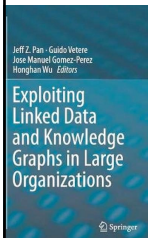


INFR11215 Knowledge Graphs

Knowledge Graph Construction

Jeff Z. Pan

<http://knowledge-representation.org/j.z.pan/>



[Reading: Pan et al., Sections 4.1 and 4.2, as well as 2.2, 2.3 and 3.2]

1

A Note on Tutorial 1

- Question 1:
 1. Judge if it is a basic fact or a complex fact (complex facts always have some unknown/unspecified entities)
 2. If it is a complex fact, decide whether it is pattern 1 (about some) or pattern 2 (about all)
- Question 2:
 1. Transform the statements into DL syntax
 2. Check what needs to be proved
 3. Check what reasoning service is needed; in this case, we need Instance Checking --- Mike : MountainClimber

2

A Note on Tutorial 1

- “Mike dislikes whatever Tony likes and likes whatever Tony dislikes”:
 1. Mike dislikes whatever Tony likes
 1. $\exists \text{like}^-. \{ \text{Tony} \} \sqsubseteq \neg \exists \text{like}^-. \{ \text{Mike} \}$
 2. And (Mike) likes whatever Tony dislikes
 1. $\neg \exists \text{like}^-. \{ \text{Tony} \} \sqsubseteq \exists \text{like}^-. \{ \text{Mike} \}$
 3. Axiom (1.1) can be rewritten into
 1. $\exists \text{like}^-. \{ \text{Mike} \} \sqsubseteq \neg \exists \text{like}^-. \{ \text{Tony} \}$
 4. Axioms (2.1) and (3.1) together are equivalent to the following axiom:
 1. $\exists \text{like}^-. \{ \text{Mike} \} \equiv \neg \exists \text{like}^-. \{ \text{Tony} \}$

Lecture Outline

- Motivation
- Overview of KGs Construction
- Detailed Discussions on Competency Questions
- Practical

THE UNIVERSITY of EDINBURGH


Knowledge Graph

- **Knowledge Graph (KG):** KB of inter-connected entities with vocabulary defined by a schema
- **Three levels of knowledge**
 - Entity
 - Triple (unit of a multi-relational graph)
 - Schema (defining the vocabulary, making KGs different from other graphs)

[Photo: <https://www.w3.org/TR/rdf11-primer/>]

Knowledge Graphs
Jeff Z. Pan





Mona Lisa
Painting by Leonardo da Vinci



The Mona Lisa is a half-length portrait painting by Italian artist Leonardo da Vinci. Considered an archetypal masterpiece of the Italian Renaissance, it has been described as "the best known, the most visited, the most written about, the most sung about, the most parodied work of art in the world". Wikipedia

Artist: [Leonardo da Vinci](#)
Dimensions: 77 cm x 53 cm
Location: [Louvre Museum](#) (since 1797)
Created: 1503
Subject: [Lisa del Giocondo](#)
Period: [Renaissance](#)
Medium: [Oil Paint](#)

People also search for [View 10+ more](#)

   
The Last Supper Leonardo da Vinci
Girl with a Pearl Earring Johannes Vermeer
Salvator Mundi Leonardo da Vinci
Lady with an Ermin Leonardo da Vinci

5

THE UNIVERSITY of EDINBURGH

Freebase

freebase™
alpha

Alan Turing : Freebase

Keyword search Freebase Search

Home Data Apps Discuss Help | Welcome back, finin. Not you? Sign out.

Alan Turing

Discuss "Alan Turing" Show Empty Fields

- Types: Person (People), Book Subject (Publishing), Computer Scientist (Computers), Deceased Person (People), Influence Node (mikelove's types), Computer Designer (Computers)
- Date of death: Jun 7, 1954
- Place of death: Manchester, England
- Cause of death: Cyanide Poisoning, Suicide
- Gender: Male
- Date of Birth: Jun 23, 1912
- Place of Birth: London
- Profession: Computer Scientist, Mathematician, Scientist, Philosopher
- Religion: Atheism
- Parents: Ethel Stony Turing, Julius Mathison Turing

Description

Alan Mathison Turing, OBE, FRS (23 June 1912 – 7 June 1954) was an English mathematician, logician, and cryptographer. Turing is often considered to be the father of modern computer science. Turing provided an influential formalisation of the concept of the algorithm and

Page History
Created by Metaweb Oct 22, 2006
Last edited by mikelove Nov 1, 2007

Books About This Topic
Cryptonomicon

Education
Princeton University • 1937 • 1938 • Ph.D.
King's College, Cambridge • 1931 • 1934

Quotations
Mathematical reasoning may be regarded...
No, I'm not interested in developing a powerful brain...
Science is a differential equation.
Religion is a boundary condition.
We can only see a short distance ahead, but we can see plenty there that needs to be done.

