

Concepção e Análise de Algoritmos

LeGrep

Turma 2 - Grupo C

José Peixoto	200603103	ei12134@fe.up.pt
Pedro Moura	201306843	up201306843@fe.up.pt

29 de Maio de 2015

Conteúdo

1	Formalização do problema	2
1.1	Dados de entrada	2
1.2	Limites de aplicação	2
1.3	Situações de contorno	2
1.4	Resultados esperados	2
2	Descrição da solução	2
2.1	Algoritmos de pesquisa exacta em strings	3
2.1.1	Naive	3
2.1.2	Baseado em autómato finito	3
2.1.3	Knuth-Morris-Pratt	3
2.1.4	Complexidade empírica	3
3	Lista de casos de utilização	3
4	Relato das principais dificuldades encontradas no desenvolvimento do trabalho	4
5	Indicação do esforço dedicado por cada elemento do grupo	4
5.1	José Peixoto	4
5.2	Pedro Moura	4

Resumo

No âmbito da unidade curricular de Concepção e Análise de Algoritmos, foi-nos proposto o desenvolvimento de uma aplicação que emulasse as funcionalidades do utilitário da linha de comandos **grep** (globally search a regular expression and print), explorando o conhecimento acerca de algoritmos para pesquisa exacta de strings.

1 Formalização do problema

1.1 Dados de entrada

- P - padrão a procurar
- $FilePath$ - caminho do ficheiro onde se pretende procurar o padrão
- T - cada linha de texto lida do ficheiro

Comandos opcionais de entrada

- $-n$, uso do algoritmo “naive” na pesquisa exacta
- $-m$, uso do algoritmo baseado em autómato finito
- $-k$, uso do algoritmo de Knuth-Morris-Pratt (pré-definido)
- $-i$, ignorar diferenças de capitalização entre caracteres
- $-v$, mostrar linhas sem correspondências
- $-h$, mostrar manual de ajuda
- $-BN$, imprimir N linhas de contexto anterior
- $-AN$, imprimir N linhas de contexto posterior
- $-CN$, imprimir N linhas de contexto anterior e N linhas de contexto posterior

1.2 Limites de aplicação

1.3 Situações de contorno

1.4 Resultados esperados

$$Output = B + T + A$$

2 Descrição da solução

É feita uma procura de correspondência de um padrão numa cadeia de caracteres ‘string de texto’ a cada linha lida de um dado ficheiro de texto.

2.1 Algoritmos de pesquisa exacta em strings

Na pesquisa de uma correspondência exacta de um dado padrão P num texto T , recorre-se a três algoritmos diferentes, escolhidos de forma opcional pelo utilizador:

2.1.1 Naive

Para um padrão de tamanho m e texto de tamanho n , encontra todos os deslocamentos s válidos, verificando desde o início do padrão a condição $P[1..m] = T[s + 1..m]$ para cada um dos $n - m + 1$ valores possíveis de s . Este procedimento implica, no pior dos casos, uma complexidade temporal de $\Theta((n - m + 1)m)$. Torna-se bastante ineficiente para padrões com tamanho grande, uma vez que se faz uma verificação de equivalência do tamanho do padrão a cada novo deslocamento.

2.1.2 Baseado em autómato finito

Cada carácter do texto é verificado apenas uma vez. A análise ao texto é feita em tempo constante $|T|$, porque se recorre à modelação de um autómato finito com um conjunto de estados Q de tamanho $|P|$. A função de transição δ possibilita a atribuição de novos estados q a cada novo carácter a lido do texto.

$$\delta(q, a) = \sigma(P_q a) \quad (1)$$

No ciclo principal verifica-se a cada novo carácter lido do texto se se atingiu o estado de aceitação e por conseguinte, uma correspondência exacta do padrão no texto.

A construção do autómato é uma primeira fase de processamento e envolve a determinação do alfabeto Σ do padrão. Na determinação da função de transição recorre-se a uma função auxiliar para o sufixo $\sigma(x)$ que retorna, para um dado conjunto de caracteres do alfabeto, Σ^* o maior comprimento (estado) que seja prefixo do padrão.

$$\sigma(x) = \max\{k : P_k \sqsubset x\} \quad (2)$$

2.1.3 Knuth-Morris-Pratt

2.1.4 Complexidade empírica

3 Lista de casos de utilização

- Multi-plataforma, testado em sistemas Linux e Windows.

4 Relato das principais dificuldades encontradas no desenvolvimento do trabalho

5 Indicação do esforço dedicado por cada elemento do grupo

5.1 José Peixoto

- Tratamento dos dados de entrada.
- Mostra dos comandos de ajuda.
- Concepção dos algoritmos “naive” e baseado em autômato finito.
- Autor do relatório.

5.2 Pedro Moura

- Concepção do algoritmo Knuth-Morris-Pratt.

Referências

- [1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, *Introduction to algorithms Third Edition*, Cambridge, MA [etc.] : The MIT Press, cop. 2009.