

Proposta para Trabalho de Formatura:
Aplicações de uma solução eficiente para o
problema do Menor Ancestral Comum

Alexandre L. J. H. Albano

IME USP 2006

Caráter

O trabalho será baseado em uma Iniciação Científica, que teve início em março de 2006, sob orientação do Prof. Dr. Alair Pereira do Lago.

Objetivos

Estudar algoritmos que se beneficiam de uma solução eficiente do problema do Menor Ancestral Comum (MAC), como a listagem de documentos e a busca de subsequência comum a duas palavras mais comprida.

Atividades já realizadas

Foi estudada uma solução eficiente do MAC e uma redução ao MAC do problema do mínimo de um vetor (permite-se pré-processamento, deseja-se consulta em tempo constante). Também foi estudada a principal estrutura de dados utilizada no algoritmo da listagem de documentos (a árvore de sufixos).

Estrutura esperada

Inicialmente, será explicado o problema do MAC, e em seguida será enunciada a existência de uma solução bastante eficiente. Esta solução não será exposta ao longo do texto: o trabalho se aterá a exibir as aplicações que se beneficiam por uma tal solução eficiente.

Em um segundo momento, será introduzido o problema de obter-se o mínimo de um vetor no seguinte contexto: pré-processamento é permitido, mas deseja-se responder a consultas em tempo constante. A seguir, será mostrado como este problema se reduz ao MAC.

Após terem sido introduzidos estes dois problemas, serão expostos dois algoritmos que precisam resolver internamente instâncias destes dois problemas. Mais especificamente, será apresentado um algoritmo que resolve o problema da listagem de documentos (que contenham determinadas palavras), e um segundo algoritmo que resolve o problema da busca de uma subsequência comum mais comprida.

Estes dois algoritmos serão devidamente motivados e terão suas aplicações explicitadas: O algoritmo apresentado para a listagem de documentos é de

veemente importância para a área de recuperação de informação, enquanto o algoritmo para a busca de uma subsequência comum mais comprida é de grande interesse para diversas áreas, como por exemplo a bioinformática, devido à existência de aplicações que comparam seqüências genéticas.

Outros algoritmos (correlatos a estes dois) poderão eventualmente ser abordados de forma breve.

Esquemáticamente, a seguinte estrutura é esperada:

- 1) O problema do menor ancestral comum (MAC)
 - 1.1) Definição
 - 1.2) Descrição da solução ótima

- 2) O problema do mínimo de um vetor
 - 2.1) Definição
 - 2.2) Redução ao MAC

- 3) Listagem de documentos
 - 3.1) Definição
 - 3.2) Motivação e aplicações
 - 3.3) Algoritmo dinámico $O(NM)$,
 - 3.2) Contraste com o problema da listagem de ocorrências
 - 3.3) Árvore de Sufixos
 - 3.4) Um algoritmo eficiente para a listagem de documentos
 - 3.4.1) Redução ao problema "colored range query"
 - 3.4.2) Solução ótima, resolvendo-se o mínimo de um vetor

- 4) Subseqüência comum máxima
 - 4.1) Definição
 - 4.2) Motivação e aplicações
 - 4.3) Algoritmo dinámico $O(NM)$, $N = |X|$, $M = |Y|$
 - 4.4) Um algoritmo $O(ND)$, $D = \text{diferenças}(X, Y)$
 - 4.4.1) Análise de caso médio
 - 4.4.2) Análise de pior caso
 - 4.4.3) Redução da complexidade de pior caso utilizando MAC

- 5) Parte subjetiva
 - 5.1) Desafios e frustrações
 - 5.2) Disciplinas do currículo que mais ajudaram neste trabalho

Cronograma

1. Estudar o artigo [Mye86]
2. Estudar o artigo [Mut02]
3. Preparar o cartaz
4. Redigir e revisar a monografia

	1	2	3	4
Julho	x			x
Agosto	x			x
Setembro		x		x
Outubro		x		x
Novembro			x	x

Referências Bibliográficas

- [Mut02] S. Muthukrishnan. Efficient algorithms for document retrieval problems. *Symposium on Discrete Algorithms*, 2002.
- [Mye86] Eugene W. Myers. An $o(nd)$ difference algorithm and its variations. *Algorithmica*, 1986.