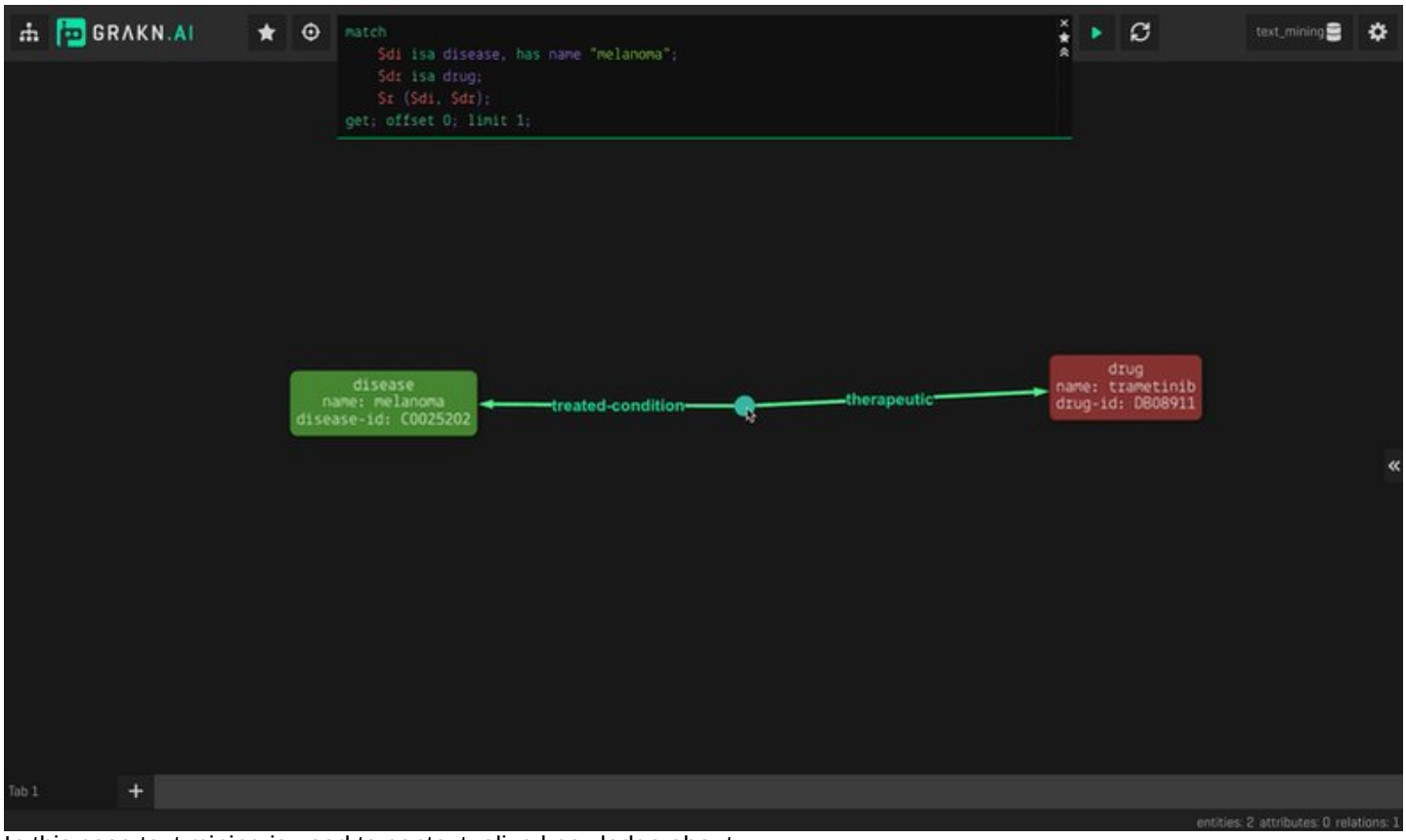
Twitter Thread by Philip Vollet

Philip Vollet
@philipvollet

Why are graphs the future of biomedical research and what is the value of NLP here?

