

Notes from the CS520: Knowledge Graph Seminar (Spring 2020)

Knowledge 520 course Graphs: How should AI explicitly represent knowledge?

Gary Berg-Cross

Focus and Motivation: What is a knowledge graph (KG)?

Motivation – KGs now widely used. Used to store info extracted by algorithms.

Motivation for the Seminar

- Knowledge Graphs are being used in
 - Web search
 - Answering questions
 - Data integration
 - Knowledge Graphs are also target of output for
 - NLP and computer vision algorithms
 - ML algorithms more generally
 - Knowledge Graphs are a topic of a major program from N
- <https://www.nsf.gov/od/oia/convergence-accelerator/Award%20listings/track-a-ia>

Outline:

CS520: Knowledge Graph Seminar Session 1 (Spring 2020)

Press **Esc** to exit full screen

Seminar Outline

Knowledge Graph

- What is it?
- How do create it?
- How do we reason with it?
- How do we use it with modern AI algorithms?
- Where is the research?

CS520: Knowledge Graph Seminar Session 1 **What is a knowledge graph?**

Denny Vrandečić Knowledge Graphs have found wide usage in many applications in various industries, for diverse research tasks, and by increasingly also by hobbyists and

student developers. In this talk we will informally introduce the ideas behind knowledge graphs, show use cases and applications, and how they have proven to be useful

We build a KG because: It is easy for knowledge to be hidden, get lost in natural language, if you have lots of text.

Need to data information that is actionable.

Intro to Wikidata which supports Wikipedia.

Shows connectivity of the graph say for cities of the world.

8 billion nodes, connected using RDF standard.

LoC also has a large KG using RDF.

More than 4800 catalogs connected. 12 million publishers of RDF. Using schema.org

So what can we do with this? Can as to give every node X connected by relation Y

Graph patterns.

With a simple query you get the geo-coordinates of every x connected by y.

Note, what we know of the world is inconsistent, so Wikidata allows this which amounts to alternate models of reality.

Summary and Outlook

- Knowledge Graphs are composable
- Allow for inferencing
- Trillions of edges readily available
- Other data sources can be lifted to graphs
- Knowledge Graphs are easily editable
- Graph patterns as a query language
- KG provide explainability
- Standards for graphs and graph patterns



Jans Aasman CEO Allegro-graph

We will discuss a Knowledge Graph technology approach that encapsulates a novel Entity-Event Model, natively integrated with domain ontologies and metadata, and dynamic ways of setting the analytics focus on all entities in the system (patient, person, devices, transactions, events, operations, etc.) as prime objects that can be the focus of an analytic (AI, ML, DL) process.

We will demonstrate core elements of a knowledge graphs --- its knowledge representation and its core principles, example of one or two knowledge graphs that show how modern knowledge graphs are a mix of graph database technology and NLP and ML techniques.

Knowledge Graph Seminar Session 1 (Spring 2020)

What is in a modern knowledge graph?

- A semantic graph database, scalable, secure, acid
- **Ontologies and taxonomies**
- Reasoning and rule based processing
- Smart integration of silos of information
- Machine Learning and Advanced Analytics
- Natural Language Processing and Text classification

And recently: more and more:

- Speech recognition
- **Chatbots.**

35:28 / 1:55:37 Scroll for details FRANZINC.

Example of a Chomsky KG demonstrates how to build a KG from text on Chomsky's

publications.


Has Gruff graph DB

Large HC an events based KB.

Knowledge Graph Seminar Session 1 (Spring 2020)

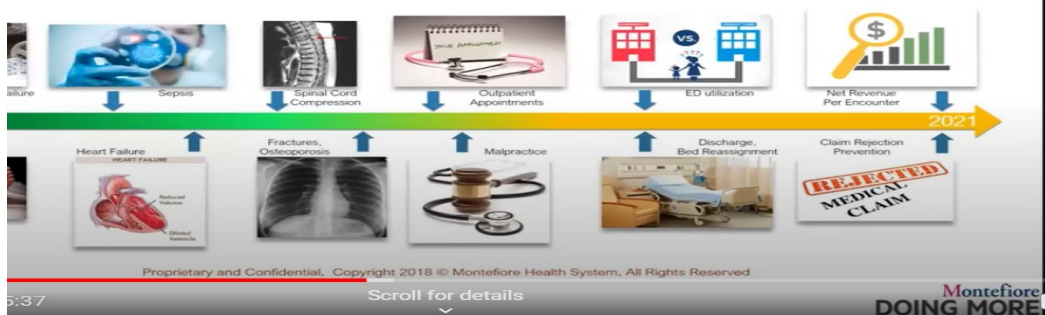
First production use: use a predictive model for respiratory failure on top our Knowledge Graph

- A random forest model detects respiratory failure up to 48 hours before the event. Faster than doctors and nurses
- The model uses 46 complex variables, a doctor in general looks at 3 to 4 variables.
 - A mix of lab results, (real time) measurements, diagnostics, other
 - Complex: give me largest difference between the value for SerumCalciumLevel and the midpoint 9.5 in the range of 0 to 20 for the last 24 hours.
- Greatly reduce unnecessary intubation with same clinical outcome
- Saves many millions per year



h Seminar Session 1 (Spring 2020) Knowledge Graph Roadmap in Montefiore

ed using grant funding (NHLBI, PCORI, and ICTR) and Intel/Franz collaboration
egrated with Epic, and hosted by MIT Data Center.
January 2017 with Respiratory Failure and Mortality Prediction → Prevention
sis, HF, Spinal Cord Compression (etc) all with **associated ROI**
and multi-disciplinary spectrum of applications confirms a **platform** approach



Turn everything
into an event.
They have the
same data shape
pattern.

Events directly connect to Entity

Press **Esc** to exit full screen

The KG provides radical simplification of the EDW schema:

we turn everything into an EVENT

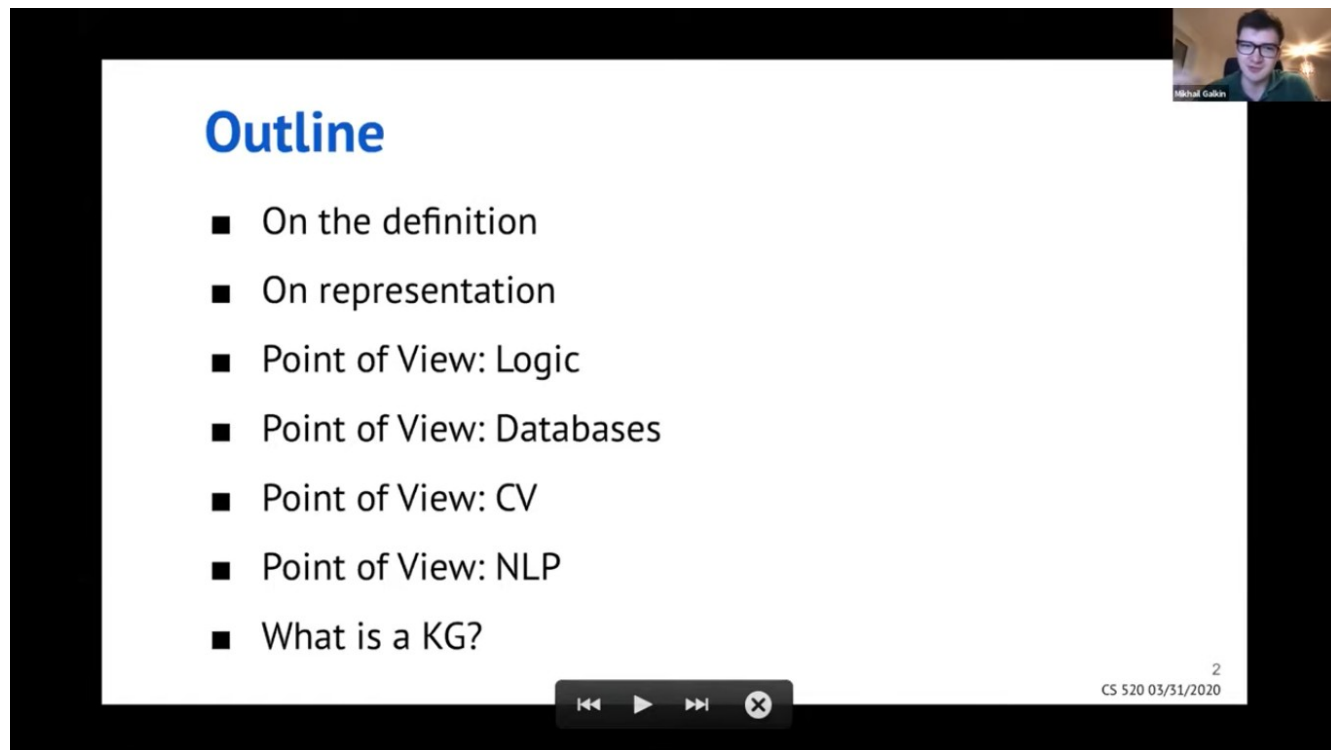
- Healthcare: everything that can happen to a patient is a time based (sub) event: Check In, Check Out, Test, Diagnosis, Procedure, Medication administration, Medication order, Sensor reading for vital signs, Invoice, Bill payment, Non-bill payment, all insurance interactions.
- Telecom: everything that happens with a telco user is a time based (sub) event: telephone call, sms, whatsapp, web site visit, location record, crm call, bill pay, non-bill pay ...
- Even your demographic features are events (names, address, etc)
- **From thousands of tables to one event table (well, graph)**

[Mikhail Galkin](#) (Dresden) Outline: In this talk we will discuss various perspectives and points of view on understanding knowledge graphs as seen from the recent ML and AI conferences [1-4]. Finally, we will identify common attributes that pertain to knowledge graphs.

The slides for the presentation are available [here](#).

1. [Knowledge Graphs at AAAI 2020](#)
2. [Machine Learning on Graphs at NeurIPS 2019](#)
3. [Knowledge Graphs at EMNLP 2019](#)
4. [Knowledge Graphs at ACL 2019](#)

What do we need to reach “a shared understanding of KG definition” Some are mathematical all are correct.



The screenshot shows a video player interface. In the top right corner, there is a small video feed of the speaker, Mikhail Galkin. The main area of the slide is white with a black border. The title 'Outline' is in blue. Below it is a list of topics, each preceded by a black square bullet point. At the bottom of the slide, there is a black bar containing a video control bar with play, pause, and close buttons, and a footer with the text 'CS 520 03/31/2020' and a small number '2'.

Outline

- On the definition
- On representation
- Point of View: Logic
- Point of View: Databases
- Point of View: CV
- Point of View: NLP
- What is a KG?

CS 520 03/31/2020 2

World models some with shared vocabularies.

How do we encode these? Symbolic (tuples) or vector (embedding in high dim space).

CS520: Knowledge Graph Seminar Session 1 (Spring 2020)

World models?

Entities and relations define our **domain of discourse**

How to encode it?

Source: <https://lod-cloud.net/>

CS 520 03/31/2020

Might be open or closed world.

On representation of Knowledge Graphs

Symbolic

Logic

DB & DI

s, p, o
 $p(s, o)$
 (s, p, o)

Open-world assumption

Closed-world assumption

Temporal / evolving

Vector

CV

NLP

$s, p, o \in \mathbb{R}^d$

7

There are different points of view like logic using logic programming. Consistent set of facts.

Or the logic of DL

Ontology is at the T box level while KG is at the A-Box level.

Dbs get integrated into a KB, gives them power.

More AI (CHH + RPNPOV: Scene graphs build related objects identified in an image or series of images.

Kgs can be built from text.

Named entity recognition (apple as a surface form label has many meanings)

Another task in KB building is relation linking.

All of these are used in NL question answering

Language models are used to predict a word in a sentence – often trained on large textbook good benchmark results

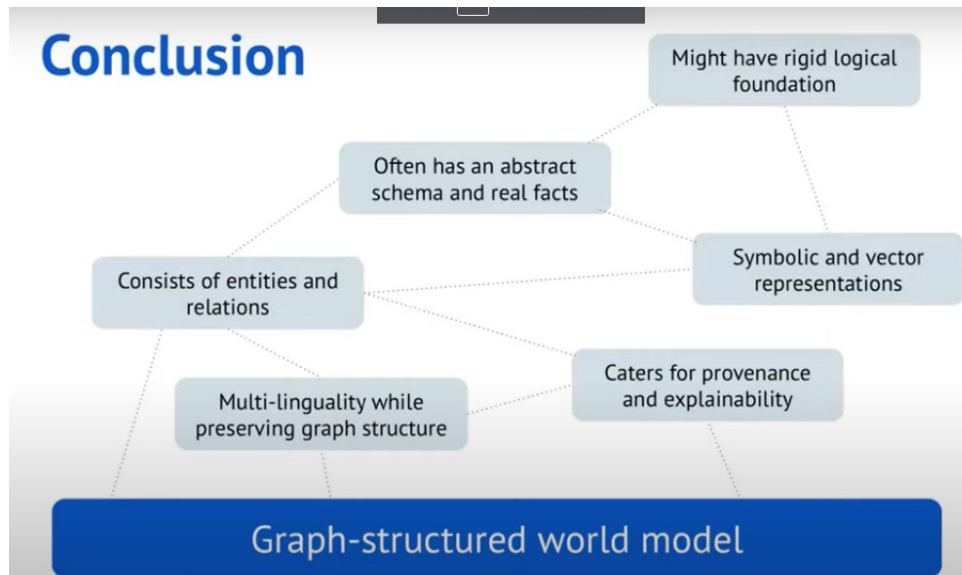
Qs

are ontologies and Kgs the same? How related?

Yes technically but not practically. You can get by w/o a real ontology like Chomsky, but for a banking application you need a strict ontology for restrictions on what you include in the KG. It forces you to be consistent.

A little top-level ontology helps find errors in a KG.

Ontology defines the set of labels used in a KG.



Do we use the ACID properties for Kgs? Not unless we are using the KG for transactions.

Limitations – rdf has limited expressiveness compared to natural language.

Time series are not well expressed in KGs.

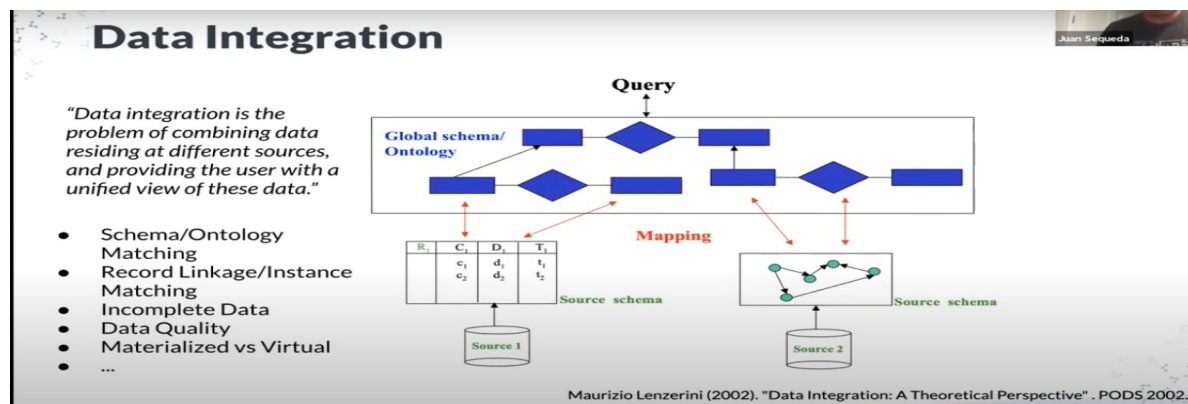
CS 520: Knowledge Graph Seminar Session 2 **How to create a knowledge graph?**

Juan Sequeda (Data.world) Data Integration has been an active area of computer science research for over two decades. A modern manifestations is as Knowledge Graphs which integrates not just data but also knowledge at scale. Tasks such as schema and ontology matching are fundamental in the data integration process. Research focus has been on studying this phenomena from a technical point of view (algorithms and systems) with the ultimate goal of automating this task.

In the process of applying scientific results to real world enterprise data integration scenarios to design and build Knowledge Graphs from enterprise databases, we have experienced numerous obstacles. In this talk, I will share insights about these obstacles. I will argue that we need to think outside of a technical box and further study the phenomena of data integration with a human-centric lens: from a socio-technical point of view.

Enterprise scenario are RDBs from enterprises to support business questions.

Integration issue.



Map data to the ontology schema. Known problems such as data quality.

Too many tables with impossible labels. Hard to join the tables since data was modeled for application and not for integration.

Large scale mapping is hard in practice, since there are many names and what do they mean?

Kbs may enable us to bridge a data meaning gap. There is a socio-tech solution with people, methods and tools. It more than linked data it is a linking of knowledge and data. Need a Knowledge Scientist role. (formerly KE)

Tools include GRA.FO.

Take away *Technology Fallacy* Kane et al.

[Chris Ré](#) Theory and systems for weak supervision If you want to build a high-quality machine learning product, build a large, high-quality training set. At first glance, this seems as useful as the statement “if you want to be rich, get a lot of money.” However, a key idea driving our work is that new theoretical and systems concepts including weak supervision, automatic data augmentation policies, and more, can enable engineers to build training sets more quickly and cost effectively.

Along with state-of-the-art results on benchmarks, these concepts have allowed our group and collaborators to build a range of state-of-the-art applications including patient-care monitoring on electronic health records, automatic triage systems for radiologists, and enabling cardiologists to spot rare abnormalities in video MRI—along with widely used products from Apple and Google. This talk describes the theoretical and systems challenges that such applications create.

Much of this work is open source and available at <http://snorkel.org> or my website.

ML Applications are Models plus (training) data plus hardware

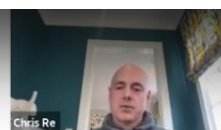
Not enough data. Usually it comes from a messy, noisy process.

Can we build a mathematical approach to clean this up?

Supervised ML is where the action is. Example of reading lung X-ray images.

Needed to create a large image DB. Takeaway prglem with image Dbs and training bases are critical (dense-net etc.)

Q on Open sources tools for inference – pellet is one.



Is Deep Learning the Answer?

This is not an easy question...

- No benchmark dataset
- Effects of data quality are unclear
- No assessment of existing algorithms
- No feedback from clinical community

...so we spent a year trying to answer it!

- Created large dataset of clinical labels
- Evaluated effect of label quality
- Work published in a *clinical journal*

Model	Test Accuracy
BOVW + KSVM	0.88
AlexNet	0.87
ResNet-18	0.89
DenseNet-121	0.91

Often: Differences in models ~ 2-3 points.

Later: Label quality & quantity > model choice.

41:06 / 1:53:27 Scroll for details

Forefront need data augmentation of accuracy. Quality of labels more important than exact models

Created datasets is slow, expensive and static which is a problematic.

So they have looked at faster dynamic ML approaches that are also cheap.

But the labels will be lower quality due to simple labeling functions.

Image classifier. They blend KG data and ML processes like looking for triples in graphs.

Snorkel uses labeling functions for extraction from text reports and they try to use deep learning model the noisy data.

DL discovers things that your function ignored. It matches results obtained over years.

Example of a use – Google's Snorkel Drybell,

Shows how SW is changing. Data problem becomes coding. You have supervision stack to build code.

Another use: Overton

CS 520: Knowledge Graph Seminar Session 2 (Spring 2020)

Snorkel: Formalizing Programmatic Labeling

Goal: Replace *ad hoc* weak supervision with a formal, unified, theoretically grounded approach for programmatic labeling

46:53 / 1:53:27

CS 520: Knowledge Graph Seminar Session 2 (Spring 2020)


Press **Esc** to exit full screen

You probably have *used it...*

Overton: A Data System for Monitoring and Improving Machine-Learned Products

Christopher Ré Apple	Feng Niu Apple	Pallavi Gudipati Apple	Charles Srisuwananukorn Apple
-------------------------	-------------------	---------------------------	----------------------------------

Migrating a Privacy-Safe Information Extraction System to a Software 2.0 Design

	Ying Sheng Google Mountain View, CA, USA yingsheng@google.com	Nguyen Vo Google Mountain View, CA, USA nguyenvo@google.com	James B. Wendt Google Mountain View, CA, USA jwendt@google.com
	Sandeep Tata Google Mountain View, CA, USA tata@google.com	Marc Najork Google Mountain View, CA, USA najork@google.com	

54:29 / 1:53:27

Jason Fries' recent work noted.

[Sharon Li's blob post](#) on text augmentation



Training Signal is key to pushing SotA

New methods for gathering signal leading the state of the art

Google AI AutoAugment: Using learned **data augmentation policies**

- **Augmentation Policies** first in Ratner et al. NIPS '17



Henry
Ehrenberg (to: Washington)

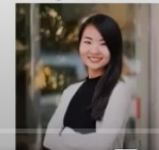


Alex Ratner
(to: Washington)



Facebook Hash tag weakly supervised pre-training

- Pre-train using a massive dataset with *hashtags*



Sharon Y. Li (to: Wisconsin)



42:08 / 1:53:27



Scroll for details





Related Work in Weak Supervision

- **Distant Supervision:** Mintz et. al. 2009, Alfonseca et. al. 2012, Takamatsu et. al. 2012, Roth & Klakow 2013, Augenstein et. al. 2015, etc.
- **Crowdsourcing:** Dawid & Skene 1979, Karger et. al. 2011, Dalvi et. al. 2013, Ruvolo et. al. 2013, Zhang et. al. 2014, Berend & Kontorovich 2014, etc.
- **Co-Training:** Blum & Mitchell 1998
- **Noisy Learning:** Bootkrajang et. al. 2012, Mnih & Hinton 2012, Xiao et. al. 2015, etc.
- **Indirect Supervision:** Clarke et. al. 2010, Guu et. Al. et. al. 2017, etc.
- **Feature and Class-distribution Supervision:** Zaidan & Eisner 2008, Druck et. al. 2009, Liang et. al. 2009, Mann & McCallum 2010, etc.
- **Boosting & Ensembling:** Schapire & Freund, Platanios et. al. 2016, etc.
- **Constraint-Based Supervision:** Bilenko et. al. 2004, Koestinger et. al. 2012, Stewart & Ermon 2017, etc.



Propensity SVMs: Joachims 17

Scroll for details



Programming Stack Supervision Stack

High-level



Low-level

Application Interfaces

Declarative Language

High-Level Language

Assembly Language

Machine Language

LFs Auto-Generated from User Behavior

LFs Compiled from Natural Language

LFs Built on Advanced Primitives

LFs Coded Directly

Individual Labels

ICLR 19.

