

A Practical Guide to Knowledge Graph Construction from Technical Short Text

What is a knowledge graph and why are they useful?

Dr Michael Stewart

ARC ITTC for Transforming Maintenance through Data Science (CTMTDS)

Tutorial at AJCAI
5 December 2022

Data Science
Transforming
Maintenance

Outline – Part 1

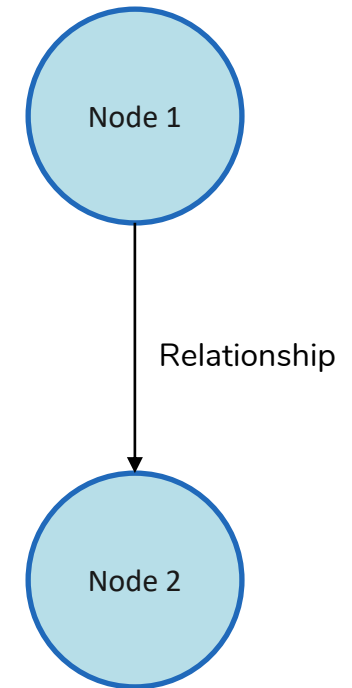
- » Intro to graph databases and knowledge graphs
- » What makes knowledge graphs so useful?
- » Knowledge graph construction from text

Outline – Part 1

- » Intro to graph databases and knowledge graphs
- » What makes knowledge graphs so useful?
- » Knowledge graph construction from text

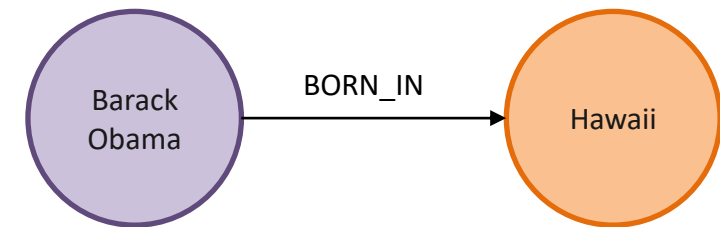
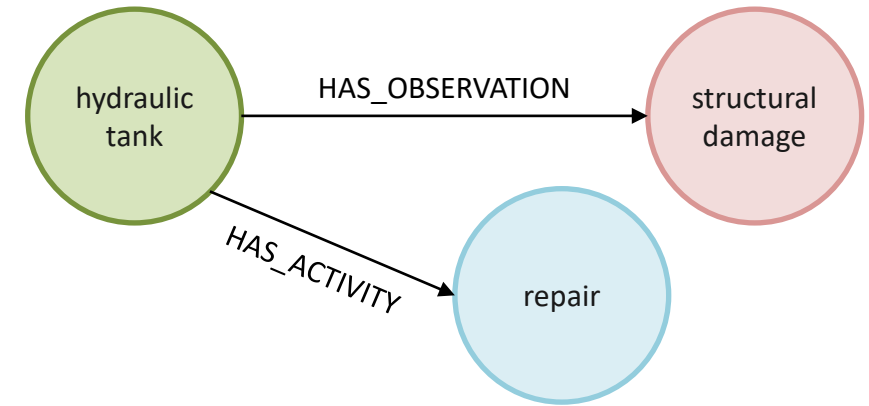
Graph Databases

- » A **graph database** models data as a **graph**, i.e. a collection of **vertices** and **edges**, also known as **nodes** and **relationships**. Edges can be **directed** or **undirected**.
- » We can model all sorts of scenarios using graphs – social networks, natural language, scientific papers, etc.
- » While **graph databases** have been around for many years, they have gained popularity in recent years thanks to the advent of **knowledge graphs**.
- » The most popular graph database is **Neo4j**, which is free to use both internally and for personal use.



Knowledge Graphs

- » A knowledge graph is a type of graph database that captures information about **entities, objects, events** or **concepts**.
- » It is comprised of *triples*, i.e. facts in the form $\langle \text{entity_1}, \text{relation}, \text{entity_2} \rangle$.
- » Knowledge graphs are often built from unstructured text and extended to incorporate **structured information** from a range of data sources.



