



# Improving Self-Organized Information Maps for Digital Libraries: A Semantic Approach and Its Evaluation

Yiling "Jennifer" Lin, Peter Brusilovsky, Daqing He  
 {yil54,peterb,dah44}@pitt.edu

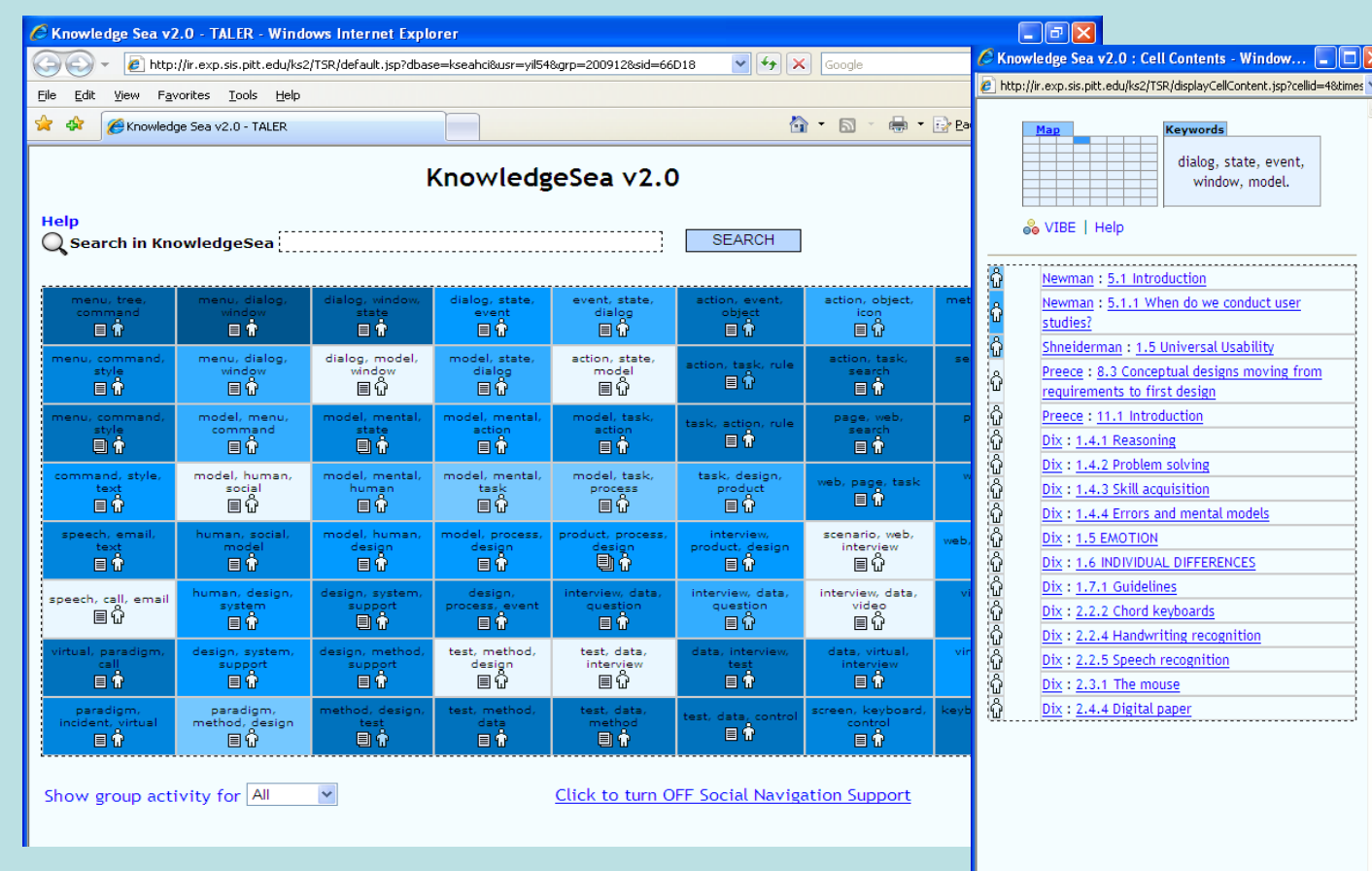
## MOTIVATION

### Semantic improvement

- Problems with keyword representation
    - Multiple word concepts, e.g. "information retrieval"
    - Synonymy, e.g. "find" and "discover"
    - Polysemy, e.g. "bank"
  - Many researchers use semantic representations to replace or augment keyword level representations in various applications
- ⇒ **The usages of semantic-rich features in constructing maps.**

