) Google Knowledge Graph (formerly known as Freebase).

TREC

Text REtrieval Conference

...to encourage research in information retrieval from large text collections.

Abbreviation	TREC
Discipline	information retrieval

Publication details

Publisher	NIST
History	1992; 32 years ago
Frequency	annual
Website	trec.nist.gov

Example

IBM Watson defeated two of Jeopardy's greatest champions in 2011

Beyond textual QA problems

- Today, we will mostly focus on how to answer questions based on **unstructured text**.

Visual QA

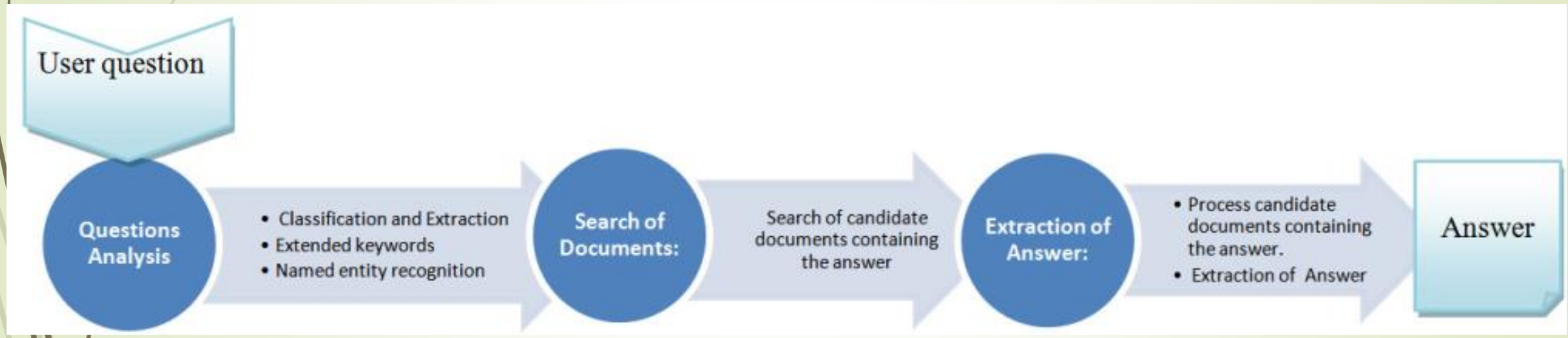
What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

(Antol et al., 2015): Visual Question Answering

Typical model

- Modules (Q processing, Doc Processing, Answer Ext)



Questions

- Factoid (simple..)
  - When, where, how many, how much, who, what
- Non-factoid (hard..)
  - Why, How

Question	Answer
Where is the Louvre Museum located?	in Paris, France
What are the names of Odin's ravens?	Huginn and Muninn
What kind of nuts are used in marzipan?	almonds
What instrument did Max Roach play?	drums
What's the official language of Algeria?	Arabic

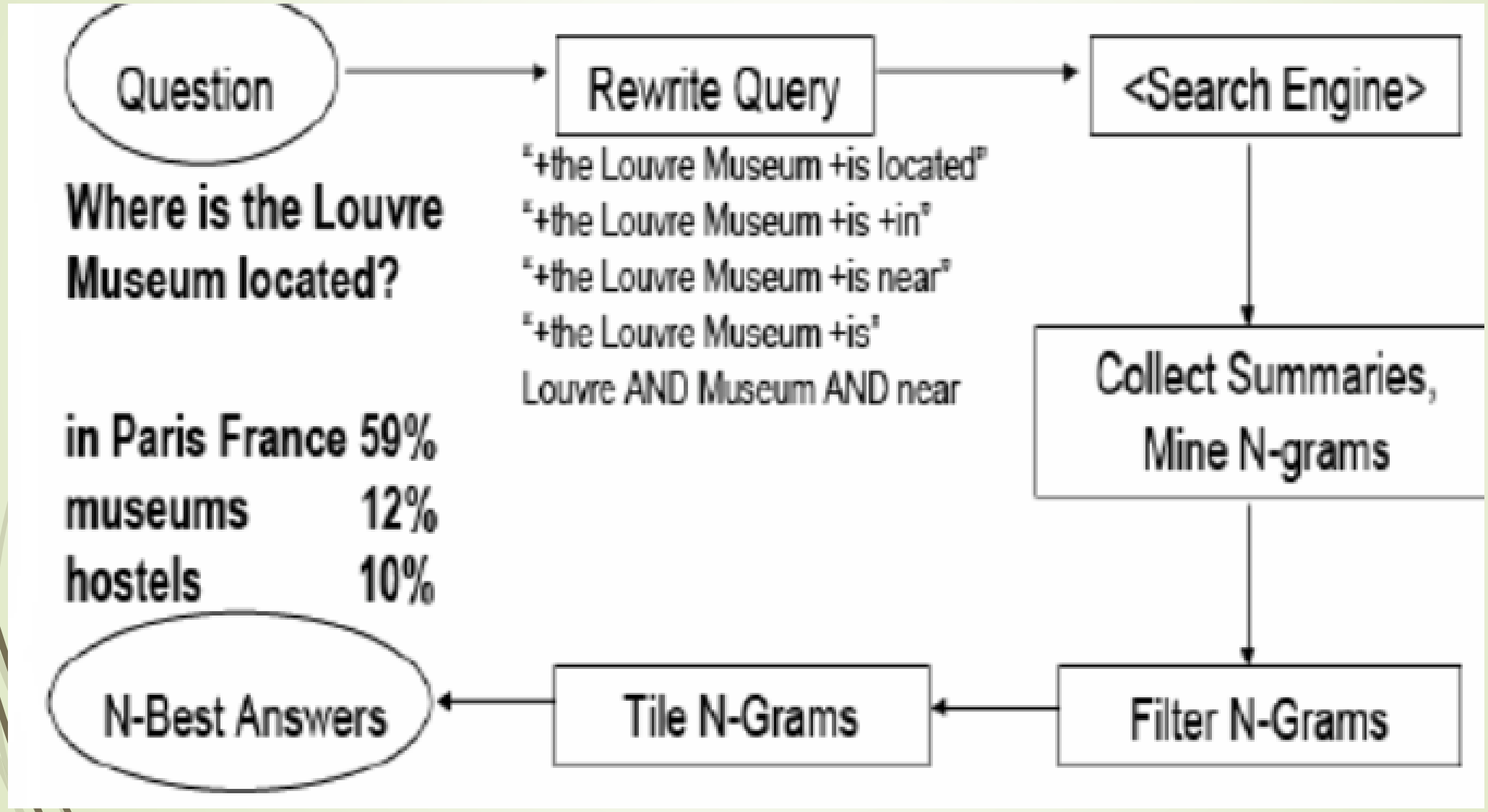
Figure 14.9 Some factoid questions and their answers.

