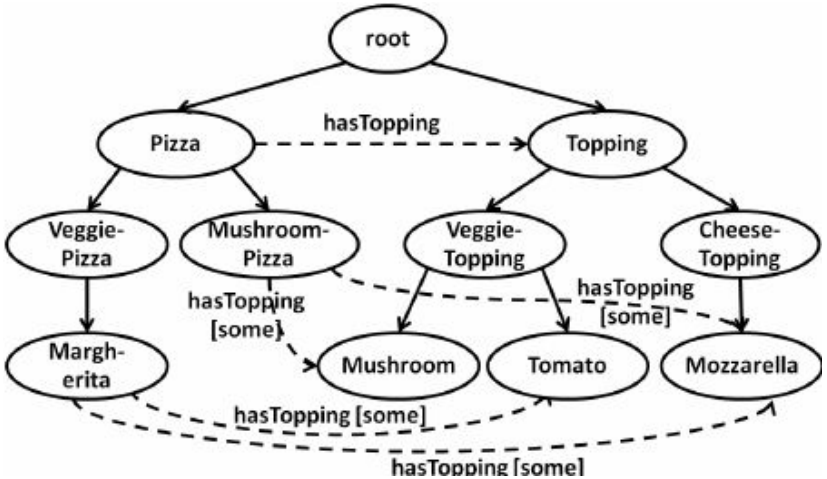
Going Meta #6: Ontology learning from graph data

We want to categorise a data set

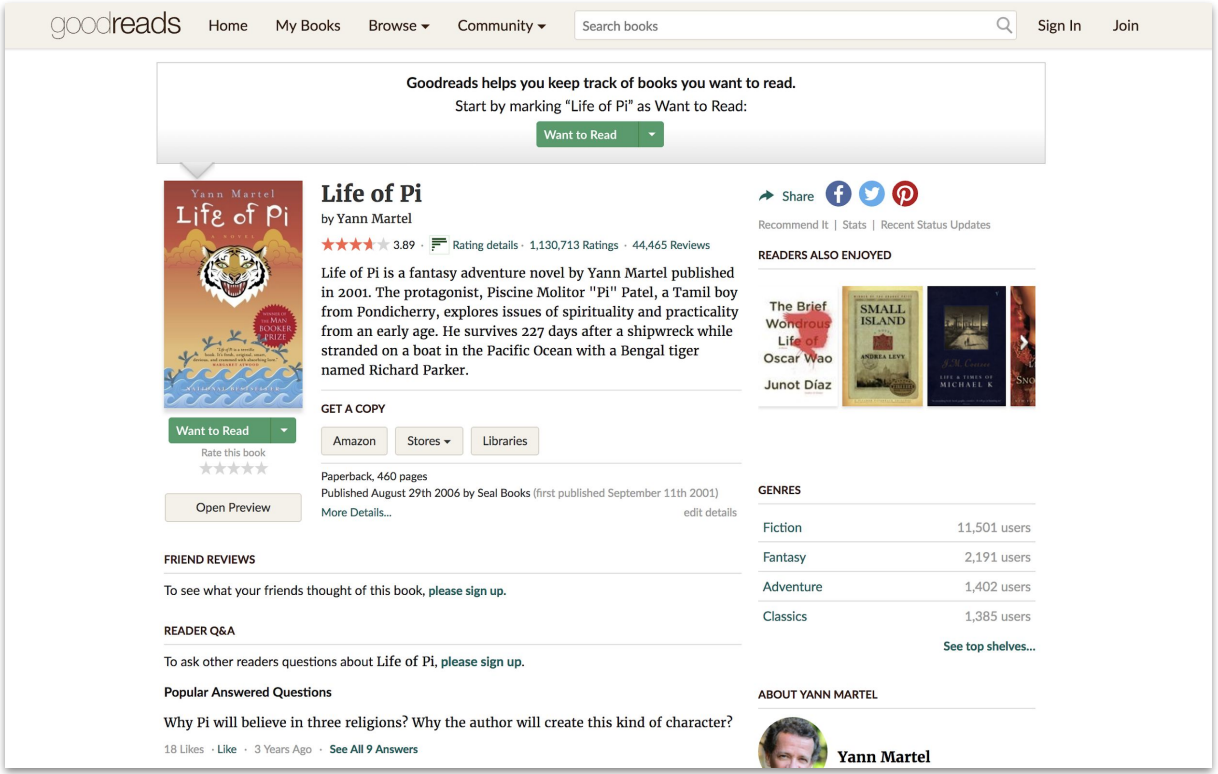
We can have domain experts manually create an ontology/taxonomy/… any classification scheme

(episode 2: semantic search)

