Knowledge Graphs: A Practical Introduction across Disciplines

Mayank Kejriwal
University of Southern California
December, 2020
About me
Agenda

Knowledge Graphs: Definitions and Examples

Landscape of Important Findings

Knowledge Graph Construction

Summary and Research Trends

Cross-disciplinary Perspectives and Applications
Agenda

Knowledge Graphs: Definitions and Examples

Landscape of Important Findings

Cross-disciplinary Perspectives and Applications

Knowledge Graph Construction

Summary and Research Trends
What is a Knowledge Graph?

Set of triples, where each triple \((h, r, t)\) represents a relationship \(r\) between head entity \(h\) and tail entity \(t\)

(Barack Obama, wasBornOnDate, 1961-08-04),
(Barack Obama, hasGender, male),
...
(Hawaii, hasCapital, Honolulu),
...
(Michelle Obama, livesIn, United States)
What is a Knowledge Graph?

Technically, a multi-relational directed labeled graph with semantics

Both edges and nodes have labels, but not all labels are equal (literals vs. identifiers)

Where do the semantics come from?
• Complex question, only starting to be understood
More on semantics

Traditionally, semantics are believed to come from ontology
• An ontology is a ‘formal, explicit specification of a shared conceptualization’ (we will go deeper into this in a while)
• In philosophy, an ontology is a ‘study of what there is’ including the study of the ‘most general features of what there is, and how the things there are relate to each other in the metaphysically most general ways’

Source: https://plato.stanford.edu/entries/logic-ontology/

More recently, in AI, we have started to recognize a more commonsense view of semantics guided by findings in linguistics and distributional semantics
Agenda

Knowledge Graphs: Definitions and Examples

Landscape of Important Findings

Knowledge Graph Construction

Summary and Research Trends

Cross-disciplinary Perspectives and Applications
Agenda

- Knowledge Graph Construction
- Knowledge Graphs: Definitions and Examples
- Landscape of Important Findings
- Summary and Research Trends
- Cross-disciplinary Perspectives and Applications
A typical KGC workflow starts from corpus acquisition and ends with applications.
INFORMATION EXTRACTION (IE)
Named Entity Recognition (NER)

contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday’s 
PaperAdvertisementSupported ORG by F.B.I. Agent Peter Strzok PERSON, 
Who Criticized Trump PERSON in Texts, Is Fired Image Peter Strzok, a top F.B.I. GPE counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President Trump PERSON were uncovered, was fired. Credit T.J. Kirkpatrick PERSON for The New York Times By Adam Goldman ORG and Michael S. Schmidt Aug PERSON. 13 CARDINAL, 2018 WASHINGTON CARDINAL — Peter Strzok PERSON, the F.B.I. GPE senior counterintelligence agent who disparaged President Trump PERSON in inflammatory text messages and helped oversee the Hillary Clinton PERSON email and Russia GPE investigations, has been fired for violating bureau policies, Mr. Strzok PERSON’s lawyer said Monday DATE. Mr. Trump and his allies seized on the texts — exchanged during the 2016 DATE campaign with a former F.B.I. GPE lawyer, Lisa Page — in PERSON assailing the Russia GPE investigation as an illegitimate “witch hunt.” Mr. Strzok PERSON, who rose over 20 years DATE at the F.B.I. GPE to become one of its most experienced counterintelligence agents, was a key figure in the early months DATE of the inquiry. Along with writing the texts, Mr. Strzok PERSON was accused of sending a highly sensitive search warrant to his personal email account. The F.B.I. GPE had been under immense political pressure by Mr. Trump PERSON to dismiss Mr. Strzok PERSON, who was removed last summer DATE from the staff of the special counsel, Robert S. Mueller III PERSON. The president has repeatedly denounced Mr. Strzok PERSON in posts on

Demo: displaCy

https://explosion.ai/demos/displacy-ent
NER workflows

Many methods proposed over the previous 3-4 decades:

- Rule-based
- Dictionary-based
- Simple machine learning
- Sequence labeling (e.g., using conditional random fields or, before that, hidden Markov models)

Today, deep learning methods designed for sequences (such as RNNs and, more recently, transformers) are state-of-the-art

Much research still remains (especially for social media!)