Topics Taxonomy

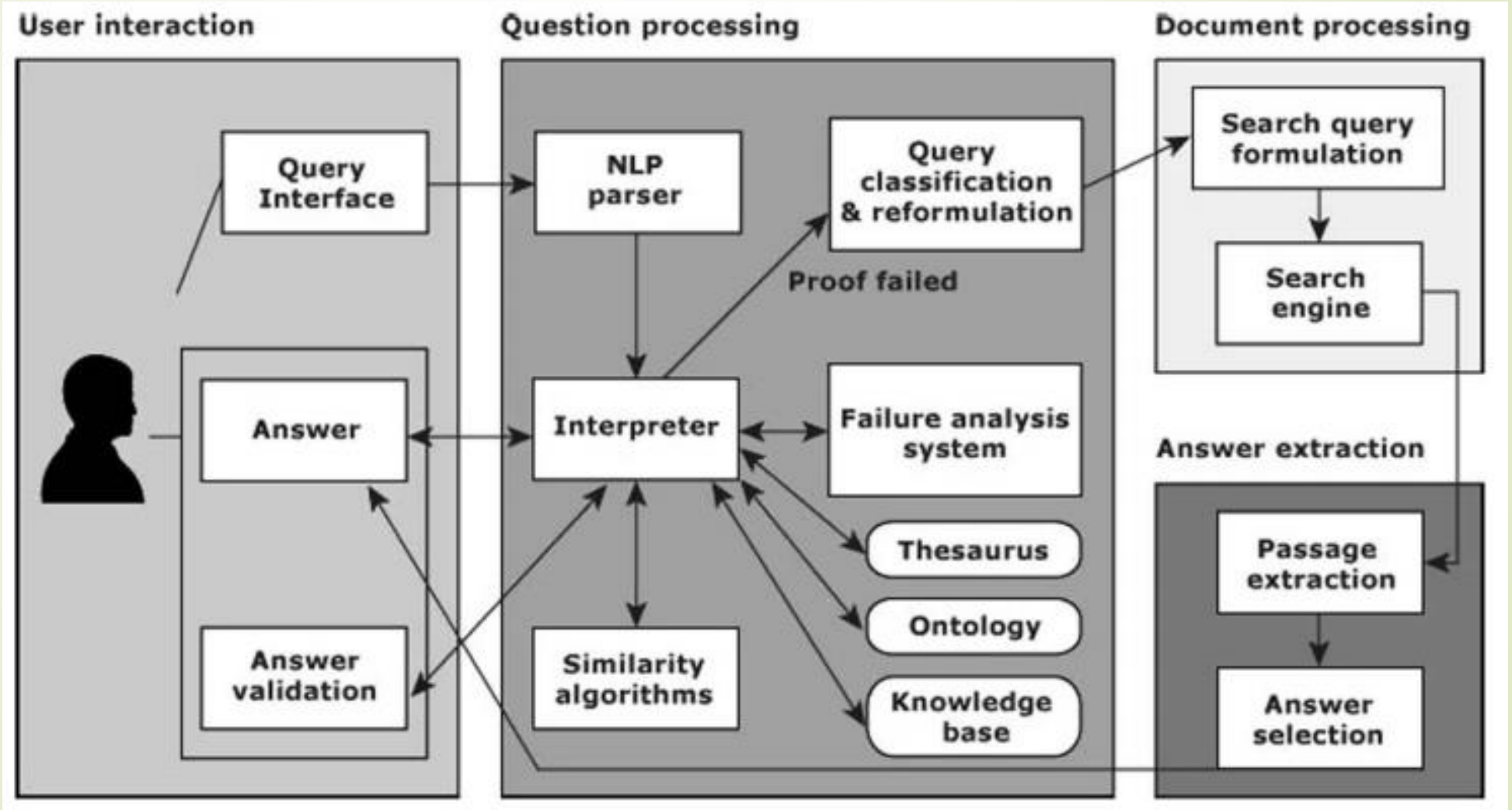
- Flat
- Hierarchical (more precise and time-consuming)