ACL 18

NeurIPS17&18, VLDB19

NuerIPS16, VLDB18

Manual



1:01:32 / 1:53:27

Scroll for details



See Snorkel.org

Xiao Ling K based construction at Apple Siri one of the goals Siri aims at is to be a *know-it-all* question answering system, capable of answering questions from hundreds of millions of users about nearly anything. The question answering system is backed by a knowledge graph that was automatically constructed from a vast number of data sources including natural language text, HTML tables, and many others.

In this talk, we will give a brief overview of automatic knowledge base construction (AKBC) at Siri and discuss two concrete problems, Wikipedia infobox extraction and entity resolution. Wikipedia infobox extraction has been somewhat considered solved in the literature. We will show why this is still a challenging problem in the AKBC context and present our work published in NAACL 2019. We will also discuss in-progress work in tackling entity resolution by embedding entities to vector spaces for matching and promising results in our preliminary experiments.

Want to answer open queries with Siri. But how is the knowledge graph to do this built?

4 sources turning into canonical triples – text, semi-structured (DBpedia infobox info), structured data and curated. These 4 triples are then fused and inference is run to add new concluded knowledge.

Problems – DBpedia has know problems like wrong anchor links.

Built Robust InfoBox Extractor (RIBE).

RIBE components – 1 Relation extraction (M R N) is an important strategy. Ex. Membership relation for Beatles.

2. Is entity linking (uses classifier).

Entity Resolution, a key problem. Need resolution of entities in many data sets around. By hand can take 6 months.

So work on identifying equivalence of 2 entities, maybe in 2 sources combined in the KG. Names are often not reliable.

Approach embed all entities in real-value vectors and then match entities in that space. Use a NN to function approximate the hand coded heuristics used to identify entities.

Entity Linking is the other key activity.

Questions

Q on open Qs...There are many. Ontology alignment is one. Needed for messy data. Is there an automated way to do this.

Still and open issue on the best way to do data augmentation.

Open Q on how to automate parts of the current methodologies to build Kgs.

Knowledge Graph Seminar Session 3

What are some advanced knowledge graphs?

Mike Tung Diffbot KB trying to be a comprehensive map of human knowledge-based. Computer scientists have long dreamed of building a "semantic web" - a machine-readable network of knowledge that can be used by intelligent systems to perform useful services. However, current approaches to building knowledge graphs that involve a human element are either narrow and vertical applications, or broad but shallow, limiting the knowledge to the popular entities. We argue that the only way to build applications powered by real-world knowledge is with fully-automated approaches that are able to acquire knowledge on demand.

Mike Tung will give an overview of the architecture behind the Diffbot Knowledge Graph, an AI-generated, production knowledge graph built from crawling the entire web. He will describe the component technologies including web-scale crawling, rendering and classification, automated visual extraction, natural language extraction of facts from text and computer vision extraction of facts from images, record-linking, knowledge fusion, and search.

Uses ML



What is Diffbot?

- VC-backed startup of 38 AI researchers, engineers and KG enthusiasts
- Our mission is to: Build the World's first Comprehensive map of Human Knowledge
- First startup in Stanford's StartX accelerator
- Backed by Andy Bechtolsheim, Sky Dayton and early backers of SpaceX, Tesla, and Tencent
- Commercially operating since 2012. 400 customers including Facebook, Amazon, NASDAQ, and Salesforce
- Profitable



Located 8 minutes from campus at SRI



2:32 / 1:50:48

Scroll for details



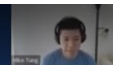
How it is built: First crawl the web, but comprehensively. They do a render of the page as if it is like a video game accessing each pixel, font types, colors, locations. It is like what a human experiences.

This data is serializes as input to their ML. Then it is classified (is it a product page or article?)

Used by their customers to mo monitor what is happening on the web.

Good precision.

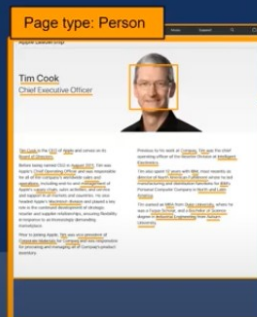
Then do visual extraction using visual features. Pick out names of product, related products, headline. You can earch in English but find articles on that topic in another language.



How the Diffbot KG is built

An overview of some key machine learning components in the KG pipeline:

1. Page type classification
2. Visual extraction
3. Natural language understanding
4. Record Linking



Tim Cook
 Title1: CEO
 Emp1: Apple
 StartDate1: 2011

Skills: sales, operations, management, supply chain, service, support
 Edu: Duke, Degree: MBA
 Edu: Auburn, Degree: BS
 Glasses: true



6:47 / 1:50:48

Scroll for details



NLP uses Open (world) relation extraction uncovers new relations not explicitly expressed in text on the web.

Image extraction works similarly and ML is then used to integrate image and text extracted knowledge.

One example used concerned knowledge about the CDC.

Created tools to clean up data from tables. For example, typos and different addresses. They enhance the data directly by linking it to their KG. Items get resolved and incomplete data gets filled in. Info from one language source can be placed in another.

There is a paper on how to populate Kgs using NLP. Free access for research groups.

CS520: Knowledge Graph Seminar Session 3 (Spring 2020)

CS520: The Diffbot Knowledge Graph

Press **Esc** to exit full screen

Ongoing Research

KnowledgeNet

A benchmark task for Knowledge base population from text

[EMNLP 2019 paper](#)

Mesquita, Cannaviccio, Schmidek, Mirza, Barbosa

27:59 / 1:50:48

Scroll for details

Other examples – multi-modal knowledge fusion of text and images released for researchers. Comes out with a probability that a fact is true.

Has to hand spam using knowledge trustworthiness (from a blog vs Wikipedia)

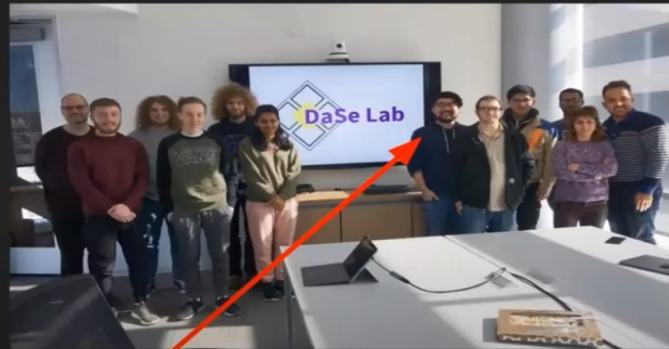
[Cogan Shimizu](#) modular ontology engineering. Knowledge graphs are poised to be significant disruptors in both private and public industry. However, as with any complex entity, developing a knowledge graph requires significant expertise and time, as well as maintenance. In order to reduce the upfront and sustained costs, it is necessary to design the schema of a knowledge graph (also known as an ontology), in a way that supports interoperability, reuse, and maintainability. This presentation will put forward the notion of pattern-based modular ontology engineering and its supporting methodology and infrastructure.

One area is praxis of Kgs.



The Praxis of Knowledge Graphs

- How are they used?
- Development Tools
- (Re)Use
- Useful semantics
- Structural and Usage Patterns
- Automatic knowledge graph construction and population



Me!

The Data Semantics Lab

www.daselab.cs.ksu.edu/

How can we develop a highly useful KG? What infrastructure is useful for this?

Modular ontology engineering is one thing. Kgs often are hard to reuse. They may lack a clear schema. Schemas are important to understand the structure in the KG.

But what constitutes a good schema? Illustrated using a University example from a Wikipedia data and what seems to be the underlying schema.

A good schema will allow adding contextual info such as the time that enrollment is true. So we need to add blank nodes for this type of context. Makes the knowledge clearer.

We also want schemas that are low cost for maintenance and adaptability Doesn't require as much expertise or time.

Should be Human readable.

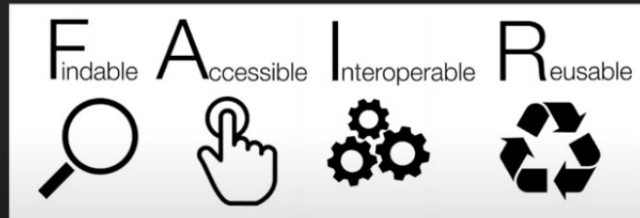
Design decisions should be clear.

Press **Esc** to exit full screen

More Properties of Good KG Schema

A Knowledge Graph schema should be easily adaptable to new or changing use-cases, reduce ambiguity, and be well documented.

- Easy to Adapt and Maintain
- Human Readable Documentation
- Rigorous Definitions



Graphic By SangyaPundir - Own work,
CC BY-SA 4.0



41:06 / 1:50:48

Scroll for details



Methods to achieve these include use of logic to express the schema with precision. For example OWL DL. Promotes the FAIR data practices.

For maintenance and adaptability we use modularity. It allows Kgs to evolve. Modules act as a bridge between human conceptualization and data. They allow a plug and play.

Methods of modular ontology engineering.

Press **Esc** to exit full screen

Modular Ontology Engineering (briefly)

A methodology for developing highly reusable knowledge graphs that emphasizes a pattern-based approach using graphical representations and systematic axiomatization [2,3,4].

- Use-case driven
- Assumes an “Empirical or Data Reality”
- Ontology Design Patterns
- Create Modules from Patterns
- Modules are the primary component

46:01 / 1:50:48

Scroll for details



They use schema diagrams to informally depict relations of classes. There are ambiguities in these which have to be overcome subsequently. What does a particular edge represent?

They use a set of 17 or so axioms (such as temporal extent) to detail the diagrams.

For example we may add a temporal role to the diagram.

List of steps such as scoping since you can't model everything.

They have come up with some tooling that plugs into Protege : CoModiDE.

See CoModiDE.com

Further Supporting Infrastructure

- OPLa Annotator
Quickly and easily add OPLa Annotations to your modules [12].
- ROWLTab
Create OWL axioms from existential rules [13].
- OntoSeer
Recommends vocabularies, classes, axioms, and ODPs to reuse [14].
- ODPReco
Recommends ODPs based on description, CQ's and axioms [15].

54:29 / 1:50:48

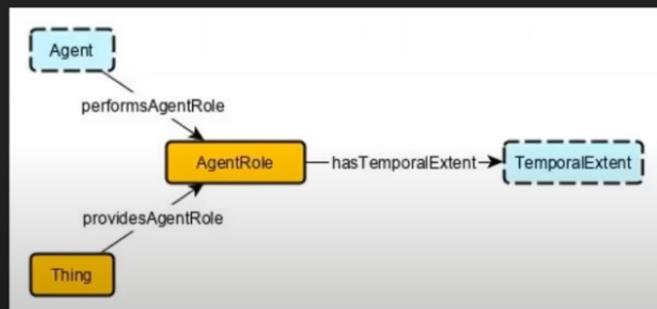
Scroll for details



Schema Diagrams

Informal, but intuitive, graphical representations depicting the relationships between classes in an ontology [5].

- Yellow Boxes := The Key Notion
- Blue Boxes := "Hidden Complexity"
- Arrows := Properties / Relations
- Open Arrows := Subclass Relation

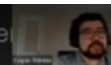


47:32 / 1:50:48

Scroll for details



A Schema Diagram for the "Agent Role" ODP

Press **Esc** to exit full screen

The Methodology

1. Define use case or scope of use cases.
2. Make competency questions while looking at possible data sources and continue scoping the problem and use-case(s).
3. Identify key notions from the data and the use case and identify which pattern should be used for each. Use “stubs” as necessary.
4. Instantiate these key notions from the pattern templates, then adapt the result as needed, to create modules. Develop the remaining modules from scratch.
5. Systematically add axioms for each module.
6. Assemble the modules and add axioms which involve several modules.
7. Reflect on all entity names and possibly improve them. Check module axioms whether they are still appropriate after putting all modules together.
8. Create OWL files.



51:25 / 1:50:48

Scroll for details



[Marie-Laure Mugnier](#) Reasoning on data with Existential Tools

Existential rules are a knowledge representation language dedicated to reasoning on data. This syntactically simple yet powerful language overcomes limitations of both Datalog, the language of deductive databases, and prominent ontological languages like description logics. After introducing existential rules and their salient properties, we will present the main paradigms to query knowledge graphs in this framework, whether these graphs are materialized or virtually defined from multiple data sources.

Use logic to represent Kgs Defined the idea of an existential rule to identify an unknown node in a set of facts.

Used a conflict of interest situation to discuss what types of knowledge is needed to answers conflict questions.

Conflict of interest rule defined with complex relations of objects – beyond unary and binary relations. We reify such n-ary relations. Allows flexible descriptions with first class objects.

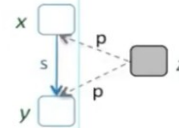
RULE APPLICATION

Press **Esc** to exit full screen

R is **applicable** to F if there is a homomorphism h from $body(R)$ to F
 ie a substitution h of the variables in $body(R)$ by terms in F such that $h(body(R)) \subseteq F$

$$R = \text{SiblingOf}(x,y) \rightarrow \exists z \text{ parentOf}(z,x) \wedge \text{parentOf}(z,y)$$

$$F = \{ \text{SiblingOf}(a,b) \}$$

$$h: \begin{matrix} x \mapsto a \\ y \mapsto b \end{matrix}$$


The **application** produces $h(head(R))$
 where a fresh variable (a null) is created for each existential variable in R

$$F' = \{ \text{SiblingOf}(a,b), \text{parentOf}(z_0,a), \text{parentOf}(z_0,b) \}$$

Research lines from investigating graph-based rules. Find all the answers to query in a KG.

THEORETICAL FOUNDATIONS

Graph-based KR

[Chein Mugnier
1992, 2009]

logical translation of graph rules

 $\forall \exists$ -rules, existential Rules [Baget+ IJCAI 2009]

Datalog+/- family [Cali+ PODS 2009]

Datalog (70-80s)

+ existential variables
in rule heads

RDFS

Lightweight Description Logics,

e.g. OWL 2 tractable profiles

More generally, Horn Description Logics

- Same logical form as « Tuple-Generating Dependencies » (TGDs)
long studied in database theory



1:10:45 / 1:50:48

Scroll for details

8



To make queries tractable they have identified some pattern of rules that are decidable.
 So we have families of queries.
 Forward chaining called Chase to rewrite (materialize) the rule.
 Backward chaining or rewriting to a first order query is not also applicable.

CS520: Knowledge Graph Seminar Session 3 (Spring 2020)

Press Esc to exit full screen

query rewriting

Finite chase

weakly-sticky

sticky-join

glut-fg

jointly-fg

weakly frontier-guarded

jointly-acyclic

weakly-acyclic

sticky

aGRD

warded

weakly-guarded

frontier-guarded

guarded

frontier-1

Datalog

linear

DL-Lite

RDFS

Scroll for details

12

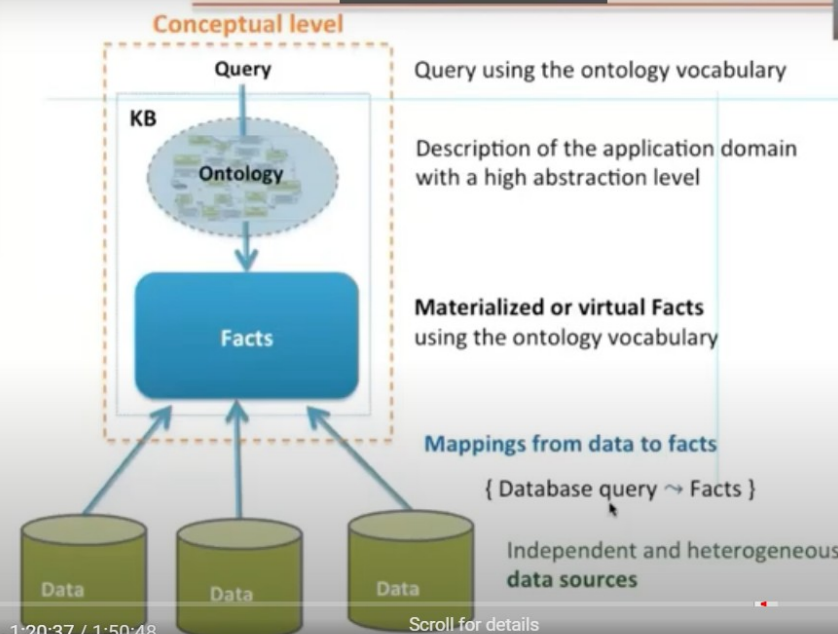
CC

Settings

Fullscreen

There is a corresponding view of complexity – such as polynomial.

Ontology Based Data Access framework. Mappings allow us to select data for our application. The framework is being extended using existential rules which allow powerful mappings that in turn find new knowledge such as missing values.



1:20:37 / 1:50:48

Scroll for details

14



Summary

- Existential rules are able to express **complex structures** and create **new objects**
- These features can be exploited for both expressing **ontological knowledge** and **integrating data**
- A wide range of rule classes offer various **expressivity/complexity** tradeoffs
- The framework has been **extended** in several ways:
 - other rules: negative constraints and Equality Generating Dependencies
 - existential rules extended to stratified negation and disjunctive heads
- Efficient **systems** are available



Maurice Mugnier – Reasoning on Data with Existential Rules

17



Applications of Existential rules:

CS520: Knowledge Graph Seminar Session 3 (Spring 2020)

SOME SYSTEMS

Press **Esc** to exit full screen

Query answering with existential rules

- VLog** — fast chase-based engine — <https://icli.inf.tu-dresden.de/web/VLog/en>
- RDFOx** fast chase-based engine <https://www.cs.ox.ac.uk/isg/tools/RDFOx/>
- Vadalog** fast chase-based engine for warranted Datalog+/- [Commercial]
- Graal** toolkit with fast query rewriting, rule set analysis, chase algorithms, ...
<http://graphik-team.github.io/graal/>

and many other usable tools developed for other purposes:

- Llunatic** data exchange and data cleaning (chase on tuple-generating dependencies)
- DLV** Answer Set Programming (DLV[∃]: for shy existential rules) [Commercial]
- ...

OBDA systems are still restricted to lightweight description logics or (extended) RDFS:
OnTop, Mastro, UltraWrap^{OBDA} ...

1:24:04 / 1:50:48 Reasoning on Data with Existential Rules

18

Questions and Answers

Q for Mike on DIFFBOT Do they use a modular approach and ontologies?

A An open answer in ML field. They can find the ontologies and may use an upper level depending on the application. The lower level needs experts.

Q Do you have a fact that people have one mother?

A This is extracted from the data.

Q on developing pattern libraries

A Being done as part of OKN funded research – emphasis on geo-information.

Q on CoVID work using Kgs.

A Mike - some work crawling to see what effect social distancing makes. Or can Kgs reveal interactions between drugs.

Some work on logistics also being done by Krzysztof Janowicz,

Q Does DIFFBOT use schema.org?

A It has access and can use it as a feature. But it is often just used as markup to enhance presence on the Web.

Knowledge Graph Seminar Session 4

What are some knowledge graph inference algorithms?

[An Hai Doan](#) Fundamental Operation for Building Kgs (Magellan Project and *Cloudmatcher* at UW)

Recent years have witnessed the emergent success of graph neural networks (GNNs) for modeling structured data. However, most GNNs are designed for homogeneous networks, in which all nodes or links have the same feature space and representation distribution, making them infeasible for representing real-world evolving heterogeneous graphs, such as knowledge graphs. In this talk, I will introduce GNN architectures that can model billion-scale heterogeneous graphs with dynamics. The focus will be on how we design the graph attention and temporal encoding mechanisms to capture the heterogeneous and dynamic natures of real-world graphs. With this, I will further discuss the strategies of pre-training such GNNs for general graph mining tasks. Finally, to handle Web-scale data, I will introduce the heterogeneous mini-batch graph sampling algorithm for efficient and scalable training. Extensive experiments show the promise of GNN pre-training for billion-scale (knowledge) graphs in practice.

Example of a customer DB. Entity **matching** (EM) is a key. Algorithms have gotten very complex.

(Magellan Project took an end-to-end approach, PyMatcher a PostgreSQL world.)

Cloudmatcher allow non-expert users to do matching by answering questions (labeling for active learning)

Two main steps to reduce the complexity are

1. blocking (with a variety of similarity measures & rules on feature vectors from tokenized attributes) to only look at likely cases and


2. matching say using random forest approach that encodes matching rules again on feature vectors.

