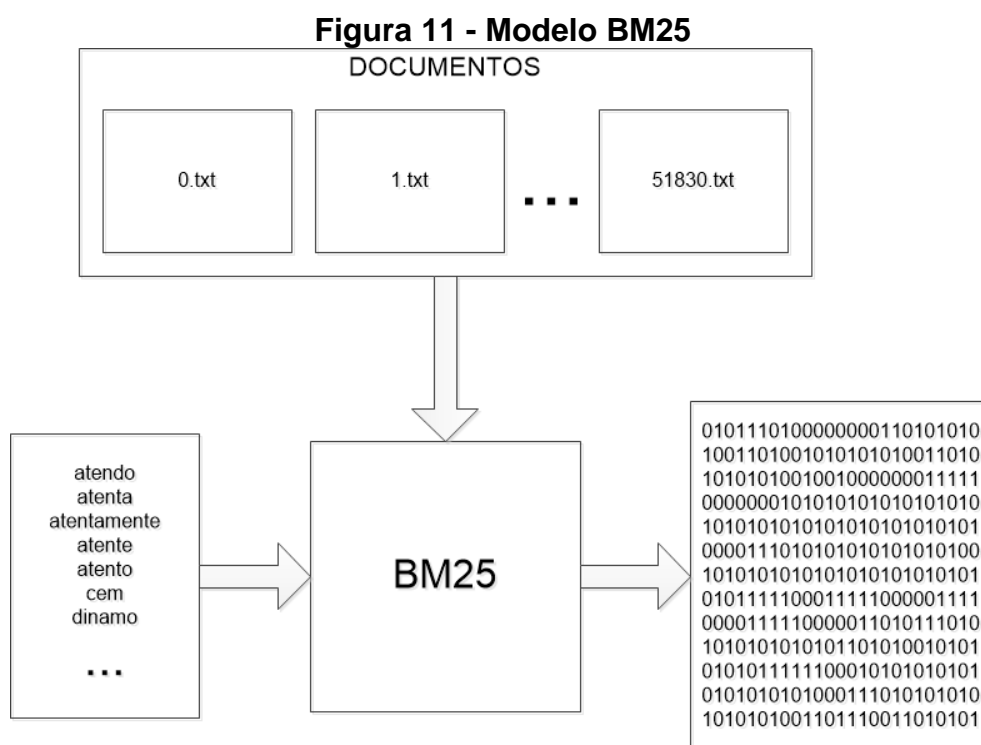


- **“COLECAO BM25”**: é o diretório onde os vetores dos documentos são armazenados;
- **“Tamanho dos documentos.txt”**: armazena a quantidade de caracteres de cada documento. A quantidade de caracteres é o tamanho de cada um deles;
- **“Tamanho medio dos documentos.txt”**: tamanho médio dos documentos calculado com base no número de caracteres;
- **“Vetores BM25.txt”**: contém o conjunto de vetores do BM25 agrupados.

Os diretórios **“DOCUMENTOS”** e **“VISA0 LOGICA”** e o documento **“Vocabulario.txt”** são compartilhados com o MPC.

A Figura 11 ilustra o processo de geração dos vetores do BM25. Primeiro os documentos e o vocabulário são carregados na memória principal do computador. Se um termo do vocabulário estiver presente em um documento, o passo seguinte é a contagem das ocorrências desse termo, se não estiver presente nenhum processamento é feito. A construção dos vetores do BM25 também pode ser interrompida assim como na construção dos vetores binários do MPC.



Fonte: Elaborado pelo autor

Quando uma consulta é enviada para o BM25, uma lista de documentos ordenados em ordem decrescente de relevância de conteúdo é feita aplicando-se a equação de similaridade do BM25.