6

Freebase and Google

Google Acquires Metaweb To Make Search Smarter

Posted Jul 16, 2010 by [Leena Rao \(@leenarao\)](#)



Next Story

Google has **bought** semantic search startup **Metaweb**, according to recent post on the search giant's blog. Terms of the deal were not disclosed.



Metaweb develops both semantic data storage infrastructure for the web, and **Freebase**, an "open, shared database of the world's knowledge". Freebase is a massive, collaboratively edited database of cross-linked data. The idea behind the product is to create a system for building the semantic web. Freebase allows anyone to contribute, structure, search, copy and use data. It sounds like Wikipedia, but instead of arranging by articles, it is more of an almanac, organized like a database, and readable by people and software. You can read our previous coverage of Freebase [here](#).

Clearly, Google is acquiring Metaweb to boost its own search offerings. Metaweb's database of tagged data will help make Google search smarter. And Freebase will be maintained as a

AdChoices

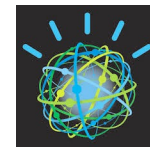
NEWSLETTER SUBSCRIPTIONS

Knowledge Graphs
Jeff Z. Pan

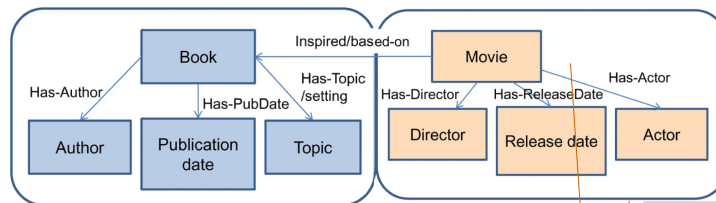
7

7

Question Answering in IBM Watson



A 1992 movie starring Anthony Hopkins was based on an 1897 book by whom?



And, interlinking frames

- About a billion frames (shallow knowledge graphs) were used in Watson Jeopardy!:

Bram Stoker's famous book on vampires was published in which year?

Knowledge Graphs
Jeff Z. Pan

8

8

THE UNIVERSITY of EDINBURGH

KG Use Cases

IBM | Cloud | Products | Solutions | Pricing | Docs | Support | Explore more

- What is a knowledge graph?
- How a knowledge graph works
- Use cases of knowledge graphs
- Related solutions
- Resources
- Take the next step

- Retail:** Knowledge graphs have been for up-sell and cross-sell strategies, recommending products based on individual purchase behavior and popular purchase trends across demographic groups.
- Entertainment:** Knowledge graphs are also leveraged for artificial intelligence (AI) based recommendation engines for content platforms, like Netflix, SEO, or social media. Based on click and other online engagement behaviors, these providers recommend new content for users to read or watch.
- Finance:** This technology has also been used for know-your-customer (KYC) and anti-money laundering initiatives within the finance industry. They assist in financial crime prevention and investigation, allowing banking institutions to understand the flow of money across their clientele and identify noncompliant customers.
- Healthcare:** Knowledge graphs are also benefiting the healthcare industry by organizing and categorizing relationships within medical research. This information assists providers by validating diagnoses and identifying treatment plans based on individual needs.

[Credit: <https://www.ibm.com/topics/knowledge-graph#Use+cases+of+knowledge+graphs>]

<https://opencourse.inf.ed.ac.uk/kg/resource-list>

Knowledge Graphs
Jeff Z. Pan

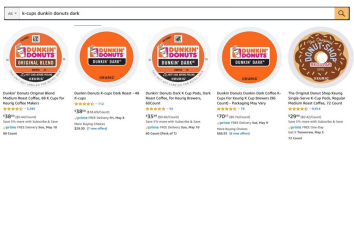

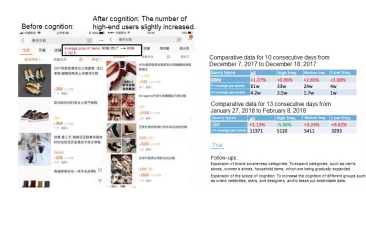
9

9

THE UNIVERSITY of EDINBURGH

Product Knowledge Graphs (PKG)