Knowledge Graph Seminar Session 4 (Spring 2020)

Press **Esc** to exit full screen

Conclusions

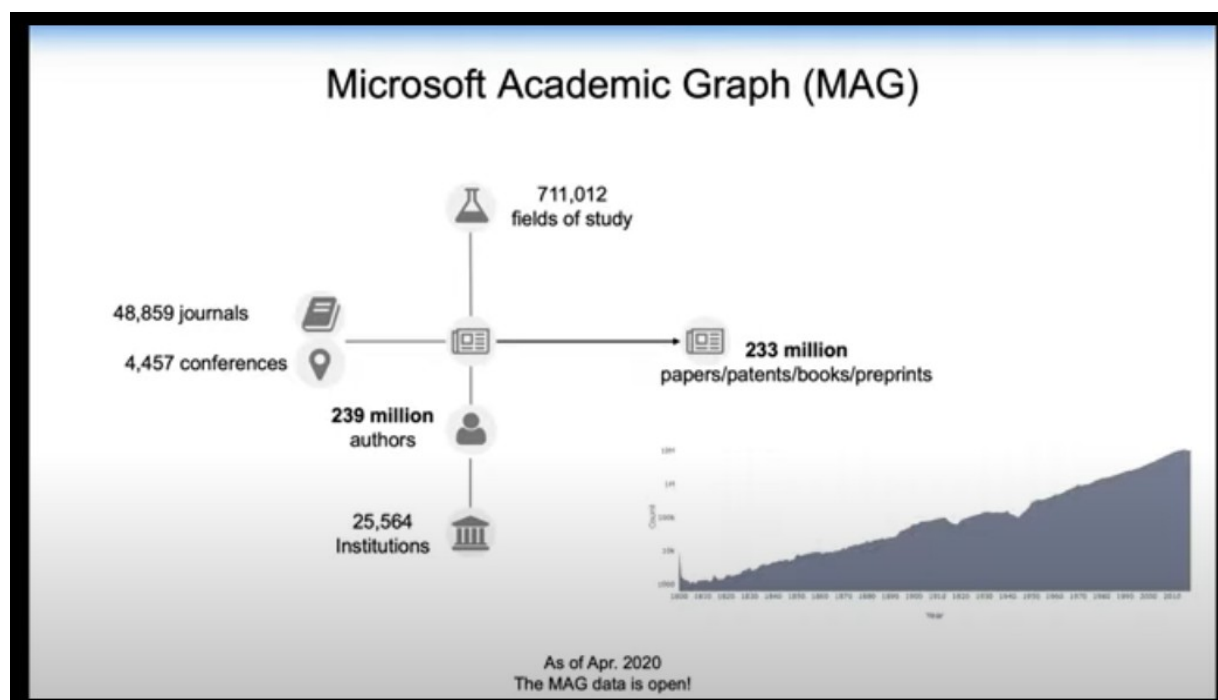
- We are developing Magellan, a “PostgreSQL” for entity matching
- Two major systems developed by 2019, with many real-world users
 - PyMatcher: open-source on-prem EM platform (Python)
 - CloudMatcher: close-source cloud-based EM platform
- CloudMatcher
 - Uses very expressive & powerful blocking and matching rules
 - Learns these rules from labeled data via active learning
 - Needs to combine ML, big data technologies, and effective user interaction
- Magellan is being commercialized
 - in GreenBay and in partnership with Informatica
- A major step forward for entity matching R&D
 - New innovative system platforms, demonstrates how ML can be used effectively
 - Research Highlight in Communications of the ACM
 - Suggests similar system-centric solutions for other data cleaning and integration tasks



46

Yuxiao Dong Learning with Academic (Domain) Knowledge Graphs

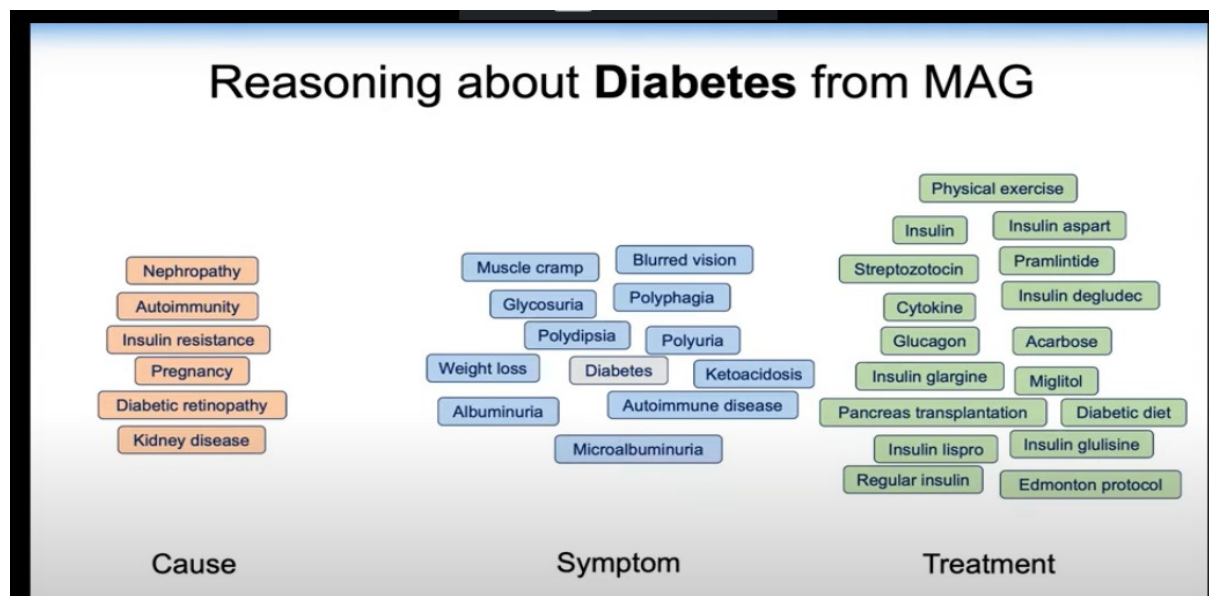
Compared to say Google Scholar they are more interested in the semantics than terms. And they are much larger.



Now we have heterogeneous Kgs (BING Academic KG etc) of entities and relations. So we need a way to identify disambiguate entities (say journal Nature or Yang Yang).

What are the topics of an article is very difficult.







They are trying to reason about Diabetes and Covid-19 with MAG.



Research - Do we need a tool for each task? Trying to pre-train the language models in NLP.

CS520: Knowledge Graph Seminar Session 4 (Spring 2020)

Tasks in Microsoft Academic Graph

Data	Tasks
 228,146,081 Papers	<ul style="list-style-type: none">• Name disambiguation• Infer spam/fake papers• Infer paper/author research topics• Infer future scientific impact• Infer collaboration and team formation• Infer future paper title?•
 230,741,184 Authors	
 664,845 Topics	
 4,406 Conferences	
 48,754 Journals	
 25,566 Institutions	

40:58 / 1:50:24 Scroll for details

Training is in 2 parts – at the model level and the task level.

At the model level they introduce more complex relations – heterogeneous graph transformer approach

For the dynamic part they use relative temporal encoding. So different timestamps can affect each other (like relative geo-positions),

Knowledge Graph Seminar Session 4 (Spring 2020)

Challenges

- Model level: Current graph neural networks are not capable enough to capture
 - The graph heterogeneity
 - The graph dynamics
- Task level: Traditional tasks are not informative enough to model the graph
 - One-relation link inference
 - Multi-relation link inference

1. Ziniu Hu, Yuxiao Dong, Kuansan Wang, Yizhou Sun. Heterogeneous Graph Transformer. WWW 2020.

Summary

As part of their Info-NCE training they introduce fake or adversarial papers (events) so that it learns to discriminate these.

Knowledge Graph Seminar Session 4 (Spring 2020)

Challenges

- Model level: Current graph neural networks are not capable enough to capture
 - The graph heterogeneity
 - The graph dynamics
- **Solution: Heterogeneous Graph Transformer**
 - A model that can learn to capture complex dependencies of entities with different types, relationships, and timestamps (without labeled data)
- Task level: Traditional tasks are not informative enough to model the graph
 - One-relation link inference
 - Multi-relation link inference
- **Solution: Event-driven Info-NCE Training**
 - A learning objective that can guide a single model to learn all the relationships simultaneously.

1. Ziniu Hu, Yuxiao Dong, Kuansan Wang, Yizhou Sun. Heterogeneous Graph Transformer. WWW 2020.

Knowledge Graph Seminar Session 4 (Spring 2020)

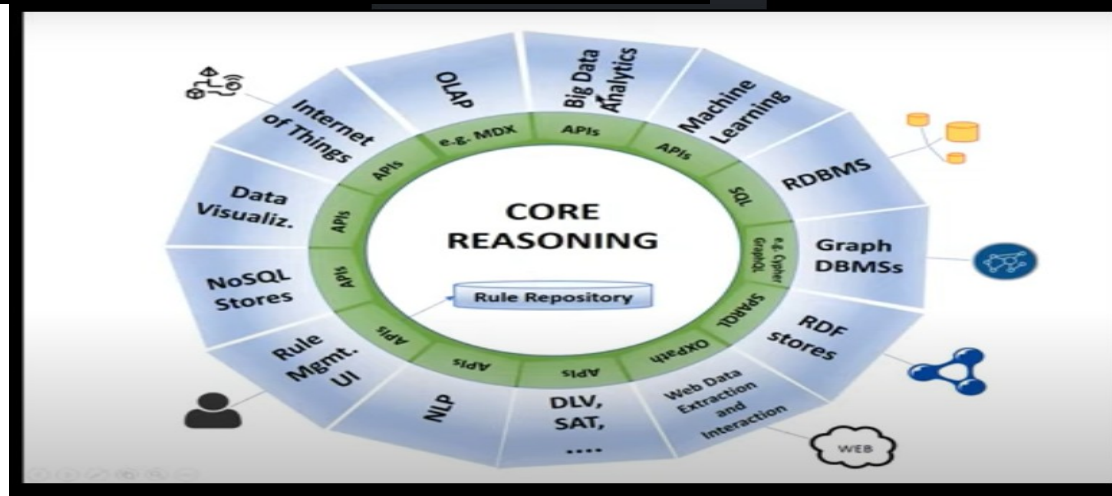
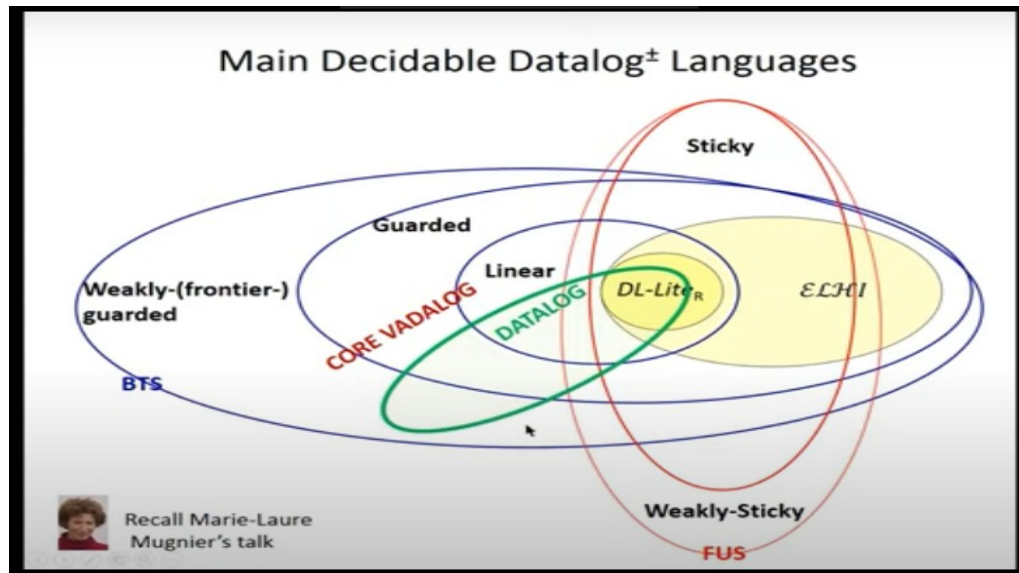
Press **Esc** to exit full screen

Summary

- **Heterogeneous Graph Transformer for Knowledge Graph Inference**
 - A model that can learn to capture complex dependencies of entities with different types, relationships, and timestamps (without labeled data)
- **Pre-Training Strategies for Heterogeneous Graph Neural Networks**
 - A learning objective that can guide a single model to learn all the relationships simultaneously
- **The Microsoft Academic Graph is Open to Download at**
 - <https://docs.microsoft.com/en-us/academic-services/graph/get-started-setup-provisioning>

Georg Gottlob VADALOG a language and system like a full Datalog with restricted use of existential quantifiers and stratified negation – it extends Datalog in a decidable way

the VADALOG Flower showing interfaces.



Note a graph DB does not suffice. There may be no reasoning there. DL does not allow us to express somethings we need that are in Dbs such as “married from x to y” knowledge in these rules

Hard to express some things like aggregate functions and recursion - X controls y if

- x directly hold over 50% of y or'
- x controls a set of companies that jointly hold over 50% of y

Clumsy to express in SQL

Other examples involve commonsense – creditworthiness if you live in multiple places.

Knowledge layers

h Seminar Session 4 (Spring 2020)

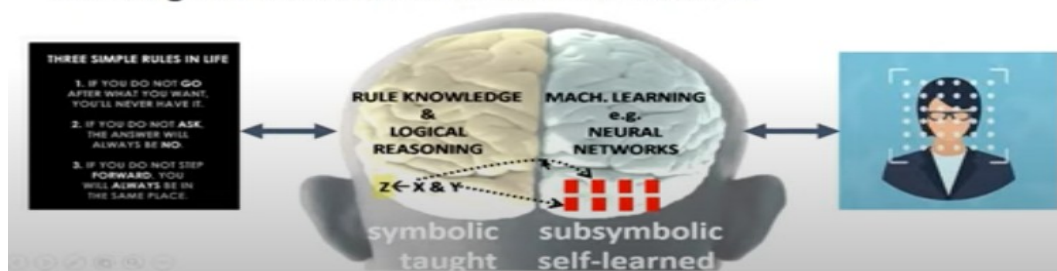
Press **Esc** to exit full screen

Knowledge Graph Management Systems

KGMS combine the power of rule-based reasoning with machine learning over Big Data:

$$\text{KGMS} = \text{KBMS} + \text{Big Data} + \text{Analytics}$$

Misusing the lateralization thesis for illustration

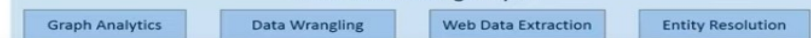


Knowledge Layers

Vertical-specific Knowledge Layers



General Knowledge Layers



Core Reasoning Engine

Strong performance and Expressiveness, Graph Navigation,
+ Integrations with Machine Learning & Enterprise Databases

Vadalog: The Core Reasoning Language

Core Vadalog = full Datalog + restricted use of \exists + stratif. negation + \perp

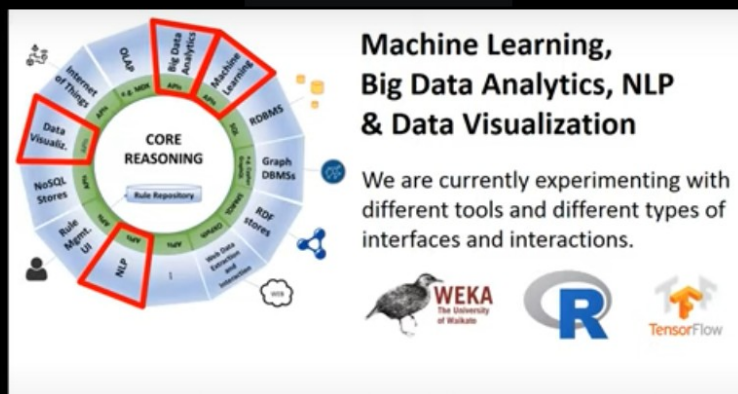
Why existential quantifiers in rule heads?

- Data exchange, data integration
- Data extraction
- Reasoning with RDF → Wikidata example
- Ontology querying (DL-Lite, EL, etc.)
- Automated product configuration
- Conceptual Modeling (e.g., UML)

Research on good algorithms to do useful thing with ML.

New project to extract data automatically **DIADEM**

CS520: Knowledge Graph Seminar Session 4 (Spring 2020)



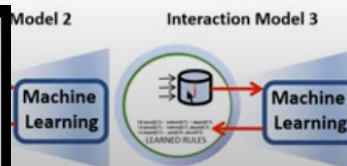
Crucial Question

We have a powerful language and system for reasoning with rules Over „Big Data“ and can interact with machine learning.

But are there actually real problems that can be solved with a reasonable number of rules?

Yes, there are many!, for example:

- Banks: Fraud detection, [current project]
- Banks: Creditworthiness
- Logistics: Supply chain risks
- Security companies: Detection of critical events
- Fully automated Web data extraction: **The DIADEM project**
- ... and 10000 more.



Questions:

For Georg have you tried using NL to express the rules? Yes, thought of this. Can express rules in SQL.

For MAG have you ever tried to couple this with rules to supplement the ML.

A. Yes, there are rules employed say in name disambiguation for ORCID IDs. Or a URL from an academic institution gives high value.

Knowledge Graph Seminar Session 5 **How to evolve a knowledge graph?**

Héctor Pérez-Urbina (google graph work) KG Evolution and Maintenance

We are usually modeling a dynamic world so things changes and you need to adapt.

Here are some things we found:

- Vocaloids –fictional characters that have albums forced them to change their assumption on what could be an artist.
- The book domain is involved. From books you may want to talk about other types of “publications”. but need to have a general concept. But the applications on top of the KB may already use a property for a book. So generalizations can cause problems.
- CEOs are modeled with a type with start and end date.
- Music albums are covered in multiple sources. The schemas differ so integration is hard.

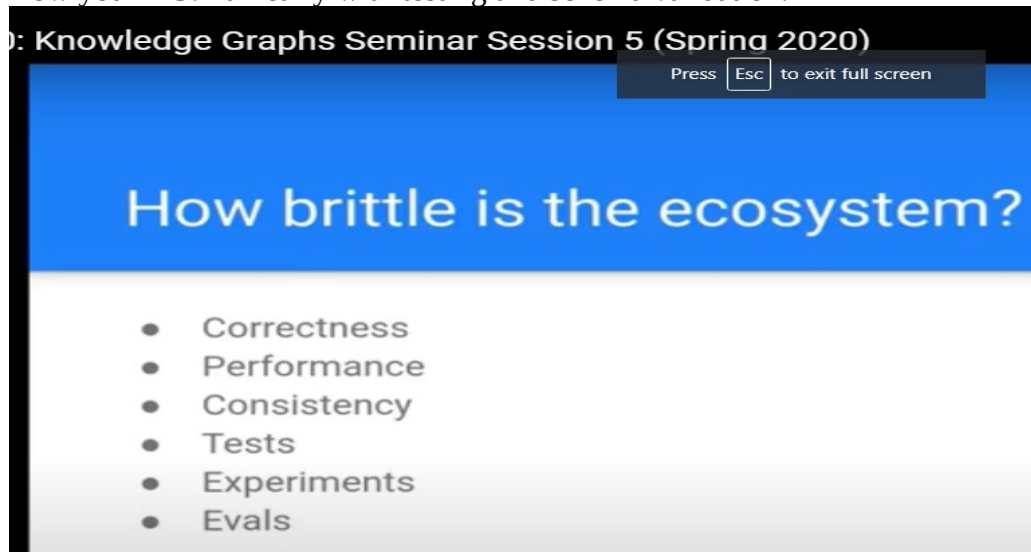
- Inference can be hard when based on wrong assumptions. Not only movies are played in a cinema. Operas got classified as movies since shown there

But modifying a KG can be easier than modifying a DB. Adding a column is harder than adding a triple. Factors you need to take into account:

- What is changing? Might be the scope. Adding some model,
- Is there data already? It might be noisy, incomplete and need to be integrated. Is it a one time migration or continually synchronized. How do you deal with retractions?
- Who is using the data? To some users people and animal are disjoint. Not to biologists. Assumptions of a model can change. Groups may be switching to a new version.
- How complex or brittle is the ecosystem? You may have many SPARQL endpoints, many models integrated and many sources. How quickly are changes to the schema made? What is done for truth maintenance? Are constraints mutable? Not allowed if it breaks the schema.

Brittleness – if billions of triples need to change to make something correct you have to take it into account.

So know your KG. Fail early with testing and schema validation.



Get support for experiments and do comprehensive evals.

You need a governance and policy to manage the KG construction and maintenance.

Automation is your friend here: pipelines, automated orchestrate schema release, help with rollbacks and emergency fixes.

Example of a perfect storm situation with corona virus where things were changing rapidly. The schema evolved quickly as different sources of data popped up. Parking lots now became medical testing areas.

They could keep up with a KG and doubt they could with Dbs.

Summing up

- Data changes all the time
- KGs are versatile, but changes might still be painful :)
- What to consider
 - What's changing?
 - Is there data already?
 - Who's using the data?
 - How complex/brittle is the ecosystem?
- We can ease the pain through testing and automation

José Manuel Gómez-Pérez KG(s) for NLP

CS520: Knowledge Graphs Seminar Session 5 (Spring 2020)

Overview of the talk

Press **Esc** to exit full screen

- **Two main questions**
 - **Why** do we need knowledge graphs (KGs) in NLP?
 - **How** can we extract information from text and KGs to build better representations for NLP?
- **Three main blocks/systems**
 - **COGITO©**: An industrial NLP system based on KGs
 - **Vecsigrafo**: Learning joint word and concept representations from text corpora and KGs
 - **Transigrafo**: Combining language models and KGs for word-sense disambiguation

27:55 / 1:50:34

Scroll for details

CC BY-NC-SA

EXPERT SYSTEM

Contrast of 2 approaches – KB and Neural

Demo of Cogito to understand text on the 2016 US election.

Extracts details on people you can drill down on and visually shows places

Demo of Vecsignafo learns words and concept embeddings in a shared vector space.

Can be used to detect bot messages rather than human.

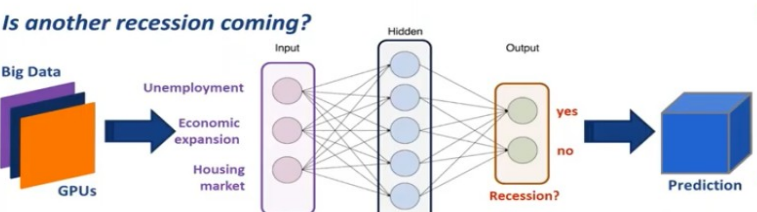
Some progress on cross lingual mapping between languages but KG(s) differ in size between 2 languages.

Also there are different strategic decisions made in the 2 different languages.

Still challenging to co-evolution tasks.

CS520: Knowledge Graphs Seminar Session 5 (Spring 2020)

The neural (data-driven) approach



Is another recession coming?

Big Data → GPUs → Unemployment, Economic expansion, Housing market → Input → Hidden → Output → Prediction

Recession? (yes/no)

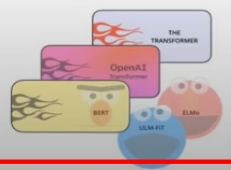
Pros

- Grounded on the data
- Broad, flexible, scalable
- State of the Art in most NLP/NLU benchmarks (QA, RTE, NLI, etc.)

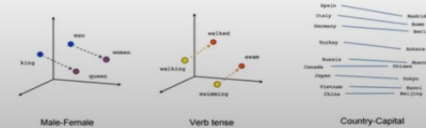
Cons

- Still a black box. Induction, not logical explanation
- Lack of true understanding of real-world semantics and pragmatics
- Training corpora needs to be carefully selected to prevent bias

Language models



Word embeddings



Male-Female, Verb tense, Country-Capital

29:03 / 1:50:34

Scroll for details

3 CC

EXPERT SYSTEM

CS520: Knowledge Graphs Seminar Session 5 (Spring 2020)

The knowledge-based approach

PROS

- Based on highly curated resources
- No need to train a model!
- Logically interpretable, explainable
- Graph structure is of direct benefit for NLP tasks like word-sense disambiguation (WSD)
- Tools available to support modeling

CONS

- Representations can be rich and deep but also rigid and brittle
- Automation can be challenging
- Well trained labor to manually encode knowledge can be expensive

30:22 / 1:50:34

Co-evolution example = concepts like financing work

CS520: Knowledge Graphs Seminar Session 5 (Spring 2020)

COGITO

- COGITO is Expert System's NLP platform
- Built on the Sensigrafo KG, which contains word definitions, related concepts and linguistic information
- Sensigrafo includes concepts, lemmas (canonical representation of a word) and relations (properties, hypernymy, polysemy, synonymy...)
- 300K concepts, 400K lemmas and 80 relation types, which render 3 million links per language
- COGITO covers the whole NLP/NLU pipeline, including Word-Sense Disambiguation, NLI, IE, Categorization...

