- **Pattern Clustering Algorithm**
    - **Hook Words and Hook Corpora**
        - Input: Corpus.
        - Output:
            - Hook words – N (100?.. 20?..) words that appear less than F_C, and more than F_B.
            - For each hook word – create hook corpus (set of contexts) of size W (= 5?..).
    - **Pattern Specification**
        - Input: Words?.. (the initial corpus again?.. the hook corpora?..)
        - output:
            - Classify each word into $HFWs$ or CWs
                - HFW – word that appears more than F_H.
                - CW – word that appears less than F_C.
            - Create patterns of the form:
                - *[Prefix]* CW_1 *[Infix]* CW_2 *[Postfix]*
                - How can we extract patterns just from words?..
    - **Discovery of Target Words**
        - Input: Hook corpora.
        - Output:
            - Pattern instances, where one CW is the hook word (of this corpus) and the other CW is the target word (not the hook).
            - Filtering the top and bottom $L\%$ of the target words (after sorting them by 'pointwise mutual information..').
    - **Pattern Clustering**
        - Input:
            - Set of patterns for each hook corpora.
            - The target words that used to extract them.
        - What to do:
            - Group patterns that extracted using the same target word.
            - Merge clusters that share more than $S\%$ of their clusters.
            - Merge pattern clusters from different hook corpora using the provided algorithm.
        - Output:
            - Set of pattern clusters, where for each cluster there are two subset, *core* patters and *unconfirmed* patters.

- **Relationship Classification**
  - **The *HITS* Measure**
    - Input:
      - Pattern clusters.
      - All pairs from the *training* and *test* sets.
    - Output:
      - The HITS values of each $(C,(w_1,w_2))$.
      - Which ∝ to use?.. (0.5?.. 0.2?..)
  - **Classification Using Pattern Clusters**
    - **Classification by cluster *HITS* values as features**
      - Input:
        - Training pairs.
        - Test pairs.
      - What to do:
        - Build feature vectors for the *training* and the *test* pairs (a feature is the *HITS* measure corresponding to a single pattern cluster).
        - Use WEKA to construct a Model and to evaluate it on the test set (we already did that in the last step, didn't we?..).
      - Output:
        - The Model?.. (is this the final output?..)

- **Results**

Corpus Size	Precision	Recall	F-Score	Accuracy
X				
Y				
Z				

  - How many sizes?..
  - How to measure these?..

- **General Questions**
  - What is the goal of this application?..
    - Relation between words?.. – i.e. – to find instances of the 7 relationships?.. (Cause-Effect, Instrument-Agency, Product-Producer, Origin-Entity, Theme-Tool, Part-Whole, and Content-Container)?..
    - The Model?..