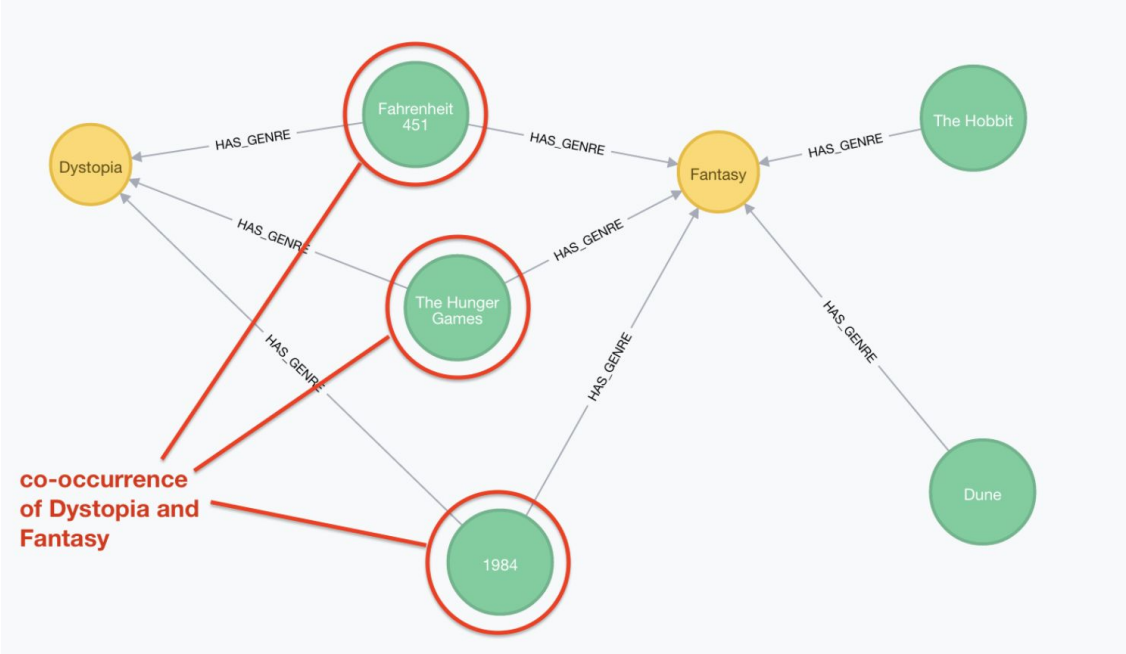


Or we can try to learn it from the data!

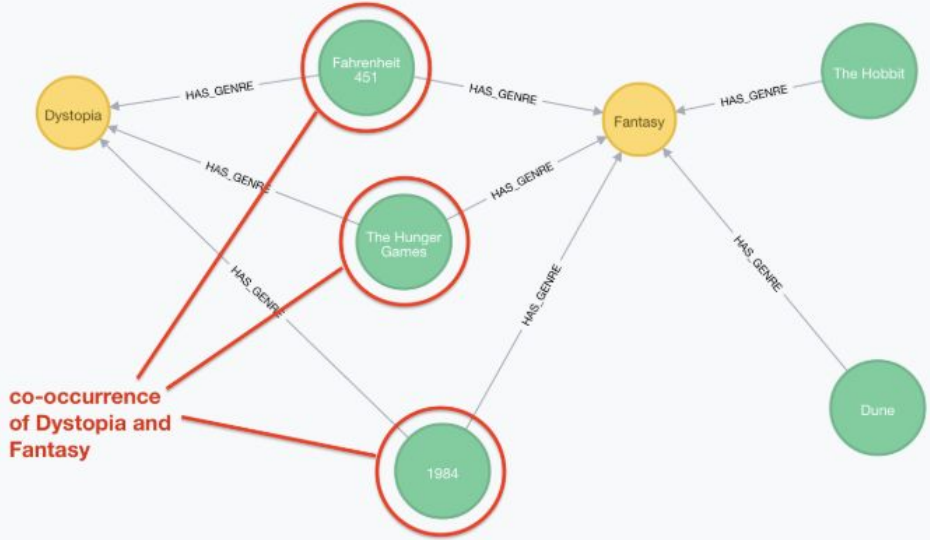
Annotated datasets: Books with multiple genres



Taxonomy learning algo: Co-occurrence



Co-oc score (directional)



COOC(Dystopia,Fantasy)
= #nodes tagged as both Dystopia AND Fantasy / #nodes tagged as Dystopia
= 3 / 3
= 1

COOC(Fantasy,Dystopia)
= #nodes tagged as both Fantasy AND Dystopia / #nodes tagged as Fantasy
= 3 / 5
= 0.6

2 simple rules

Rule 1: If all items tagged as A are also tagged as B and all items tagged as B are also tagged as A, then the categories defined by A and B are equivalent

Rule 2: If all items tagged as A are also tagged as B (but not all B's are tagged as A) we can “derive” that tag A defines a category “narrower than” the one defined by tag B (subcategory?)