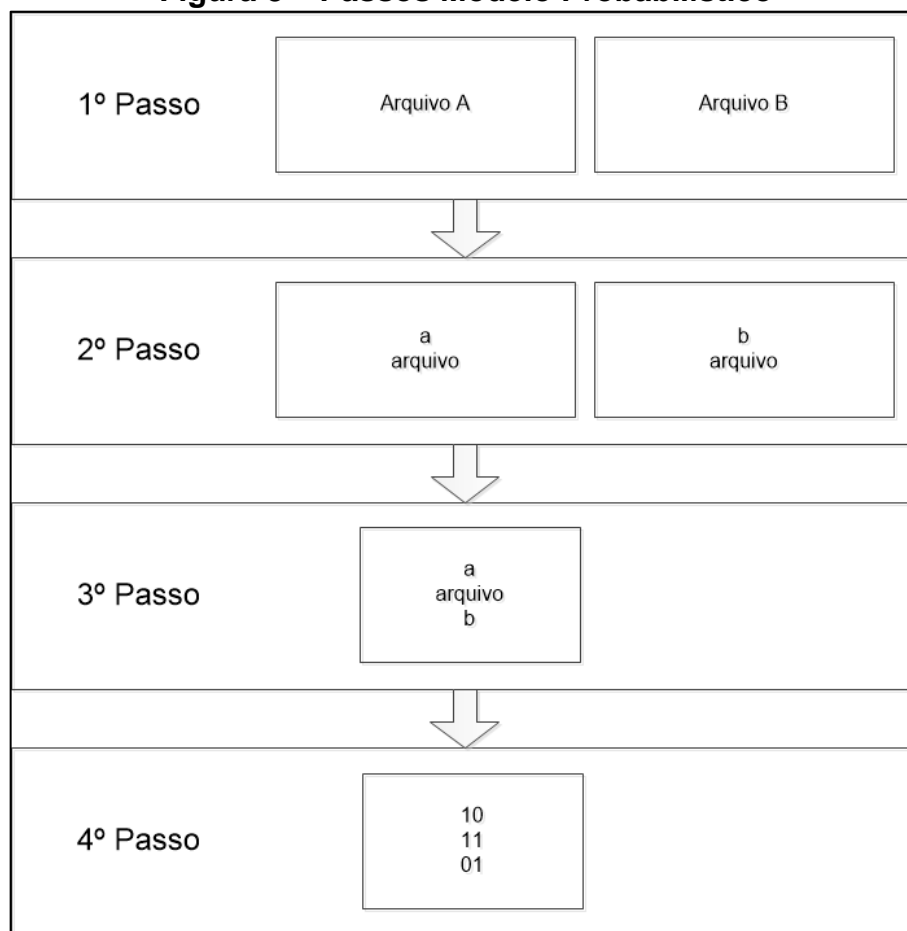


Figura 8 – Passos Modelo Probabilístico

Fonte: Elaborado pelo autor

A Figura 8 ilustra os passos utilizados para processar os documentos utilizados no MPC, nela um documento com o texto “Arquivo A” e outro com o texto “Arquivo B” passam por etapas, ou passos, que tornam a recuperação possível. O primeiro passo é a obtenção e seleção dos documentos da base; o segundo é formado pela normalização do texto e pela criação dos documentos de visão lógica (coleção de termos sem repetição); o terceiro passo é o processamento das visões lógicas, que gera um único documento de termos únicos, que seria o vocabulário da coleção; o quarto e último passo é a criação dos vetores binários para cada documento, cada vetor binário indica a presença (1) ou a ausência (0) dos termos do vocabulário em cada documento.

Um vetor é uma estrutura de dados composta de dados de mesmo tipo alocados sequencialmente (SILVA, 2012).

Deve-se atentar que para o funcionamento correto do conjunto formado por Interpretador e MPC, as consultas devem ser habilitadas na inicialização do AC. É importante dizer também que este componente tem uma coleção de vetores