

# Lecture 19: Web Question Answering

## Background: New trends in web search

- Despite the name *information retrieval*, it's not really *information* retrieval. It's more of *document* retrieval.
- New common theme in search engines: doing graph search over structured knowledge rather than traditional text search.
  - More semi-structured information embedded in web pages
- Move to mobile favors a move to speech, which favors *natural language information search* - proving importance of NLU and **QA**.
- Toward more intelligent agents: two goals
  1. Things not strings
  2. Inference not search

## Comparing 3 QA approaches

### Knowledge-based approaches (Siri): current trend

- Build a semantic representation of the query: times, dates, locations, entities, numeric quantities
- Map from this semantics to query structured data or resources: geospatial databases, ontologies, etc.

### Text-based QA

- **Question processing**: detect question type, answer type, focus, relations; formulate them as queries for the search engine
- **Passage retrieval**: Retrieve ranked documents; break into suitable passages and rerank.
- **Answer processing**: Extract candidate answers; rank candidates using evidence from *relations* in the text & external sources

### Hybrid

- Build a *shallow* semantic representation of the query
- Generate answer candidates using IR methods: augmented with ontologies and semi-structured data
- Score each candidate using *richer knowledge sources*: geospatial databases, temporal reasoning, taxonomical classification

## Learning actions from web usage logs

- Bing: Towards actions
  - Recognize entity in query
  - (associated) (potential user) actions easily accessible (in the interface)
  - Click through experience can now leverage strongly-typed identifier
  - Brokered actions (one click conversions)

## Entity disambiguation and linking

- Entities need to get identified and disambiguated
  - Named entity recognition
  - Entity linking (“Wikification”)