- Rule-based
- Machine Learning based
  - Decision Tree
  - Naïve Bayes
  - SVM

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

Closed Domain Question Answering Systems (ex. Baseball)



Other Categorizations from KB point of view

- Graph-based
- Semantic Parsing
  - The system should **comprehend!**

subject	predicate	object
Ada Lovelace	birth-year	1815

"When was Ada Lovelace born?" → `birth-year (Ada Lovelace, ?x)`
"What is the capital of England?" → `capital-city(?x, England)`

How many people survived the sinking of the Titanic?	(count (!fb:event.disaster.survivors fb:en.sinking.of.the.titanic))
How many yards longer was Johnson's longest touchdown compared to his shortest touchdown of the first quarter?	ARITHMETIC diff(SELECT num(ARGMAX(SELECT)) SELECT num(ARGMIN(FILTER(SELECT))))

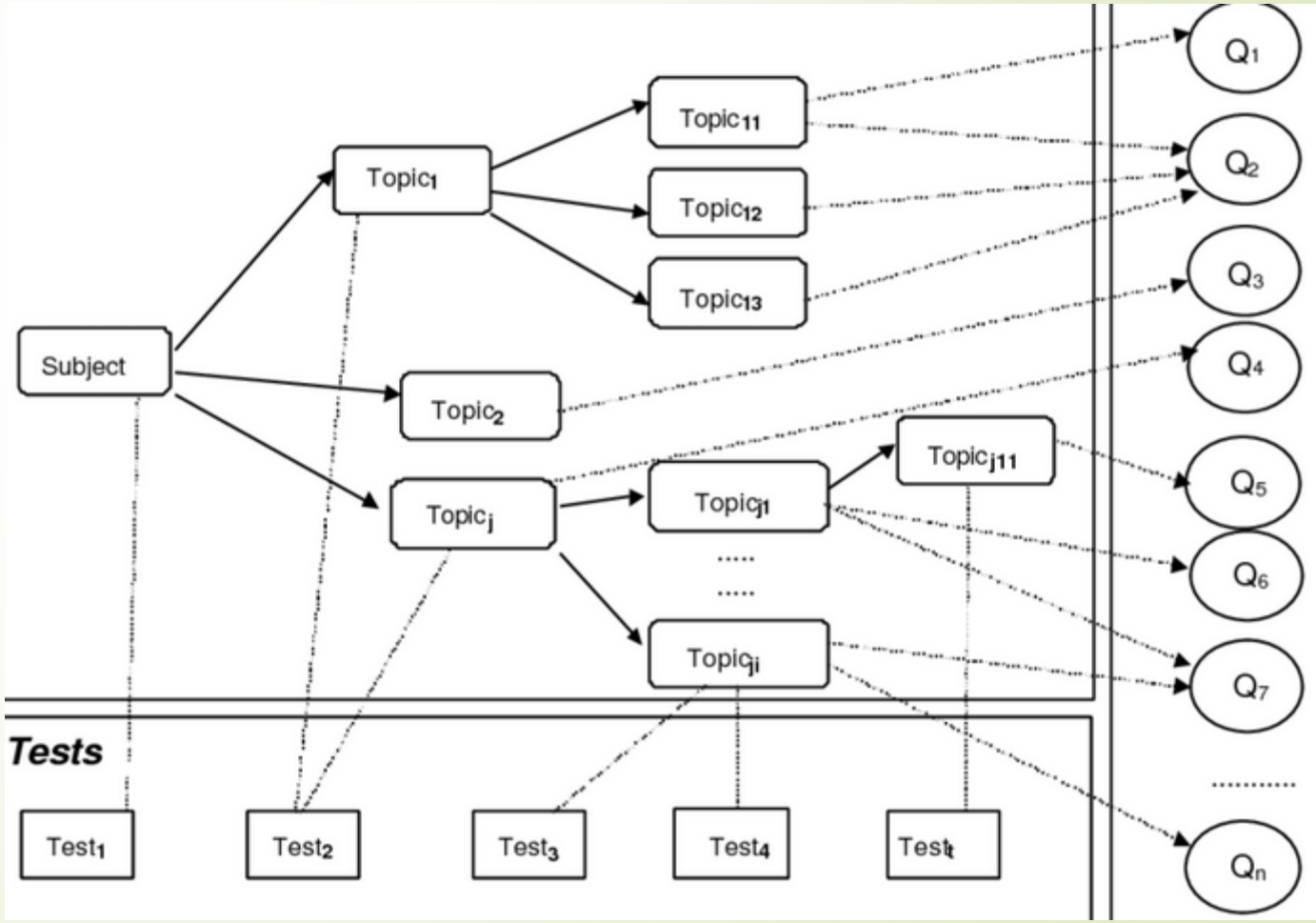
$$\mathbf{m}_r = \text{BERT}_{\text{CLS}}([\text{CLS}]q_1 \cdots q_n[\text{SEP}])$$

$$s(\mathbf{m}_r, r_i) = \mathbf{m}_r \cdot \mathbf{w}_{r_i}$$

$$p(r_i|q_1, \cdots, q_n) = \frac{\exp(s(\mathbf{m}_r, r_i))}{\sum_{i=1}^{N_R} \exp(s(\mathbf{m}_r, r_i))}$$

