CS-481 — Artificial Intelligence: Natural Language Understanding

Programming Assignment: Lexical Chains

You are to write a program that takes a text file as input and outputs the set of lexical chains in the text, using the method discussed in class (refer to the slides). Use single words and don't worry about extracting nouns that are phrases (such as "White House" or "lexical chain"). Download Wordnet and (optional) a library for accessing it from Java, Ruby, or Python. (If you find and use another Wordnet library, just document it.)

1. Create lexical chains based on the lexical relationships of synonymy, antonymy, and one level of hyper/hyponymy. In other words, if a noun is related by one of these relationships to an existing lexical chain, it should be added to it, otherwise it should start a new chain.

   Output should be in the form:

   ```
   Chain 1: girl(2), woman(1), women(3), person(1), she(4)
   Chain 2: hat(2), skirt(1), pants(2), clothes(1), shirt(2)
   ...
   ```

   Each chain lists the words it contains, with the number of word occurrences in parentheses.
2. Evaluate the results (qualitatively) and comment on how they might be improved.
3. Implement one of the following enhancements:
    1. Use more lexical relationships, including sibling hypo/hypernymy and meronym relationships.
    2. Only allow a noun to enter a lexical chain if it is within a threshold number of words (this should be a parameter) to the closest element of the chain.
    3. As you add elements to a chain, restrict the senses (synsets) of the nouns in the chain to those consistent with all nouns in the chain (i.e., just those senses related by the given lexical relations).
    4. Find and use phrasal nouns, not just single words. You can do this by finding short sequences of words that have their own entries in Wordnet.
4. Evaluate the results of your enhanced system, and compare them to the previous system.
5. Extra credit: Use the lexical chains to automatically create a summary of the input article.

Links:

Wordnet
Java library
Ruby library
Python library (integrated NLP toolkit)

---

argamon
Last modified: Thu Apr 22 17:07:53 CDT 2010