

## 6

### Conclusão

Como é cediço, o crescimento denso e desorganizado da *WWW* enseja a formação de uma estrutura extremamente complexa, pois congrega milhões de participantes em tempo real, todos potenciais criadores de novas informações e conhecimento.

Cuida-se, a *WWW*, de uma estrutura global consistente em verdadeira rede de dados interconectados composta por páginas e hiperlinks, e que tem sido objeto de intenso estudo, especialmente no tocante à organização da vasta informação nela contida.

A presente tese, como se vê, insere-se no contexto desse estudo, pois objetiva, através da utilização do modelo aqui exposto, denominado *XHITS*, melhorar os resultados obtidos por qualquer usuário após busca parametrizada na rede.

O *XHITS* proposto neste trabalho, como exposto ao longo dos capítulos, traduz uma generalização do *HITS* desenvolvido por Jon Kleinberg, pois procura inserir na estrutura desse modelo outras tantas categorias de páginas *Web*, além das então utilizadas por esse autor, *hubs* e *authorities*, de modo a, dessa forma, viabilizar nova metodologia de busca e retorno da informação contida na rede.

Forte em tal premissa, o *HITS* é aqui reescrito em termos de matrizes de bloco, o que culmina em um sofisticado mecanismo para estender o modelo de Kleinberg.

Com base nessa nova concepção, são exploradas as ideias de *hubs* e autoridades e, assim, o modelo é estendido em  $k$  outras categorias.

Em seguida, é desenvolvida a ideia de estrutura de influência, sendo ela responsável por revelar a propagação em hiperlink,  $F$ , e o reforço de categorias,  $B$ , que faz o ajuste fino do modelo e o torna independente da consulta.

Ainda, é explorado o conceito de reforço simétrico, em que  $B = F^T$ , bem como provada a convergência do modelo.

Entretanto, mesmo com  $B = F^T$  todos os elementos da matriz  $F$  necessitam ser definidos. Então, é proposto um processo de aprendizagem baseado em Decomposição de Valores Singulares para obter os melhores valores para os elementos de  $F$ .

Para a realização dos testes necessários, utilizamos o cenário das competições *TREC Web Track 2009* e *TREC Web Track 2010*, não somente pela posição de referência em RI que tais competições ocupam, mas também porque a coleção de dados ofertada, *ClueWeb09*, figura como a primeira que permite bom desempenho aos algoritmos baseados apenas em hiperlinks.

E, após os inúmeros experimentos aqui realizados, como oportunamente detalhados em capítulo próprio, chegamos a resultados que têm o condão de mostrar o potencial do modelo *XHITS* à luz do escopo proposto.

Como dito, treze grupos diferentes participaram da *Web Track 2009* e o desempenho do *XHITS*, em comparação, é equivalente ao do competidor da sexta posição, utilizando a base de 2010 como treinamento. Ainda, dos vinte grupos diferentes que participaram da *Web Track 2010*, o desempenho do *XHITS* é, em comparação, equivalente ao do competidor da sexta posição, ou quarta posição se alterada a métrica de ordenação.

Os resultados aqui obtidos através da utilização do modelo *XHITS*, que apenas se vale da informação proveniente de hiperlink para classificar as páginas, revela ainda que a coleção *ClueWeb09* contempla estrutura densa e pode ser explorada profundamente.

Destarte, a tese desenvolvida, em resumo, contribui fortemente para o modelo *HITS* generalizado, dando ensejo ao *XHITS*, através da inserção de  $k$  novas categorias, assim permitindo que sejam extraídas mais informações das estruturas de hiperlink das páginas *Web* e melhorando as suas classificações.

O presente trabalho viabiliza, pois, um processo de aprendizagem eficiente e capaz de ajustar os pesos gerados pelas inserções das novas  $k$  categorias, mostrando que o *XHITS* é uma proposta competitiva.

## 6.1

### Trabalhos Futuros

À luz de todo o estudo aqui desenvolvido, podemos indicar a existência de diversas direções interessantes para trabalhos futuros, tais como:

- propor a transformação do modelo *XHITS* de forma que ele fique inteiramente independente de consulta, ou seja, conseguir classificar as páginas *Web* sem precisar construir o grafo induzido;
- propor um processo de aprendizagem mais eficiente, que permita o treinamento do modelo em menos tempo, assim amplificando a análise da qualidade do *XHITS* no tocante à classificação das páginas *Web*;
- realizar um estudo sobre a qualidade das classificações das páginas *Web* geradas pelo *XHITS* a partir do julgamento de relevância exposto por usuários através de modelos estatísticos;
- fazer um estudo sobre as métricas de avaliação de desempenho com o escopo de verificar se estão as mesmas adequadas ao *XHITS*;
- e, por fim, empreender um estudo sobre o número de categorias de páginas *Web* ideal a ser utilizado no modelo *XHITS* com o fito de melhorar a respectiva classificação.