- Many Companies produced PKGs:**
 - Amazon (2018)
 - Walmart (2019)
 - Jingdong (2019)
 - Microsoft (2019)
 - Alibaba (2020)
- Key applications**
 - Improving search
 - Improving recommendation
 - Question answering
 - Shopping guide
 - Product classification
 - Product completion

Knowledge Graphs
Jeff Z. Pan

10

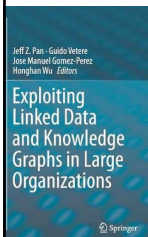
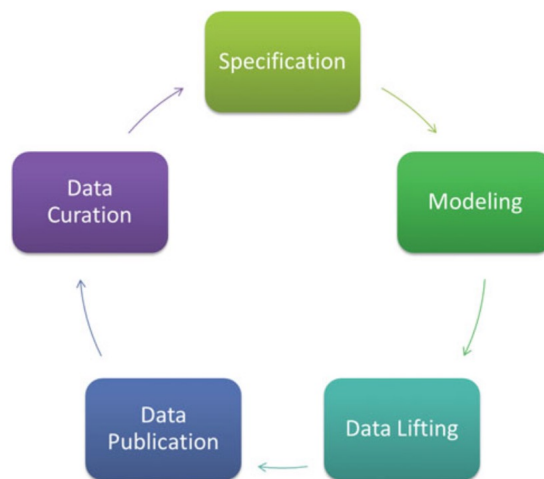
10

Lecture Outline

- Motivation
- Overview of KG Construction
- Detailed Discussions on Data Lifting

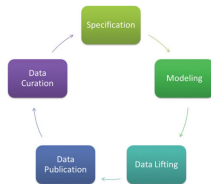
11

Knowledge Construction and Maintenance (KCM) Lifecycle



12

KCM Lifecycle: Specification



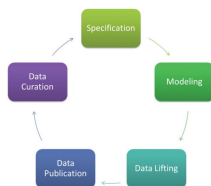
- Requirement: **competency questions**
 - for specifying what the target KG should cover
- Data source identification and documentation
 - internal vs external
- Data source analysis
 - identify entities and relations covered
 - identify their schema
- URI/IRI design
 - use meaningful URIs
 - separate schema (*ontology*) and data (*resource*) URIs/IRIs, like in DBpedia

Knowledge Graphs
Jeff Z. Pan

13

13

KCM Lifecycle: Schema Modelling



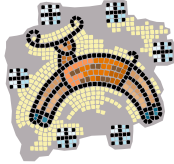
- **Generate authoring tests from requirements**
- KG schema: **Ontology**
- Test Driven Ontology Construction
 - try to reuse existing schema and vocabulary for passing the authoring tests
 - Check e.g. existing repositories such as LOV and semic.eu
 - Further improve the KG to pass all the **authoring tests**

Knowledge Graphs
Jeff Z. Pan

14

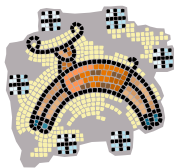
14

Constructing an Ontology (from scratch)



- Join a small group (2-3 people)
 - Take out a piece of paper
- We are going to
 - Construct an animal ontology
 - To index a children's book of animals
 - Key dimensions for competency questions
 - Where they live
 - What they eat
 - How dangerous they are
 - How big they are

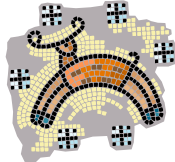
Collect Important Concepts



- Dog
- Cat
- ...

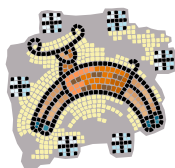
- Draw a small box in your paper
- Write down up to 10 important concepts
- Discuss in your group to agree upon the set of important concepts

Extend the Concepts 'Laddering'



- Take a group of things and ask
 - What they have in common
 - What other 'siblings' there might be
- Example:
 - Cat, Dog => Mammal

Identify Connections



- Choose concepts that can be defined based on others
 - Distinguish **atomic concepts** from **definable concepts**
 - Be aware of the level of details that one needs
- Identify relations
 - Such as "eats" ...