Other kinds of IE: Relation Extraction

Source: Stanford TACRED
Other kinds of IE: Open Information Extraction

### Is IE a solved problem?

Table 2: Main Results on Testing Set: $F_1$ Score (Precision/Recall) (in %)

<table>
<thead>
<tr>
<th>Method</th>
<th>CoNLL03 (81.13/63.75)</th>
<th>Tweet (90.11/91.10)</th>
<th>OntoNote5.0 (84.59/87.88)</th>
<th>Webpage (50.07/54.76)</th>
<th>Wikigold (55.40/54.30)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Entity Types</strong></td>
<td>4</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>KB Matching</td>
<td>71.40</td>
<td>35.83(40.34/32.22)</td>
<td>59.51(63.86/55.71)</td>
<td>52.45(62.59/45.14)</td>
<td>47.76(47.90/47.63)</td>
</tr>
<tr>
<td><strong>Fully-Supervised</strong> (Our implementation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RoBERTa</td>
<td>90.11(89.14/91.10)</td>
<td>52.19(51.76/52.63)</td>
<td>86.20(84.59/87.88)</td>
<td>72.39(66.29/79.73)</td>
<td>86.43(85.33/87.56)</td>
</tr>
<tr>
<td>BiLSTM-CRF</td>
<td>91.21(91.35/91.06)</td>
<td>52.18(60.01/46.16)</td>
<td>86.17(85.99/86.36)</td>
<td>52.34(50.07/54.76)</td>
<td>54.90(55.40/54.30)</td>
</tr>
<tr>
<td><strong>Baseline</strong> (Our implementation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BiLSTM-CRF</td>
<td>59.50(75.50/49.10)</td>
<td>21.77(46.91/14.18)</td>
<td>66.41(68.44/64.50)</td>
<td>43.34(58.05/34.59)</td>
<td>42.92(47.55/39.11)</td>
</tr>
<tr>
<td>AutoNER</td>
<td>67.00(75.21/60.40)</td>
<td>26.10(43.26/18.69)</td>
<td>67.18(64.63/69.95)</td>
<td>51.39(48.82/54.23)</td>
<td>47.54(43.54/52.35)</td>
</tr>
<tr>
<td>LRNT</td>
<td>69.74(79.91/61.87)</td>
<td>23.84(46.94/15.98)</td>
<td>67.69(67.36/68.02)</td>
<td>47.74(46.70/48.83)</td>
<td>46.21(45.60/46.84)</td>
</tr>
<tr>
<td><strong>Other Baseline</strong> (Reported Results)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KALM$\dagger$</td>
<td>76.00 (84.11/68.61)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>ConNET$\circ$</td>
<td>75.57 (84.11/68.61)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Our BOND Framework</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>75.61(83.76/68.90)</td>
<td>46.61(53.11/41.52)</td>
<td>68.11(66.71/69.56)</td>
<td>59.11(60.14/58.11)</td>
<td>51.55(49.17/54.50)</td>
</tr>
<tr>
<td>BOND</td>
<td>81.48(82.05/80.92)</td>
<td>48.01(53.16/43.76)</td>
<td>68.35(67.14/69.61)</td>
<td>65.74(67.37/64.19)</td>
<td>60.07(53.44/68.58)</td>
</tr>
</tbody>
</table>

**Source:** Liang et al. BOND: BERT-Assisted Open-Domain Named Entity Recognition with Distant Supervision. KDD Conference. 2020.
"I had no idea I was getting in so deep," says Mr. Kaye, who founded Justin in 1982. Mr. Kaye had sold Capetronic Inc., a Taiwan electronics Maker, and retired, only to find he was bored. With Justin, he began selling toys and electronics made mostly in Hong Kong, beginning with Mickey Mouse radios. The company has grown -- to about 40 employees, from four initially, Mr. Kaye says. Justin has been profitable since 1986, adds the official, who shares his office...
(nw/wsj/2418)

Sources:
https://aryamccarthy.github.io/wiseman2016learning/
(Wiseman, Rush, and Shieber, 2016) at NAACL

Source:
A typical KGC workflow starts from corpus acquisition and ends with applications.

