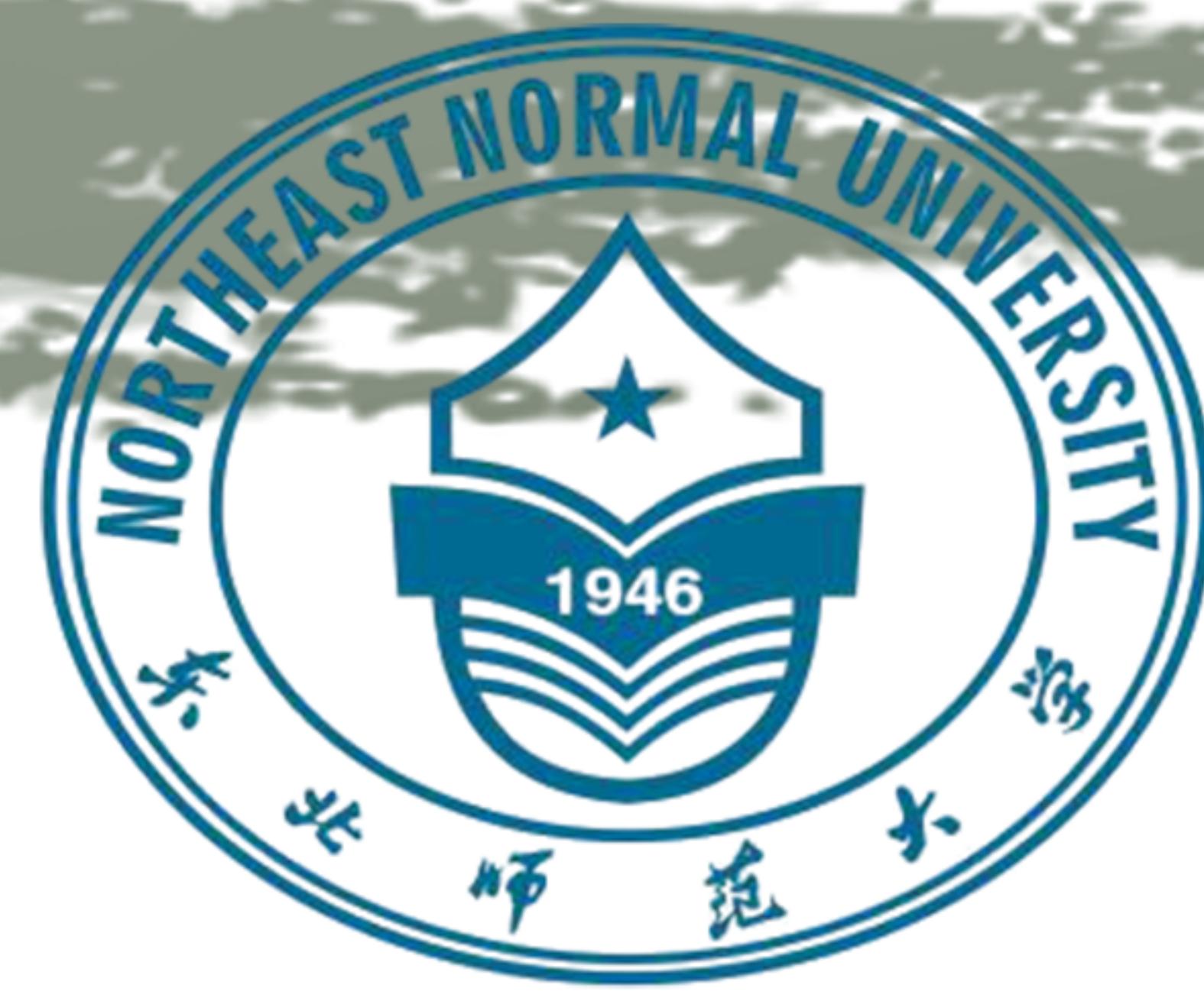


An Integrated Method to Build Read-Aiding E-Books Based on Text Mining and Interactive Aesthetics

Weijun Li (liwj531@nenu.edu.cn), Fukai Yang, Ruibing Jia, Renmin Li, Minghao Yin

We propose a four-stage method to build read-aiding e-books directly from text, integrating the technologies in text mining and the ideas in interactive aesthetics. By applying this method, we manage to identify and present multiple complex relationships in the classical Chinese novel *Romance of the Three Kingdoms*, providing readers with vivid scenes and rich interactions for better comprehension.