"Chancellor Merkel is making a last-ditch attempt to persuade EU member states to back her coronavirus deal as Tuesday's key meeting vote looms closer."

- Word: making, lemma: make, concept: en#73741
- Meaning of back: part of a garment (en#13189), body part (en#100133379), to support (en#66955)

Natural language text is read and analyzed

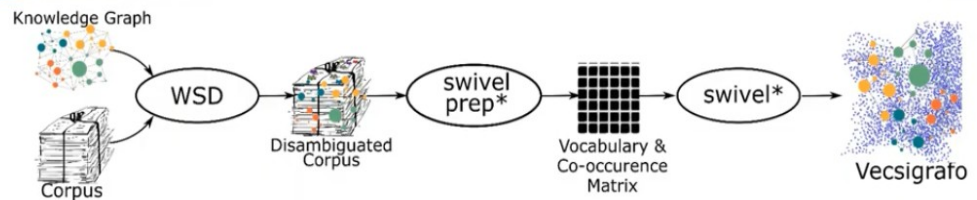
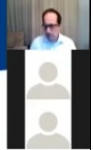
- Sentence splitting / parsing**
 - Divide text into words
- Morphological analysis**
 - Understand language forms (e.g. Verb conjugations)
- Sentence / logical / grammatical analysis**
 - Understand how words relate to other words, and what their function is in the sentence
- Semantics analysis / disambiguation**
 - Understanding sentences and texts as a whole, taking into account words synonyms, context, plausibility

...and then translated into a conceptual map

32:30 / 1:50:34

Vecsigrafo: Crossing the NLP Chasm

Press Esc to exit full screen

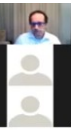


- **Vecsigrafo jointly learns word and concept embeddings** in a **shared vector space**
- Unlike knowledge graph embeddings, Vecsigrafo combines **corpus-based and graph-based** approaches **to build enriched word representations**
- **Extends (*) the Swivel algorithm** with both lexical and semantic entries as part of the vocabulary
 - The corpus is lemmatized following different tokenization strategies, disambiguated and expanded
 - Word and semantic (KG lemma and concept) embeddings are then jointly learnt

Denaux R and Gomez-Perez JM. 2019. **Vecsigrafo: Corpus-based Word-Concept Embeddings**. *Semantic Web Journal* (2019), 68–96. <https://doi.org/10.3233/SW-190361>

Co-evolution support & examples

Press Esc to exit full screen



- Supported co-evolution tasks include
 - Identifying KG modeling gaps
 - Suggesting merge of equivalent concepts
 - Detecting disambiguation errors
- More examples available in backup slides

"Financing" (EN→ES)

@	lemma/syncon	cosim	comment
1	financing	0.96	lemma
2	finance	0.85	lemma
3	funding	0.80	lemma
4	en#178501: adverb for financing	0.79	
5	en#75764: verb for fund, finance	0.76	
6	es#126922: noun financiación,	0.75	synonym



"Scrap value" (EN→ES)

@	lemma/syncon	cosim	comment
1	salvage value	0.92	lemma
2	scrap value	0.90	lemma
3	replacement cost	0.72	
4	en#57338: replacement cost	0.72	
30	en#195309: reduced price, sale	0.61	
??	precio de compra	0.48	
??	es#20836: cambio, valor comercial	0.23	



"PYME" (ES→EN)

@	lemma/syncon	cosim	comment
1	es#92662: pequeña y mediana empresa PYME	0.99	
2	en#2739337:SME	0.81	synonym
3	sme	0.81	synonym
4	mediana empresa	0.79	narrower
5	es#307734: mediana empresa	0.78	narrower



13

Transigrafo Decoupling linguistic a(how sentence are built and KR (concept model) allows advance

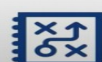
Capturing relational knowledge from word embeddings



Can we set up a pipeline to evaluate how well pre-trained word embeddings capture a set of relations?



- Assume neural models are indeed very good at finding patterns in data
- Use hand-crafted KGs as silver-standard to avoid having to create custom/limited datasets
- Focus on lexico-semantic relations
- Use both lemmas and concepts to study effect of polysemy and abstraction capabilities of embedding models



Transigrafo demo for WSD

Press **Esc** to exit full screen

Sentence to display: 10 111

Token to display: 10 24

☐ About (How do I interpret this?)

☐ How do I use this?

☐ Limitations/Improvements

Precomputed TX Disambiguator Explorer

Sentence #91, token #19:

Either: copy can make the proteins needed to control cell growth, so for cancer to arise, both copies must be inactivated.

Mismatched disambiguation

	Cogito 14.2
glosses	an initiation or reproduction of an original: "she made a copy."
main_lemma	copy
pos	NOUN
id	14962
lemmas	['copy']

	tx 3
glosses	Duplication, as of a nucleic acid, by copying from a molecule.
main_lemma	replication
pos	NOUN
id	101312073
lemmas	['replication', 'copy']
sense	lem_copyiem01312073
confidence	0.425665009021709

	tx 2
glosses	the production of many copies of a section of DNA, naturally -
main_lemma	gene amplification
pos	NOUN
id	127486
lemmas	['gene amplification', 'amplification']
sense	lem_gene_amplificationem127486
confidence	0.425665009021709

- Based on cogito disambiguation instead of a manually curated corpus
- See this as: *how much does Cogito and BERT agree?* rather than which one is better at disambiguating
 - Green is **agreement**, italics **disagreement**
- Model trained on 37K sentences from the SemCor corpus, disambiguated using Cogito 14.2
- 112 sentences from SemEval All
- Limitations
 - Biased training set
 - Evidence limited to the senses in the relatively small training set

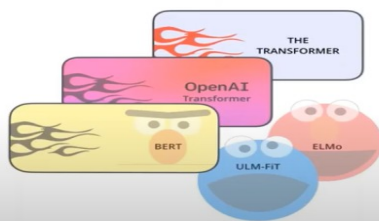
58:03 / 1:50:34

Scroll for details

17



KGs and Transformers

Press **Esc** to exit full screen

- The field of NLP has advanced substantially with the advent of large-scale language models such as ELMo, ULMFit, and particularly transformers like GPT, BERT and RoBERTa
- Trained to perform various language prediction tasks such as predicting a missing word or the next sentence
- Large amounts of text used for training, e.g. BERT trained on Wikipedia + the Google Book Corpus of 10,000 books
- They can also be fine-tuned for other language prediction tasks, such as question-answering, natural language inference, paraphrasing, or named entity recognition
- Transformers actually contain linguistic and word knowledge. Are they the basis for the next-generation KGs?
- How can transformers and KGs interplay, if possible at all?

53:02 / 1:50:34

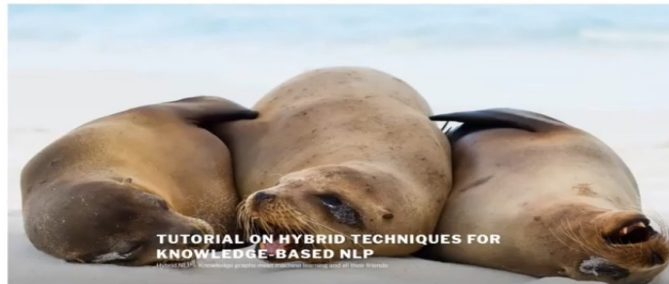
Scroll for details

15



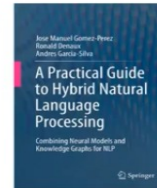
Summing up sources

Resources: Book and tutorial



<http://hybridnlp.expertsystemlab.com/tutorial>

<https://www.springer.com/gp/book/9783030408177>



2020. X, 265 p. 32 illus., 9 illus. in color.
Printed book
hardcover
€ 119,99 | £ 109,99 | \$ 149,99
Pb € (D) 128,39 | € (A) 131,99 |
CHF 141,50
eBook
€ 89,99 | £ 87,50 | \$ 109,00
Pb € (D) 96,29 | € (A) 96,29 |
CHF 113,00
Available from your library or
springer.com/ship
MyCopy®
Printed eBook for just
€ | \$ 24,99
springer.com/mycopy

 Springer

Lifelong 40% discount for authors



Special offer / Get 20% off the printed book
Use the following token on Springer.com
0817FKP2eZncst / Valid Apr 20, 2020 - May 31, 2020

A Practical Guide to Hybrid Natural Language Processing

Combining Neural Models and Knowledge Graphs for NLP

- Provides readers with a practical guide to hybrid approaches to natural language processing involving a combination of neural methods and knowledge graphs
- Includes a comprehensive set of experiments and exercises to illustrate the ideas described
- All the examples and exercises proposed in the book are available as executable Jupyter notebooks in a GitHub repository

This book provides readers with a practical guide to the principles of hybrid approaches to natural language processing (NLP) involving a combination of neural methods and knowledge graphs. To this end, it first introduces the main building blocks and then describes how they can be integrated to support the effective implementation of real-world NLP applications. To illustrate the ideas described, the book also includes a comprehensive set of experiments and exercises involving different algorithms over a selection of domains and corpora in various NLP tasks. Throughout, the authors show how to leverage complementary representations stemming from the analysis of unstructured text corpora as well as the entities and relations described explicitly in a knowledge graph, how to integrate such representations, and how to use the resulting features to effectively solve NLP tasks in a range of domains. In addition, the book offers access to executable code with examples, exercises and real-world applications in key domains, use disinformation analysis and machine reading comprehension of scientific literature. All the examples and exercises proposed in the book are available as executable Jupyter notebooks in a [...]

Order online at springer.com / or for the Americas call (800) 777-4638 / or email us at: customerservice@springer.com / For outside the Americas call +49 (0) 6223 7445 4000 / or email us at: customerservice@springernature.com
The first € price and the \$ and £ price are net prices, subject to local VAT. Prices indicated with [D] include VAT for books; the €(D) includes VAT for Germany, the €(A) includes 20% for Austria. Prices indicated with [E] include VAT for electronic products; 30% for Germany, 20% for Austria. All prices exclusive of carriage charges. Prices and other details are subject to change without notice. All errors and omissions excepted. [D] no discount for mycopy

Part of **SPRINGER NATURE**

Mike Uschold (briefed us) Obliterate Silos with KG

Silos are a big problem in any enterprise. They arise from limitations in relational database technology including the lack of explicit semantics. We describe how to avoid silos using ontologies and knowledge graphs. We show how it works in practice and illustrate with case studies. We warn against the use of these newer technologies to gain a local advantage in an organization but ultimately recreating silos across the wider enterprise. The use of an enterprise ontology as a schema to populate an RDF-based knowledge graph opens the door to removing silos and never creating them again. The technology is mature and ready for prime time.

The slides for the presentation are available [here](#).

Knowledge Graphs Seminar Session 6: How do users interact with knowledge graphs?

Amit Prakash (early BING work) Search and AI Driven Analysis (not really knowledge graph oriented)

System called ThoughtSpot to help people get to the facts. Most systems are static, data driven and broken.

Not good for answering a sequence of questions. And people aren't knowledge experts and don't know how to ask and design a questions to system.

Most systems can't handle Big Data.

520: Knowledge Graphs Seminar Session 6 (Spring 2020)

Press **Esc** to exit full screen

We've made fact-finding and distribution easy via search & AI

	KNOWN QUESTIONS	UNKNOWN QUESTIONS
PUSH	<ul style="list-style-type: none">Report DistributionBots	<ul style="list-style-type: none">AI-driven insightsScheduledContinuous, Self-driving
PULL	<ul style="list-style-type: none">Search to analyze	<ul style="list-style-type: none">Search & auto-analyzePervasive and contextual

Anyone can find their own insights using search, or get automated insights before you have to ask

Scroll for details

different:

How it is

520: Knowledge Graphs Seminar Session 6 (Spring 2020)

Press **Esc** to exit full screen

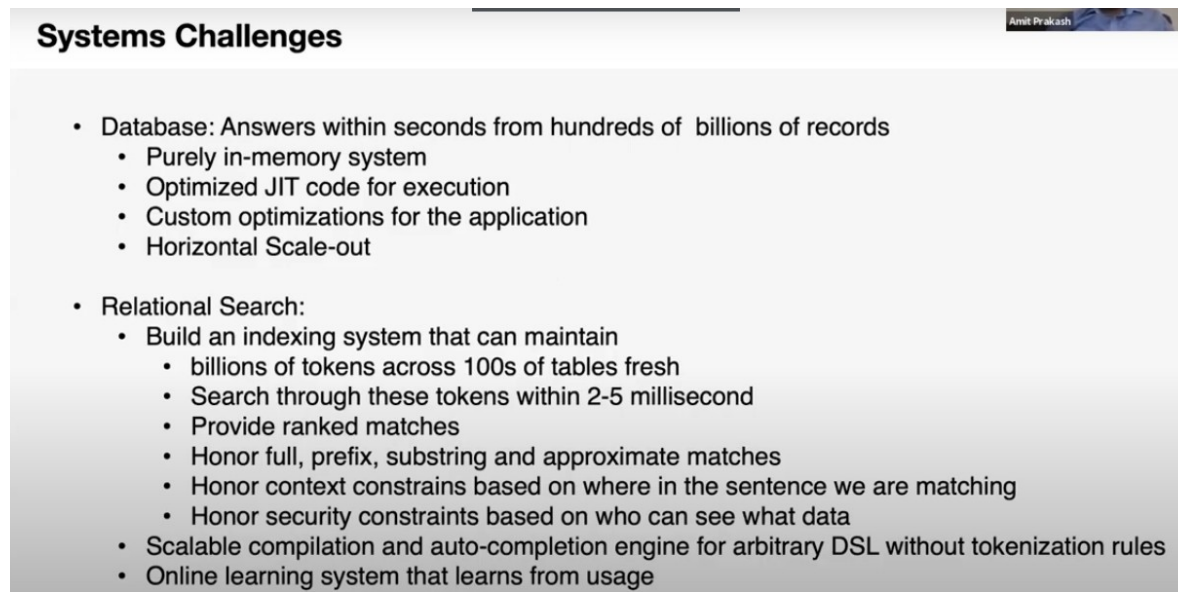
What makes ThoughtSpot different?

Simple	Smart	Fast
Fast and easy access to data for everyone	Automated data discovery	Performance at Enterprise Scale

A quick Demo

Search IQ is a product in development that might use a KG.

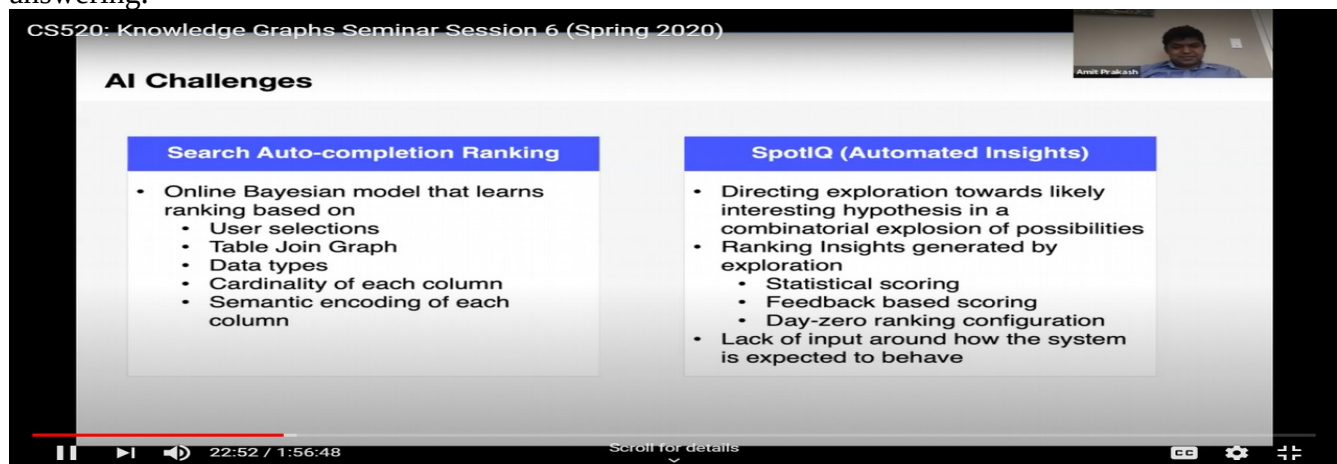
You can teach the system as it goes along so it gets smarter.



Systems Challenges

- Database: Answers within seconds from hundreds of billions of records
 - Purely in-memory system
 - Optimized JIT code for execution
 - Custom optimizations for the application
 - Horizontal Scale-out
- Relational Search:
 - Build an indexing system that can maintain
 - billions of tokens across 100s of tables fresh
 - Search through these tokens within 2-5 millisecond
 - Provide ranked matches
 - Honor full, prefix, substring and approximate matches
 - Honor context constraints based on where in the sentence we are matching
 - Honor security constraints based on who can see what data
 - Scalable compilation and auto-completion engine for arbitrary DSL without tokenization rules
 - Online learning system that learns from usage

Search IQ involves NL (using a NN, parse tree and pattern matching) and real world knowledge. How to bootstrap it with knowledge and how can it fail gracefully letting the user know it has trouble answering.



CS520: Knowledge Graphs Seminar Session 6 (Spring 2020)

AI Challenges

Search Auto-completion Ranking	SpotIQ (Automated Insights)
<ul style="list-style-type: none">• Online Bayesian model that learns ranking based on<ul style="list-style-type: none">• User selections• Table Join Graph• Data types• Cardinality of each column• Semantic encoding of each column	<ul style="list-style-type: none">• Directing exploration towards likely interesting hypothesis in a combinatorial explosion of possibilities• Ranking Insights generated by exploration<ul style="list-style-type: none">• Statistical scoring• Feedback based scoring• Day-zero ranking configuration• Lack of input around how the system is expected to behave

22:52 / 1:56:48 Scroll for details

This provides an opportunity to teach the system with a user feedback “, vs “what’s the longest movie name?”

Another example is Qs with best and worse. How to sort these for users? Worst should be last?

There work includes:

- mining entities from text
- context and
- how to support transfer learning with ML technology.

CS520: Knowledge Graphs Seminar Session 6 (Spring 2020)

Press **Esc** to exit full screen

SearchIQ: Smart Engine on top of Thoughtspot's Relation Search

What are the hottest Adidas products

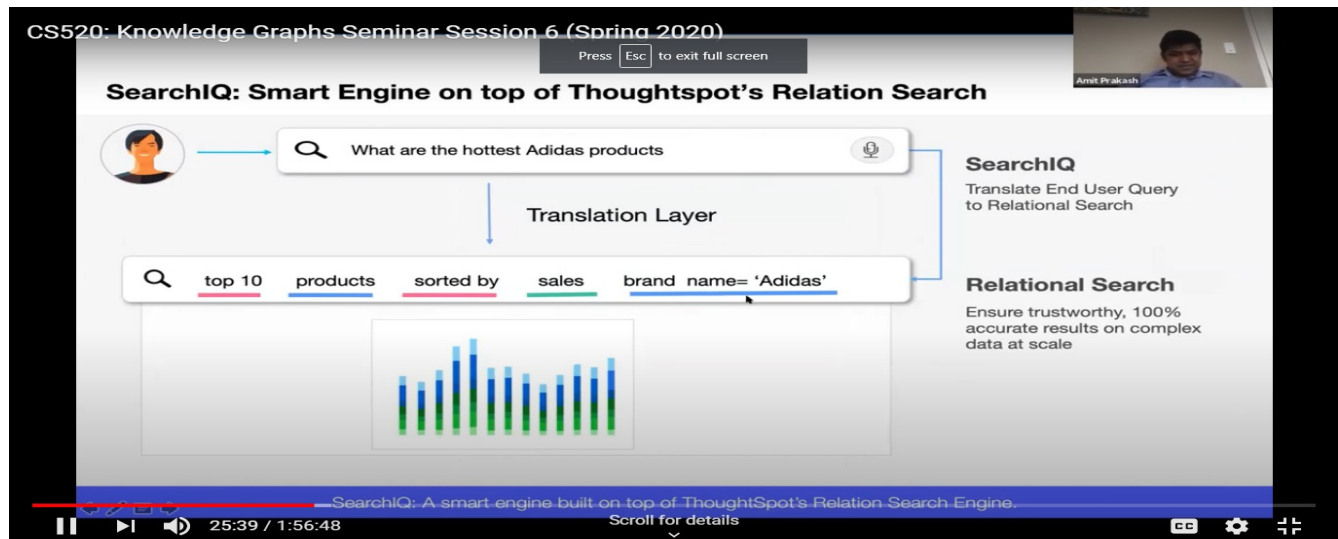
Translation Layer

top 10 products sorted by sales brand name= 'Adidas'

SearchIQ
Translate End User Query to Relational Search

Relational Search
Ensure trustworthy, 100% accurate results on complex data at scale

SearchIQ: A smart engine built on top of ThoughtSpot's Relation Search Engine.
Scroll for details