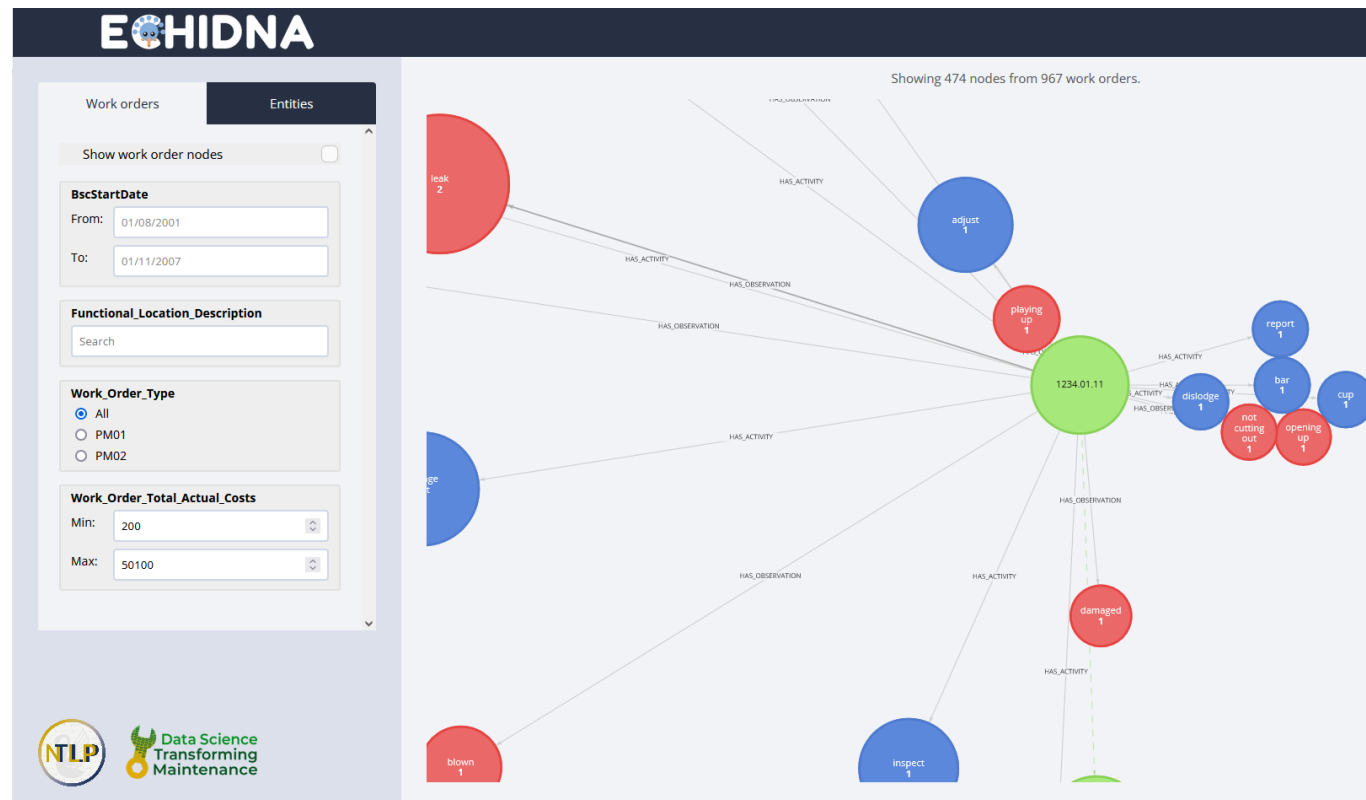
Examples of Knowledge Graphs in use today

Google's search functionality is made possible via a knowledge graph.