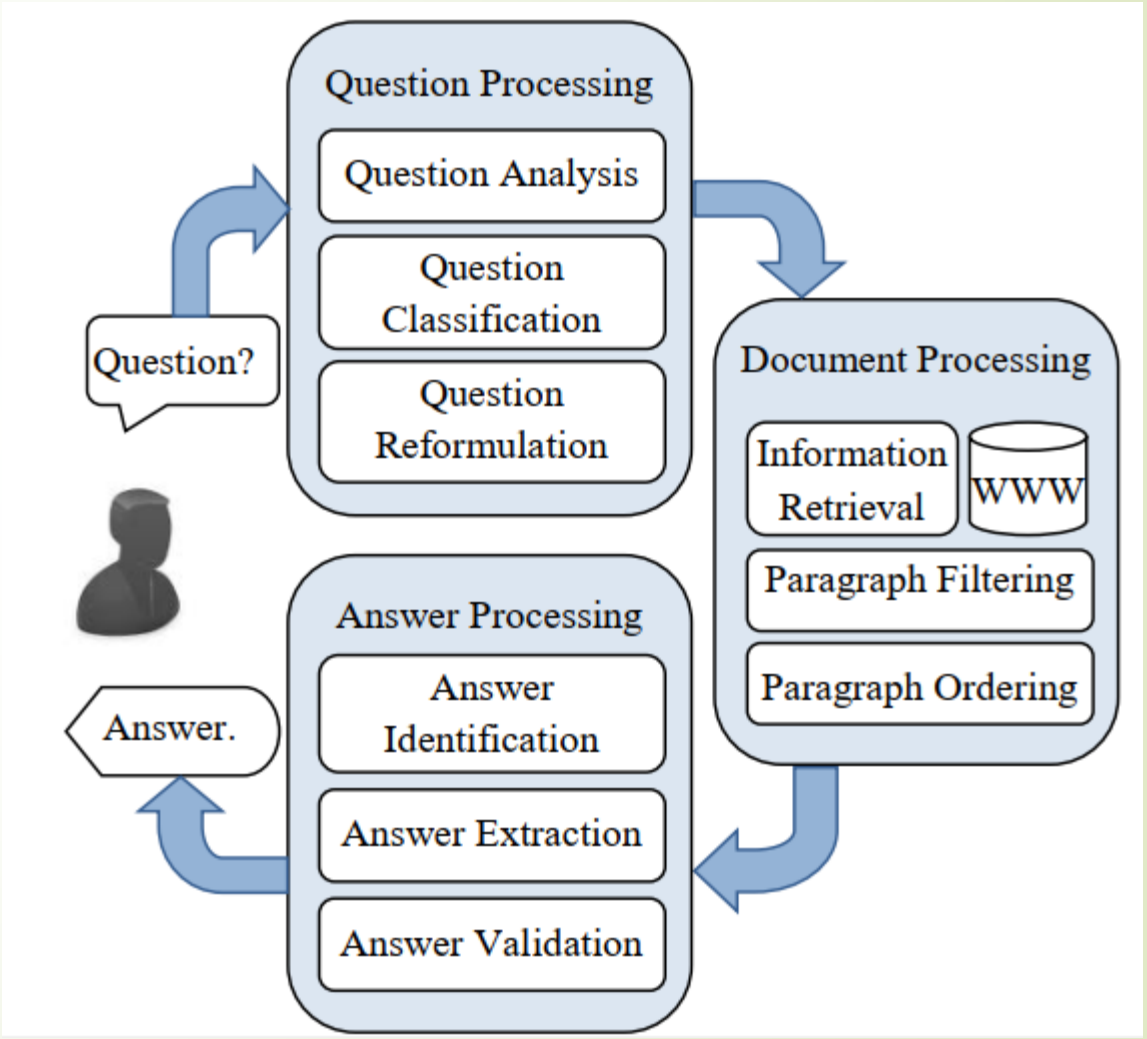
Knowledge base

- Structured
- Unstructured

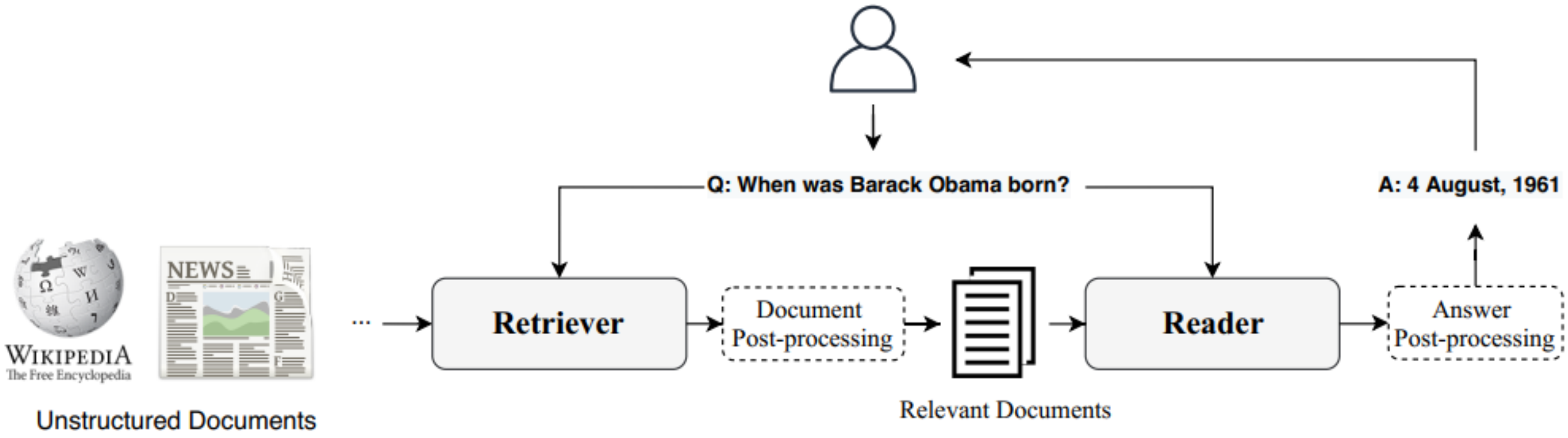


Architecture of a QAS

- Keyword, convert to Machine Understandable
