Universidade de Brasília

Faculdade do Gama – Engenharia de Software

Aluno: Greg Ouyama Martins

Orientador: Sérgio Antônio Andrade de Freitas

**Trabalho de Conclusão de Curso 2**

**Busca e Ordenação de Informação Relevante Utilizando Recuperação de Informação Web**

**Resumo**

A busca de informação teve seu princípio através de bibliotecas, buscando o que o usuário viria a necessitar através de buscas, por meio de, por exemplo, cartões de catálogos, categorizando os livros por título, por autor, ano ou editora dos livros. Com o avanço da tecnologia, ocorreu a automação deste processo de busca fazendo com que esse tipo de busca fosse realizada através de um computador.

Juntamente com o avanço tecnológico, ocorreram três fatores para que houvesse a popularização da internet para o avanço nas buscas Web, que se consideradas os mecanismos de buscas disponíveis atualmente, utilizam-se de indexações muito parecidas com as que eram utilizadas por bibliotecários no passado. Dentre estes fatores, suas causas são o barateamento do acesso a várias fontes de informações, o avanço de todos os tipos de comunicações digital, promovendo o acesso rápido e fácil à rede e por fim a liberdade de publicar qualquer informação que se julgar útil [1].

Entretanto com o grande volume de informação disponível, nem sempre é fácil encontrar que se procura, alguns julgam até mesmo impossível, tornando assim a atividade de busca cansativa e trabalhosa. Para tratar este problema existem estudos e implementação a respeito da ordenação de informação obtida através da recuperação de informação. Através de estudos a cerca de algoritmos dinâmicos e estáticos de ordenação da recuperação da informação [2][3], o objetivo deste trabalho é analisar o que existe tratando de ordenação da recuperação de informação e como a engenharia de software pode contribuir com a ordenação da recuperação de informação.

Para realizar essa análise, serão utilizados um framework de um motor de busca completo, com todos os componentes necessários para seu funcionamento, incluindo a implementação de um algoritmo de ordenação, em java, chamado Lucene, e a biblioteca JUNG que auxilia na implementação de *network graphs*, porém será utilizada para implementar perfis dentro dos algoritmos de medição de importância, como *PageRank* e HITS.

**Bibliografia**

[1] Baeza – Yates, R.; Ribeiro – Neto, B. Modern Information Retrieval, ACM press, 1999.

[2] Aditi Sharma, Nishtha Adhao and Anju Mishra. Article: A Survey: Static and Dynamic Ranking. International Journal of Computer Applications 70(14):7-12, May 2013. Published by Foundation of Computer Science, New York, USA.

[3] Sharma, Dilip et al. Aticle: Comparative Analysis of Web Page Rank Algorithms. International Journal on Computer Science and Engineering 02(08):2670-2676, November 2010. Published by Engg Journals Publications, Chennai, Tamil Nadu, India.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Greg Ouyama Martins

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Sérgio Antônio Andrade de Freitas