Corpus (usually documents, but also webpages, tables, reviews, social media...)

Corpus acquisition

Information extraction

Co-reference resolution

KG Construction

KG Completion

Entity Resolution

Knowledge Graph Embeddings

Applications

KG
KNOWLEDGE GRAPH COMPLETION
Entity Resolution

Algorithmically identifying and linking/grouping different manifestations of the same real-world object

Problem has existed for 50 years in many communities (databases, graphs, networks, tables...)

**Source:** Entity Resolution: Tutorial. Getoor and Machanavajjhala. VLDB, 2012
In the world of knowledge graphs
Representation Learning on Knowledge Graphs aka Knowledge Graph Embeddings

Knowledge graph embeddings:
- TransE, H…
- Neural tensor networks
- Graph convolutional networks (or their variants)
- Matrix factorization
- …
KGEs (results)

Useful resources:

- **OpenKE:**
  [http://139.129.163.161/index/toolkits#pretrained-embeddings](http://139.129.163.161/index/toolkits#pretrained-embeddings)

- **StarSpace:**
  [https://github.com/facebookresearch/StarSpace](https://github.com/facebookresearch/StarSpace)

- Recent transformer-based models could potentially be adapted, including BERT and RoBERTa:

<table>
<thead>
<tr>
<th>Method</th>
<th>WN18</th>
<th>FB15k</th>
<th>Filtered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>H@10</td>
<td>MRR</td>
</tr>
<tr>
<td>SE (Bordes et al., 2011)</td>
<td>1011</td>
<td>68.5</td>
<td>-</td>
</tr>
<tr>
<td>Unstructured (Bordes et al., 2012)</td>
<td>315</td>
<td>35.3</td>
<td>-</td>
</tr>
<tr>
<td>SME (Bordes et al., 2012)</td>
<td>545</td>
<td>65.1</td>
<td>-</td>
</tr>
<tr>
<td>TransH (Wang et al., 2014)</td>
<td>401</td>
<td>73.0</td>
<td>-</td>
</tr>
<tr>
<td>TransR (Lin et al., 2015b)</td>
<td>238</td>
<td>79.8</td>
<td>-</td>
</tr>
<tr>
<td>CTTransR (Lin et al., 2015b)</td>
<td>231</td>
<td>79.4</td>
<td>-</td>
</tr>
<tr>
<td>KG2E (He et al., 2015)</td>
<td>342</td>
<td>80.2</td>
<td>-</td>
</tr>
<tr>
<td>TransD (Ji et al., 2015)</td>
<td>224</td>
<td>79.6</td>
<td>-</td>
</tr>
<tr>
<td>ippTransD (Yoon et al., 2016)</td>
<td>283</td>
<td>80.5</td>
<td>-</td>
</tr>
<tr>
<td>TransSparse (Ji et al., 2016)</td>
<td>223</td>
<td>80.1</td>
<td>-</td>
</tr>
<tr>
<td>TATEC (García-Durán et al., 2016)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NTN (Socher et al., 2013)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DISTMULT (Yang et al., 2015)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ComplEx (Trouillon et al., 2016)</td>
<td>-</td>
<td>-</td>
<td>0.587</td>
</tr>
<tr>
<td>HolE (Nickel et al., 2016b)</td>
<td>-</td>
<td>-</td>
<td>0.616</td>
</tr>
<tr>
<td>RESCAL (Nickel et al., 2011) [*]</td>
<td>-</td>
<td>-</td>
<td>0.603</td>
</tr>
<tr>
<td>TransE (Bordes et al., 2013) [*]</td>
<td>-</td>
<td>-</td>
<td>0.351</td>
</tr>
<tr>
<td>STTransE (Nguyen et al., 2016b)</td>
<td>-</td>
<td>217</td>
<td>80.9</td>
</tr>
<tr>
<td>rTransE (García-Durán et al., 2015)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PTransE (Lin et al., 2015a)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GAKE (Feng et al., 2016b)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gaifman (Niepert, 2016)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hiri (Liu et al., 2016)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NLFeat (Toutanova and Chen, 2015)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TEKE.H (Wang and Li, 2016)</td>
<td>-</td>
<td>127</td>
<td>80.3</td>
</tr>
<tr>
<td>SSP (Xiao et al., 2017)</td>
<td>-</td>
<td>168</td>
<td>81.2</td>
</tr>
</tbody>
</table>
Other proposals: knowledge graph identification using probabilistic soft logic