... and Now We Have

- **Living Thing**

- Animal
 - Mammal
 - Cat
 - Dog
 - Cow
 - Person
 - Fish
 - Carp
 - Goldfish
- Plant
 - Tree
 - Grass
 - Fruit

- **Modifiers**

- domestic
 - Pet
 - Farmed
 - Draft
 - Food
- Wild
- Health
 - Healthy
 - Sick
- Gender
 - Male
 - Female
- Age
 - Adult
 - Child

- **Relations**

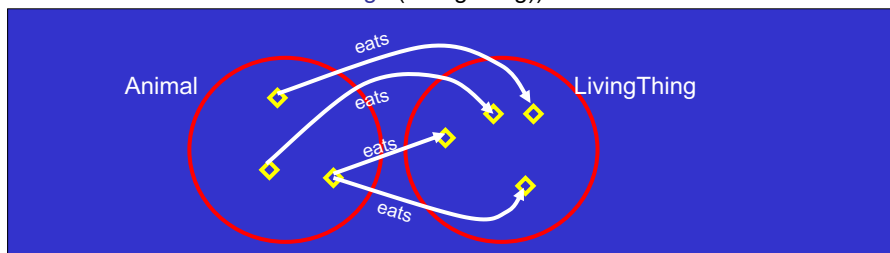
- eats
- owns
- parent-of
- ...

- **Definable**

- Carnivore
- Herbivore
- Child
- Parent
- Mother
- Father
- Food Animal
- Draft Animal

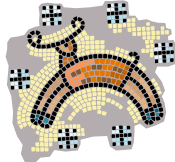
Identify Domains and Ranges for Relations

- Animal *eats* Living thing
 - ObjectProperty (*eats* domain (Animal)
range (LivingThing))



- “If an individual has an eats relationship then it must be an instance of the class Animal”
- “If an individual is a value of the eats relationship then it must be an instance of the class LivingThing”

Clarify the Definitions



- A 'Parent' is an animal that is the parent of *some* other animal
 - Class (Parent complete intersectionOf (Animal restriction (parent_of someValuesFrom (Animal))))
- What does it mean?
 1. If an individual is an instance of the class **Animal** and has some parent_of relationship with some instance of the **Animal** class, then it is an instance of the **Parent** class (*sufficient condition*)
 2. If an individual is an instance of the **Parent** class, then it is an instance of the class **Animal** and has some parent_of relationship with some instance of the **Animal** class (*necessary condition*)

Lecture Outline

- Motivation
- Overview KG Construction
- Detailed Discussions on Data Lifting

THE UNIVERSITY of EDINBURGH

KCM Lifecycle: Data Lifting

Knowledge Extraction:

- Extractor for texts
- Extractor for tables
- Type classifier

Knowledge Base

- beliefs
- candidate facts

Knowledge Integrator

Subsystem Components

- CPL
- CSEAL
- CMC
- RL

Data Resources (e.g., corpora)

Knowledge Graphs
Jeff Z. Pan

23

23

THE UNIVERSITY of EDINBURGH

How is DBpedia Constructed?

Property Value

dbpedia:Planet/density	• 5514.0
dbpedia:Planet/maximumTemperature	• 329.65
	• 330.0
dbpedia:Planet/meanRadius	• 6371.0
dbpedia:Planet/meanTemperature	• 288.0
	• 288.15
dbpedia:Planet/minimumTemperature	• 183.95
	• 184.0
dbpedia:Planet/surfaceArea	• 1.4894E8
	• 3.61132E8
	• 5.10072E8

Physical characteristics

- Mean radius: 6 371.0 km (3 958.8 mi)^[R]
- Equatorial radius: 6 378.1 km (3 952.2 mi)^[R]
- Polar radius: 6 356.8 km (3 949.9 mi)^[R]
- Flattening: 0.003 3528^[R]
- 1/298.257 222 101 (ETRS89)
- Circumference: 40 075.017 km equatorial (24 901.481 mi)^[R]
- 40 007.86 km meridional (24 866.73 mi)^[R]^[12]
- Surface area: 510 072 000 km² (196 940 000 sq mi)^[12]^[10]^[4]
- 148 940 000 km² land (57 510 000 sq mi; 29.2%)
- 361 132 000 km² water (139 430 000 sq mi; 70.8%)
- Volume: 1.083 21 × 10¹² km³ (2 599 76 × 10¹¹ cu mi)^[R]
- Mass: 5 972 37 × 10²⁴ kg (1.316 68 × 10²⁴ lb)^[R]
- (3.0 × 10^{−4} M_J)
- Mean density: 5.514 g/cm³ (0.1962 lb/cu in)^[R]
- Surface gravity: 9.807 m/s² (1 g; 32.18 ft/s²)^[R]
- Moment of inertia factor: 0.3307^[17]
- Escape velocity: 11.186 km/s^[R]
- (40 270 km/h; 25 020 mph)
- Sidereal rotation period: 0.997 269 68 d^[R]
- (23h 56m 4.105s)
- Equatorial rotation velocity: 0.4651 km/s^[R]
- (1 674.4 km/h; 1 040.4 mph)
- Axial tilt: 23.439 2811^[R]
- Albedo: 0.367 geometric^[R]
- perihelion
- Satellites: 1 natural satellite: the Moon