- Filtering and ordering
- Validation
- Target: achieve Human understanding

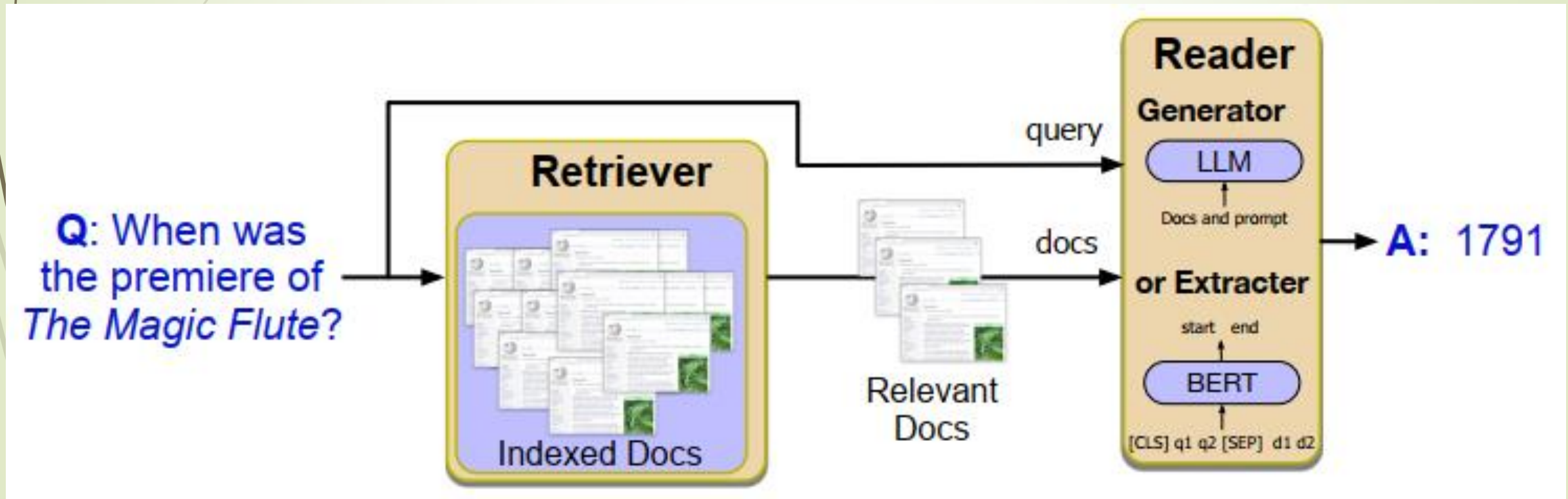


Open Domain QAS



Open Domain QAS

- Both parts may use ML techniques



Retriever Part..

