

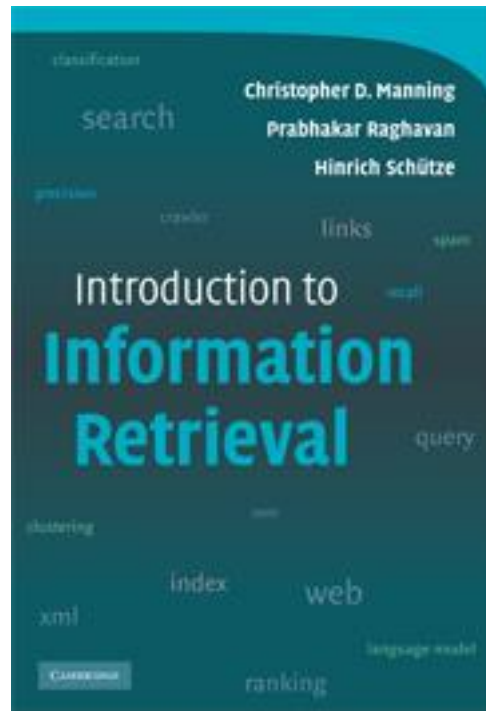
# Missing in IIR

CS5154/6054

Yizong Cheng

11/17/2022

# IIR Has Been there since 2008



# Additions to CS276 (2019) by Chris Manning

- Decision tree for document classification
- Word representations/neural embedding
- Learning to rank
- Question answering

Christopher Manning

Pandu Nayak

# Introduction to **Information Retrieval**

CS276: Information Retrieval and Web Search

Lecture 19: Web Question Answering

# “Information retrieval”

---

The name **information retrieval** is standard, but as traditionally practiced, it's not really right

All you get is **document retrieval**, and beyond that the job is up to you

# Web Search in 2025?

---

The web, it is a changing.

What will people do in 2025?

- Type key words into a search box?
- Use the Semantic Web?
- Ask questions to their computer in natural language?
- Use social or “human powered” search?

# The role of knowledge bases

---

- Google Knowledge Graph
- Facebook Graph Search
- Bing's Satori
- Things like Wolfram Alpha

Common theme: Doing graph search over structured knowledge rather than traditional text search

# Mobile

Move to mobile favors a move to **speech** which favors **natural language information search**

- Will we move to a time when over half of searches are spoken?





# Towards intelligent agents

---

Two goals

- Things not strings
- Inference not search

# Two paradigms for question answering

---

- Text-based approaches
  - TREC QA, IBM Watson, DrQA
- Structured knowledge-based approaches
  - Apple Siri, Wolfram Alpha, Facebook Graph Search

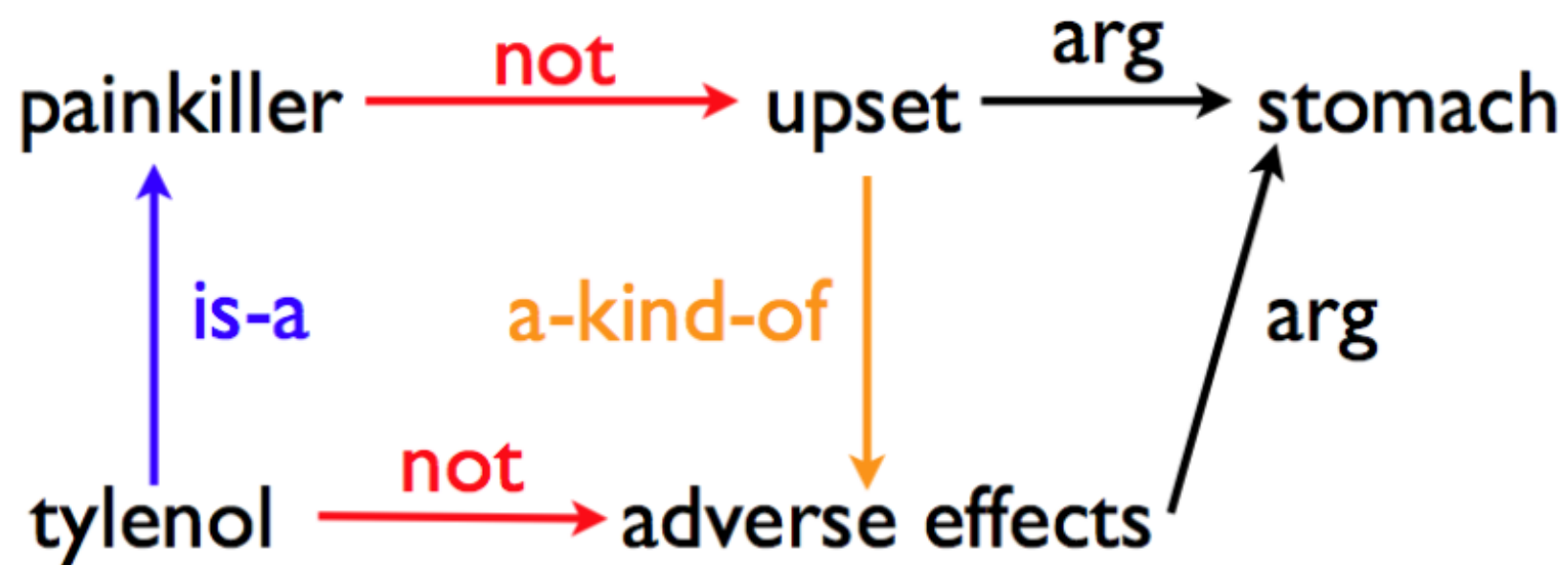
(And, of course, there are hybrids, including some of the above.)

