Mining Knowledge Graphs from Text

WSDM 2018

Jay Pujara and Sameer Singh
Introducing Presenters

Jay Pujara: Research Scientist at USC/ISI

Sameer Singh: Assistant Professor at UCI
Tutorial Overview

https://kgtutorial.github.io
Tutorial Overview

https://kgtutorial.github.io

Part 1: Knowledge Graphs
Tutorial Overview

https://kgtutorial.github.io

Part 1: Knowledge Graphs

Part 2:
Knowledge Extraction
Tutorial Overview

Part 1: Knowledge Graphs

Part 2: Knowledge Extraction

Part 3: Graph Construction

https://kgtutorial.github.io
Tutorial Overview

https://kgtutorial.github.io

Part 1: Knowledge Graphs

Part 2: Knowledge Extraction

Part 3: Graph Construction

Part 4: Critical Analysis
Tutorial Overview

Part 1: Knowledge Graphs

Part 2: Knowledge Extraction

Part 3: Graph Construction

Part 4: Critical Analysis

https://kgtutorial.github.io
Tutorial Overview

https://kgtutorial.github.io

Part 1: Knowledge Graphs

Part 2: Knowledge Extraction

Part 3: Graph Construction

Part 4: Critical Analysis
Tutorial Outline

1. Knowledge Graph Primer [Jay]
2. Knowledge Extraction Primer [Jay]
3. Knowledge Graph Construction
   a. Probabilistic Models [Jay]
   b. Embedding Techniques [Sameer]

Coffee Break

4. Critical Overview and Conclusion [Sameer]
What if I have a question?
Tutorial Overview

Part 1: Knowledge Graphs

Part 2: Knowledge Extraction

Part 3: Graph Construction

Part 4: Critical Analysis
Knowledge Graph Primer

TOPICS:

What is a Knowledge Graph?
Why are Knowledge Graphs Important?
Where do Knowledge Graphs come from?
Knowledge Representation Choices
Problem Overview
Knowledge Graph Primer

TOPICS:

**What is a Knowledge Graph?**

**Why are Knowledge Graphs Important?**

**Where do Knowledge Graphs come from?**

**Knowledge Representation Choices**

**Problem Overview**
What is a knowledge graph?
What is a knowledge graph?

• Knowledge in graph form!
What is a knowledge graph?

• Knowledge in graph form!

• Captures entities, attributes, and relationships
What is a knowledge graph?

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities
What is a knowledge graph?

• Knowledge in graph form!

• Captures entities, attributes, and relationships

• Nodes are entities

• Nodes are labeled with attributes (e.g., types)
What is a knowledge graph?

- Knowledge in graph form!

- Captures entities, attributes, and relationships

- Nodes are entities
- Nodes are labeled with attributes (e.g., types)
- Typed edges between two nodes capture a relationship between entities
Example knowledge graph

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities
- Nodes are labeled with attributes (e.g., types)
- Typed edges between two nodes capture a relationship between entities
Knowledge Graph Primer

TOPICS:

What is a Knowledge Graph?

Why are Knowledge Graphs Important?

Where do Knowledge Graphs come from?

Knowledge Representation Choices

Problem Overview
Why knowledge graphs?

- Humans:
  - Combat information overload
  - Explore via intuitive structure
  - Tool for supporting knowledge-driven tasks

- AIs:
  - Key ingredient for many AI tasks
  - Bridge from data to human semantics
  - Use decades of work on graph analysis
Applications 1: QA/Agents

What's the weather like at the Ritz Carlton hotel
Tap to Edit

It should be nice in Ritz Carlton hotel today... up to 71°F:

WEATHER

Marina del Rey
Sunny
Chance of Rain: 0%
High: 71°F Low: 53°F

4 PM 66
5 PM 64
6 PM 62

who is playing in this year's super bowl

Super Bowl LII

NFL - Today, 3:30 PM

Philadelphia Eagles
New England Patriots

Watch on: NBC

All times are in Pacific Time
Applications 2: Decision Support
Applications 3: Fueling Discovery

**beatles (musicartist)**

Literal strings: BEATLES, Beatles, beatles

**Help NELL Learn!**

NELL wants to know if these be
If they are or ever were, click thumbs-up. Or

- beatles is a musical artist
- beatles is a musician in the genre classic pop (musicgenre)
- beatles is a musician in the genre pop (musicgenre)
- beatles is a musician in the genre rock (musicgenre)
- beatles is a musician in the genre classic rock (musicgenre)
Knowledge Graphs & Industry