Examples of ontological constraints

1. \( DOM(R, L) \neg REL(E_1, E_2, R) \Rightarrow LBL(E_1, L) \)
2. \( RNG(R, L) \neg REL(E_1, E_2, R) \Rightarrow LBL(E_2, L) \)
3. \( INV(R, S) \neg REL(E_1, E_2, R) \Rightarrow REL(E_2, E_1, S) \)
4. \( SUB(L, P) \neg LBL(E, L) \Rightarrow LBL(E, P) \)
5. \( RSUB(R, S) \neg REL(E_1, E_2, R) \Rightarrow REL(E_1, E_2, S) \)
6. \( MUT(L_1, L_2) \neg LBL(E, L_1) \Rightarrow \neg LBL(E, L_2) \)
7. \( RMUT(R, S) \neg REL(E_1, E_2, R) \Rightarrow \neg REL(E_1, E_2, S) \)

Solution: Knowledge Graph Identification (KGI)

- Performs graph identification:
  - entity resolution
  - node labeling
  - link prediction
- Enforces ontological constraints
- Incorporates multiple uncertain sources
A typical KGC workflow starts from corpus acquisition and ends with applications.

Corpus (usually documents, but also webpages, tables, reviews, social media...)

KG Construction
- Information extraction
- Co-reference resolution
- ...

KG Completion
- Entity Resolution
- Knowledge Graph Embeddings
- ...

KG

Applications
Open-source KGs that have been built
Agenda

Knowledge Graphs: Definitions and Examples

Landscape of Important Findings

Knowledge Graph Construction

Summary and Research Trends

Cross-disciplinary Perspectives and Applications
CROSS-DISCIPLINARY PERSPECTIVES: WEB AND INFORMATION RETRIEVAL
Google Knowledge Graph
Domain-specific search (DSS)

The Massive YouTube Ecosystem

- 56 Countries
- 2nd Largest
- 60 Years

- One Billion
- 100 Hours

- 17% Mobile
- Hundreds of Millions

- 25% Mobile
- 1 Billion

BloomReach surveyed 2,000 CONSUMERS

Where do consumers start product searches?

- 44% go directly to Amazon first

- 21% start at a specific retailer

- 34% use top search engines like Google, Bing and Yahoo!

source: https://photos.prnewswire.com/prnfull/20151006/274273-INFO
Emerging opportunities for DSS

Fighting human trafficking

Predicting cyberattacks

Stopping Penny Stock Fraud

Accurate geopolitical forecasting

Emerging opportunities for DSS

Fighting human trafficking

Predicting cyberattacks

Stopping Penny Stock Fraud

Accurate geopolitical forecasting
DARPA/IARPA programs

DARPA Memex

IARPA Hybrid Forecasting Competition

DARPA AIDA

DARPA Causal Exploration

DARPA LORELEI

IARPA CAUSE
Research Question

General Search  Google Knowledge Graph

DSS  Domain-Specific Knowledge Graphs

How do we construct domain specific knowledge graphs over web data for powerful DSS applications?
Knowledge Graphs for DSS

Relevance Model
http://site1/...
http://site2/...
estort, incall, ...

