

W5. Check List — Junyi

▼ how to create kg from text? → information extraction:

▼ entity extraction - how?

▼ sequence labeling

▼ first, last, interval word in the entity, word not in the entity, single word entity

▼ language model

▼ rule-based

▼ relationship extraction - how?

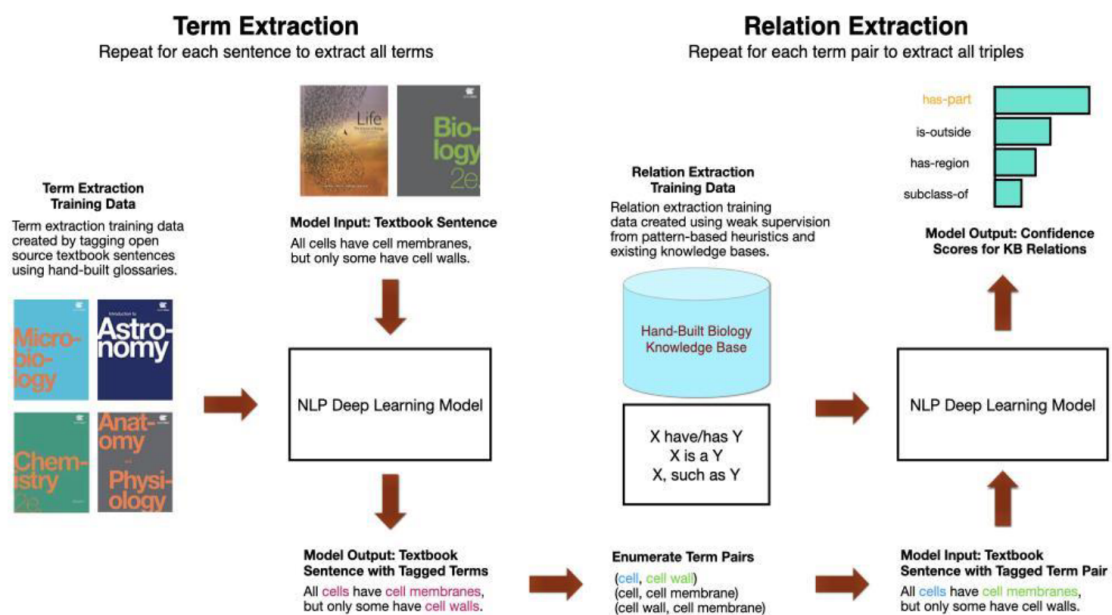
▼ syntactic patterns (/ rule-based)

▼ supervised learning → language model

▼ open information extraction → does not rely on a designed set of relations

▼ entity resolution - how?

▼ The application of intelligent textbook:



▼ quest for meaning in KG

- ▼ logical meaning
 - ▼ The nodes are generics as opposed to concrete entities (person, etc.)
 - ▼ This is more than a directed labeled graph (intro new objects)
- ▼ semantic meaning
 - ▼ define has part and has function
 - ▼ meaning of structure and function: Structure and function are correlated at all levels of biological organization
- ▼ computational meaning
 - ▼ Identify the requirements in terms of a set of questions
- ▼ Entity extraction:
 - ▼ training data: glossary of the textbook
 - ▼ a good lexicon is essential for term extraction
 - ▼ illustration
- ▼ relation extraction:
 - ▼ get the training data: pre-existing KB, distant supervision, weak supervision
 - ▼ steps
 1. define a set of label functions
 2. apply each of there label functions to every training instance
 3. aggregate these sets of labels into a single label for each instance
 - a. hard labels: majority votes
 - b. soft labels: combine label functions based on estimated reliability to get a probability distribution across relations
 - ▼ illustration
- ▼ KG authoring:
 - ▼ plus human review and KG validation