- Google Knowledge Graph
  - Google Knowledge Vault
- Amazon Product Graph
- Facebook Graph API
- IBM Watson
- Microsoft Satori
  - Project Hanover/Literome
- LinkedIn Knowledge Graph
- Yandex Object Answer
- Diffbot, GraphIQ, Maana, ParseHub, Reactor Labs, SpazioDati
Knowledge Graph Primer

TOPICS:

What is a Knowledge Graph?
Why are Knowledge Graphs Important?
Where do Knowledge Graphs come from?
Knowledge Representation Choices
Problem Overview
Where do knowledge graphs come from?
Where do knowledge graphs come from?

- Structured Text
  - Wikipedia Infoboxes, tables, databases, social nets
Where do knowledge graphs come from?

- **Structured Text**
  - Wikipedia Infoboxes, tables, databases, social nets

- **Unstructured Text**
  - WWW, news, social media, reference articles
Where do knowledge graphs come from?

• Structured Text
  ◦ Wikipedia Infoboxes, tables, databases, social nets

• Unstructured Text
  ◦ WWW, news, social media, reference articles

• Images
Where do knowledge graphs come from?

• Structured Text
  ◦ Wikipedia Infoboxes, tables, databases, social nets

• Unstructured Text
  ◦ WWW, news, social media, reference articles

• Images

• Video
  ◦ YouTube, video feeds
Knowledge Graph Primer

TOPICS:

What is a Knowledge Graph?
Why are Knowledge Graphs Important?
Where do Knowledge Graphs come from?
Knowledge Representation Choices
Problem Overview
Knowledge Representation

• Decades of research into knowledge representation

• Most knowledge graph implementations use RDF triples
  • <rdf:subject, rdf:predicate, rdf:object> : r(s,p,o)
  • Temporal scoping, reification, and skolemization...

• ABox (assertions) versus TBox (terminology)

• Common ontological primitives
  • rdfs:domain, rdfs:range, rdf:type, rdfs:subClassOf, rdfs:subPropertyOf, ...
  • owl:inverseOf, owl:TransitiveProperty, owl:FunctionalProperty, ...
Semantic Web

- Standards for defining and exchanging knowledge
  - RDF, RDFa, JSON-LD, schema.org
  - RDFS, OWL, SKOS, FOAF

- Annotated data provide critical resource for automation

- Major weakness: annotate everything?
Information Extraction from Text

• Focus of this tutorial!

• Answer to the knowledge acquisition bottleneck

• Many challenges:
  • chunking
  • polysemy/word sense disambiguation
  • entity coreference
  • relational extraction
Knowledge Graph Primer

TOPICS:

What is a Knowledge Graph?
Why are Knowledge Graphs Important?
Where do Knowledge Graphs come from?
Knowledge Representation Choices
Problem Overview
What is a knowledge graph?

• Knowledge in graph form!

• Captures entities, attributes, and relationships

• Nodes are entities
• Nodes are labeled with attributes (e.g., types)
• Typed edges between two nodes capture a relationship between entities
Basic problems
Basic problems

- **Who** are the entities (nodes) in the graph?
Basic problems

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
Basic problems

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
- **How** are they related (edges)?
Basic problems

- **Who** are the entities (nodes) in the graph?
- **What** are their attributes and types (labels)?
- **How** are they related (edges)?
Knowledge Graph Construction
Two perspectives

Knowledge Extraction

- **Who** are the entities (nodes) in the graph?
  - Named Entity Recognition
  - Entity Coreference

- **What** are their attributes and types (labels)?
  - Named Entity Recognition

- **How** are they related (edges)?
  - Relation Extraction
  - Semantic Role Labeling

Graph Construction

- **Who** are the entities (nodes) in the graph?
  - Entity Linking
  - Entity Resolution

- **What** are their attributes and types (labels)?
  - Collective Classification

- **How** are they related (edges)?
  - Link Prediction
Tutorial Outline

1. Knowledge Graph Primer [Jay]
2. Knowledge Extraction Primer [Jay]
3. Knowledge Graph Construction
   a. Probabilistic Models [Jay]
   b. Embedding Techniques [Sameer]
4. Critical Overview and Conclusion [Sameer]