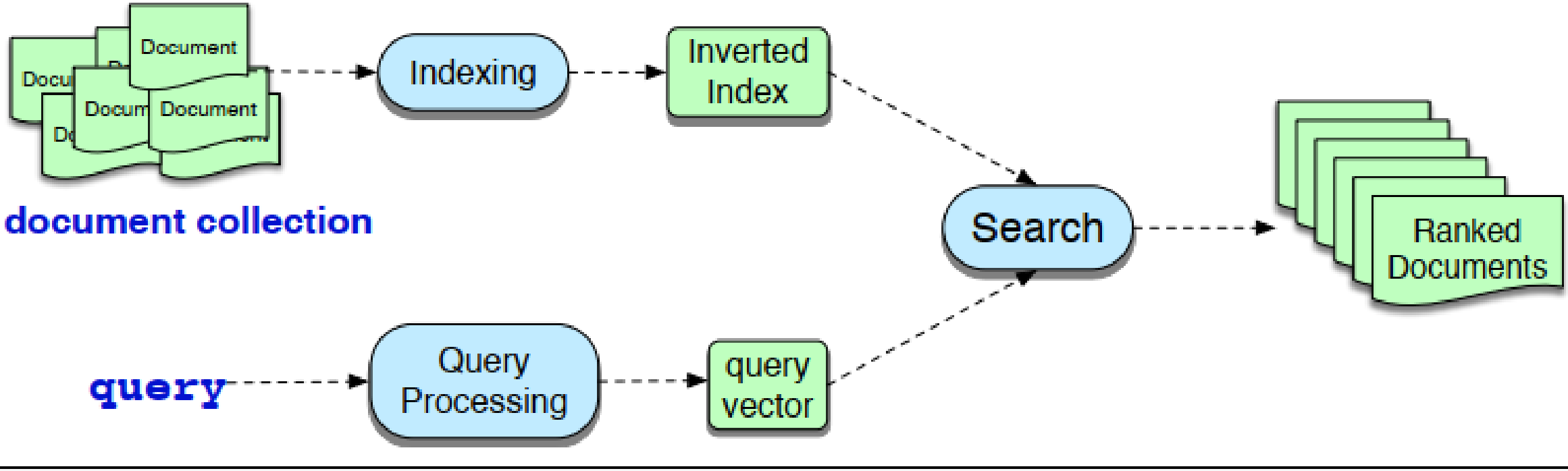


Figure 14.1 The architecture of an ad hoc IR system.

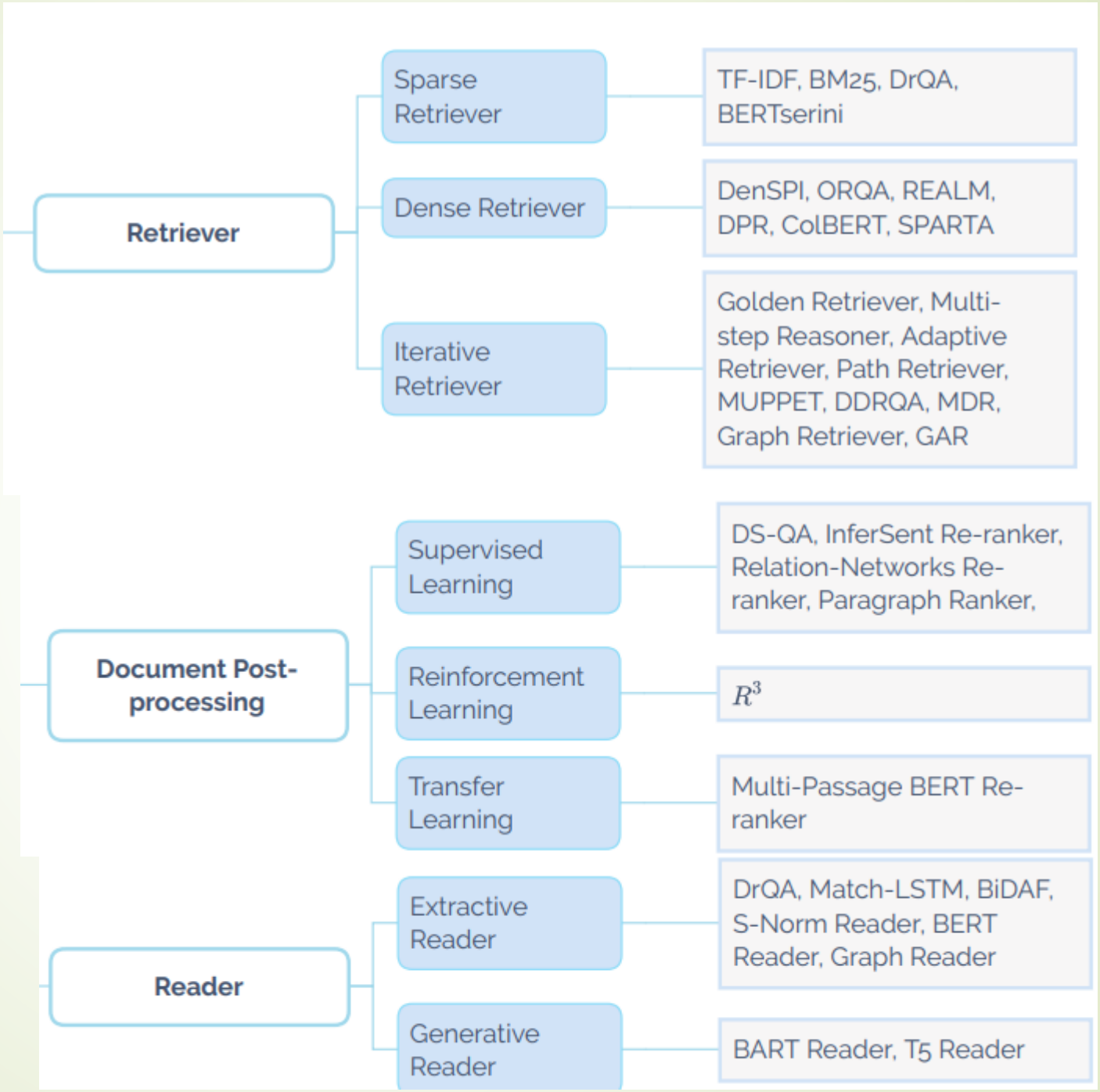
Retriever Part..

