

SETI@ONI (Mihai Pătrașcu) – Ideea de rezolvare

O metodă de a rezolva această problemă este să se ordoneze toate substringurile de lungime 16 din mesajul care trebuie analizat. Apoi, pentru fiecare cuvânt din dicționar se fac două căutări binare. Sunt necesare însă următoarele două rafinări ale ideii:

- sortarea tuturor substringurilor este consumatoare de timp. Dacă se folosesc metode precum quicksort sau heapsort, este probabil ca programul să nu treacă toate testele. Cea mai eficientă metodă de sortare în cazul de față, care se încadrează fără probleme în timp, este radixsort.
- nu se poate ține un tabel cu substringurile după ce au fost sortate. În cel mai bun caz trebuie să reținem o referință care fiecare substring (de exemplu, un pointer sau un indice), dar fiindcă pot fi peste 100,000 de substringuri, referința trebuie să fie de minim 24 de biți, ceea ce face tabloul de dimensiune inacceptabilă. Soluția constă în a împărți arbitrar mesajul în bucăți de 32,768 de caractere, și de a aplica algoritmul menționat pentru fiecare bucată. Bineînțeles, trebuie să ne asigurăm că algoritmul ia în calcul și secvențele conținute în două bucăți alăturate.

Testarea

Test	N	Sursa text	M	Sursa cuvinte	Timp**
0	1,111	"mcmcmcmc..."	10	sucesiuni de 'm' și 'c'	0.061
1	300	"aaa..."	1,000	30% secvențe de 1-15 a-uri 70%: 13 'a' și o literă aleatoare	0.075
2	500	"A Connecticut Yankee in King Arthur's Court" (by Mark Twain)	5,200	fiecare literă din alfabet de 100 de ori	0.104
3	500	"A Connecticut Yankee in King Arthur's Court" (by Mark Twain)	32,000	fiecare grup de 4 litere din text	0.254
4	1,000	"A Connecticut Yankee in King Arthur's Court" (by Mark Twain)	5,099	dicționarul englez distribuit cu Unix V5	0.115
5	1,500	"A Connecticut Yankee in King Arthur's Court" (by Mark Twain)	5,099	dicționarul englez distribuit cu Unix V5	0.170
6	2,000	"A Connecticut Yankee in King Arthur's Court" (by Mark Twain)	5,099	dicționarul englez distribuit cu Unix V5	0.223
7	2,047	"The Adventures of Tom Sawyer" (by Mark Twain)	32,000	grupuri de câte 15 litere din text	0.860
8	2,047	"The Adventures of Tom Sawyer" (by Mark Twain)	32,000	câte 15 litere din text, plus o literă aleatoare la sfârșit	0.826
9	2,047	aleator: 5% 'c', 5% 'm', 90% 'P'	32,000	1% litera 'p'; 1% litera 'P'; 98% câte 16 litere din text	0.907

** - Coloana **Timp** se referă la timpul de rulare a soluției oficiale.