The screenshot shows a Google search for 'perth'. The search bar at the top contains 'perth' with a search icon. Below the search bar, navigation tabs for 'All', 'Maps', 'News', 'Images', 'Videos', and 'More' are visible, along with 'Settings' and 'Tools'. The results indicate 'About 190,000,000 results (0.84 seconds)'. The 'Top stories' section features three video thumbnails with captions: 'PerthNow: Perth traffic: Car pile-up causes commuter chaos on Kwinana Freeway northbound' (23 hours ago), 'PerthNow: Perth weather: Bureau of Meteorology warns of huge rain falls while Cyclones Odette...' (2 days ago), and 'PerthNow: Prince Harry and Meghan say Prince Philip to 'be greatly missed'' (12 hours ago). A 'View all' button is below these stories. The 'People also ask' section lists four questions: 'What is Perth best known for?', 'Is Perth an expensive city?', 'Which is better Perth or Melbourne?', and 'Is Perth better than Sydney?'. Below this is a link to 'https://perth.wa.gov.au' with the heading 'Welcome to the City Of Perth | City of Perth' and a brief description. On the right, a knowledge panel for 'Perth' is displayed, showing a city skyline and a map. The panel includes the text 'Perth, capital of Western Australia, sits where the Swan River meets the southwest coast...' and lists key facts: Area (6,418 km²), Founded (12 June 1829), Weather (28 °C, Wind E at 23 km/h, 31% Humidity), Local time (Saturday 2.25 pm), Population (1.985 million (2016) United Nations), and Established (4 June 1829). It also has a 'Plan a trip' section with 'Things to do', '3-star hotel averaging \$110, 5-star averaging \$216', and 'Upcoming events'.

Examples of Knowledge Graphs in use today

Echidna: A graph-based tool for visualising maintenance work orders



<https://nlp-tlp.org/echidna/>

Companies using Knowledge Graphs



Other examples of real-world knowledge graphs

- » There are many real-world KGs available, for example:
 - » **Wikidata**: The large-scale (700m + triples) KG behind Wikipedia.
 - » **Freebase**: Massive (3b+ triples) KG used by Google.
 - » **YAGO**: A huge semantic knowledge base derived from Wikipedia, Wordnet, and Geonames.
 - » **Semantic Scholar**: A large KG of scientific literature.

There is an excellent list of real-world knowledge graphs available at <https://github.com/totogo/awesome-knowledge-graph>.

Outline – Part 1

- » Intro to graph databases and knowledge graphs
- » What makes knowledge graphs so useful?
- » Knowledge graph construction from text

**“The search for information takes
14-30 percent of the engineers’
time.”**

Deloitte, “Wisdom of Enterprise Knowledge Graphs”

What makes knowledge graphs so useful?

Natural format for capturing knowledge in unstructured text

- » Over **80% of data** in organisations is **unstructured**, and therefore inaccessible.
- » **Noisy, unstructured text** is present in many domains, e.g. maintenance work orders, doctor's notes, safety records, and so on.
- » For example, consider **maintenance work orders**:

replace seal on pump

fix a/c too hot

- » Graph databases provide the means to **unlock the knowledge** captured within unstructured data and combine this with knowledge held within the structured data.

What makes knowledge graphs so useful?

Data integration

- » Most companies maintain numerous databases each containing valuable knowledge.
- » This data is often **unconnected** due to differences in formats and structure.
- » Each separate database only captures a fragment of knowledge held within a company.
- » Knowledge graphs provide the facility to integrate the data into one complex graph.