### Human-centric evaluation

- Quality issues
    - Classification: the quality of map topology
    - User-driven navigation: the quality from a user's point of view and cost issue.
  - More studies have adopted SOM for navigational support
- ⇒ **The evaluation of the map quality from a navigational point of view.**



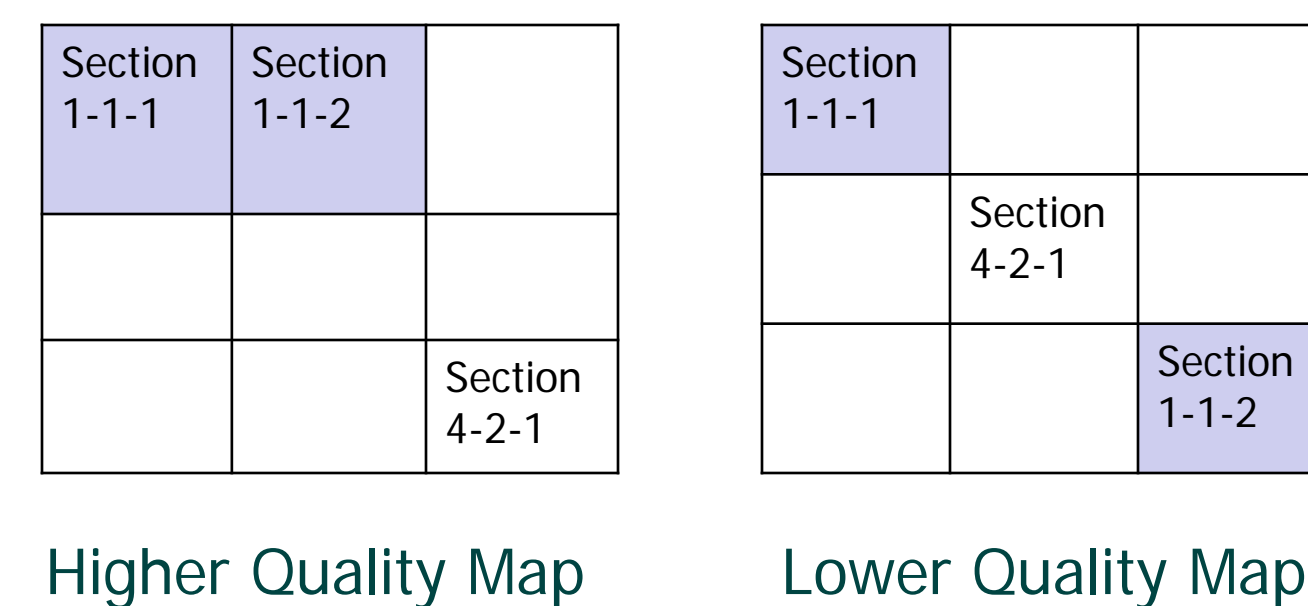
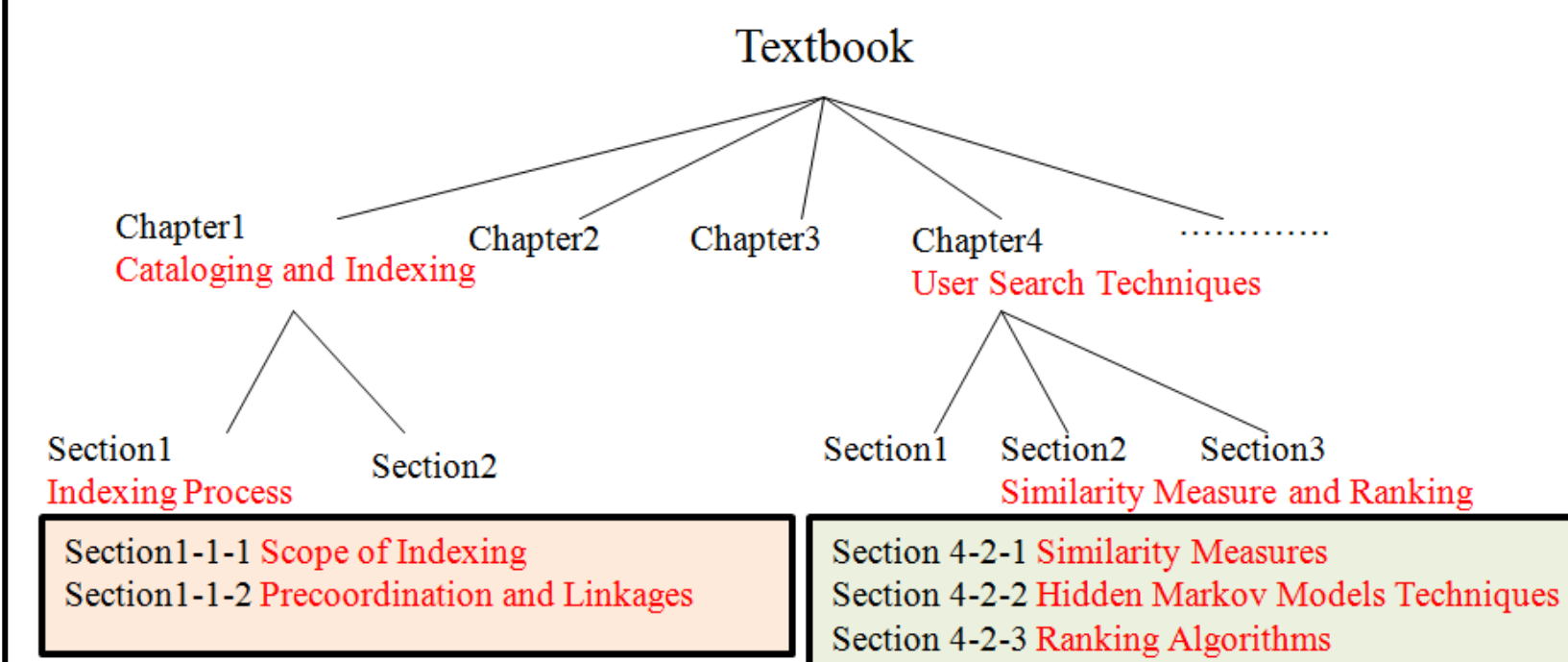
Knowledge Sea II