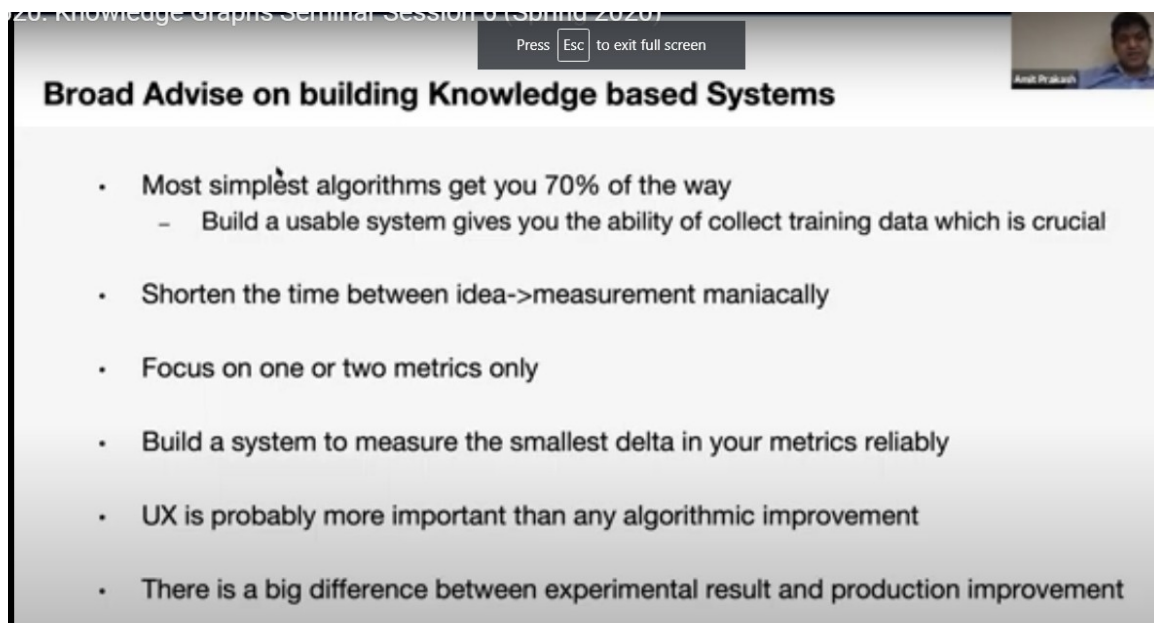
Take aways:

CS520: Knowledge Graphs Seminar Session 6 (Spring 2020)

Press **Esc** to exit full screen

Broad Advice on building Knowledge based Systems

- Most simplest algorithms get you 70% of the way
 - Build a usable system gives you the ability of collect training data which is crucial
- Shorten the time between idea->measurement maniacally
- Focus on one or two metrics only
- Build a system to measure the smallest delta in your metrics reliably
- UX is probably more important than any algorithmic improvement
- There is a big difference between experimental result and production improvement

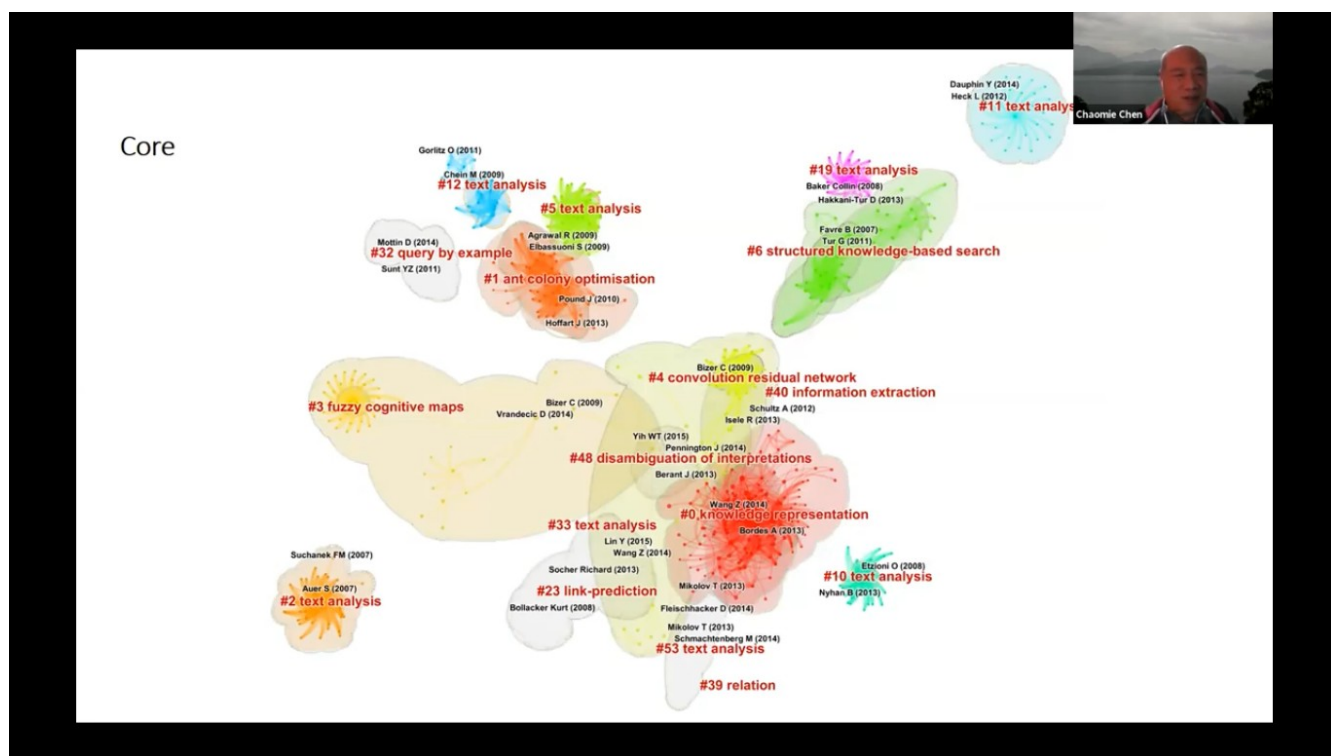


Chaomei Chen Making sense of a field of research?

What has been done in KG research? Graphical view when you search the field

Lens.org and CiteSpace - We can see labels on the clouds of areas that involve KG(s). A core area is knowledge representation.

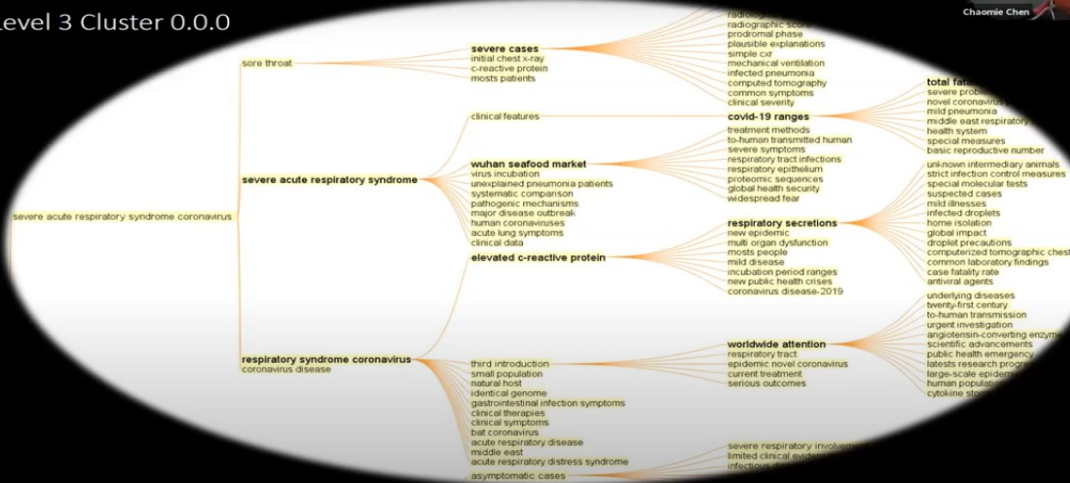
We see some of the neighborhoods - like CRN and text analysis.



You can also look at how the core has changed over time. KG and corona virus used as example to see clusters of concepts and articles on these topics such as vaccine development.

There are 2 or 3 levels of detail that you can go to and get there interactively.

Level 3 Cluster 0.0.0



1:00:13 / 1:56:48

Scroll for details



References

Press **Esc** to exit full screen

Related Topics

Cascading Citation Expansion

- Chen, C., Song, M. (2019) Visualizing a Field of Research: A Methodology of Systematic Scientometric Reviews. PLoS One 14 (10), e0223994.

Epistemic and Other Types of Uncertainties

- Chen, C., Song, M., Heo, G. E. (2018) A Scalable and Adaptive Method for Finding Semantically Equivalent Cue Words of Uncertainty. Journal of Informetrics 12 (1), 158-180.
- Chen, C., Song, M. (2017) Representing Scientific Knowledge: The Role of Uncertainty. Springer.

Leilani Gilpin topic Explaining Explanations (XAI)

Did an large scale literature review.

Imprecise explanations where and error occurs.

Why we need explanations? When systems break why and where did this happen.

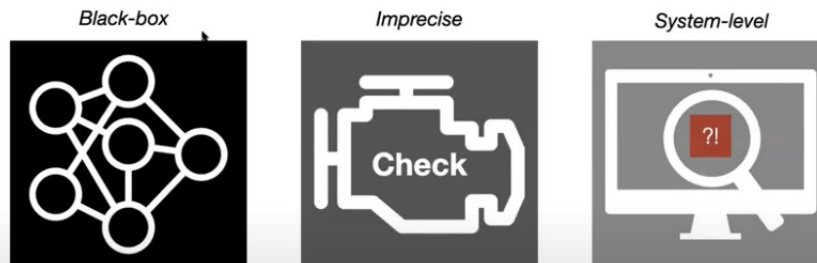
Once we have autonomy on a complex system we can have problems like update OS doesn't allow us to access files.

And the systems lack commonsense. Need for better decisions.

Explanations in AI has been around, but now a big field with deep nets.

Press **Esc** to exit full screen

Explaining Explanations



Leilani H. Gilpin - MIT

Leilani H. Gilpin

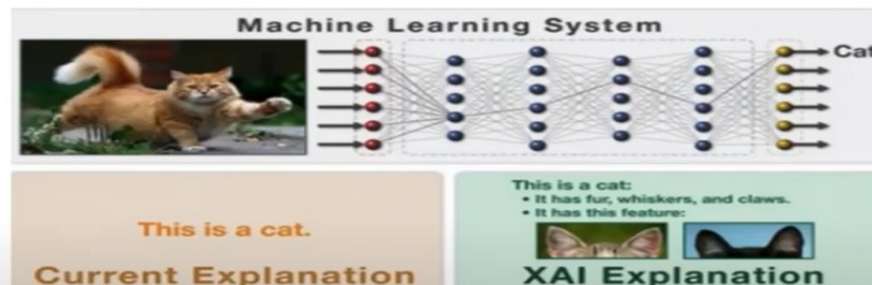
MIT Explaining Explanations
Scroll for details

1:02:29 / 1:56:48

CC ⚙️ 📐

Can't talk about features.

What is Explainability?



From Darpa XAI

Explanations have to be deeper and give new insights

Philosopher's view:

"Explanations...express answer to not just any questions but to questions that present the kind of intellectual difficulty..."

—Sylvain Bromberger, *On What We Know We Don't Know*

Deep NNs are everywhere now including self driving cars, Go and classification in medical tasks.

So we are used to them but they aren't very understandable.

One problem is that there isn't good theory on what goes on or should in the middle layers of a DNN. So when they fail we don't understand why.

A definition of explanation:

Knowledge Graphs Seminar Session 6 (Spring 2020)

Press **Esc** to exit full screen

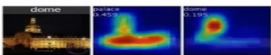

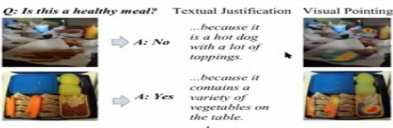
Definitions

- Explainability \neq Interpretability
- **Interpretability** describes the internals of a system that is *understandable* to humans.
- **Completeness** describes operation in an *accurate* way.
- An explanation needs **both**.

it is not the same as interpretability which is about a model like a decision tree.

It also needs to be complete.

What we Have

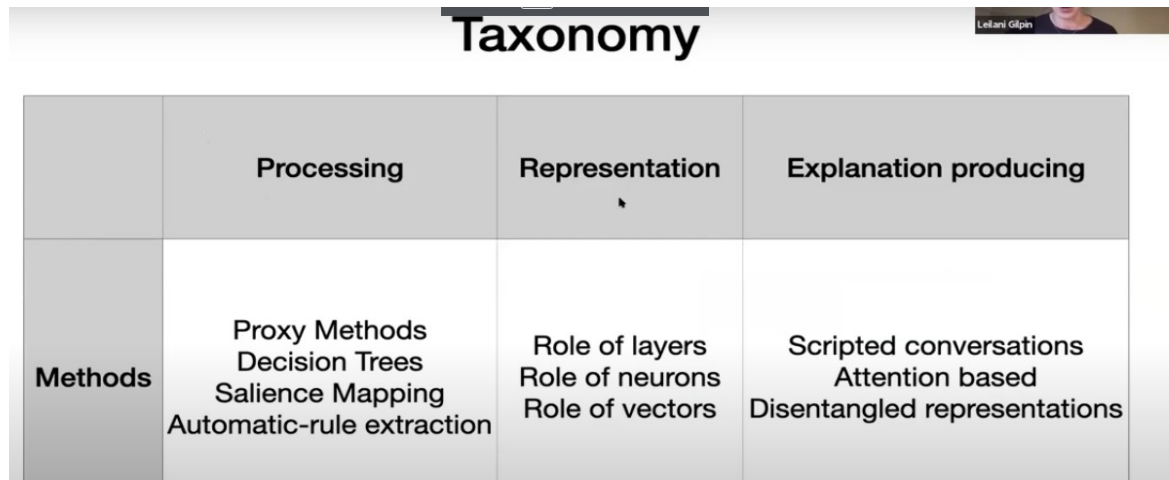
Visual cues	Role of individual units	Attention based
		
Interpretable, not complete	Complete, not interpretable	Interpretable, not complete

L.H. Gilpin MIT Explaining Explanations 9

We don't have good explanation (what is being explained) and this matters say in a fraud detection or errors in cancer classifier – turns out it was largely based on image accuracy.

Visual cues are explaining a different thing than role of individual units (representation).

A



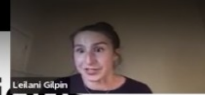
The image is a screenshot of a presentation slide titled "Taxonomy". It features a table with three columns: "Processing", "Representation", and "Explanation producing". The first row of the table lists these categories. The second row, labeled "Methods" in the first column, lists specific techniques under each category: Proxy Methods, Decision Trees, and Saliency Mapping for Processing; Role of layers, Role of neurons, and Role of vectors for Representation; and Scripted conversations, Attention based, and Disentangled representations for Explanation producing.

	Processing	Representation	Explanation producing
Methods	Proxy Methods Decision Trees Saliency Mapping Automatic-rule extraction	Role of layers Role of neurons Role of vectors	Scripted conversations Attention based Disentangled representations

Taxonomy from 87 papers on explanations

Example of methods and what they do well and badly.

DeepRed as an example, a saliency map looks like a decision tree.



Examples of Explained Representations

Network Dissection: Quantifying Interpretability of Deep Visual Representations

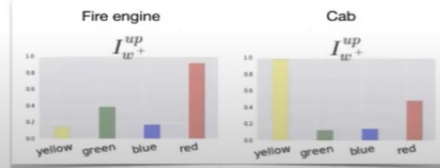
David Bau*, Bolei Zhou*, Aditya Khosla, Aude Oliva, and Antonio Torralba
CSAIL, MIT
{davidbau, bzhou, khosla, oliva, torralba}@csail.mit.edu



D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, "Network dissection: Quantifying interpretability of deep visual representations," in *Computer Vision and Pattern Recognition*, 2017.

Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)

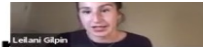
Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler,
Fernanda Viegas, Rory Sayres



Kim, Been, et al. "Tcav: Relative concept importance testing with linear concept activation vectors." *arXiv preprint arXiv:1711.11279* (2017).

VQA and explanation by default.

Examples that Produce Explanations



Multimodal Explanations: Justifying Decisions and Pointing to the Evidence

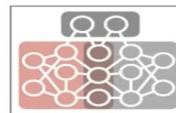
Dong Huk Park¹, Lisa Anne Hendricks¹, Zeynep Akata^{2,3}, Anna Rohrbach^{1,3},
Bernt Schiele³, Trevor Darrell¹, and Marcus Rohrbach⁴

¹EECS, UC Berkeley, ²University of Amsterdam, ³MPI for Informatics, ⁴Facebook AI Research



Park, Dong Huk, et al. "Multimodal Explanations: Justifying Decisions and Pointing to the Evidence." 31st IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[1] L.H. Gilpin. Explaining possible futures for robust autonomous decision-making. Proceedings of the AAAI Fall Symposium on Anticipatory Thinking, 2019.
[2] L.H. Gilpin et al. Anomaly Detection Through Explanations. Under Review.

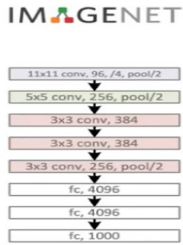


The best option is to veer and slow down. The vehicle is traveling too fast to suddenly stop. The vision system is inconsistent, but the lidar system has provided a reasonable and strong claim to avoid the object moving across the street.


So

These approaches are good, but we need to understand it better. Say how we come up with justifications.


The More Complex (Deeper) The Deeper the Mystery




AlexNet (2012)
8 layers; acc 84.7%



VGG (2014)
19 layers; acc 91.5%



GoogLeNet (2015)
22 layers; acc 92.2%



ResNet (2016)
152 layers; acc 95.6%

L.H. Gilpin
MIT Explaining Explanations
23

For visual cues you ask how complete it is.

For Individual units it is completion on other tasks

For attention based you need to work with users.

Recommendations – need to understand what is being explained.

CS520: Knowledge Graphs Seminar Session 6 (Spring 2020)

Press **Esc** to exit full screen

Taxonomy

	Processing	Representation	Explanation producing
Methods	Proxy Methods Decision Trees Salience Mapping Automatic-rule extraction	Role of layers Role of neurons Role of vectors	Scripted conversations Attention based Disentangled representations
Evaluation	Completeness to model Completeness on a substitute task	Completeness on a substitute task Detect biases	Human evaluation Detect biases

Leilani Gilpin

1:27:12 / 1:56:48

MIT Explaining Explanations
Scroll for details

25

How explanations can help

Challenges with metrics and standards – what makes a good explanation.

We don't have a fixed data set to test on and gauge metrics. Metrics are usually fuzzy and based on information feedback from evaluations.

Need benchmarks for mission critical apps. How to design scenarios

But How Can Explanations Help?

Leilani Gilpin

Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning

Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter and Lalana Kagal
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
{lgilpin, davidbau, bzy, abajwa, specter, lkagal}@mit.edu

- Ex-post-facto
- Static
- Dynamic
- Self-explaining architectures.

Current work on explanatory anomaly detection.

Thinking about dynamic explanations and architectures to support adaptive learning.

Explanatory Anomaly Detection

Adaptable self-explaining architectures

1. Hierarchy of overlapping self-explaining committees.
2. Continuous interaction and communication.
3. When failure occurs, a story can be made, combining the member's explanations.

[1] L.H. Gilpin. Explaining possible futures for robust autonomous decision-making. Proceedings of the AAAI Fall Symposium on Anticipatory Thinking, 2019.
[2] L.H. Gilpin et al. Anomaly Detection Through Explanations. Under Review.

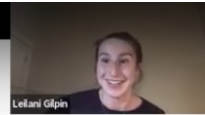
L.H. Gilpin MIT Explaining Explanations 28

Local sanity test to constantly explain what it is doing and learn from mistakes.

We need to make a NN explain itself.

Contributions and Future Work

- A taxonomy and best practices for explanations via completeness and interpretability
 - What [part or parts] is being explain?
- Future directions
 - How can a network explain itself?
 - How to incorporate explainable methods?
 - Is there a provable trade-off between completeness and interpretability?
 - What explanations are best suited for policy?
 - See our follow-up paper: "Explaining explanations to society"



Questions

Are human explanations complete or even accurate?

Can't peer into brain to check. But even if we don't know how we arrived at an explanation we can supplement these with commonsense. I feel hot and similar to how I felt when sick.

Can we construct a complete model of a research domain?

A. Globally this is hard and maybe philosophical. We get local views.

Q. A system that reaches a human interpretable explanation...are we losing somethings?

A. There is a tradeoff but it is open we don't know and don't have a base to test.

Q Do we look at counterfactual explanations? I didn't make a left turn. What would have happened?

A. Yes we are looking at that. Example, "You were denied because you didn't make a payment in 2012."

Q. Do Q and A systems handle counterfactuals?

A. Not yet.

Q what is an explanation is a lie? Say it reads a false billboard.

A. We based some on physics. But we need user feedback.

Q. What level of explanation is useful?

A. we have very verbose ones, but it is user sensitive. Some need terse answers.

Q. Is Ontological precision/validity used?

A, Mostly not. We have a traditional KB, haven't looked at faults. We do some as part of reasoning, some generalization from embeddings of words.

Q. Different approaches to handle open vs. open situations

CS520: Knowledge Graphs Seminar Session 7

Implementing Knowledge Graphs in Products

What are some prelevant graph engines in industry?

Philip Rathle (Neo4j) Property Graphs (Neo4j browser)and Graph Algorithms

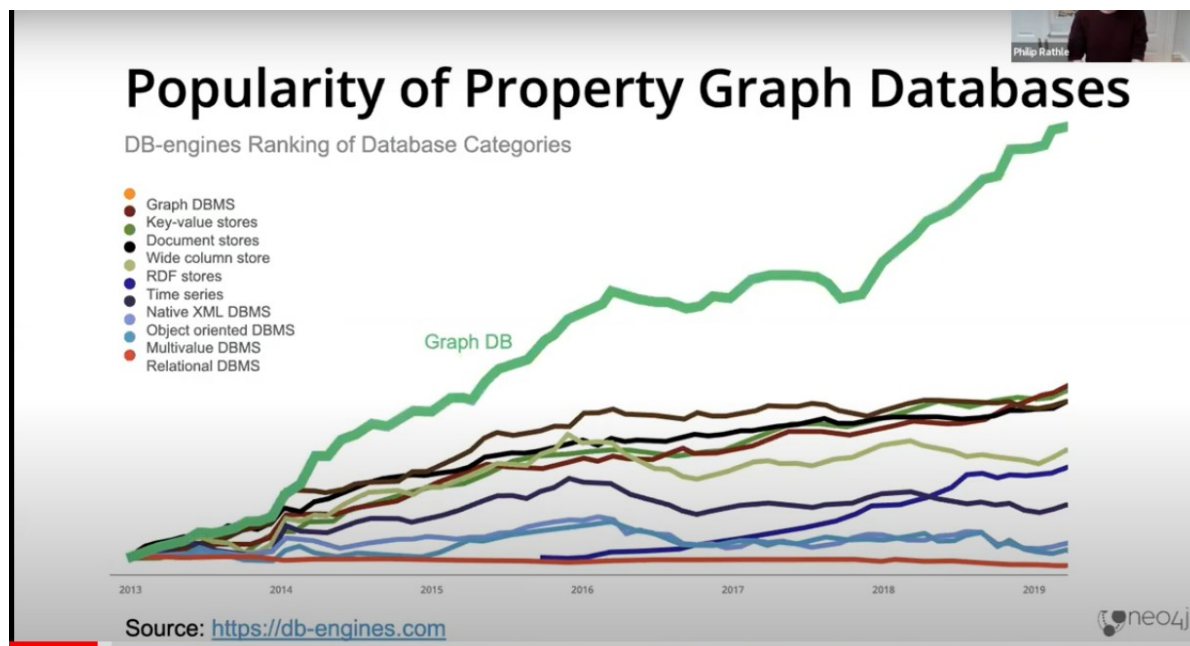
NASA and Ebay Conversational systems, or German Center for Diabetes Research (covidgraph.org), ICIJ on Panama papers of 2.6 TB of 11.5 million documents curated into a KG.

