

Marco Serpa Molinaro

Improved Approximations for the k -Hotlink
Assignment Problem and for Binary
Searching in Trees

MsC Thesis

Thesis presented to the Post-graduate Program in Computer
Science of the Computer Science Department, PUC-Rio as
partial fulfillment of the requirements for the degree of Master
in Computer Science

Adviser: Prof. Eduardo Sany Laber

Rio de Janeiro
April 2008

Marco Serpa Molinaro

**Improved Approximations for the k -Hotlink
Assignment Problem and for Binary
Searching in Trees**

Thesis presented to the Post-graduate Program in Computer
Science of the Computer Science Department, PUC-Rio as
partial fulfillment of the requirements for the degree of Master
in Computer Science. Approved by the following commission:

Prof. Eduardo Sany Laber

Adviser

Department of Computer Science — PUC-Rio

Prof. Marcus Poggi de Aragão

Departamento de Informática — PUC-RIO

Prof. Valmir Carneiro Barbosa

COPPE — UFRJ

Prof. Jayme Luiz Szwarcfiter

COPPE — UFRJ

Prof. José Eugenio Leal

Head of the Science and Engineering Center — PUC-Rio

Rio de Janeiro — April 28, 2008

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Marco Serpa Molinaro

Marco Molinaro graduated from Pontifícia Universidade Católica do Rio de Janeiro in Computer Engineering. During his undergraduate years, he participated in multiple research projects funded by CNPq. For the duration of his Masters, he was supported by an academic scholarship and performed theoretical research focusing on approximation algorithms.

Ficha Catalográfica

Molinaro, Marco Serpa

Improved approximations for the k-Hotlink Assignment Problem and for binary searching in trees / Marco Serpa Molinaro; adviser: Eduardo Sany Laber. - 2008.

89 f. : il. (col.) ; 30 cm

Dissertação (Mestrado em Informática)-Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia.

1. Informática – Teses. 2. Algoritmos aproximativos. 3. Hotlinks. 4. Busca. 5. Árvores. I Laber, Eduardo Sany. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Acknowledgments

To my adviser Professor Eduardo Laber, for his patience and dedication during this time. He has been an amazing source of inspiration, technical knowledge and encouragement.

To my family, for giving me the freedom to pursue the professional track I wanted.

To my colleagues at PUC–Rio, for both stimulating conversations about various scientific topics and for their soothing words during hard times.

To professors José Correa and Marcos Kiwi for invaluable discussions on the problems presented in this thesis.

To my professors at PUC–Rio, in particular to Marcus Poggi and Ruy Milidiú, for their constant guidance and support since my undergraduate days.

To CNPq, CAPES and PUC–Rio, for the financial support.

Abstract

Molinaro, Marco Serpa; Laber, Eduardo Sany. **Improved Approximations for the k -Hotlink Assignment Problem and for Binary Searching in Trees**. Rio de Janeiro, 2008. 89p. MsC Thesis — Department of Computer Science, Pontifícia Universidade Católica do Rio de Janeiro.

Here we present a study on two optimization problems in trees: the k -Hotlink Assignment Problem and the problem of Binary Searching in Trees. As a result, we obtain improved approximation algorithms for both problem. The k -Hotlink Assignment Problem can be defined as follows. Let $G = (V, E)$ be a directed acyclic graph representing a web site, where nodes correspond to pages and arcs to hyperlinks. In this context, hotlinks are defined as shortcuts (new arcs) added to web pages of G in order to reduce the time spent by users to reach their desired information. Here we consider the problem where G is a rooted directed tree and the goal is minimizing the expected time spent by users by assigning at most k hotlinks to each node. For the most studied version of this problem where at most one hotlink can be assigned from each node, we prove the existence of an FPTAS, improving upon the constant factor algorithm recently obtained in [Jacobs, WADS 2007]. In addition, we develop the first constant factor approximation algorithm for the most general version where k hotlinks can be assigned from each node.

In the second part of this work, we consider the problem of computing efficient strategies for searching in trees. As a generalization of the classical binary search for ordered lists, suppose one wishes to find a (unknown) specific node of a tree by asking queries to its arcs, where each query indicates the endpoint closer to the desired node. Given the likelihood of each node being the one searched, the objective is to compute a search strategy that minimizes the expected number of queries. Practical applications of this problem include file system synchronization and software testing. Here we present a linear time algorithm which is the first constant factor approximation for this problem. This represents a significant improvement over previous $O(\log n)$ -approximation.

