

## FATORES COMUNS MAIS LONGOS E SUPERMAXIMAIS

Y. KOHAYAKAWA

Data de entrega: 1/7/2013 (23:55)

## 1. INTRODUÇÃO E DEFINIÇÕES BÁSICAS

Este EP trata de processamento de cadeias de caracteres (*strings*). Supomos fixo um *alfabeto*  $\Sigma$ , cujos membros chamamos de *letras* (um *alfabeto* é simplesmente um conjunto finito). Nossas cadeias de caracteres, ou *palavras*, sobre  $\Sigma$  são sequências finitas de letras. O conjunto das palavras sobre  $\Sigma$  é denotado  $\Sigma^*$ . O comprimento  $|s|$  de uma palavra  $s \in \Sigma^*$  é o número de letras que a compõe. Dadas duas palavras  $s$  e  $t \in \Sigma^*$ , a *concatenação*  $st$  destas palavras é uma nova palavra sobre  $\Sigma$ ; claramente  $|st| = |s| + |t|$ . A palavra vazia (aquela de comprimento 0) é denotada por  $\lambda$ .

## 2. FATORES COMUNS

Uma palavra  $f$  é um *fator* de uma palavra  $s$  se  $s = s_1fs_2$  com  $s_1$  e  $s_2 \in \Sigma^*$ . O conjunto dos fatores de uma palavra  $s$  é denotado  $\text{Fat}(s)$ . Os fatores comuns de duas palavras  $s$  e  $t$  são os membros de  $\text{Fat}(s) \cap \text{Fat}(t)$ . Uma *testemunha* para o fator comum  $f \in (\text{Fat}(s) \cap \text{Fat}(t)) \setminus \{\lambda\}$  é um par  $(i, j)$  de inteiros tal que  $s = s_1fs_2$ ,  $t = t_1ft_2$ ,  $|s_1| = i - 1$  e  $|t_1| = j - 1$ . A testemunha  $(i, j)$  também será chamada de uma *repetição* de  $f$ .

**Fatores comuns maximais.** Uma repetição  $(i, j)$  de  $f \in \text{Fat}(s) \cap \text{Fat}(t)$  ( $f \neq \lambda$ ) é *maximal à direita* se  $s_2 = \lambda$  ou  $t_2 = \lambda$  ou se  $s_2 \neq \lambda$  e  $t_2 \neq \lambda$  e a primeira letra de  $s_2$  é diferente da primeira letra de  $t_2$ . Analogamente, a repetição  $(i, j)$  de  $f$  é *maximal à esquerda* se  $s_1 = \lambda$  ou  $t_1 = \lambda$  ou se  $s_1 \neq \lambda$  e  $t_1 \neq \lambda$  e a última letra de  $s_1$  é diferente da última letra de  $t_1$ . Uma repetição é *maximal* se ela for ambos maximal à direita e maximal à esquerda. Um *fator comum maximal* é um fator comum que admite uma repetição maximal. Denotemos por  $M(s, t)$  o conjunto dos fatores comuns maximais do par de palavras  $(s, t)$ .

**Fatores comuns supermaximais.** Um fator  $f$  de  $s \in \Sigma^*$  é *próprio* se  $f \neq \lambda$  e  $f \neq s$ . Dado um  $M \subset \Sigma^*$ , seja  $S(M) = \{m \in M : \text{não existe } m' \in M \text{ tal que } m \text{ é fator próprio de } m'\}$ . Vamos chamar os membros de  $S(M(s, t))$  de fatores comuns *supermaximais* de  $s$  e  $t$ .

## 3. SEU PROGRAMA

Seu EP5 deve ter como entrada um par de palavras  $(s, t) \in \Sigma^* \times \Sigma^*$  e deve ter como saída o máximo dos comprimentos dos fatores comuns de  $s$  e  $t$ , isto é  $\max\{|f| : f \in \text{Fat}(s) \cap \text{Fat}(t)\}$ . Ademais, seu programa deve também listar todos os fatores comuns supermaximais de  $s$  e  $t$ .

**Instâncias.** Você deve supor que seu EP5 será tipicamente executado com palavras  $s$  e  $t$  bastante longas. Por exemplo,  $s$  e  $t$  poderiam ser as sequências completas de nucleotídeos que compõe um cromossomo de um organismo simples. Por exemplo, veja o cromossomo III do *C. elegans* em [http://www.ncbi.nlm.nih.gov/nuccore/NC\\_003281.8?save=on&report=fasta&from=1&to=13783681](http://www.ncbi.nlm.nih.gov/nuccore/NC_003281.8?save=on&report=fasta&from=1&to=13783681) (14MB).

**Instruções adicionais.** Seu EP deve ter as seguintes formas de uso. A chamada

```
prompt$ ep5 -fs.in -ft.in
```

significa que as palavras de entrada, digamos  $s$  e  $t \in \Sigma^*$ , estão nos arquivos `s.in` e `t.in`. A chamada acima deve ter como saída o máximo dos comprimentos dos fatores comuns de  $s$  e  $t$ . Ademais, a chamada

```
prompt$ ep5 -fs.in -ft.in -s
```

deve listar os elementos de  $S(M(s, t))$ , em ordem *militar semi-reversa*: as palavras de  $S(M(s, t))$  devem ser listadas em ordem decrescente de comprimento, com as palavras de um mesmo comprimento listadas em ordem alfabética. Formalmente,  $s$  vem antes de  $t$  na ordem militar semi-reversa se  $|s| > |t|$  ou  $|s| = |t|$  e  $s$  precede  $t$  na ordem alfabética (naturalmente, supomos  $\Sigma$  ordenado).

Segue um pequeno exemplo: suponha que  $s = \text{abcbccbcbccba}$  e  $t = \text{accbaacbaccacbb}$ . Devemos ter, então,

```
prompt$ ep5 -fs.in -ft.in
```

```
4
```

```
prompt$ ep5 -fs.in -ft.in -s
```

```
ccba
```

```
bcc
```

```
prompt$
```

**Observações.** Seguem as observações usuais.

1. *Este EP é estritamente individual.* Programas semelhantes receberão nota 0.
2. Seja cuidadoso com sua programação (correção, documentação, apresentação, clareza do código, etc), dando especial atenção a suas estruturas de dados. A correção será feita levando isso em conta.
3. Comparem entre vocês o desempenho de seus programas.
4. Entregue seu EP no Paca.
5. Não deixe de incluir em seu código um *relatório* para discutir seu EP: discuta as estruturas de dados usadas, os algoritmos usados, etc. *Se você escrever claramente como funciona seu EP, o monitor terá pouca dificuldade em corrigi-lo, e assim você terá uma nota mais alta.* (Se o monitor sofrer para entender seu código, sua nota será baixa.)

*Observação final.* Envie dúvidas para a lista de discussão da disciplina.

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, RUA DO MATÃO 1010, 05508-090 SÃO PAULO, SP

Endereço eletrônico: [yoshi@ime.usp.br](mailto:yoshi@ime.usp.br)