People consider many types of graphs to be a KG

10

<https://neo4j.com/graphconnect-2018/session-topics/?topic=Knowledge%20Graphs>

Evidence for property graph popularity – motivated by data storage & management interests, also querying and interest of developers and applications. This is a different vision than RDF, PG nodes and relationships have internal structure unlike with RDF.



Nodes
and

Relations are synonymous with vertex and edge.

In PG RDB rows become nodes and join tables become a relationship.

An open language based on patterns is Cypher.

CS520: Knowledge Graphs Seminar Session 7 (Spring 2020)

Press **Esc** to exit full screen

Philip Rathke

Query: Who loves this person named Ann?

SPARQL

```
prefix ms: <http://myschma.me/>
prefix rdf: <http://www[...]#>

SELECT ?who
{
  ?a rdf:type ms:Person .
  ?a ms:name ?asName .
  FILTER regex(?asName,'Ann')
  ?who ms:likes ?a .
}
```

Cypher

```
MATCH (who)-[:LOVES]->(a:Person)
WHERE a.name CONTAINS 'Ann'
RETURN who
```

neo4j

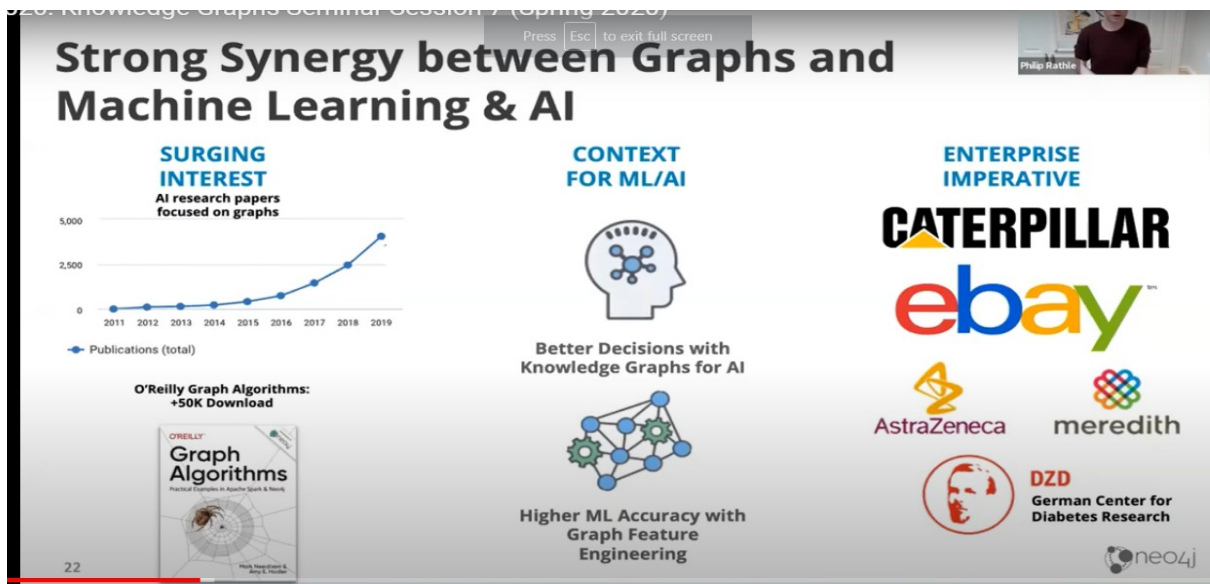
17:49 / 1:53:16

Scroll for details

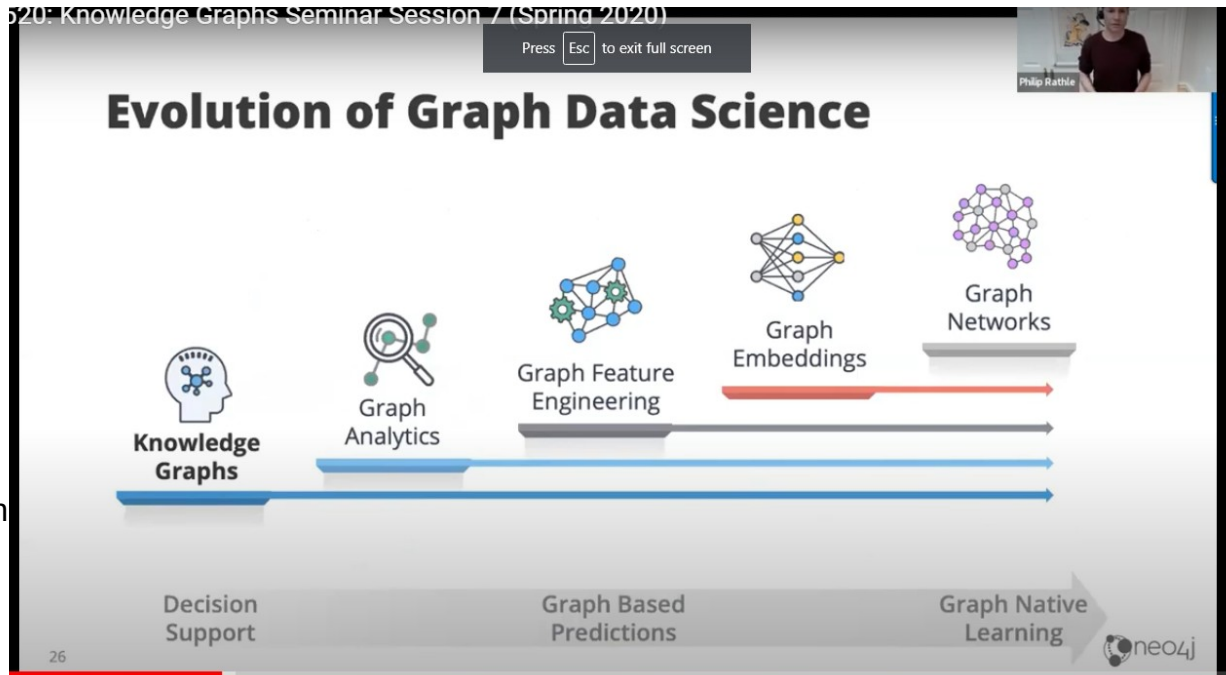
CC ⚙️ 🔍

Big space for this work between AI and KG – graph data scienc. We learn about a node in a network by the nodes around it.

[illegible]

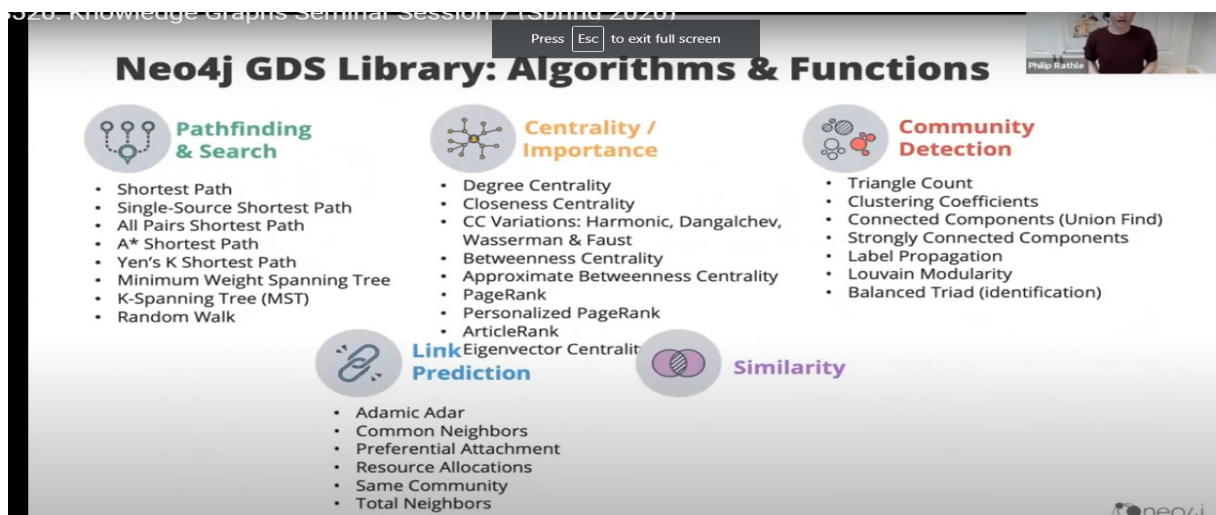


We see evolution in areas like graph feature engineering.



The graph

library with lots of algorithm and functions. There is an ecosystem see neo4j.com



A variety of customer use cases discussed with benchmark results for queries.

Brad Bebee Putting data into context with Amazon Neptune

It's just a graph so just make it so.



Background included 2014 Blazegraph which was picked by Wikipedia like other customers like linking things and for applications. You can make new relations easily.

Graphs make sense many places.


The Q becomes what type of system should I use to manage the data. SQL had problems expressing some things and for joins. RDBs aren't optimized for graphs. You often have to make big changes to add relations.

CS520: Knowledge Graphs Seminar Session 7 (Spring 2020)


Graph use cases



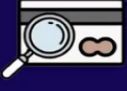
Social networking




Recommendations




Knowledge graphs



Fraud detection



Life Sciences



Network & IT operations

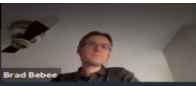
Connected data

- Navigate (variably) connected structure
- Filter or compute a result on the basis of the *strength*, *weight*, or *quality* of relationships

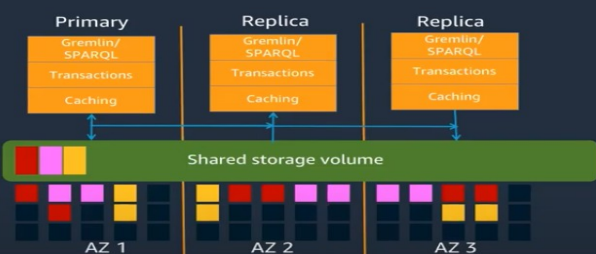
CS520: Knowledge Graphs Seminar Session 7 (Spring 2020)

Press **Esc** to exit full screen

Distributed storage architecture



- Performance, availability, durability
- Scale-out replica architecture
- Shared storage volume with 10-GB segments striped across hundreds of nodes
- Data are replicated 6 times across 3 AZs
- Hotspot rebalance, fast database recovery
- Log applicator embedded in storage layer



Delivered as a managed service

© 2020, Amazon Web Services, Inc. or its Affiliates.

42:53 / 1:53:16

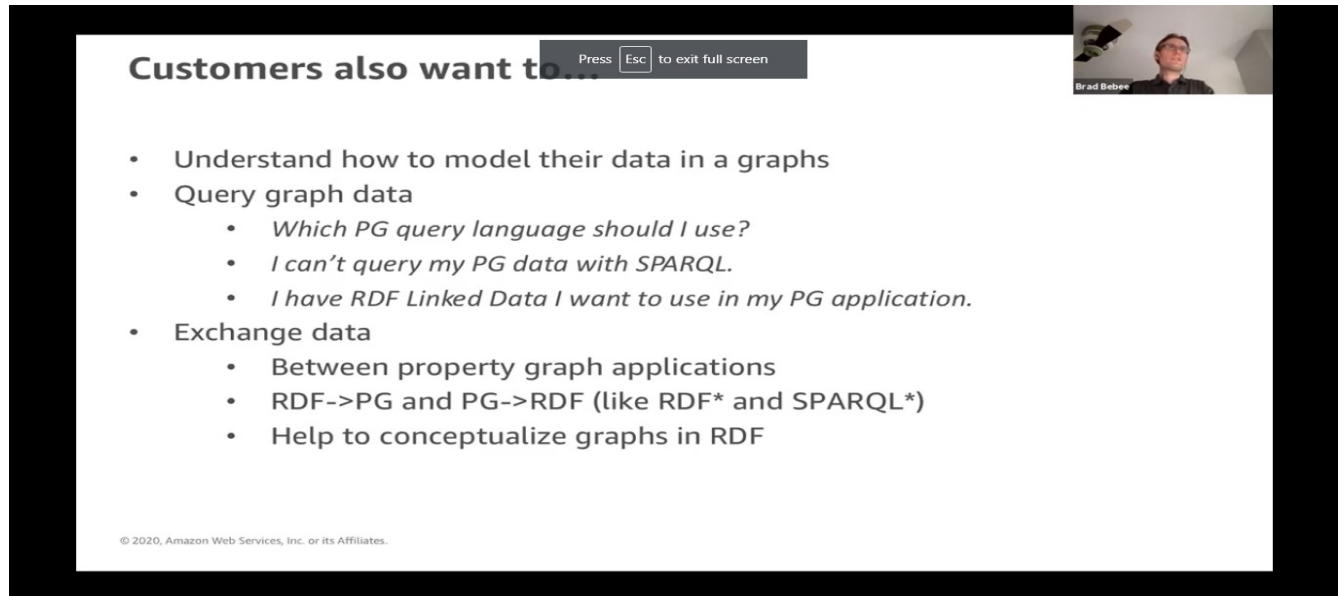
Scroll for details

But users are not typically thinking of different graph models like RDF or property graphs.

They want us to make this easier and not have to worry about the model (may affect tools?)

So they provide both with their architecture- they use a quad format (S-P-O, graph) to store data, They use several indices such as S-P-O, graph or S-P-graph-O to make retrieval fast.

What customers want:



Customers also want to...

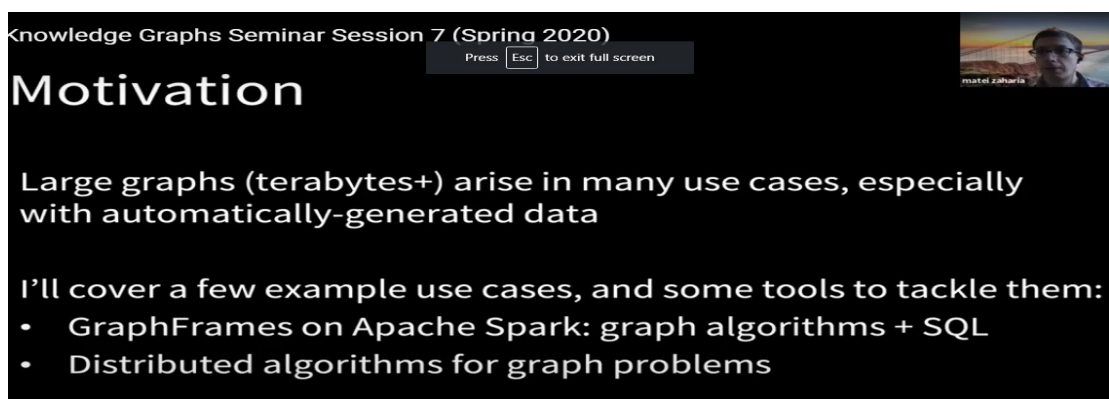
- Understand how to model their data in a graphs
- Query graph data
 - *Which PG query language should I use?*
 - *I can't query my PG data with SPARQL.*
 - *I have RDF Linked Data I want to use in my PG application.*
- Exchange data
 - Between property graph applications
 - RDF->PG and PG->RDF (like RDF* and SPARQL*)
 - Help to conceptualize graphs in RDF

© 2020, Amazon Web Services, Inc. or its Affiliates.

Question – what is the different PG or RDA and how to chose?

A. Not for customers but scale issues with billions of nodes. If building a graph standardization efforts will move on and build support for one language.

Matei Zaharia (open source databricks -platform on top of clouds) Large scale analytics with Apache Spark.



Knowledge Graphs Seminar Session 7 (Spring 2020)

Motivation

Large graphs (terabytes+) arise in many use cases, especially with automatically-generated data

I'll cover a few example use cases, and some tools to tackle them:

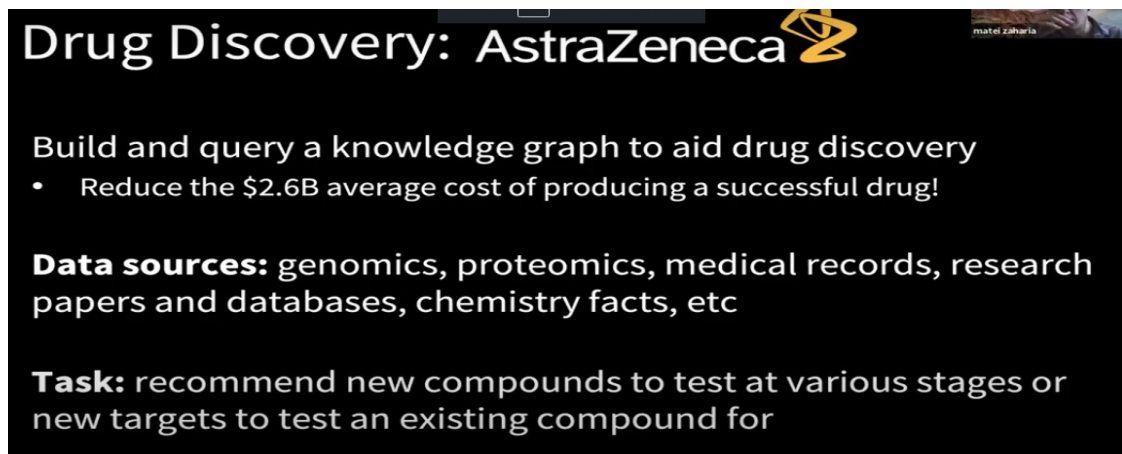
- GraphFrames on Apache Spark: graph algorithms + SQL
- Distributed algorithms for graph problems

Cool applications with analytic automated generated data.

Use cases: FINRA regulatory organization to detect illegal automated trades.

Large volume, saved data and look for patterns.

Drug discovery is another example:



Drug Discovery: AstraZeneca

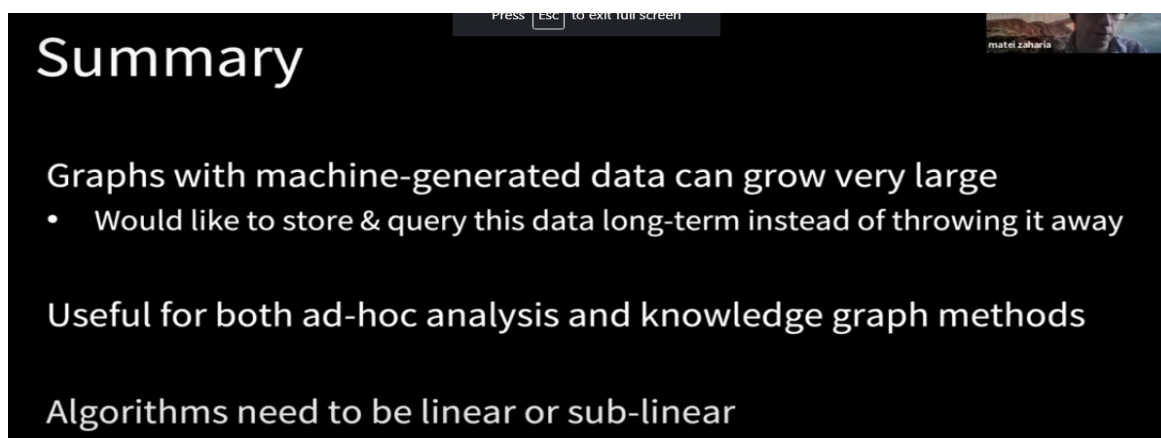
Build and query a knowledge graph to aid drug discovery

- Reduce the \$2.6B average cost of producing a successful drug!

Data sources: genomics, proteomics, medical records, research papers and databases, chemistry facts, etc

Task: recommend new compounds to test at various stages or new targets to test an existing compound for

Use case summary:



Summary

Graphs with machine-generated data can grow very large

- Would like to store & query this data long-term instead of throwing it away

Useful for both ad-hoc analysis and knowledge graph methods

Algorithms need to be linear or sub-linear

Tools & APIs – used GraphX now using GraphFrames package now. It implements graph computations on top of a SQL engine.

Table Fields provide info about a vertex or edge/relation (like a PG) Allows you to do SQL query.

Motif finding also allows Neo4J pattern matching using search for structures based on something similar to the Cypher language. Use together with filtering in airport flight example,

Questions – what is the role of ML in these use cases?

A. Often part of the workflow as needed by the client. Often this is part of the pattern match. A graph output can be sent to a ML application.

Neptune may use the entire KG to train a ML application.

Q of suitability of Dbs vs Kgs. Some differences on scaling maintenance and use of schema free approaches. Varies by application too differences in handling time-series where RDBs are good.

Q. Do you worry at the ontological level?

A Neptune sees a need for ontologies. They support triples from ontology for truth maintenance. It is a bit like hot sauce. A little is good but you can have too much that unbalances things.

Q Q what rules operate?

Neo4J stores an RDF version separately. Often the rules are hard to maintain.

Q. what types of visualization is used to represent a KG.

A. It is used to understand the KG. With no schema this (neo4j Browser) is helpful.

Visualizing certain workflows is helpful. Customers want to see the KG in the context of their analytic tool.

Q. What are the next set of features they are working on?

Neo4J just released a suite has reactive programming and security and works with Bloom.

Neptune wants to bring in AI and making it easier to use.

Apache SPARK wants low cost data versioning and better checking of data quality. Putting lots of effort into performance.

What is the role of knowledge graphs in machine learning?