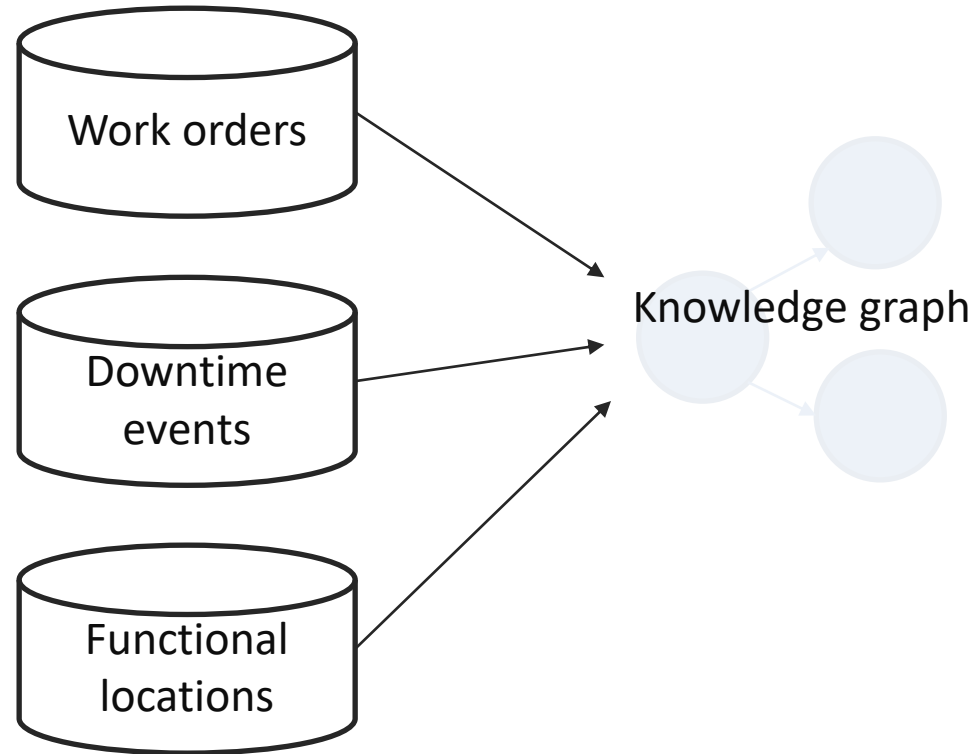


Photo found at Unsplash.com

What makes knowledge graphs so useful?

Data integration

Example – Maintenance Domain



What makes knowledge graphs so useful?