← Infobox from Wikipedia

Knowledge Graphs
Jeff Z. Pan

24

24

DBpedia Knowledge Graph: Construction

- Mapping Wikipedia Infobox template elements to elements in DBpedia elements

- Mappings are created in a community-driven process

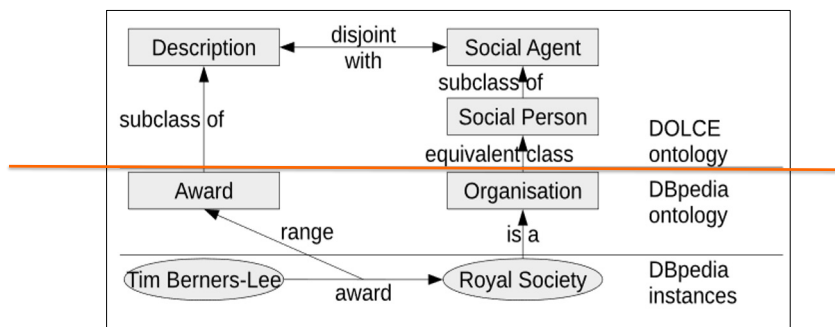
- Example: suppose we consider some wikipedia pages about actors

```

{{TemplateMapping
| mapToClass = Actor
| mappings =
  {{ PropertyMapping | templateProperty =
  name | ontologyProperty = foaf:name }}
  {{ PropertyMapping | templateProperty =
  birth_place | ontologyProperty =
  birthPlace }}
  
```

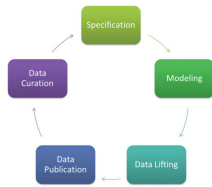
- dbpedia:Vince_Vaughn rdf:type dbpedia-owl:Actor .
- dbpedia:Vince_Vaughn foaf:name "Vince Vaughn"@en .
- dbpedia:Vince_Vaughn dbpedia-owl:birthPlace dbpedia:Minneapolis

Reasoning with Schema: Why?



Without DOLCE Lite ontology, a total of 97,749 statements (0.65% of all statements) was found to be inconsistent, with 630 different axioms involved in the corresponding explanations. With DOLCE Lite ontology, this number increases to 3,654,255 statements (24.36% of all statements), with 1,467 axioms involved in the corresponding explanations

KCM Lifecycle: Data Lifting

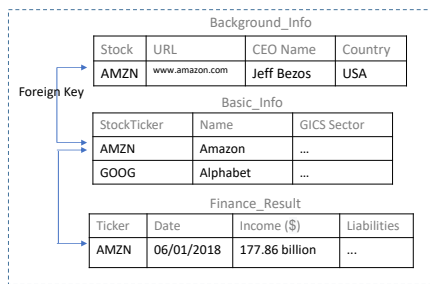


Student ID	Name	take-course
p001	John	cs3019
p002	Tom	cs3023

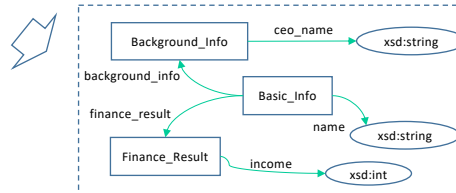
- [csd:p001 rdf:type csd:Student .]
- [csd:p002 rdf:type csd:Student .]
- [csd:p001 csd:name "John" .]
- [csd:p002 csd:name "Tom" .]
- [csd:p001 csd:take-course csd:cs3019 .]
- [csd:p002 csd:take-course csd:cs3023 .]

27

Lifting Tables to KG



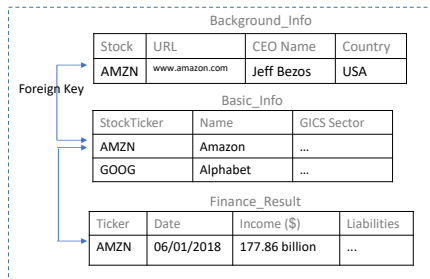
- **Table to RDF Mapping:**
 - Table and column to class
 - Inter-column relation to data or object property
 - Cell to entity



Shortcomings: rely on table headers, often with incomplete semantics!