Jure Leskovec Reasoning in KG using Embeddings

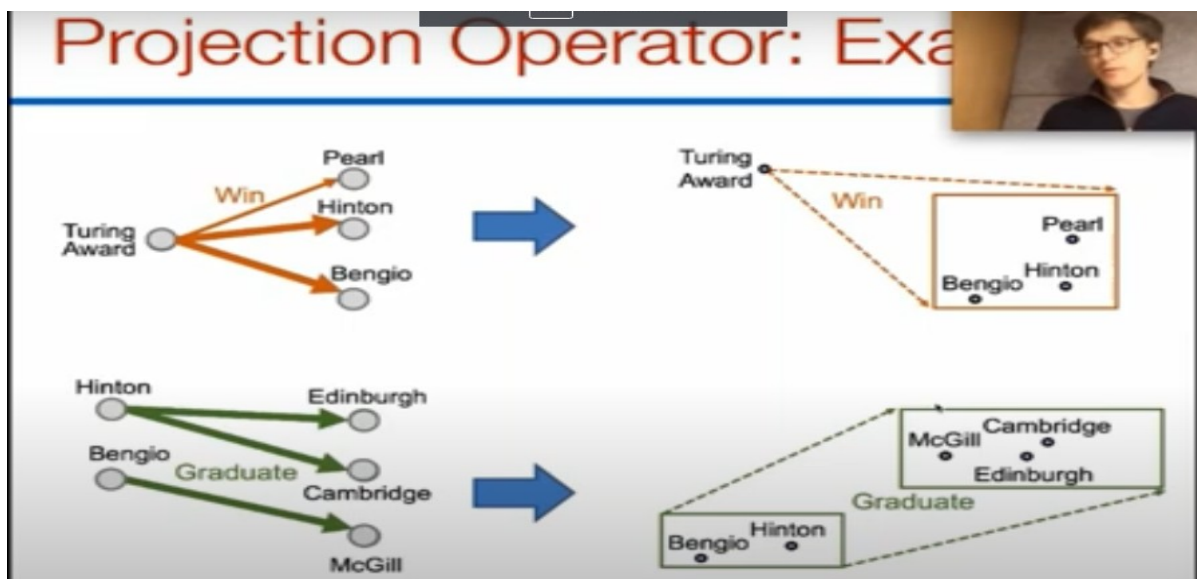
knowledge graphs are heterogeneous graphs where facts are represented as triples (h,r,t) but are often incomplete.

Can we use ML to complete them, say with linked predictions (but that takes time since matching doesn't scale) or multi-hop, logical queries?

But this is hard because a KG is large, noisy and incomplete, often with uncertainties and people want fast query times.

Solution to map the KG to an embedding space. For queries we translate to a spatial concept and do spatial search, spatial operators like overlap for a space close the query space.

They represent the query as a box and look for intersections using a projection operation.



Challenge handling arbitrary and/or queries- so it scales.

Graphs Seminar Session 8 (Spring 2020)

Press Esc to exit full screen

Disjunctive Normal Form

- Any query with AND and OR can be transformed into equivalent **Disjunctive Normal Form** (disjunction of conjunctive queries).

Original Computation Graph

Converted Computation Graph

June Leskovec, Stanford University

Training approach positive with answers in the KB and negative with random items.

Does the method generalize to hand new queries on missing info?

Train with at least one missing edge (incomplete knowledge) relation.

Test on Freebase

Query2Box: Summary

- Query2Box:**
 - Embed the query as a box
 - Logical operations become spatial operations
- Composability of queries:**
 - Generalize well to unseen, extrapolated queries
 - Explicitly training for composability is important
- Instance vs. multi-hop generalization

Future work is aimed at handling negation and uncertainty.


Also interested in developing a open graph benchmark DB to test ML with ML.

Interest is 3 sizes in science domain.

Press Esc to exit full screen


Open Graph Benchmark

- On-going effort for large-scale realistic benchmark datasets for graph ML.



OPEN GRAPH BENCHMARK

Webpage: <https://ogb.stanford.edu/>
Github: <https://github.com/snap-stanford/ogb>
Paper: <https://arxiv.org/abs/2005.00687>



Luna Dong (Amazon Product graph team) Is there a natural synergy of ML and KG?

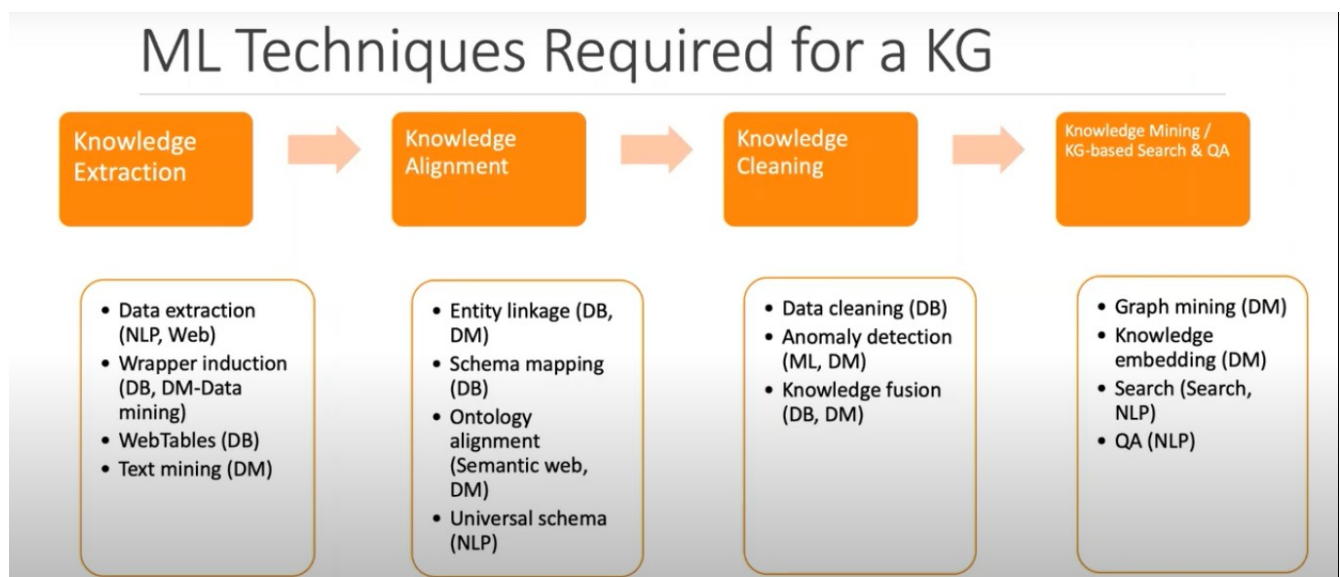
What is the Role of KG in ML

KG is a gold mine for ML but also a testbed

The more knowledgeable the more intelligent seems obvious and visa versa.

So knowledgeable returns better search results, can be an assistant, makes better suggestions,

So need to absorb K from sources (extraction), align them, clean up and mine them



10 challenges

1. Vast Unknown Space Size means you may not know all the types and relations beforehand, they change and how to organize them.

They use mining to help organize the taxonomy

2. Data is not well structured ,some is in titles.

Important info also in images. Do NER or data extraction.

3. Different ways that entities are identified heterogeneity. Need to align the data by exploring the neighborhood. Might do knowledge embedding and mapping,
4. So much data with different layouts– not just text, some semi-structured from Dbs.

How to extraction? Often hidden patterns such as headings, where relation is, attributes and dates,

CS520: Knowledge Graphs Seminar Session 8 (Spring 2020)

Press **Esc** to exit full screen

Luna Dong

Challenge 4: Data, Data, Everywhere

□ Key intuition: Hidden visual patterns across websites

- Multi-model information extraction
- GNN-based embedding
- OpenIE on semi-structured data
- WebTable extraction

노다지
A Bonanza (Nodaji)

1961년 · 대한민국 · 127분 · 1961-06-01 (개봉)

제작사 화성영화주식회사

감독 정창화

출연 김승호 황해, 엄영란 조미령, 허장강 더보기

스크랩하기

Lockard et al., Ceres: Distantly supervised relation extraction from the semi-structured web. VLDB, 2018.

5. What data is trustworthy? For cleanup look at redundancy, in same embedding space?

May try link prediction. Or try consistency of data fusion.

CS520: Knowledge Graphs Seminar Session 8 (Spring 2020)

Press **Esc** to exit full screen

Luna Dong

Challenge 6: A Working Solution May NOT Scale Up

□ Key intuition: one size fits all

- Conditioned attention in Neural Net
- Data augmentation

6. What if it doesn't scale, because each category uses a different model with different vocabularies.

7. training data for so many cases. A variety of learning approaches on labels

8. How to use a KG – find the fact in a KG? Entities may not be searched by name.

Can you find related items or do query refinement or involve user feedback.

9, Gaining insight from the graph facts. Do random walks

10. can you explain it? Interpretability. Still working on this

Take Aways

- ❑ KG is both a goldmine and a testbed for ML
- ❑ Building and using a KG involves tons of research problems, need techniques from many research areas
- ❑ At Amazon, we are building a Product Graph and using web knowledge extraction to populate Alexa knowledge

Robert Hoffman (IHMC) challenges for XAI

Emphasis on Psychological challenges of XAI.



State of the art??? Really challenges because we borrow terms from ordinary language or science to apply elsewhere - these may be orthogonal. It misleads many. - cross disciplinary confusion. (neural nets are not like neurons or brains)

Interpretations get used as explanations.

There is little deep investigations into explanations.

It requires more than a good interface.

The traps of Goofball science.

Explanation is a process not just a matter of diagrams or a statement.

The user needs to explore the solution and explanatory space. Why does it get things right or wrong?

AI systems make errors that humans would never make..due to CSK. They often don't have human concepts. And users often have different views than a designer.

Example of foot recognition and old wine in new bottles. KG is an example along with concept maps and intelligent tutoring systems. Need dialog more than just text.

Where is the knowledge in a bubble diagram?

Users have views:

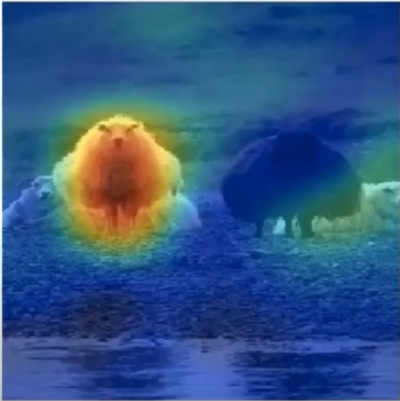
CS520: Knowledge Graphs Seminar Session 8 (Spring 2020)

Press **Esc** to exit full screen

Robert Hoffman

1. The State-of-the-Art

- Word Abuse
 - "Heat Maps"
 - "Neural Nets"
- Cognitive Anthropomorphism
 - "attention"
 - "saliency"
 - "seeing"
- Fuzzy/Goofy Concepts
 - Transparency
 - Simulatability
 - Semantic interpretability



A heatmap visualization of a cat's face, showing areas of high saliency (yellow/orange) and low saliency (blue/green). The cat is looking towards the left.

1:03:46 / 1:48:47

Scroll for details

Copyright R. Hoffman, May 2020. All rights reserved.

Triggers.

How does it work, what can it achieve?



3. Some Solutions: Types of Explanations



Global Explanations ask,
“How does it work?”

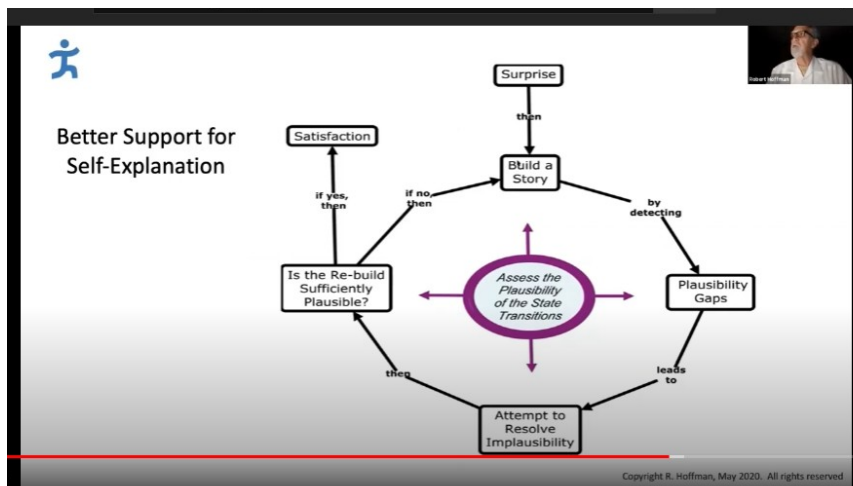
Global “Triggers” for Explanation

- How do I use it?
- How does it work?
- What do I do if it gets it wrong?
- How do I avoid the failure modes?
- What does it achieve? What can’t it achieve?
- How much effort will this take?

Copyright R. Hoffman, May 2020. All rights reserved

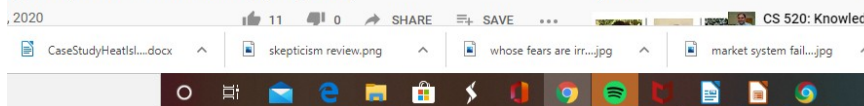
Are there solutions? Better support for self explanation using what ML provides as a mere clue or e idea of stories and plausibility for complex system. What an ML knows may be a metaphor-like thing.

List of solutions that really are just help -graphics etc.



dge Graphs Seminar Session 8 (Spring 2020)

Up next



Do we just need more data? Spoofing needed too.



3. Some Solutions

Press **Esc** to exit full screen



Spoofing as a Tool for Self-Explanation

Exploration by using filters in image processing tools

Crop out or camouflage regions or features that have a human semantics to them

“Method of Difference” (e.g., Bird + Pizza)

Performance on foils and filtered foils

Method of Maximum Distance (e.g., contrast sets)

Copyright R. Hoffman, May 2020. All rights reserved

He likes shifting the paradigm as a solution

Work context defines the task. (Cognitive work analysis)

And we need to involve the end user earlier in the process:



3. Some Solutions:

Work Systems Approach

Current Paradigm	A Better Paradigm
Researcher defines the task	The work context defines the user's activities
Researchers put themselves in the shoes of the user, and assume what the user needs	Cognitive work analysis reveals what the user really needs
The work is thought of as a set of tasks	In complex cognitive work systems, people do NOT conduct "tasks"
Researcher builds a model of a task	Cognitive work analysis leads to a model of the work
The task is considered independently of the work context	"Contextual design"

Copyright R. Hoffman, May 2020. All rights reserved

CS520: Knowledge Graphs Seminar Session 8 (Spring 2020)

Up next

AUTOPLAY ☐

447 views • May 20, 2020

11

0

SHARE

SAVE

...

CS 520: Knowledge Graph

CaseStudySeaLev...docx

CaseStudyHeatIsl...docx

skepticism review.png

whose fears are irr...jpg

market system fail...jpg

Show all

X

Search for anything


10:57 AM
6/16/2020

Need a more integrated approach and not a process control but one of interdependent computing.


Need more integration with experimental psychology.

The right vs wrong way to get Psych involved:

S520: Knowledge Graphs Seminar Session 8

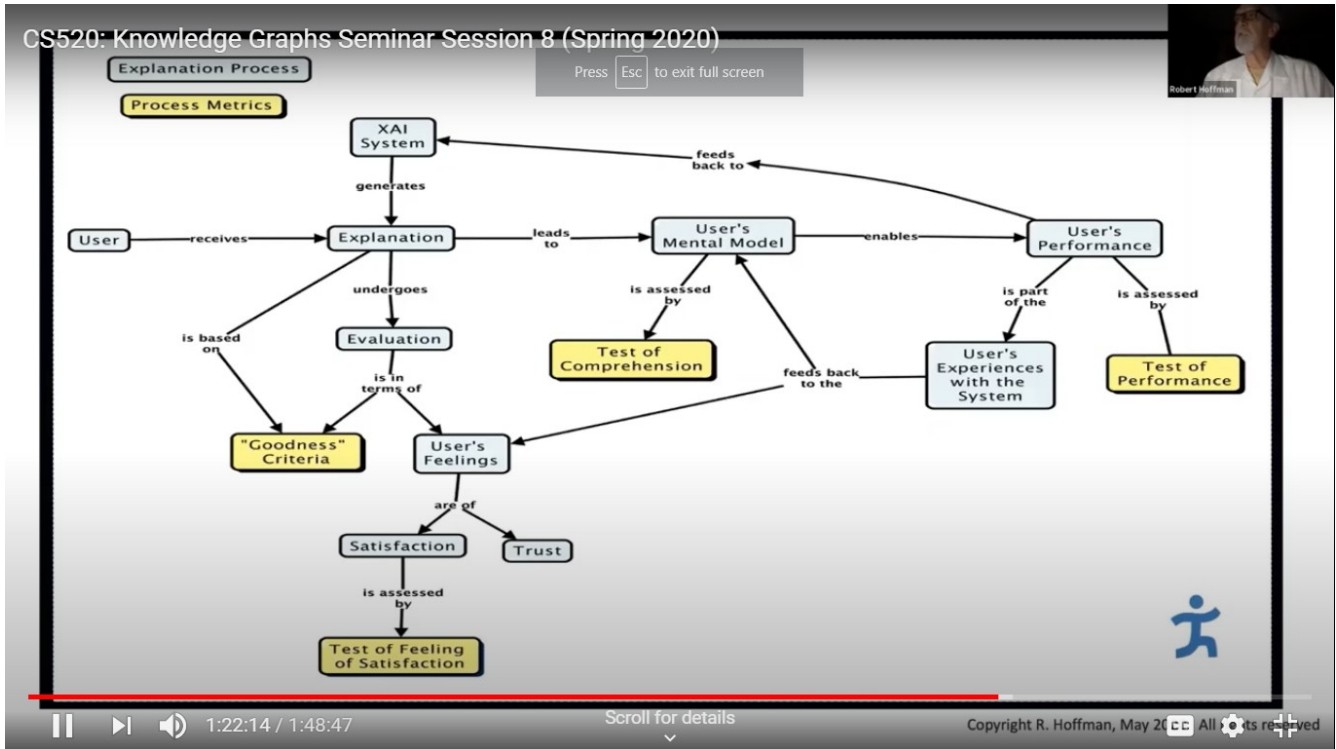


3. Some Solutions



WRONG WAY	RIGHT WAY
Large- n studies	Targeted small scale, small-n studies
Love affair with Mechanical Turk	You can find out more about user cognition and user requirements by one-on-one with 5 people than you can by collecting tons of numbers from 200 people
Complex factorial experiments	Start with simple pilot studies to test the research methods, materials, and procedures
Mindless statistics; worship of $p < .01$ Worship of Cohen's alpha, etc.	Regard statistical analysis as a set of exploratory tools, not a way of automating the process of scientific inference.
Search for the holy grail of "statistical significance"	Evaluate for <u>practical</u> significance
Assume that users digest and understand the explanations	Study user's mental models. Odds are the users will not think about any of this in the same way you do!
No control conditions	What should be controlled?

Task is low on ecological relevance	Tasks are highly ecologically relevant, and ecologically salient
Little or no user testing or usability testing; what testing there is, is not scientifically robust	Tests for understandability, learnability usability, usefulness, and trustworthiness are a part of the design process
System is evaluated using CS analytic methods only (e.g., Monte Carlo runs to determine model fits)	Experimentation is in the authentic work context; actual user learning and performance are empirically investigated
To-be-created system is thought to be a "decision aid" but ends up really being a process control system	Human-AI is an interdependence relationship; The AI complements how humans think (Human-Centered Computing)

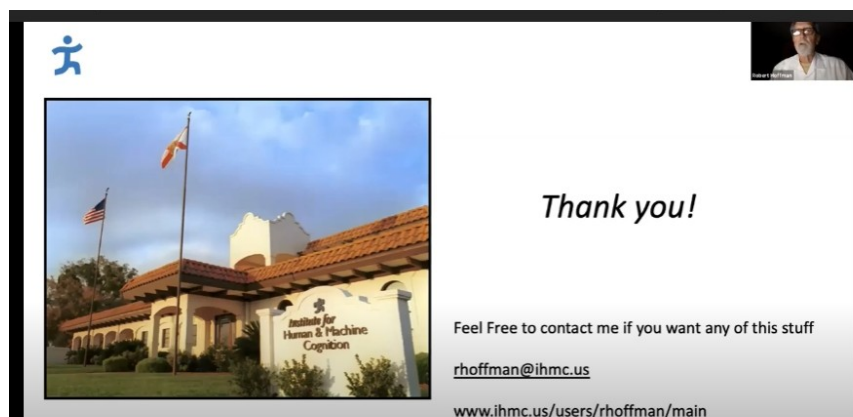


DARPA XAI approach.

Questions

We are trying to build a conceptual map, but can't do it an edge at the time...

A. What we are talking about is more like John Sowa's concept diagrams/maps.



Q for Luna – trying to extract from things loosely designed. Can you ask for more cooperation and get more structure?

A, Yes, but we would get more abuse to make products stand out. They may provide small misinformation. We do want to get sellers and buyers engaged to describe better products.

Q How do you decide what is a new product?

A. We have entity linkage system and use matching even images. We ask sellers to look first at product types.(self driving curation too._

Q. What is the role of explanation in Amazon system?

A. It can improve recommendations, relevant but perhaps different.

Q What do you think about realist ontologies like BFO? Do the high-level distinctions play a role

A. In Amazon they thought it an easy part as a backbone, but much more semantics to it.

Q Jure- What do you do when an answer doesn't make sense to the user?

A. The model isn't perfect. It is statistical.

Q for Robert, as a Psychologist, what problem can we help solve...we have been doing this for a while.

A. It depends on what the project defines as a problem. How do you measure making explanations better? It might be help with user terminology which is often mentalistic.

Q. Will your KG ever be done?

A. Some think it a 100 year project...it is hard, but is why they are working on a self driving approach to add new types by itself...includes a human in the loop.

The DARPA rapid knowledge project ran into problems with self guided approaches. Cyc may also be considered a 100 year project that ran into hard problems.

S520: Knowledge Graphs Seminar Session 9

What are some high value use cases of knowledge graphs?

Jay Yu Intuit KG Uses case

Intuit has an AI driven platform to provide good financial advice and explanation.