At the moment, structured knowledge is back in fashion, but it may or may not last

painkiller

upset

stomach



# “Things, not strings”

---

| From          | To                 | Requires                             |
|---------------|--------------------|--------------------------------------|
| Term          | Concept            | Parsing, disambiguation, coreference |
| Term identity | Entailment         | Concept relations                    |
| Co-occurrence | Syntactic relation | Document structure, parsing          |
| Term index    | Semantic index     | Concept disambiguation, inference    |

# 3 approaches to question answering:

## Knowledge-based approaches (Siri)

---

- 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 (Wikipedia infoboxes, dbPedia, WordNet, Yago)
  - Restaurant review sources and reservation services
  - Scientific databases
  - Wolfram Alpha

# Text-based (mainly factoid) QA

---

- QUESTION PROCESSING
  - Detect **question type**, **answer type**, focus, relations
  - Formulate queries to send to a search engine
- PASSAGE RETRIEVAL
  - Retrieve ranked documents
  - Break into suitable passages and rerank
- ANSWER PROCESSING
  - Extract candidate answers (as named entities)
  - Rank candidates
    - using evidence from relations in the text and external sources

## Hybrid approaches (IBM Watson)

---

- 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



O'REILLY®

# Knowledge Graphs

Data in Context for  
Responsive Businesses

Jesús Barrasa, Amy E. Hodler  
& Jim Webber

REPORT

Synthesis Lectures on  
Data, Semantics, and Knowledge

SYNTHESIS  
COLLECTION OF TECHNOLOGY

Aidan Hogan · Eva Blomqvist · Michael Cochez ·  
Claudia d'Amato · Gerard de Melo · Claudio Gutierrez  
· Sabrina Kirrane · José Emilio Labra Gayo · Roberto  
Navigli · Sebastian Neumaier · Axel-Cyrille Ngonga  
Ngomo · Axel Polleres · Sabbir M. Rashid · Anisa Rula  
· Lukas Schmelzeisen · Juan Sequeda · Steffen Staab ·  
Antoine Zimmermann

