

Erick Brener Moreira

O objetivo foi aplicar conceitos de orientação à objetos para encapsular todas as possíveis entidades que serão obtidas a partir do texto.

Em síntese, foram criados três objetos que juntos formam todas as partes do texto:

- Word - objeto de palavra
  - uma String para armazenar a palavra,
  - um contador para armazenar quantas vezes a palavra apareceu no texto,
  - um armazenador da distância para os casos em que a palavra se repete dentro de um parágrafo.
- Phrase - objeto de uma frase
  - uma String para armazenar a frase completa,
  - uma lista de palavras (Word) para armazenar todas as palavras daquela frase,
  - uma lista de palavras de conexão (Stop Word) para armazenar todos os conectores daquela frase.
- Paragraph - objeto de um parágrafo
  - uma String para armazenar o parágrafo completo,
  - um armazenador da primeira linha do parágrafo dentro do texto completo,
  - uma lista de frases (Phrase) para armazenar todas as frases do parágrafo,
  - uma lista de repetições (Word) para armazenar todas as palavras que se repetiram mais de uma vez dentro do parágrafo.

Por conta dessa estrutura, para obter todas as informações de um texto foi necessário iterar sobre cada linha.

Logo, o fluxo fica da seguinte forma:

1. O texto é lido de um arquivo txt e suas linhas são inseridas em uma lista de Strings.
2. Acontece uma iteração para capturar cada parágrafo, adicionando cada linha dentro de um buffer. Quando uma linha estiver vazia, significa que há um início ou fim de parágrafo.
3. Ao encontrar um fim de parágrafo, há um tratamento para agrupar as frases que podem ter sido separadas na lista inicial.

- a. Nesse tratamento, as frases são criadas e adicionadas na lista de frases do parágrafo. Assim como o parágrafo inteiro é concatenado.
4. A partir dessa tratativa, todas as palavras de cada frase desse parágrafo são tratadas e inseridas nas listas de palavras de cada frase.
  - a. Dentro desse tratamento, também é realizada a contagem de palavras internas de cada frase.
5. Após essas inserções, é realizada a leitura e junção das palavras em nível de frase, ou seja, para cada palavra que se repetiu é realizada a inserção dentro das repetições do parágrafo.
  - a. Também é calculada a distância dessa repetição.
6. Já na última etapa do fluxo, que é apenas para apresentação, todas as palavras de cada frase dentro dos parágrafos são inseridas em uma mesma lista, contadas e dispostas em ordem alfabética. Assim como, a apresentação de todas informações importantes: linha inicial do parágrafo, número de sentenças, número de palavras e conectores de cada sentença, as repetições do parágrafo que incluem a palavra e a distância entre elas, e por fim, todas as palavras e conectores contadas e ordenadas em ordem alfabética.

Devido ao fato da estrutura escolhida foram necessárias diversas iterações, devido à disposição e encapsulamento das informações. Dito isso, é possível afirmar que a complexidade total apenas para obter todas as informações foi de: On3 ou On4.