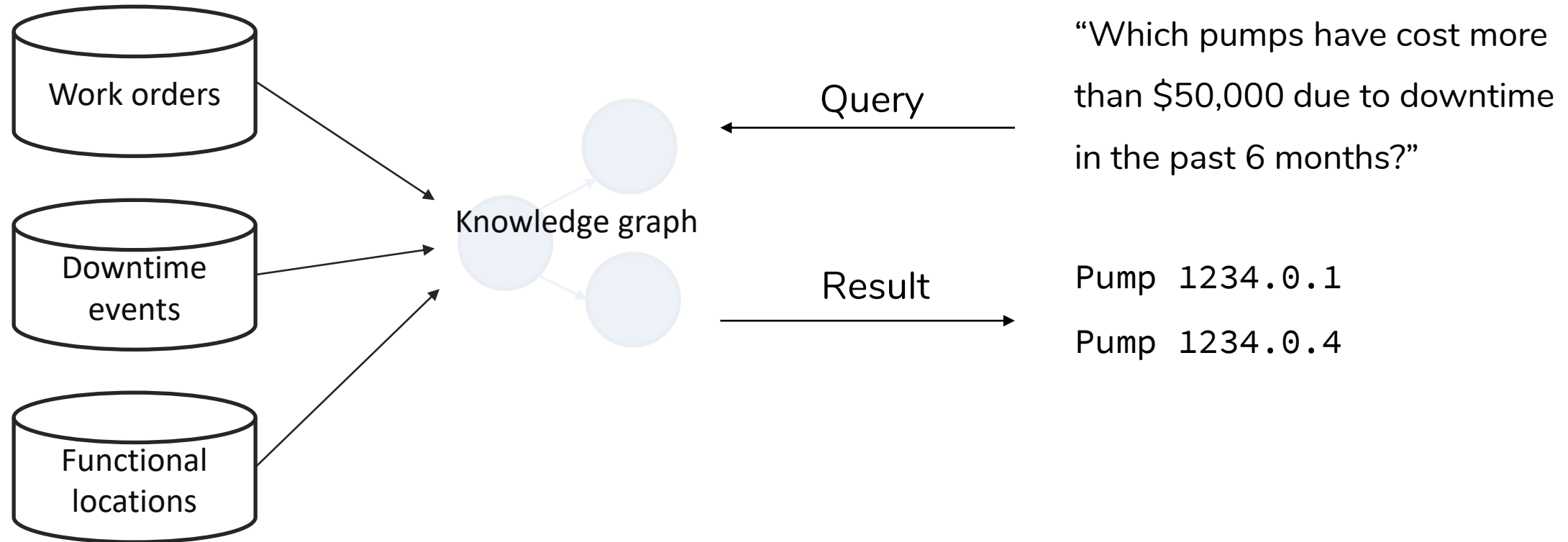
Facilitating decision making

- » Knowledge graphs facilitate **complex decision making** supported by collective knowledge from a range of domains, including unstructured data.
- » They go hand in hand with machine learning, facilitating a range of machine learning opportunities such as **question answering, recommendation systems, supply chain management**.
- » They are also excellent for increasing **data accessibility**, providing domain experts with the means to quickly ask questions across many different datasets.

What makes knowledge graphs so useful?

Facilitating decision making

Example – Maintenance Domain

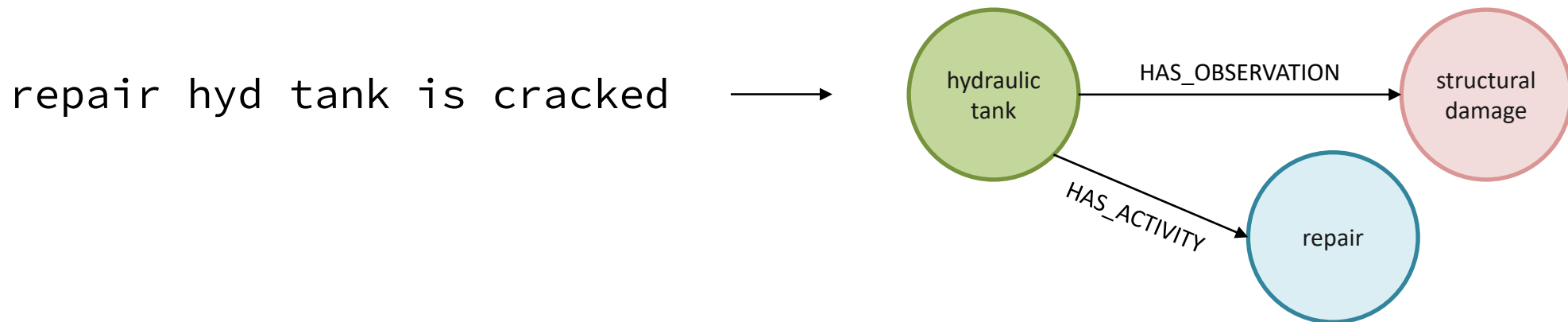


Outline – Part 1

- » Intro to graph databases and knowledge graphs
- » What makes knowledge graphs so useful?
- » Knowledge graph construction from text

Knowledge Construction from Text (KGC)

- » **Knowledge Graph Construction (KGC)** from text is the process of automatically constructing knowledge graphs from unstructured text.



- » We may then incorporate **structured data** into our knowledge graph such as costs, dates, and so on.

Knowledge Construction from Text (KGC)

- » To build knowledge graphs from unstructured text we must employ **Natural Language Processing** (NLP) or **Technical Language Processing** (TLP) techniques, which we will demonstrate in the following session.
- » Most approaches to knowledge graph construction from text are pipeline-based and include **three core components**:
 - » Entity extraction
 - » Relation extraction
 - » Entity linking
- » **Lexical normalisation** (i.e. text cleaning) is also an important technique for technical language.
- » Recent research is trending towards **end-to-end** KGC, i.e. performing all tasks in a single model.

