

Discussão

Note que tanto o BoW quanto a consulta são passados por pré-processamentos e retiradas de stopwords afim de gerar um grau de similaridade mais preciso, trazendo aqueles documentos que de fato tem mais relevância com a consulta.

Ocorre que quando utilizado o Stemming alguma das consultas tiveram resultados diferentes, isso ocorre porque essa 'radicalização' das palavras, ou seja, transformar uma palavra em sua forma raiz, como já comentado no PDF anterior sobre seu modo de funcionamento, pode agregar termos. Imagine que em um documento/consulta existem muitas palavras derivadas de um mesmo radical: amor. Exemplo: amando, amaria, amado, amei etc.

Quando aplicado o Stemming todos esses termos se transformarão em ama, com um TF maior seu grau de similaridade também será afetado, explicando o porquê do rankeamento diferente.