Schema
eye color, hair color, ethnicity, content, vendor, phone

Corpus
url-1: HTML-1
url-2: HTML-2
url-3: HTML-3

Knowledge Graph Construction

Knowledge Graph

GUI

Investigative Search Engine

SPARQL
ranked entities, attributes, aggregations

forest of SBT queries
scored entities
Domain-specific Insight Graphs
Many examples in industry and non-profit
Commercial domains: Amazon Product Graph

Mission: To answer any question about products and related knowledge in the world

Source: Dong, Luna. Building a Broad Knowledge Graph for Products. Keynote at ICDE. 2019
Another example: COVID-19

Fighting COVID-19 with Knowledge Graphs

National Science Foundation awards funding for a semantic integration platform

"The project will be based on our knowledge graph prototype linking information about pathogens, health data, and environmental indicators and enabling cross-domain inferencing," said Peter Rose, director of SDSC’s Structural Bioinformatics Laboratory and principal investigator (PI) for the project, called ‘COVID-19-Net: Integrating Health, Pathogen and Environmental Data into a Knowledge Graph for Case Tracking, Analysis, and Forecasting. “Such a graph lets researchers trace the spread of the coronavirus in different geographic conditions, focusing on specific virus strains and transmissions.”
Other COVID-19 KG examples

Source: CovidGraph


Source: Verizon Media
https://github.com/yahoo/covid-19-dashboard
CROSS-DISCIPLINARY PERSPECTIVES: SEMANTIC WEB
What is (or even isn’t) a domain?

Some dictionary definitions

(Merriam Webster) A sphere of knowledge, influence or activity
(Oxford) A specified sphere of activity or knowledge

Specifying the sphere

Rules
Scope (e.g., the legal system)
Syllabi (for classrooms)
Examples

How do domain experts specify the sphere?

Examples
Ontology
Modeling domains: Ontologies

What is an ontology?

Formal, explicit specification of a shared conceptualization

- Machine readable
  - Concepts, properties, functions, axioms are explicitly defined

- Consensual knowledge
  - Abstract model of some phenomena in the world

Examples of ontologies

Agency domain

Friend-of-a-friend
Ontologies are big in Science
An RDF graph is a set of triples, where each triple is of the form (subject, predicate, object):

- Subjects must be URIs (technically, *internationalized resource identifiers*, in practice, just Uniform Resource Locators)
- Predicates (also called ‘properties’) must be URIs
- Objects can be either URIs of literals (strings, numbers, dates…)

In the Semantic Web, RDF is the ‘building block’ of higher order vocabularies (such as RDF Schema and OWL) that can be used to represent ontologies
Example of RDF KG

As a graph

```
http://xmlns.com/foaf/0.1/Person
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://www.example.org/~joe/contact.rdf#joesmith
http://xmlns.com/foaf/0.1/homepage
http://xmlns.com/foaf/0.1/mbox
http://xmlns.com/foaf/0.1/givenname
mailto:joe.smith@example.org
```

As a set of triples

```
http://www.example.org/~joe/contact.rdf#joesmith
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://xmlns.com/foaf/0.1/Person

http://www.example.org/~joe/contact.rdf#joesmith
givenname "Joe"
```
Web Ontology Language (OWL)

OWL builds on RDF (and another layer called RDF Schema or RDFS) to provide a systematic vocabulary for defining ontologies.

Because OWL builds on RDF, every OWL ontology is also an RDF graph, but not necessarily vice-versa.

