**Reading Summary**

**Shuning Chen**

**Andrew ID: shuningc**

**Chapter 2.4**

In this section, solutions for dealing with phrase queries are discussed.

One of the techniques is biword index. When trying to search for a two-word phrase, we can compare it to combination of two words. Usually, parsing of speech text will be done before comparison. Nouns and prepositions will be separated and as such longer word sequences can be formed. The disadvantage is that not every word sequence contains two words. If you want to search a one-word term, you have to compare it with every two-word sequence. And some terms containing greater than two words might also appear. One word or more words sequence can be implemented similar to biword index. However, the most frequent terms are still consisted of one word.

Another technique is positional index. When storing posting lists, we not only store the document IDs, we but also store frequency and position information. To parse searching results containing same words scattered at different places, we use positional index to identify them.

For both solutions, storage consumption increases. When using positional index, the disk space taken will increase as the increase of average document size. Even though the documents would be compressed, the space consumption is still 2 to 4 times the size of the single documents storage.

Combination of the two techniques is available. Next word index is a strategy combining the two. The position information and phrase will both be recorded. It takes less time to get the results and accordingly, more disk space required.