- Should return **Relevant** documents
- Boolean Model (using and-or of keywords)
- Vector space Model (in vector space, cosine similarity and Euclidian distance)
- Probabilistic Model (ex. Naïve bayes)

THE FIRST three categories focus on tokens and processing

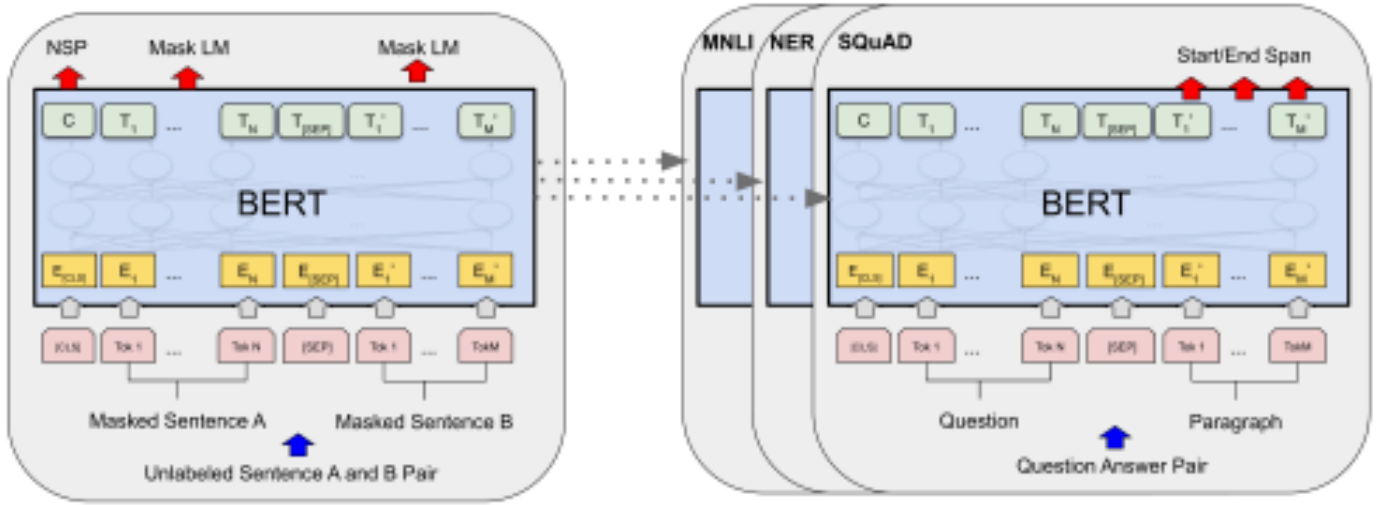
- Language Model (ex. BERT)
  - Using semantics, synonyms, context

- **I like apples**
- **I like apple MacBook's**


Retriever
Sparse Retriever
TF-IDF, BM25, DrQA, BERTserini
Dense Retriever
DenSPI, ORQA, REALM, DPR, ColBERT, SPARTA
Iterative Retriever
Golden Retriever, Multi-step Reasoner, Adaptive Retriever, Path Retriever, MUPPET, DDRQA, MDR, Graph Retriever, GAR
Document Post-processing
Supervised Learning
DS-QA, InferSent Re-ranker, Relation-Networks Re-ranker, Paragraph Ranker,
Reinforcement Learning
R^3
Transfer Learning
Multi-Passage BERT Re-ranker
Reader
Extractive Reader
DrQA, Match-LSTM, BiDAF, S-Norm Reader, BERT Reader, Graph Reader
Generative Reader
BART Reader, T5 Reader

BERT for reading comprehension

- BERT is a deep bidirectional Transformer encoder pre-trained on large amounts of text (Wikipedia + BooksCorpus)
- BERT is pre-trained on two training objectives:
  - Masked language model (MLM)
  - Next sentence prediction (NSP)
- $\text{BERT}_{\text{base}}$ has 12 layers and 110M parameters, $\text{BERT}_{\text{large}}$ has 24 layers and 330M parameters