RDF Schema Features:
- `class` (`Thing`, `Nothing`)
- `rdfs:subClassOf`
- `rdfs:Property`
- `rdfs:subPropertyOf`
- `rdfs:domain`
- `rdfs:range`
- `rdfs:individual`

Property Restrictions:
- `Restriction`
- `onProperty`
- `allValuesFrom`
- `someValuesFrom`

Class Intersection:
- `intersectionOf`

Datatypes
- `xsd:datatypes`

(In)Equality:
- `equivalentClass`
- `equivalentProperty`
- `sameAs`
- `differentFrom`
- `AllDifferent`
- `distinctMembers`

Restricted Cardinality:
- `minCardinality` (only 0 or 1)
- `maxCardinality` (only 0 or 1)
- `cardinality` (only 0 or 1)

Versioning:
- `versionInfo`
- `priorVersion`
- `backwardCompatibleWith`
- `incompatibleWith`
- `DeprecationClass`
- `DeprecationProperty`

ObjectProperty
- `DatatypeProperty`
- `inverseOf`
- `TransitiveProperty`
- `SymmetricProperty`
- `FunctionalProperty`
- `InverseFunctionalProperty`

Header Information:
- `Ontology`
- `imports`

Annotation Properties:
- `rdfs:label`
- `rdfs:comment`
- `rdfs:seeAlso`
- `rdfs:isDefinedBy`
- `AnnotationProperty`
- `OntologyProperty`

https://www.w3.org/TR/owl-features/
Reasoning over knowledge graphs

<table>
<thead>
<tr>
<th>Assertions</th>
<th>DL-axioms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every animal which can fly has wings</td>
<td>$\phi_1: \text{Fly} \sqsubseteq \text{HasWing}$</td>
</tr>
<tr>
<td>Every animal which eats fish is a piscivore</td>
<td>$\phi_2: \exists \text{Eat.Fish} \sqsubseteq \text{Piscivore}$</td>
</tr>
<tr>
<td>tweety is not an abnormal bird or cannot fly</td>
<td>$\phi_3: (\neg \text{AbnBird} \sqcup \neg \text{Fly})(\text{tweety})$</td>
</tr>
<tr>
<td>ursidae eats salmon</td>
<td>$\phi_4: \text{Eat}(\text{ursidae}, \text{salmon})$</td>
</tr>
<tr>
<td>salmon is some fish</td>
<td>$\phi_5: \text{Fish}(\text{salmon})$</td>
</tr>
<tr>
<td>ursidae is not a piscivore</td>
<td>$\phi_6: \neg \text{Piscivore}(\text{ursidae})$</td>
</tr>
</tbody>
</table>

Source: Bergman. Platforms and Knowledge Management. 2018
Example tool for reasoning and ontologies: Protege

Source: https://protege.stanford.edu/
Putting it all together: Semantic Web Layer Cake
CROSS-DISCIPLINARY PERSPECTIVES: KNOWLEDGE DISCOVERY & DATA MINING
Knowledge Graph in Personal Assistant

Source: Dong, Luna. Building a Broad Knowledge Graph for Products. Keynote at ICDE. 2019
**Others**

**Scientific Text Mining**

**Question Answering**

**Recommendation Systems**

**Summarization**

**Truth/fact-checking**
Agenda

Knowledge Graphs: Definitions and Examples

Landscape of Important Findings

Knowledge Graph Construction

Summary and Research Trends

Cross-disciplinary Perspectives and Applications
Open Knowledge Network (OKN)

Technology companies develop proprietary knowledge networks as key business technologies today. However, because these networks are proprietary and expensive to construct, government, academia, small businesses, and nonprofits do not have access to them. In contrast, an open knowledge network (OKN) would be available to all stakeholders, including the researchers who will help push this technology further. An OKN requires a nonproprietary, public–private development effort that spans the entire data science community and will result in an open, shared infrastructure.