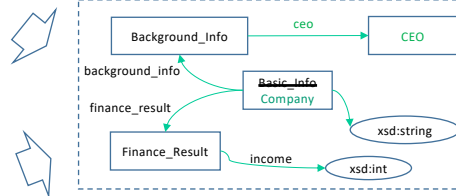
28

Lifting Tables to KG



• **Table to RDF Mapping:**

- Table and column to class
- Inter-column relation to data or object property
- Cell to entity



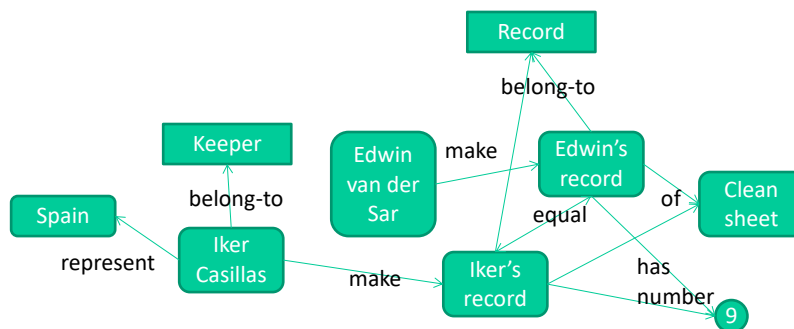
• **Semantic annotation:**

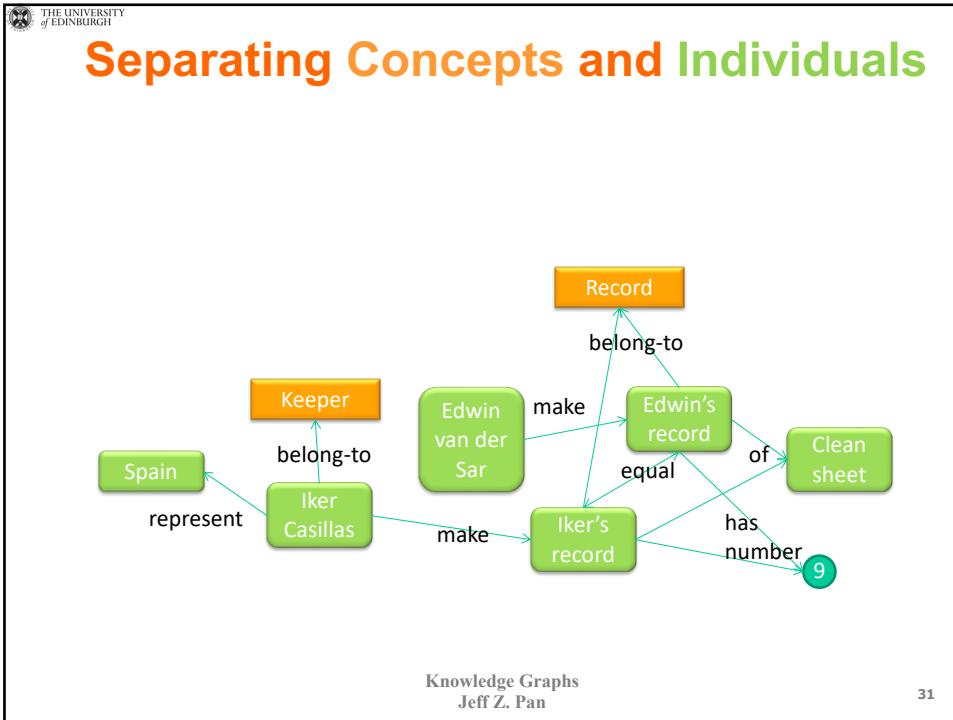
- Name of Basic_Info by class **Company** and object property **hasCompany**;
- Lead to more semantics: Company is a sub-class of Organization; the definition of hasCompany ...

