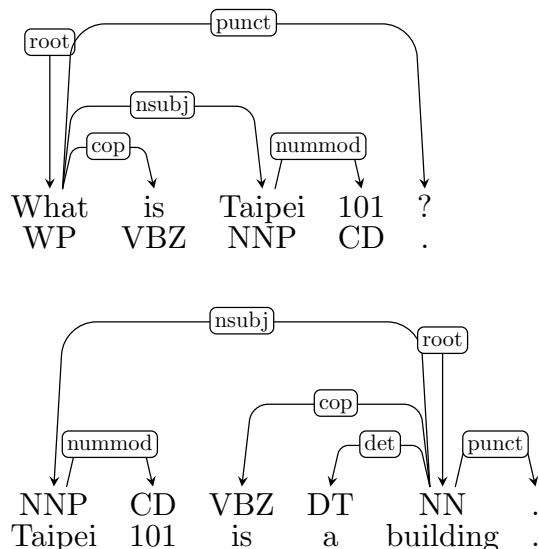


Learning Dependency-Based IR



Abstract

I explore IR in natural language using logical dependency structures. I implement weighted IR for answering forms of wh-queries in natural language. I apply my method to an academic Wikipedia dataset (Carnegie Mellon, University of Pittsburgh).

12/8/15

Methods

Preparation of Data

1. Parse documents into logical dependency structures.
2. (e.g. $\lambda x. \text{ROOT}(n_0) \wedge x(n_i) \wedge \text{nsubj}(n_0, n_i)$)

Learning Step

1. Implemented question/answer type clustering.
 1. (only *wh*-queries supported);
2. Use database of question-answer pairs to discover mappings from clustered question types to candidate answer forms.
3. (e.g. "What is Taipei 101?" \Rightarrow **w** "Taipei 101 is __" | ...)

IR Step

1. Apply QA mappings to natural language queries.
2. Return weighted results (answer forms for the query).

Results

Empirical

1. Achieved 20% accuracy on what-question (N=10) set.
2. Success: What is the middle pedal on a piano?
 1. Answer: Sustain pedal.
3. Challenges (For future work):
 1. What to do if no rules fire?
 2. How to handle quantified queries (e.g. How many _?)
 3. How to generalize to unseen data (i.e. inference)?
 4. Implementing coreference & synonym resolution?

Alan Zaffetti