Case Study : Romance of the Three Kingdoms

Introduction

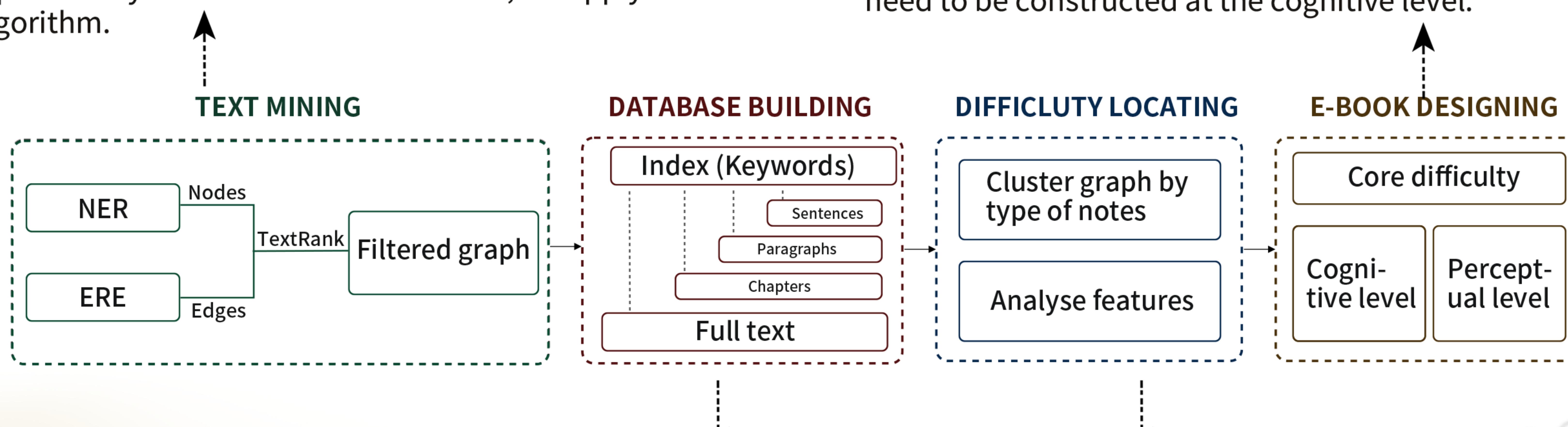


Complex plots and obscure language in books are usually unfriendly for readers, and text mining techniques such as extracting keywords and relationships are helpful in solving this problem. In order to reduce the cognitive load of the readers and build an immersive interactive environment that helps readers understand the contexts of the text, we propose a universal method for building interactive e-books based on text mining technologies and interactive aesthetics.

The Design Method

We utilize the techniques of named entity recognition and entity relationship extraction, to build a graph of items (as nodes) and their relationships (as edges). To avoid overwhelming the reader with potentially excessive amount of items, we apply the TextRank algorithm.

The designing of e-books follows the principles of interactive aesthetics and should consider both cognitive and perceptual levels. Interfaces for multisensory interactions is provided at the perceptual level, while immersive interactive scenarios need to be constructed at the cognitive level.



For the convenience of finding all corresponding text associated with each keyword, we need to build a database from the original text. It preserves all hierarchical information like chapters, sections and paragraphs, and uses the set of keywords to construct inverted indices.

After extracting keywords, we need to find out the core difficulty in a specific text. Note that each node (or extracted item) in the graph has an attribute indicating their type, such as name of characters, places, or time of events. Thus we could cluster the graph by these types and analyze their features to grab such features semiautomatically.