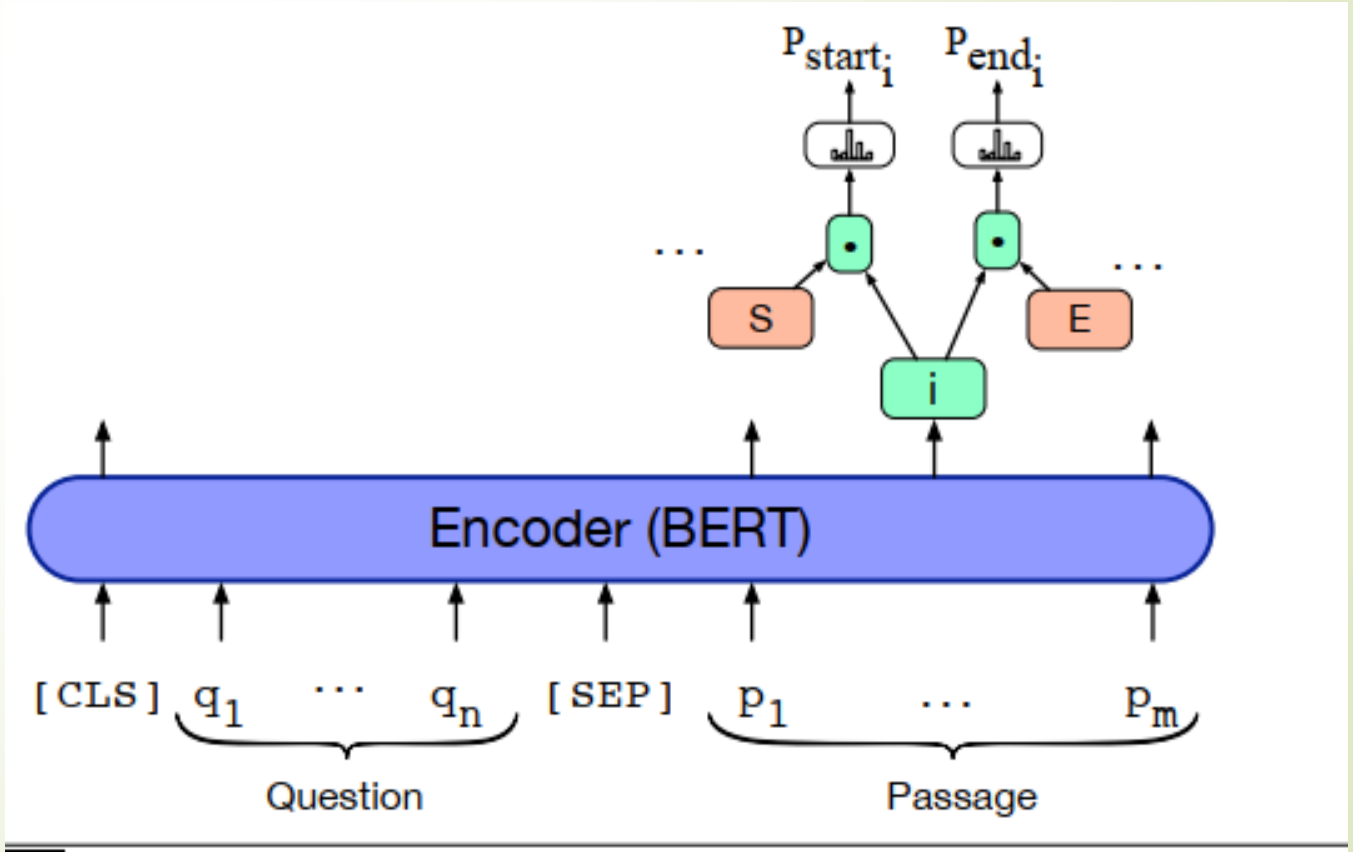


Pre-training Fine-Tuning

An encoder model (using BERT) for span-based question answering from reading-comprehension-based question answering tasks

$$P_{\text{start}_i} = \frac{\exp(S \cdot p'_i)}{\sum_j \exp(S \cdot p'_j)}$$

$$P_{\text{end}_i} = \frac{\exp(E \cdot p'_i)}{\sum_j \exp(E \cdot p'_j)}$$



Stanford question answering dataset (SQuAD)

- 100k annotated (**passage, question, answer**) triples
- Large-scale **supervised** datasets are also a key ingredient for training effective neural models for reading comprehension!
  This is a limitation— not all the questions can be answered in this way!
- Passages are selected from English **Wikipedia**, usually 100~150 words.
- Questions are **crowd-sourced**.
- Each answer is a short segment of text (or span) in the passage.
- SQuAD still remains the most popular reading comprehension dataset; it is "almost solved" today and the state-of-the-art exceeds the estimated human performance.

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, Dangerously in Love (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".
Q: "In what city and state did Beyoncé grow up?" A: "Houston, Texas"
Q: "What areas did Beyoncé compete in when she was growing up?" A: "singing and dancing"
Q: "When did Beyoncé release Dangerously in Love?" A: "2003"

Figure 14.11 A (Wikipedia) passage from the SQuAD 2.0 dataset (Rajpurkar et al., 2018) with 3 sample questions and the labeled answer spans.

SQuAD limitations

- SQuAD has a number of other key limitations too:
  - Only span-based answers (no yes/no, counting, implicit why)
  - Questions were constructed looking at the passages
    - Not genuine information needs
    - Generally greater lexical and syntactic matching between questions and answer span than you get IRL
  - Barely any multi-fact/sentence inference beyond coreference

Evaluation Metrics of Answer

1. Relevance (of answer)
2. Correctness (of answer)
3. Conciseness (shortness of answer)
4. Completeness
5. Justification
   - *Correct answer: 1 and 2*
   - *Inexact: 3 and 4*
   - *Unsupported: 5*

- Precision, Recall, F-measure
- MRR (mean reciprocal rank)
- Confidence Weighted Score

- Precision =number of correct answers /number of questions answered
- Recall =number of correct answers /number of questions to be answered
- F measure =2* (Precision * recall) /Precision +recall (harmonic mean)


THE
END