What makes a knowledge graph a “knowledge” graph?

- » There is not yet a consensus on a formal definition of a knowledge graph.
- » Wu et al.[1] describe a KG as having three essential components:
 - » **Concepts** (an entity, attribute, or a fact)
 - » **Relations**
 - » **Background knowledge** about concepts and relations
- » **Background knowledge** differentiates Knowledge Graphs from text graphs or data graphs.

Wu, X., Wu, J., Fu, X., Li, J., Zhou, P., & Jiang, X. (2019, November). Automatic knowledge graph construction: A report on the 2019 ICDM/ICBK contest. In *2019 IEEE International Conference on Data Mining (ICDM)* (pp. 1540-1545). IEEE

Text graph vs data graph vs knowledge graph

Work order

repair hyd tank is cracked

engine wont start

a/c blowing hot air

engin u/s

Text graph vs data graph vs knowledge graph

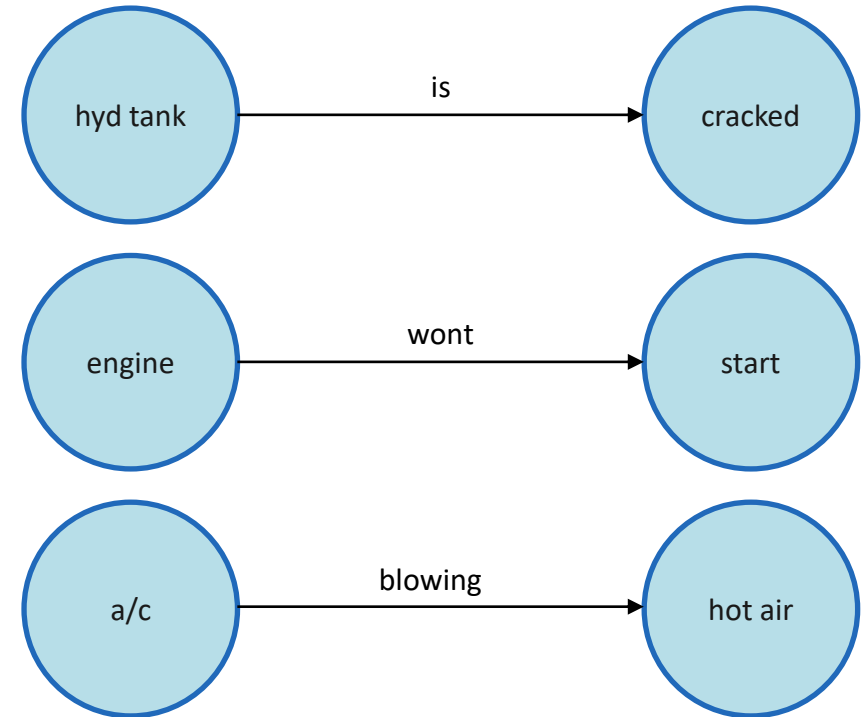
Work order

repair hyd tank is cracked

engine wont start

a/c blowing hot air

engin u/s



Text graph vs data graph vs knowledge graph

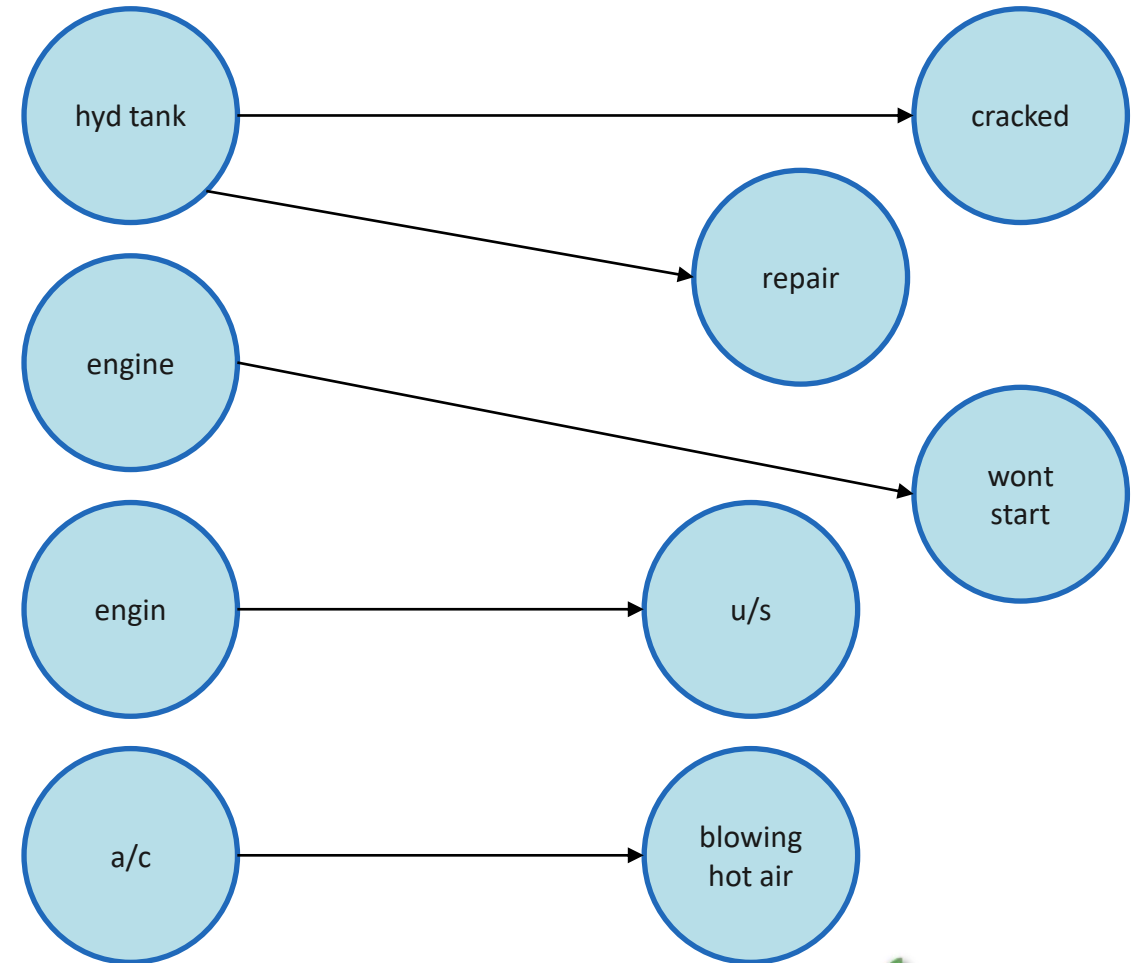
Work order

repair hyd tank is cracked

engine wont start

a/c blowing hot air

engin u/s



Text graph vs data graph vs knowledge graph

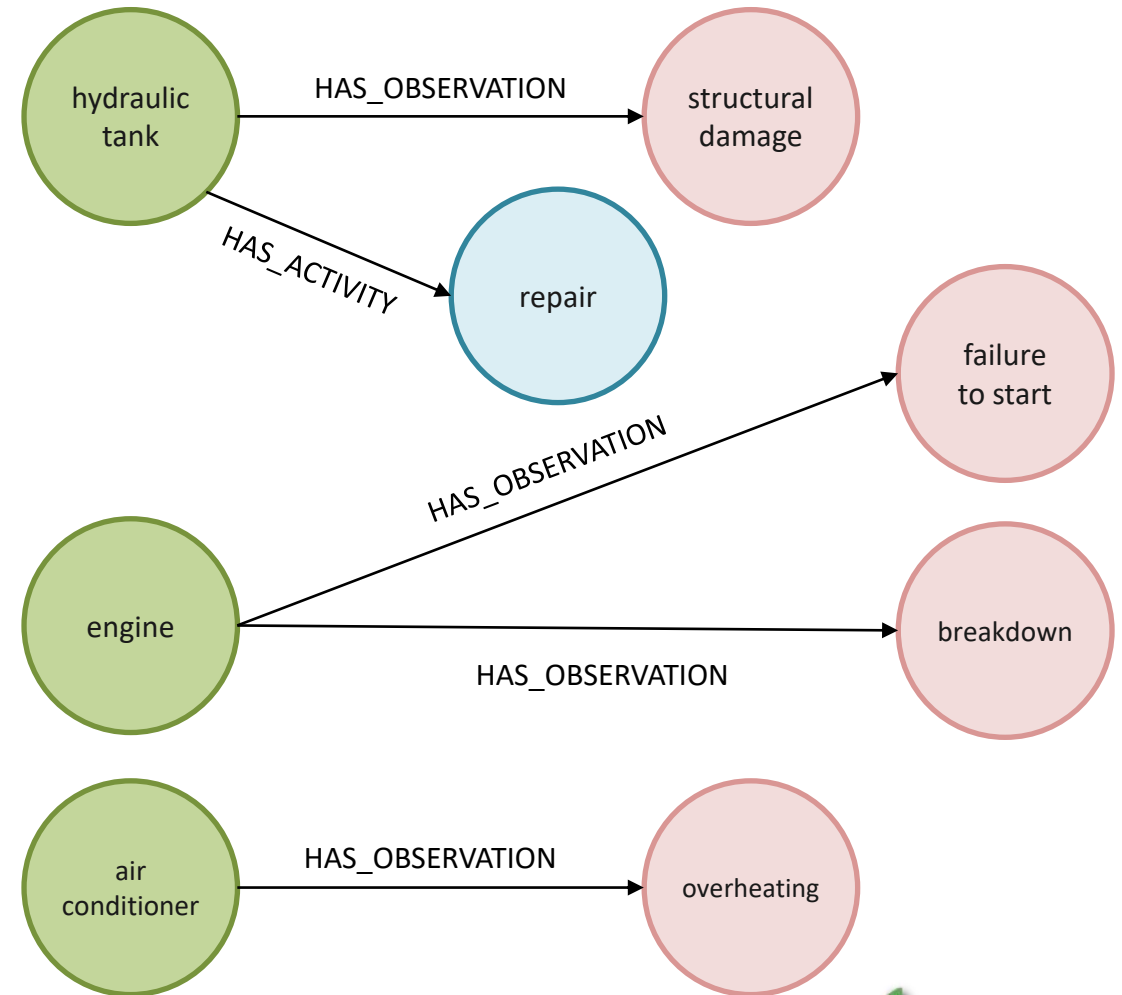
Work order

repair hyd tank is cracked

engine wont start

a/c blowing hot air

engin u/s



Conclusion – Part 1

- » Knowledge graphs are a powerful tool as they are able to **unlock knowledge** captured within the vast **unstructured text** present in many organisations.
- » Graphs excel when dealing with **highly connected data**.
- » They are also the perfect tool for **bringing data together** from across a range of areas in a business.

Conclusion – Part 1

- » We will now demonstrate the process of **Knowledge Graph Construction** via a Jupyter notebook walkthrough.
- » We will introduce the key concepts behind **information extraction** (i.e. constructing knowledge graphs from text) – **lexical normalisation**, **entity recognition**, and **relation extraction**.
- » Towards the end of the tutorial we will look at how the knowledge graph can be **queried** in Neo4j in order to easily access important knowledge captured within the graph.

Questions

Email: michael.stewart@uwa.edu.au

CTMTDS

<https://maintenance.org.au>

UWA NLP-TLP Group

<https://nlp-tlp.org>

Echidna – Demo

<https://nlp-tlp.org/echidna>

Redcoat – Demo

<https://nlp-tlp.org/redcoat>