# Knowledge Graphs

 Springer

|          |                                      |           |
|----------|--------------------------------------|-----------|
| <b>4</b> | <b>Deductive Knowledge</b>           | <b>47</b> |
| 4.1      | Ontologies                           | 48        |
| 4.1.1    | Interpretations and Models           | 49        |
| 4.1.2    | Ontology Features                    | 51        |
| 4.1.3    | Entailment                           | 56        |
| 4.1.4    | If-Then vs. If-and-Only-If Semantics | 56        |
| 4.2      | Reasoning                            | 57        |
| 4.2.1    | Rules                                | 57        |
| 4.2.2    | Description Logics                   | 60        |
| <b>5</b> | <b>Inductive Knowledge</b>           | <b>67</b> |
| 5.1      | Graph Analytics                      | 67        |
| 5.1.1    | Techniques                           | 69        |
| 5.1.2    | Frameworks                           | 70        |
| 5.1.3    | Analytics on Data Graphs             | 74        |
| 5.1.4    | Analytics with Queries               | 75        |
| 5.1.5    | Analytics with Entailment            | 76        |
| 5.2      | Knowledge Graph Embeddings           | 76        |
| 5.2.1    | Tensor-Based Models                  | 78        |
| 5.2.2    | Language Models                      | 88        |
| 5.2.3    | Entailment-Aware Models              | 90        |
| 5.3      | Graph Neural Networks                | 91        |
| 5.3.1    | Recursive Graph Neural Networks      | 91        |
| 5.3.2    | Non-Recursive Graph Neural Networks  | 95        |
| 5.4      | Symbolic Learning                    | 96        |
| 5.4.1    | Rule Mining                          | 97        |
| 5.4.2    | Axiom Mining                         | 100       |
| 5.4.3    | Hypothesis Mining                    | 102       |

|          |                                     |            |
|----------|-------------------------------------|------------|
| <b>6</b> | <b>Creation and Enrichment</b>      | <b>105</b> |
| 6.1      | Human Collaboration                 | 105        |
| 6.2      | Text Sources                        | 106        |
| 6.2.1    | Pre-Processing                      | 106        |
| 6.2.2    | Named Entity Recognition (NER)      | 107        |
| 6.2.3    | Entity Linking (EL)                 | 107        |
| 6.2.4    | Relation Extraction (RE)            | 108        |
| 6.2.5    | Joint Tasks                         | 109        |
| 6.3      | Markup Sources                      | 109        |
| 6.3.1    | Wrapper-Based Extraction            | 110        |
| 6.3.2    | Web Table Extraction                | 111        |
| 6.3.3    | Deep Web Crawling                   | 111        |
| 6.4      | Structured Sources                  | 111        |
| 6.4.1    | Mapping from Tables                 | 112        |
| 6.4.2    | Mapping from Trees                  | 114        |
| 6.4.3    | Mapping from Other Knowledge Graphs | 114        |
| 6.5      | Schema/Ontology Creation            | 115        |
| 6.5.1    | Ontology Engineering                | 115        |
| 6.5.2    | Ontology Learning                   | 116        |

File Edit View History Bookmarks Tools Help

Wikidata

https://www.wikidata.org/wiki/Wikidata:Main\_Page

English Not logged in Talk Contributions Create account Log in

Main Page Discussion Read View source View history Search Wikidata

**WIKIDATA**

Main page  
Community portal  
Project chat  
Create a new item  
Recent changes  
Random item  
Query Service  
Nearby  
Help  
Donate

Lexicographical data  
Create a new Lexeme  
Recent changes  
Random Lexeme

collaborative

open

# Welcome to Wikidata

the free knowledge base with 100,391,724 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

free


multilingual

Want to help translate? Translate the missing messages.

Welcome!

Learn about data

FileEditViewHistoryBookmarksToolsHelp

 Wikidata Query Service


×

+


←→↻🏠

🔒 https://query.wikidata.org

☆🛡️⬇️S🦊☰


 Do you need help creating a query? You can build queries without having to write SPARQL in the new [Query Builder](#).


×

 Wikidata Query Service










📁 Examples

Query Builder

 Help ▾

 More tools ▾

🌐 English



1 #Most popular tonality

2 **SELECT** ?tonalityLabel (COUNT(?tonalityLabel) as ?count)

3 **WHERE**

4 {

5   ?work wdt:P826 ?tonality.

6   **SERVICE** wikibase:label { bd:serviceParam wikibase:language "en" }

7 }

8 **GROUP BY** ?tonalityLabel

9 **ORDER BY** DESC(?count)

| File Edit View History Bookmarks Tools Help   |       |
|---|-------|
| Wikidata Query Service  |       |
| https://query.wikidata.org/#%23Most popular tonality%0ASELECT %3FtonalityLabel (COU |       |
| 44 results in 664 ms  |       |
| Code Download Link  |       |
| tonalityLabel   | count |
| C major   | 571   |
| D major   | 487   |
| F major   | 462   |
| G major   | 439   |
| B-flat major  | 412   |
| E-flat major  | 379   |
| D minor   | 295   |
| A major   | 269   |
| C minor   | 269   |
| A minor   | 246   |