Keywords

Approximation Algorithms. Hotlinks. Search. Trees.

Resumo

Molinaro, Marco Serpa; Laber, Eduardo Sany. **Improved Approximations for the k -Hotlink Assignment Problem and for Binary Searching in Trees**. Rio de Janeiro, 2008. 89p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese apresenta um estudo sobre dois problemas de otimização em árvores: o *k -Hotlink Assignment Problem* e o problema de busca binária em árvores. Como resultado, obtemos algoritmos aproximativos melhorados para ambos os problemas.

O *k -Hotlink Assignment Problem* pode ser definido da seguinte maneira. Seja $G = (V, E)$ um dígrafo acíclico representando um web site, onde nós correspondem a páginas e arcos correspondem a hyperlinks. Nesse contexto, hotlinks são definidos como atalhos (novos arcos) adicionados a páginas em G de modo a encurtar o tempo gasto por usuários para alcançar as informações desejadas. Aqui consideramos o problema onde G é uma árvore enraizada e o objetivo é minimizar o tempo médio gasto por usuários adicionando no máximo k hotlinks a cada nó. Para a versão mais estudada desse problema onde no máximo um hotlink pode ser atribuído a cada nó, nós provamos a existência de um FPTAS, uma melhora sobre a aproximação de fator constante obtida recentemente em [Jacobs, WADS 2007]. Além disso, desenvolvemos a primeira aproximação de fator constante para a versão mais geral onde k hotlinks podem ser atribuídos a cada nó.

Na segunda parte deste trabalho, consideramos o problema de computar eficientemente estratégias para buscar em árvores. Como uma generalização da clássica busca binária em listas ordenadas, suponha que deseja-se encontrar um nó específico (e desconhecido) de uma árvore através de consultas a seus arcos, onde cada consulta indica qual extremidade do arco está mais próxima ao nó desejado. Dada a probabilidade de cada nó ser aquele procurado, o objetivo é encontrar uma estratégia que minimize o número esperado de consultas. Aplicações práticas desse problema incluem sincronização de sistemas de arquivo e criação de testes de software. Apresentamos um algoritmo que executa em tempo linear e que provê a primeira aproximação de fator constante para esse problema. Esse resultado representa uma significativa melhora sobre a $O(\log n)$ -aproximação conhecida anteriormente.

Palavras-chave

Algoritmos Aproximativos. Hotlinks. Busca. Árvores.

Contents

I	k-Hotlink Assignment Problem	9
1	Introduction	10
2	Preliminaries	15
3	An Exact Exponential Algorithm for the 1-Hotlink Assignment Problem	18
3.1	Solving 1-HAPG	20
4	An FPTAS for the 1-Hotlink Assignment Problem	24
4.1	Proof of Theorem 2	26
5	Constant Factor Approximation for the k -Hotlink Assignment Problem	37
5.1	Upper Bound	38
5.2	Entropy Lower Bound	39
5.3	Alternative Lower Bound	42
5.4	Approximation Guarantee	45
II	Binary Searching in Trees	47
6	Introduction	48
7	Preliminaries	51
8	Algorithm for Binary Searching in Trees	54
8.1	Upper Bound	56
8.2	Entropy Lower Bound	58
8.3	Alternative Lower Bounds	59
8.4	Approximation Guarantee	60
8.5	Efficient Implementation	61
	Bibliography	64
A	Hotlink Assignment Problem	68
A.1	Preliminary lemmas	68
A.2	Proofs of lemmas used in Theorem 1	72
A.3	Proof of Lemma 4	76
A.4	Constant factor approximation for the k -Hotlink Assignment Problem	78
	<i>Solving k-HAP with internal weights</i>	78

Approximate no-leaf assignment	78
Algorithm for k-HAP with internal weights	82
<i>Proof of Lemma 10</i>	83
B Binary Searching in Trees	85
B.1 Searching in path-like trees	85
B.2 Proofs of lemmas	86
Contents	