Intuit explains why it is engaging intelligent assistant how it came up with a return using KG for reasoning

The screenshot displays the TurboTax Assistant interface. On the left, a sidebar shows navigation options: Tax Home, Documents, 2019 TAXES, My Info, Federal, State, Review, File, Upgrade, and Tax Tools. The main content area is titled "Wages and income" and shows a summary of tax calculations. It indicates a Federal Tax Due of \$12,743 and a CA Tax Due of \$8,171. Below this, a table lists income sources: "Job (W-2) INTUIT INC" with a value of \$137,483.00, and "Interest on 1099-INT Wells Fargo" with a value of \$1,299.00. An "Add more income" button is at the bottom left. On the right, a "TurboTax Assistant" chat window is open, displaying a conversation. The assistant explains the \$12,743.00 tax liability, breaking it down from the total tax owed (\$24,554) minus credits (\$9,811) and other factors. It offers further assistance with buttons like "Explain why I owe the IRS \$12743.0", "Explain why I have to pay to file my taxes with TurboTax", and "Explain why I owe the IRS".

Category	Amount
FEDERAL TAX DUE (in progress)	\$12,743
CA TAX DUE (in progress)	\$8,171

Income Source	Amount
Job (W-2) INTUIT INC	\$137,483.00
Interest on 1099-INT Wells Fargo	\$1,299.00

AI and KG have their own role and KG is the foundation of their data.

How they use a KG...the US tax code is difficult so there is lots to encode supporting thousands of calculations. An example is for eligibility.

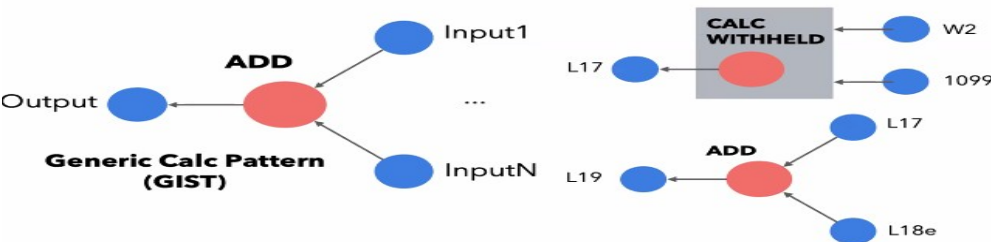
Limits of procedural approach -be explicit so can provide explanations.

Be declarative and you can compose things into a larger graph.

Develop patterns for the numerous calculations. These are generic patterns for calculations that can be stitched together. An analysis of the graphs allows you to explain the calculation. You can reason over the graph and identify if information is missing.

Knowledge Graph approach

17	Federal income tax withheld from Forms W-2 and 1099	
19	Add lines 17 and 18e. These are your total payments	
20	If line 19 is more than line 16, subtract line 16 from line 19. This is the amount you overpaid	



intuit

PATENTED INTUIT INNOVATION

(c)2020 Intuit Inc. All rights reserved. 14

Several examples of how they build their KG and why it is better than a procedural approach. As discussed by others they use various approaches to merge data and support entity resolution which might need a new relation.

The allow users to provide feedback

Comparison between two approaches

	PROCEDURAL	KNOWLEDGE GRAPH
Logic	Procedural	Declarative
Execution	Sequential	Dependency-driven
Modularity	By best practice	By design
Explainability	Blackbox	Built-in
Personalization	Hard-coded, limited	Dynamic, automatic
Automation	Not manageable	Manageable
Testability	Harder	Easier
Optimized for	Exceptional complex logic	Most tax logic
Target Developer	Engineers	Domain experts

intuit

(c)2020 Intuit Inc. All rights reserved. 20

Key takeaways

- Knowledge Graph is a **natural fit** for many use cases.
- Knowledge Graph can be used to **model logic (code)**, beyond data.
- Knowledge Graph can be **automatically created/enriched via AI**.
- Knowledge Graph **makes Intuit products smarter** with tangible customer benefits: **More Money, No Work** and **Complete Confidence**.

Intuit has been leveraging and innovating on Knowledge Graphs to bring tangible benefits to our customers!

intuit

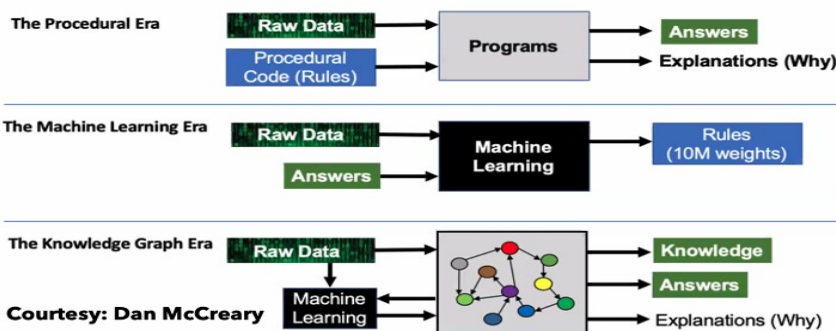
(c)2020 Intuit Inc. All rights reserved. 35

How to think about Knowledge Graphs.

Zoom Webinar

Recording

Food for thought: KGs are the core of third era of computing



"Knowledge Graphs are at the core of the third era of computing where we use machine learning to continuously enrich our shared knowledge"

<https://medium.com/@dmccreary/knowledge-graphs-the-third-era-of-computing-a8106f343450>

intuit

(c)2020 Intuit Inc. All rights reserved. 36

Apoorv Saxena(JP Morgan) Application of Knowledge Graphs

A few use cases and challenges. Finance and Banking. KG used for fraud and risk. The KG allows them to see how fraud clusters grow over time.

We see how a KG is used to answer questions in this Table. Includes recommendation systems.

Applications of Knowledge Graphs | May 2020

Use case I: Company Knowledge Graph

The types of questions we are trying to answer through the Company Knowledge Graph:

- **Link traversal:** "If Boeing goes into financial troubles, who are their vendors and suppliers? Which of them are JPMC clients? Are any of them applying for a loan? How much percentage of their revenue comes from Boeing?"
- **PageRank application:** "Which startups have attracted the most influential investors?"
- **Weak link detection:** "Which company is a weak link in the Airline industry supply chain? Is there a single company that connects groups of companies in the supply chain?"
- **Graph embedding & Node similarity:** "Which nodes are most similar to Okta?"
- **Link prediction:** "Which companies might have relationships in the future?"
- **Community detection:** "Which set of investors co-invest with a high degree?"

Problems are similar to previously described – integrating various sources of data and the problem of unique names, NER and extracting information. Dynamic data caused challenges of temporal nature, authoritative sources and keeping up to date.

There are regulations on access and who can see what.

Many business users are not sophisticated so they have trouble with queries with Ciper. They use NL queries that get converted using BERT.

Applications of Knowledge Graphs | May 2020

Lessons learned on building knowledge Systems

- Knowledge is with the business / domain owners and not the data scientist / engineer – Need to build tools for the domain expert (often non-technical staff) and not for the data engineer / scientists
- Knowledge capture is an iterative process and not one time step
- Need to come up with 1or 2 key metrics to optimize knowledge system performance
- Building good interpretability and easy feedback is key to building trust
- Good training data beats better models, reducing training data cost is key to success in the long run

David Newman Wells Fargo High Value Use Cases (also FIBO)

Talking more conceptually than operationally.

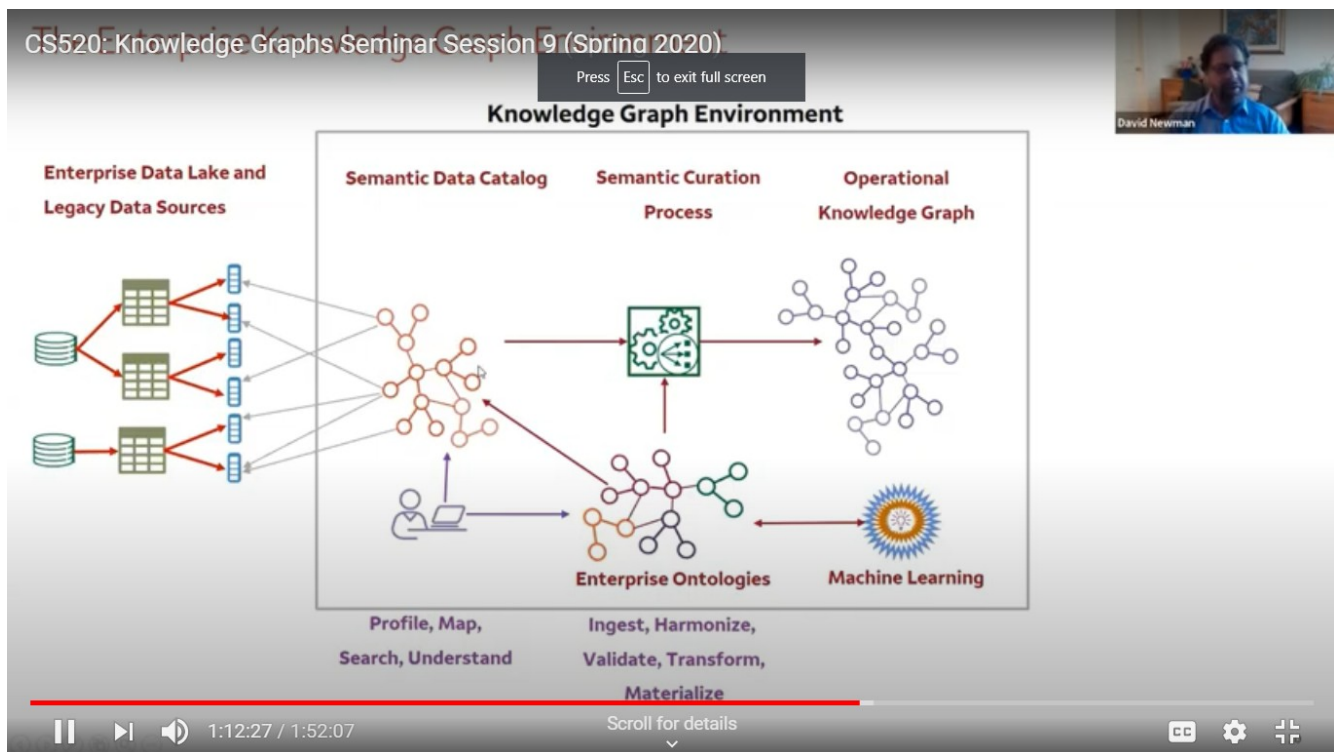
KG as a paradigm shift. How to improve our metadata and reduce data wrangling.

How to handle data silos? The data paradigm is old and we need to move on from the high costs and risks. KG provides a strategic solutions with a common semantic data

model understandable by humans and machines.

There is KG trajectory from data to semantic web now to KG 3.0 with machine understanding using ML.

Groups like EDM provide a start on data for open, reusable enterprise KGs – also usable as a



reference model since they are based on an enterprise ontology.

The idea of a semantic data catalog to help semantic search and queries for understandings.

There is a methodology that goes along with this using ML for link prediction, resolution, embedding

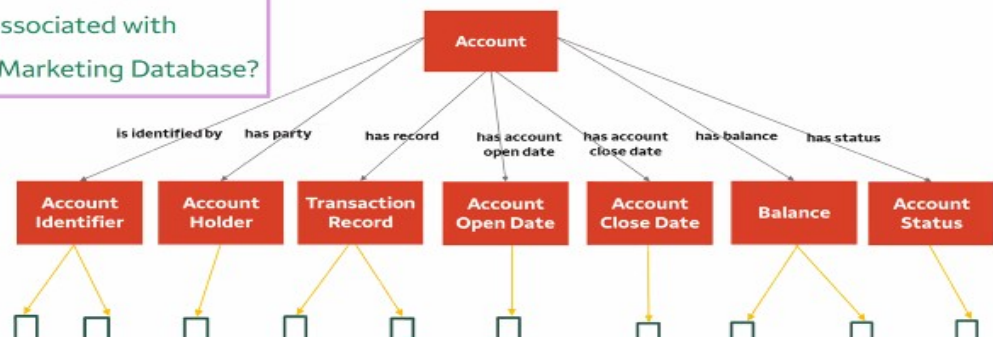
A Semantic Data Catalog is an Important Capability within an Enterprise Knowledge Graph Environment

What are Key Features of a Semantic Data Catalog?

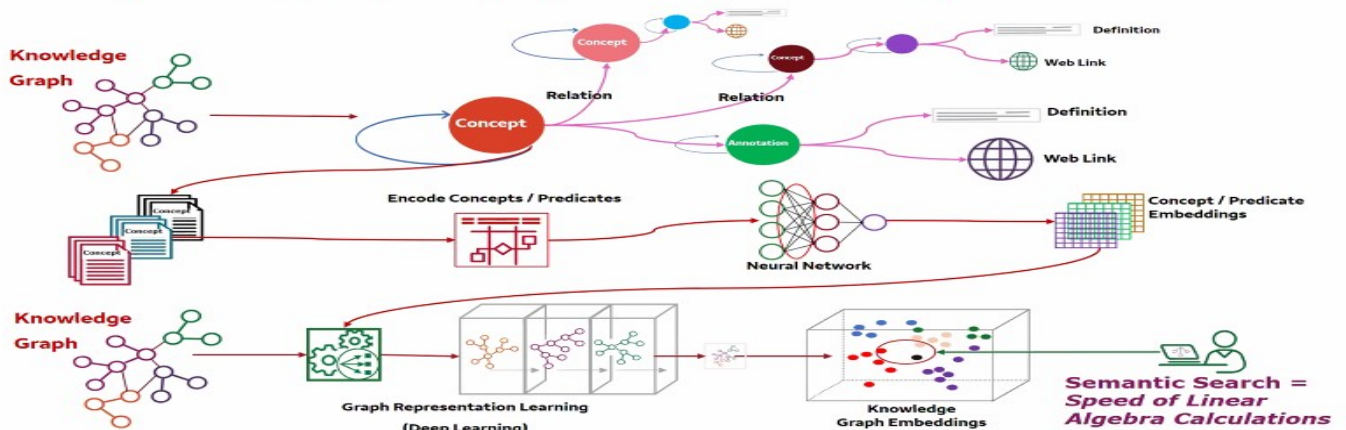


Concepts to Data Assets Capability

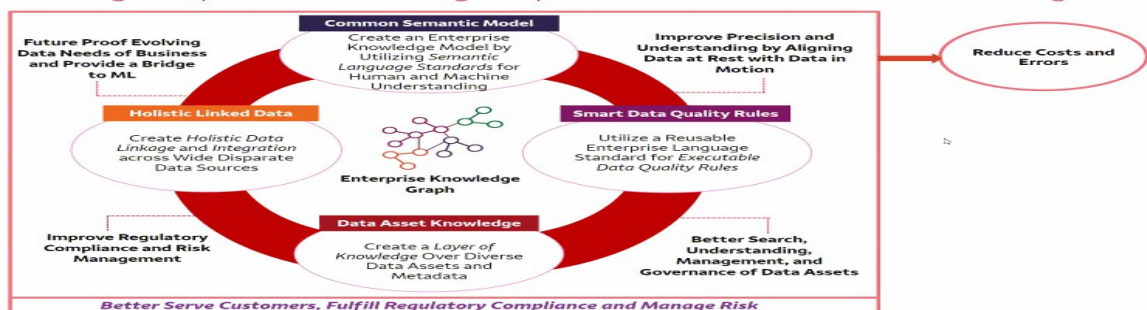
User Question: What physical elements are associated with *account* in the Marketing Database?



Training Knowledge Graph Language and Graph Embeddings for Semantic Search



Knowledge Graph Provides Strategic Capabilities to Remediate These Challenges



Questions:

How are you three groups sharing data?

A. it is the common data model such as the EDM council is discussing with others.

There are standards for sharing some of this. Some governments provide the standard such as what “money” means. KG will be the next step. We need education for users to understand a KG more.

Embeddings can also be used to guide this sharing.

Q On the idea of a computation graph. The level is at the same as the computation itself. The explanation of a computation can be so complex as to not be understandable.

A. Right. We use simple examples. The IRS publishes regulations for things in XML schema, but eventually the IRS will have to define this by say a graph.

Q The future?

A. FIBO will be crowdsourced supported by a community process.

A Jay, the future of KB is good together with ML. We need to make these open and understandable.

Apoorv Kbs will provide a base for ML.

CS 520: Knowledge Graphs Seminar Session 10

What are some open research questions on knowledge graphs (discussed during chat at June 10th session)

Richard Socher Multi-Hop Knowledge Graph Reasoning with Deep Reinforcement Learning (1 part of the future)

[Slides](#)

Real world KG are important. But in the long term, how much can they capture? They are

incomplete (who is a leader? Changes over time.) What about hypothetical?

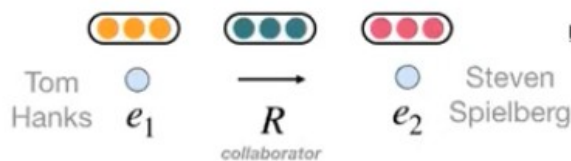
Can be useful for Covid-19 and products, orders.

We need to be more complete in filling in connections and reinforcement learning can help.

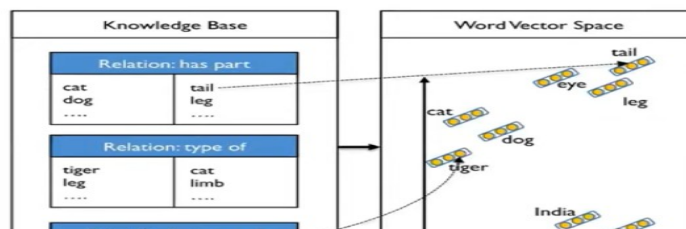
We need interpretable and complete, accurate models.

Simple and efficient Approach – embed entities as vectors into a NN architecture. You train on these.

Knowledge Graph Embeddings



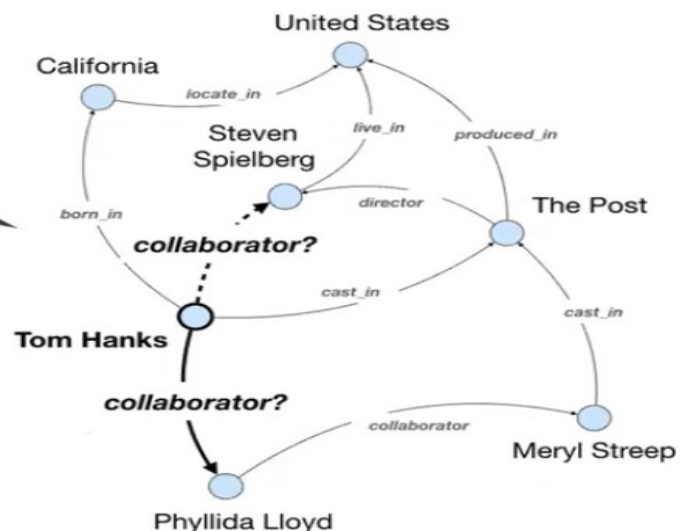
Neural Tensor Network



Contribution 2: Compositional Representation of Knowledge Base Entities

Multi-Hop Reasoning Models

Reasoning over discrete structures



Which directors has Tom Hanks collaborated with?

Interpretable with inference w/o lots of feature analysis

CS 520: Knowledge Graphs Seminar Session 10 (Spring 2020)

Interpretable Results

2019 Knowledge Graph Conference
Knowledge Graphs and AI: The Future of Financial Data
COLUMBIA SPS
20:07

NEXT (SHIFT+N)
Knowledge Graphs and AI: The Future of
Reasoning with Reward Shaping (Lin et. al. 2018)

47

15:41 / 1:50:45

Scroll for details

<https://www.youtube.com/watch?v=25UlgiiYqsE>

Key takeaways

Take-aways

- Practical knowledge graphs are incomplete and require automatic completion
- KG embeddings are effective approaches for recovering missing facts, but lack interpretability
- Multi-hop KG inference is interpretable but less accurate
- Multi-hop KG inference with embedding-based reward shaping combines the best of both approaches
- The method can potentially be extended to multi-hop reasoning over joint KG and text to better support downstream AI systems

Mark Musen (Stanford) What Do Knowledge Graphs Really Know?

Slides

Since the earliest days of AI, there has been a struggle to define what we really mean by knowledge and how to determine when a computational agent displays intelligence. Allen Newell had considerable influence in asking the AI community to think about knowledge in behavioral terms. He argued that knowledge representations do not manifest intelligence until some process comes along and operates on them. In that perspective, knowledge graphs are not intrinsically knowledgeable at all. We need agents to do something with our graphs to cause intelligent actions. With all the excitement about knowledge graphs, we've been focussing intently on the graphs. In the next round of research, we need to be thinking more about the knowledge.

Knowledge graphs give us the ability to represent unimaginable numbers of entities in the world and the relationships among them.

It's time to apply the same kind of principled thinking that has led us to the development of knowledge graphs to the construction of intelligent agents that can demonstrate, through their behaviors, what our graphs really know.

RV Guha Title: DataCommons

Slides

Publicly available data from open sources are a vital resource for students and researchers in a variety of disciplines. Unfortunately, processing these datasets to make them useful --- scraping, cleaning, normalizing, joining --- is tedious, error prone and has to be repeated by every group. DataCommons attempts to alleviate some of this pain by synthesizing a single Knowledge Graph from many different data sources. It links references to the same entities (such as cities, counties, organizations, etc.) across different datasets to nodes on the graph, so that users can access data about a particular entity aggregated from different sources. Like the Web, the DataCommons graph is open - any user can contribute data or build applications powered by the graph. We are jump-

starting the graph with data from publicly available sources such as CDC, Census, BLS, FBI, etc. and are looking to engage with the academic community to take it further.

Commonsense idea of KG

Biomedical Data Commons Datasets

- CDC - 500 Cities	- Disease Ontology*	- SIDER*
- CDC - Diabetes Atlas	- FDA - Pharmacologic Class*	- SPOKE*
- CDC - Wonder	- GTEx	- UCSC Genome Browser
- ChEMBL*	- ENCODE	- US Census - SAHIE
- ClinVar	- Entrez Gene	- UniProt*
- dbSNP	- MeSH*	- WHO - COVID-19 Cases + Deaths
- Dartmouth Medicare Atlas	- NY Times - COVID-19 Cases + Deaths	- WHO - ICD-10 Codes

*Source: UCSF SPOKE

Ending Thoughts

Ending thoughts

- Vocabulary creep: Google KG, Cyc, Wikidata all have many tens of thousands of 'schema' terms. Human language manages with few thousand terms. How do we bring the compositionality of NL to KR?

I

- The problems of old AI haven't been solved. They just can't be expressed clearly in today's ML formalisms. e.g., World's tallest mountain yesterday?

Questions and discussion

Language models can generate ahead.

ML and NLP can capture commonsense and generalize beyond what is read.