[Credit: J Chen]

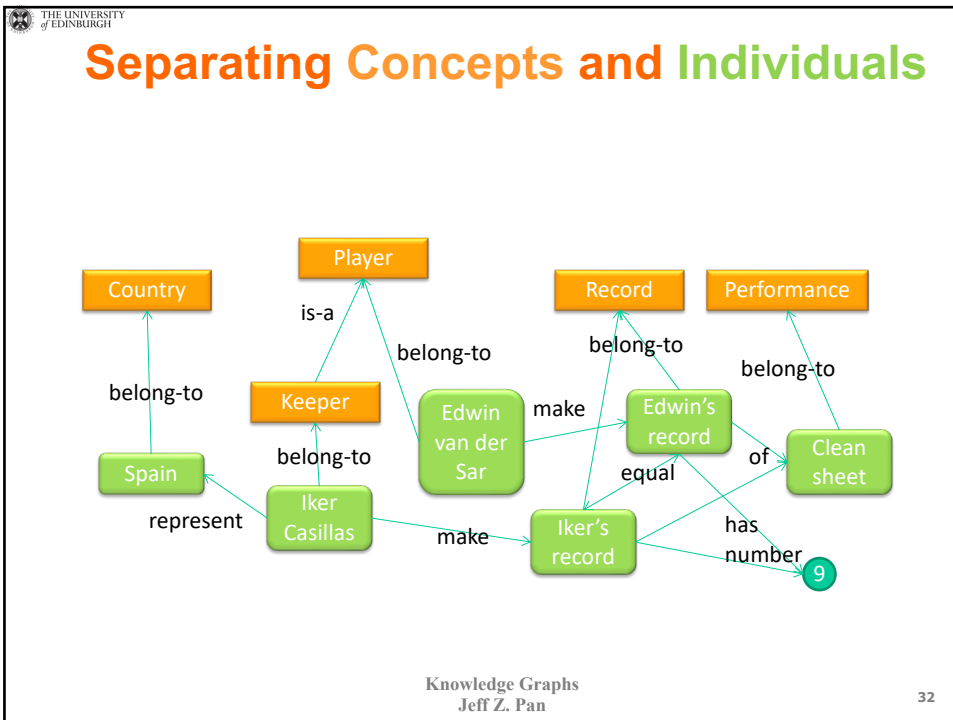
Texts to Semantic Network

"Spain keeper Iker Casillas equalled Edwin van der Sar's record of nine clean sheets."