Knowledge Graphs for Social Good (KGSG)

Best Practices, Methods, and Challenges - Held May 4th, 2020 at KGC 2020

Applying knowledge graphs for social good

JUNE 26, 2020

Knowledge Graphs for Social Good: An Entity-centric Search Engine for the Human Trafficking Domain

Publisher: IEEE  Cite This  PDF

Mayank Kejriwal; Pedro Szekely  All Authors

Information Sciences Institute
Knowledge, semantics and context: what are they and how do we better define/represent them?
Explainable AI

On The Role of Knowledge Graphs in Explainable AI

Freddy Lecue, a,b
  a CoreAI, Thales, Montreal, Canada
  b WIMMICS, Inria, Sophia Antipolis, France
  E-mail: freddy.lecue@inria.fr

https://towardsdatascience.com/knowledge-graphs-for-explainable-ai-dcd73c5c016
WRAPUP
What is a Knowledge Graph?

• Set of triples, where each triple \((h, r, t)\) represents a relationship \(r\) between head entity \(h\) and tail entity \(t\)

(Barack Obama, wasBornOnDate, 1961-08-04),
(Barack Obama, hasGender, male),
...
(Hawaii, hasCapital, Honolulu),
...
(Michelle Obama, livesIn, United States)
What is a Knowledge Graph?

Technically, a multi-relational directed labeled graph with semantics.

Both edges and nodes have labels, but not all labels are equal (literals vs. identifiers).

Where do the semantics come from?
- Complex question, only starting to be understood.
More on semantics

Traditionally, semantics are believed to come from ontology
• An ontology is a ‘formal, explicit specification of a shared conceptualization’ (we will go deeper into this in a while)
• In philosophy, an ontology is a ‘study of what there is’ including the study of the ‘most general features of what there is, and how the things there are relate to each other in the metaphysically most general ways’

Source: https://plato.stanford.edu/entries/logic-ontology/

More recently, in AI, we have started to recognize a more commonsense view of semantics guided by findings in linguistics and distributional semantics
A typical KGC workflow starts from corpus acquisition and ends with applications.
Named Entity Recognition (NER)

contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday’s PaperAdvertisementSupported ORG by F.B.I. Agent Peter Strzok PERSON, Who Criticized Trump PERSON in Texts, Is Fired Image Peter Strzok, a top F.B.I. GPE counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President Trump PERSON were uncovered, was fired. Credit T.J. Kirkpatrick PERSON for The New York Times By Adam Goldman ORG and Michael S. Schmidt Aug PERSON. 13 CARDINAL, 2018 WASHINGTON CARDINAL — Peter Strzok PERSON, the F.B.I. GPE senior counterintelligence agent who disparaged President Trump PERSON in inflammatory text messages and helped oversee the Hillary Clinton PERSON email and Russia GPE investigations, has been fired for violating bureau policies, Mr. Strzok PERSON’s lawyer said Monday DATE. Mr. Trump and his allies seized on the texts — exchanged during the 2016 DATE campaign with a former F.B.I. GPE lawyer, Lisa Page — in PERSON assailing the Russia GPE investigation as an illegitimate “witch hunt.” Mr. Strzok PERSON, who rose over 20 years DATE at the F.B.I. GPE to become one of its most experienced counterintelligence agents, was a key figure in the early months DATE of the inquiry. Along with writing the texts, Mr. Strzok PERSON was accused of sending a highly sensitive search warrant to his personal email account. The F.B.I. GPE had been under immense political pressure by Mr. Trump PERSON to dismiss Mr. Strzok PERSON, who was removed last summer DATE from the staff of the special counsel, Robert S. Mueller III PERSON. The president has repeatedly denounced Mr. Strzok PERSON in posts on

Other kinds of IE: Relation Extraction

Source: Stanford TACRED
In the world of knowledge graphs
KGEs (results)

Useful resources:

- OpenKE: http://139.129.163.161/index/toolkits#pretrained-embeddings
- StarSpace: https://github.com/facebookresearch/StarSpace
- Recent transformer-based models could potentially be adapted, including BERT and RoBERTa: https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/
Open-source KGs that have been built
Many applications and open research areas!

Information retrieval

Semantic Web

Recommender systems

Knowledge discovery/data mining

?
Numerous surveys, some more technical/field-specific


Upcoming: