

Técnicas de spamdexing têm "assombrado" os motores de busca por mais de uma década e ainda são um problema hoje em dia. Muitas técnicas baseadas em conteúdo para detectar esses métodos já foram propostas na literatura, mas a pesquisa sobre técnicas de detecção de spamdexing baseadas em links é recente. Este trabalho tem como foco o estudo e implementação de duas dessas técnicas baseadas em links, medindo suas qualidades com métricas de Recuperação de Informações (RI) bem conhecidas, e comparando seus resultados.

A ideia geral por de trás de cada uma dessas técnicas é criar um algoritmo de ranking que avalia as páginas em um grafo Web de acordo com algumas propriedades. Após, um classificador é criado, se utilizando dos resultados destes algoritmos junto com um pequeno conjunto de hosts pré-rotulados por humanos como sendo spam ou não spam. Com este classificador, o restante das páginas presentes no grafo são classificadas e, com as métricas de RI, esta classificação é então avaliada. Além disso, experimentos com a combinação de ambas as técnicas também foram realizados.

Os resultados obtidos foram, para todos os casos, muito satisfatórios, principalmente quando os resultados de ambos os algoritmos são combinados em um único classificador. Com isso, podemos concluir que, embora o problema gerado pelos métodos de spamdexing ainda não possa ser cem por cento eliminado, muito já pode ser feito para melhorar a resposta dos motores de busca para as consultas submetidas por usuários.