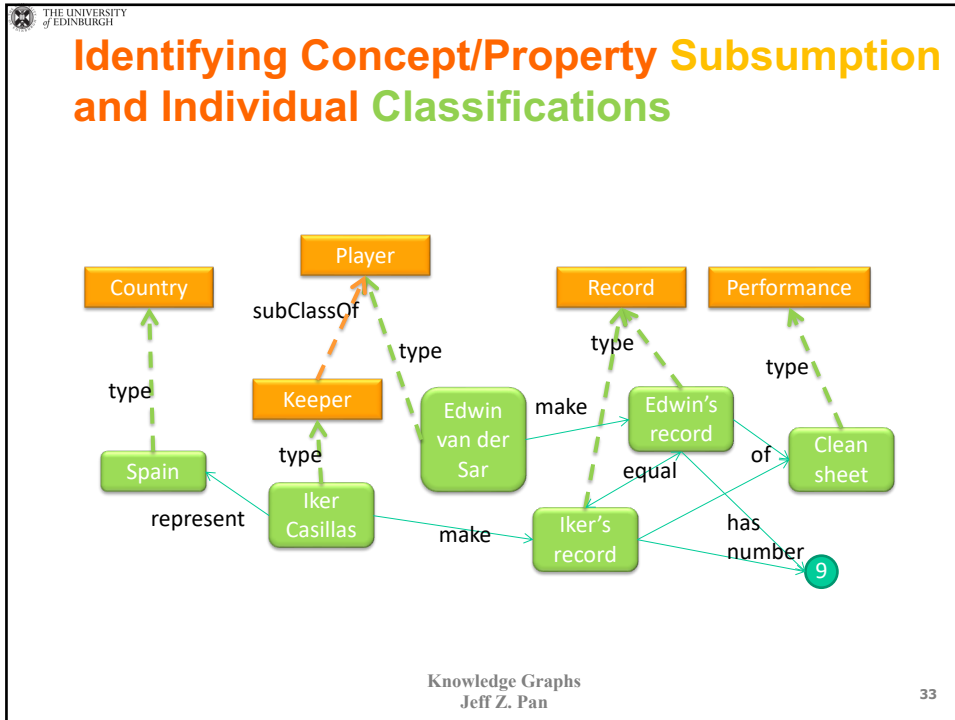




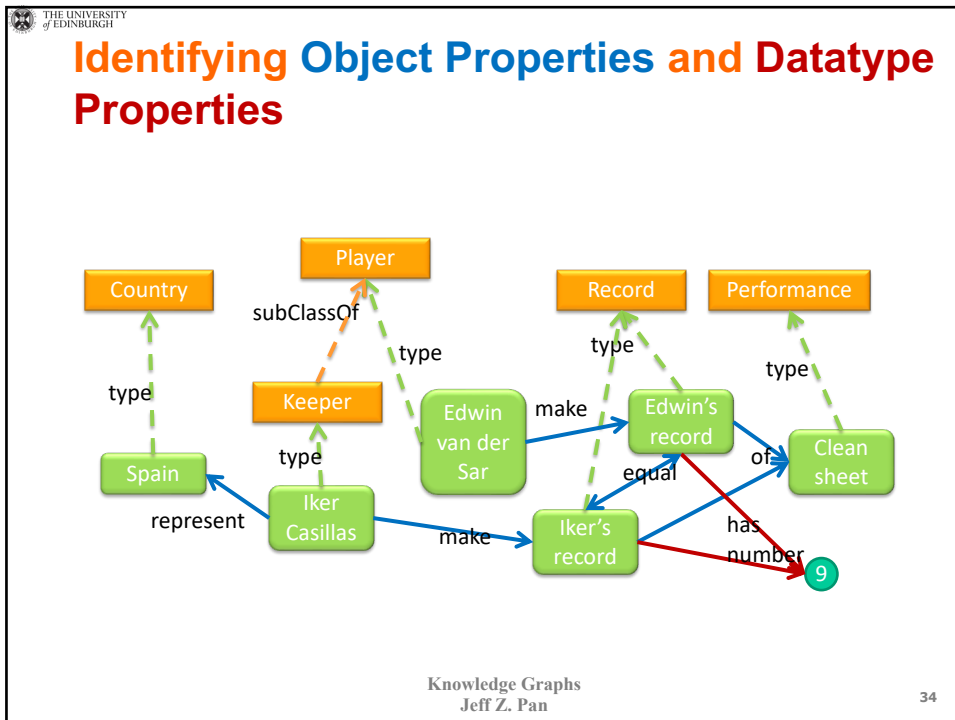
31



32



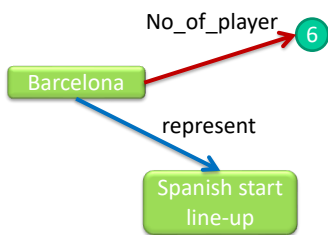
33



34

A Process of Refinement

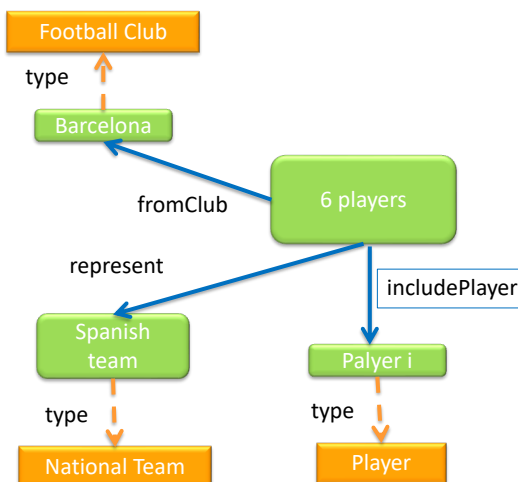
- The procedure of constructing an ontology (KG schema, TBox) is also a procedure in which you examine and refine your semantic networks!
- A typical example: multi-entity relationships
 - E.g. “Six players from Barcelona were in the Spanish starting line-up”
 - This is a complex relationship involving a “Country’s team”, a “club team”, and “6” players
 - In semantic networks, one may draw a figure like this or similar:



- This is not semantically precise:
 - Barcelona has 6 players?
 - No, it has 6 players in the Spanish start line-up
 - Barcelona represents Spain?
 - No, its 6 players represent Spain in the start line-up
- This relationship makes sense only when you take all these entities into account!

A Process of Refinement

- An alternative representation
 - there could be others
- Constructing ontology helps you re-evaluate and refine your semantic network.



Lecture Outline

- Motivation: KG is widely used in downstream applications
- Introduction: Knowledge Construction and Maintenance Lifecycle
- Focus: Modelling and Data Lifting
- Exercises (Next time we introduce Description Logics in details)
 - $\text{Young} \sqcap \text{Person}$
 - $\exists \text{has_pet. Young}$
 - `Class (Parent complete
intersectionOf (Animal
restriction (parent_of
someValuesFrom (Animal))))`