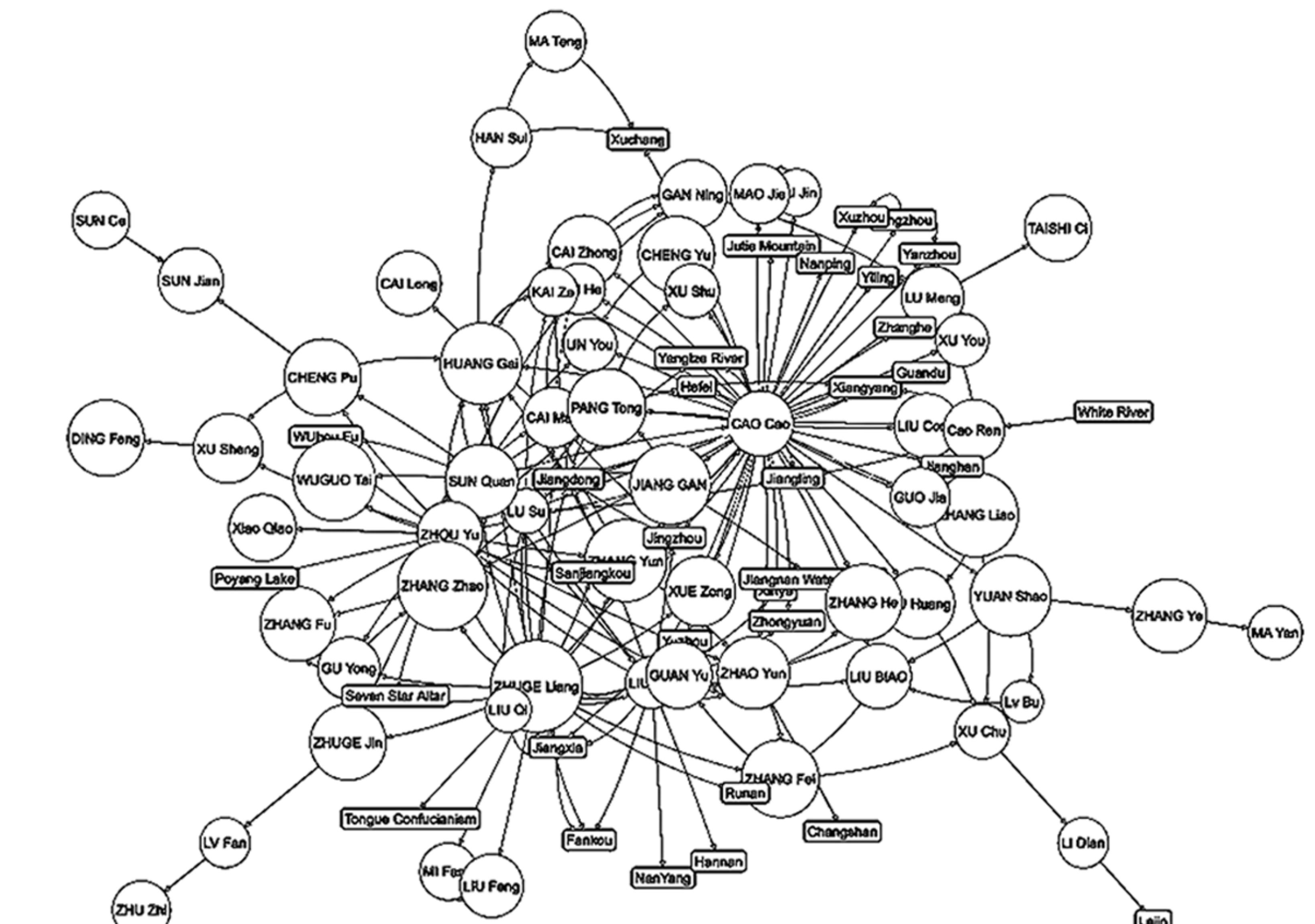


Figure above is the preliminary visualization after going through first three stages, showing that our work could extract the relationship between the characters and the places well.

We set a starting point for reading and generate reading paths based on the original text. Users can click on the nodes to view the original text. The links between one node and its related places and characters are highlighted upon clicking. By referring to the prompted reading paths, users can easily decide what to read next.