## METHOD

### Semantic Representations

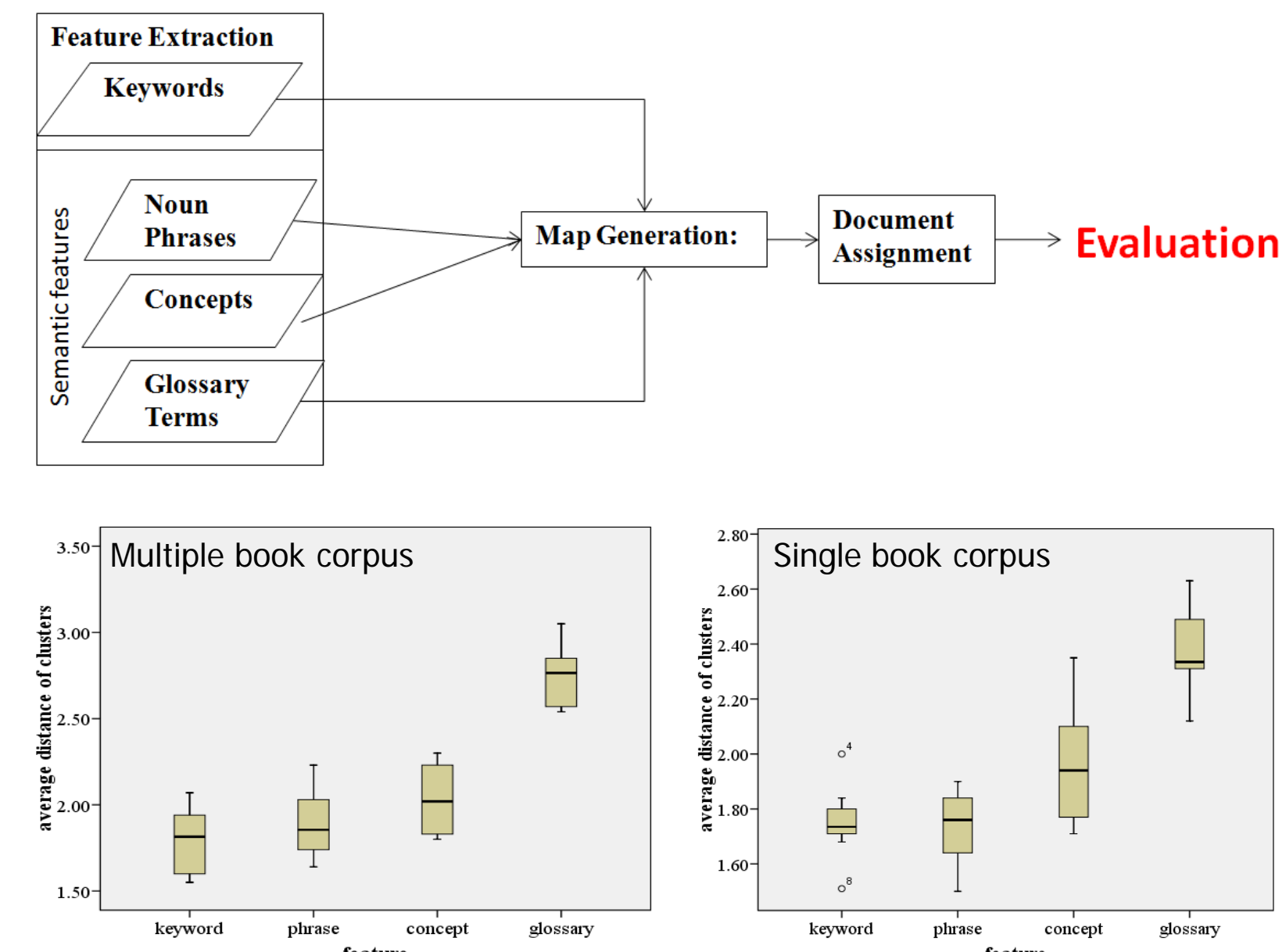
Semantic Level	Feature	Definition	Tool
Low	Keyword	A single term with special significance in the corpus e.g. neural	Indri
	Noun phrase	A phrase which its root is a noun or a pronoun, optionally with modifiers e.g. neural network algorithm	Arizona Noun Phraser
	Concept	A general idea derived from particular instances or occurrences e.g. neural network	Yahoo Term Extraction Web Service
High	Glossary term	A particular meaning in a specific domain which the author wants to make explicitly clear by including it in the glossary of his/her text e.g. neural network	Textbook