A small case study about:

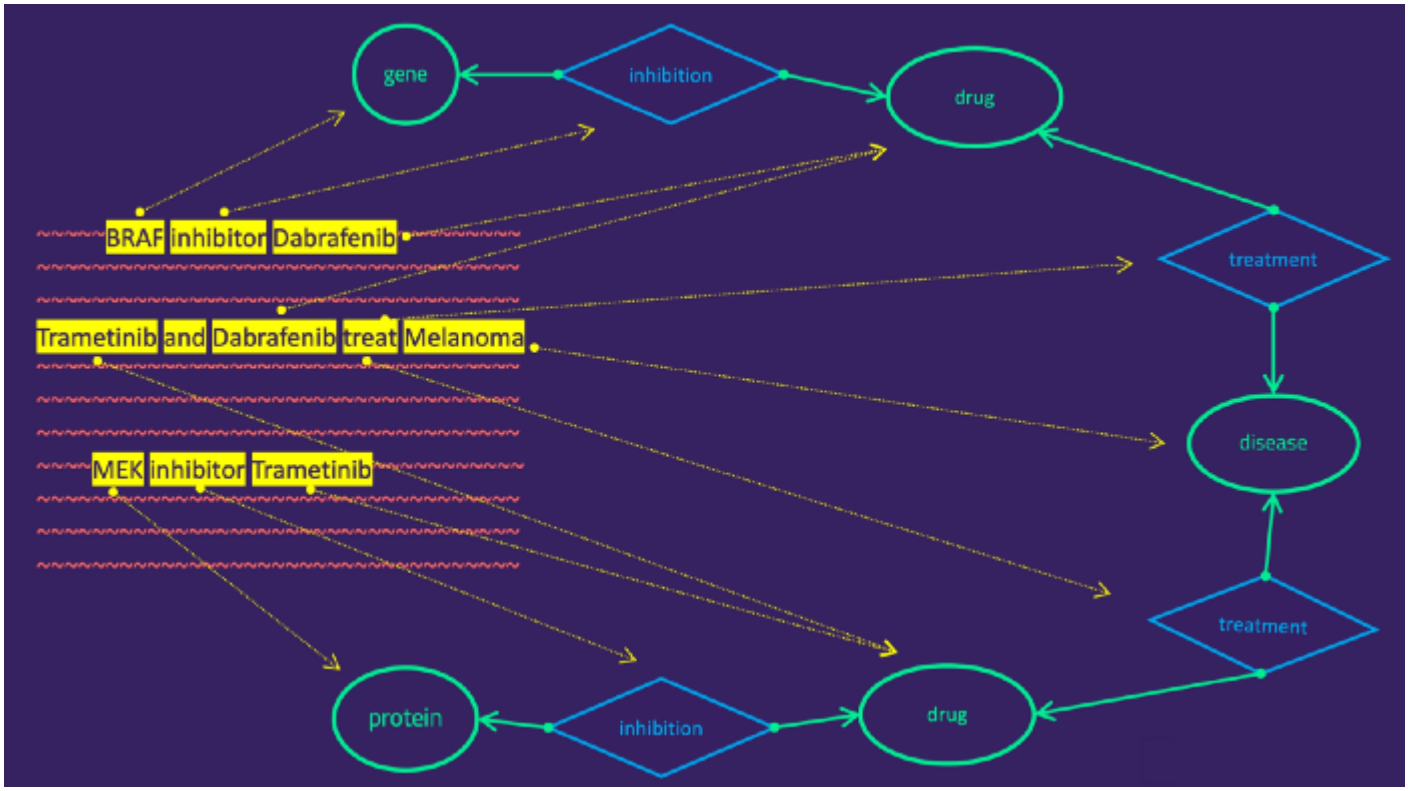
How to speed up drug discovery with knowledge graphs and discover potential cures for diseases



In this case text mining is used to contextualize knowledge about:

- Genes
- Compounds
- Diseases
- Adverse drug effects

- Receptor bindings



Which text types are processed here? Medical literature, patient notes, electronic health records, clinical reports etc.

But how to start?

First you need to identify the different entities such as compounds, diseases, adverse drug effects and receptor bindings.

This is achieved through Natural Language Processing (NLP) and there are suitable pre-trained models for processing biomedical, scientific or clinical text like scispaCy

@spacy_io models for processing biomedical, scientific or clinical text
https://t.co/1EPFZCFwoc

Another library which is specialized in biomedical text is Spark NLP

@JohnSnowLabs

https://t.co/EYM8lIyuUp

The next challenge is to extract the different relations! Diseases are related to genes which are related to receptors and compounds can bind to these receptors.

Sounds simple at first but there are several problems that need to be solved

Problems to solve

1. Difficult to ingest and integrate complex networks of text mined outputs
2. Difficult to contextualize knowledge extracted from text with existing knowledge
3. Difficult to investigate insights in a scalable and efficient way

Fortunately, Grakn solves all our problems!

@GraknLabs

How it works is explained here: https://t.co/BCGDuWrVqA

To understand how NLP and graphs are used to link medical knowledge I recommend this talk on text mining and drug discovery at Novartis

Not quite up to date but aged very well

Connecting the Dots in Early Drug Discovery at Novartis
https://t.co/QNJt6q5IsU