### Evaluation Method

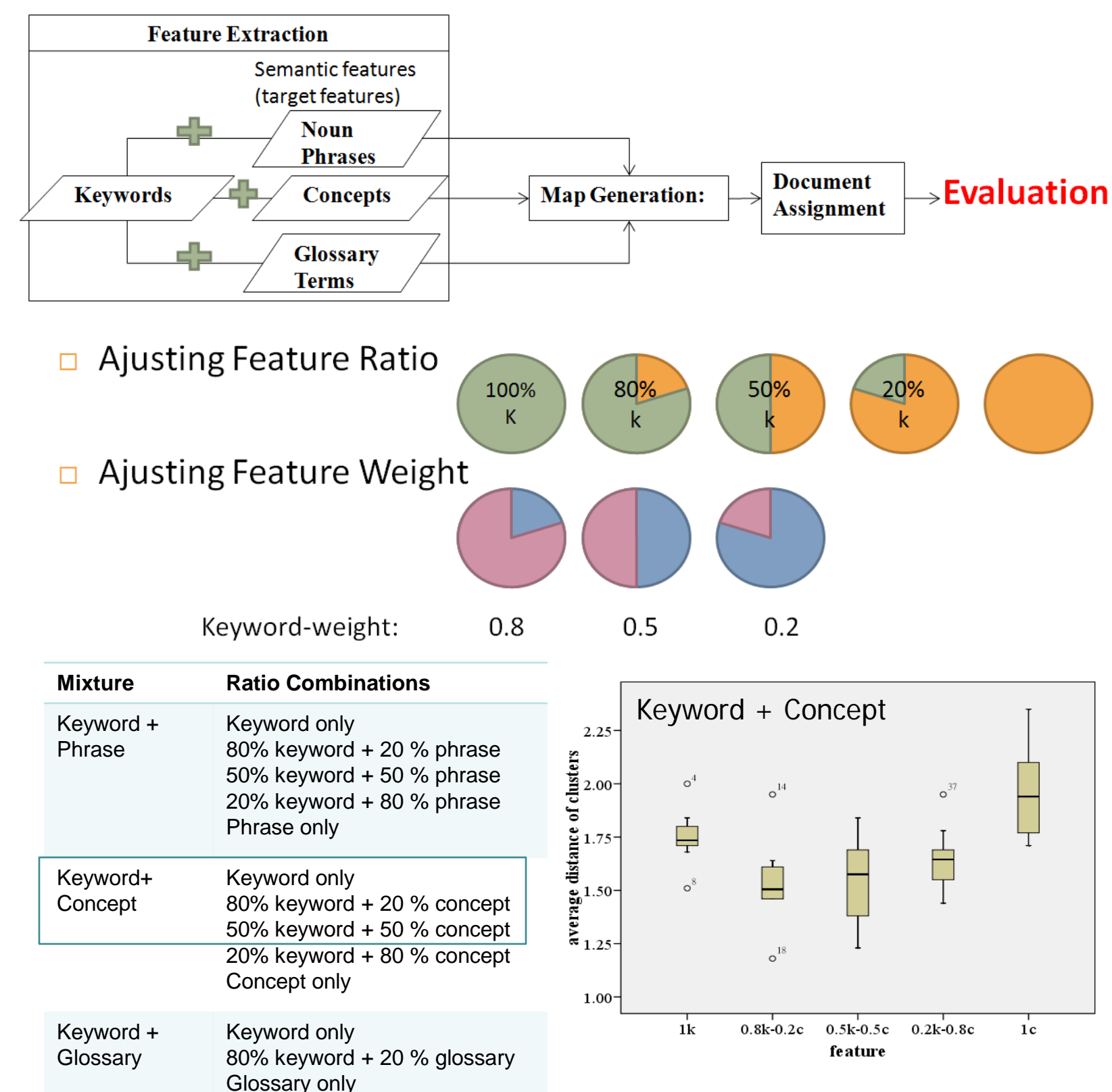


## EVALUATION

### Individual feature analysis



### Feature mixture analysis



## RESULTS

- The maps in the feature mixture adjusted with feature ratios were significantly better than the maps with the keyword only feature in the single book corpus.
- Based on the performance of the different feature ratios, the mixtures with the higher keyword ratio (up to 0.8) produced the better performance.
- The maps with keyword only outperformed the maps with any single semantic feature.
- Adjusting feature weights did not help to enhance the performance.
- Semantic features had no effect on the multiple book corpus.
- Semantic representations need to be consistent across the entire corpus in order to be effective for producing higher quality maps.

## LIMITATION

- Semantic improvement - Manual concept extraction might provide better quality semantic representation.
- Evaluation method - Different evaluation method might provide different results.

## FUTURE WORK

- Explore the reason for the success of the feature mixture approach
- Corpus consistence issues
- Terminology issues

