

Smith-Watermanov algoritam s linearnom memorijskom složenosti

Autori: Vlatka Pavišić, Janko Sladović,
Andrija Stepić, Marko Vrljičak

Uvod

- Smith-Watermanov algoritam traži lokalno poravnanje dvije sekvence
 - Kvadratna memorijska i kvadratna vremenska složenost
- Moguće je napraviti verziju algoritma koja koristi kvadratnu vremensku složenost, ali linearnu memorijsku složenost
 - Hirschbergov algoritam

1. Korak – lokalno poravnanje

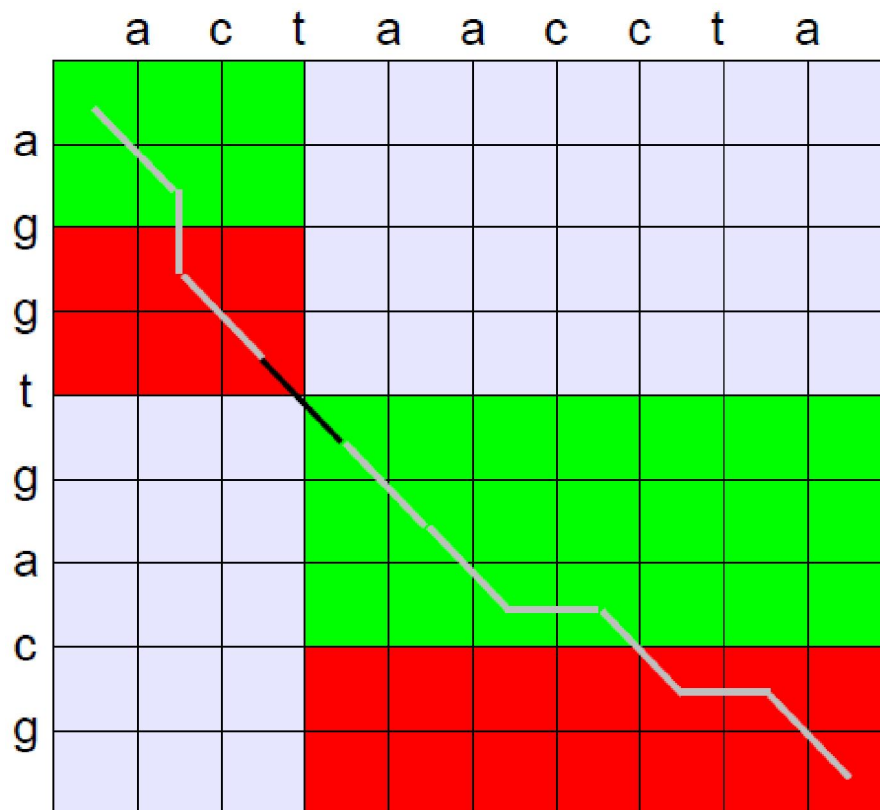
- Koristimo Smith-Watermanov algoritam kako bismo našli lokalno poravnanje dva niza
- Dovoljno je u svakom trenutku pamtiti samo dva reda tablice (prošli i trenutni)
 - Memorijska složenost tako ostaje linearna
- Nakon dva provođenja Smith-Watermanovog algoritma dobivamo optimalno globalno poravnanje

2. Korak – Hirschbergov algoritam

- Nastao iz Needleman-Wunsch algoritma
 - Sličan Smith-Wateranovom algoritmu, ali traži globalno, ne lokalno poravnanje
- Djeluje na principu „podijeli pa vladaj” kako bi se izbjegla kvadratna memorijska složenost
 - Rekurzivan algoritam
- U svakom koraku tražimo središnji brid kako bismo podijelili niz na dva podniza

Traženje središnjeg brida

- Tražimo brid koji optimalno povezuje dva središnja reda tablice



Testiranje

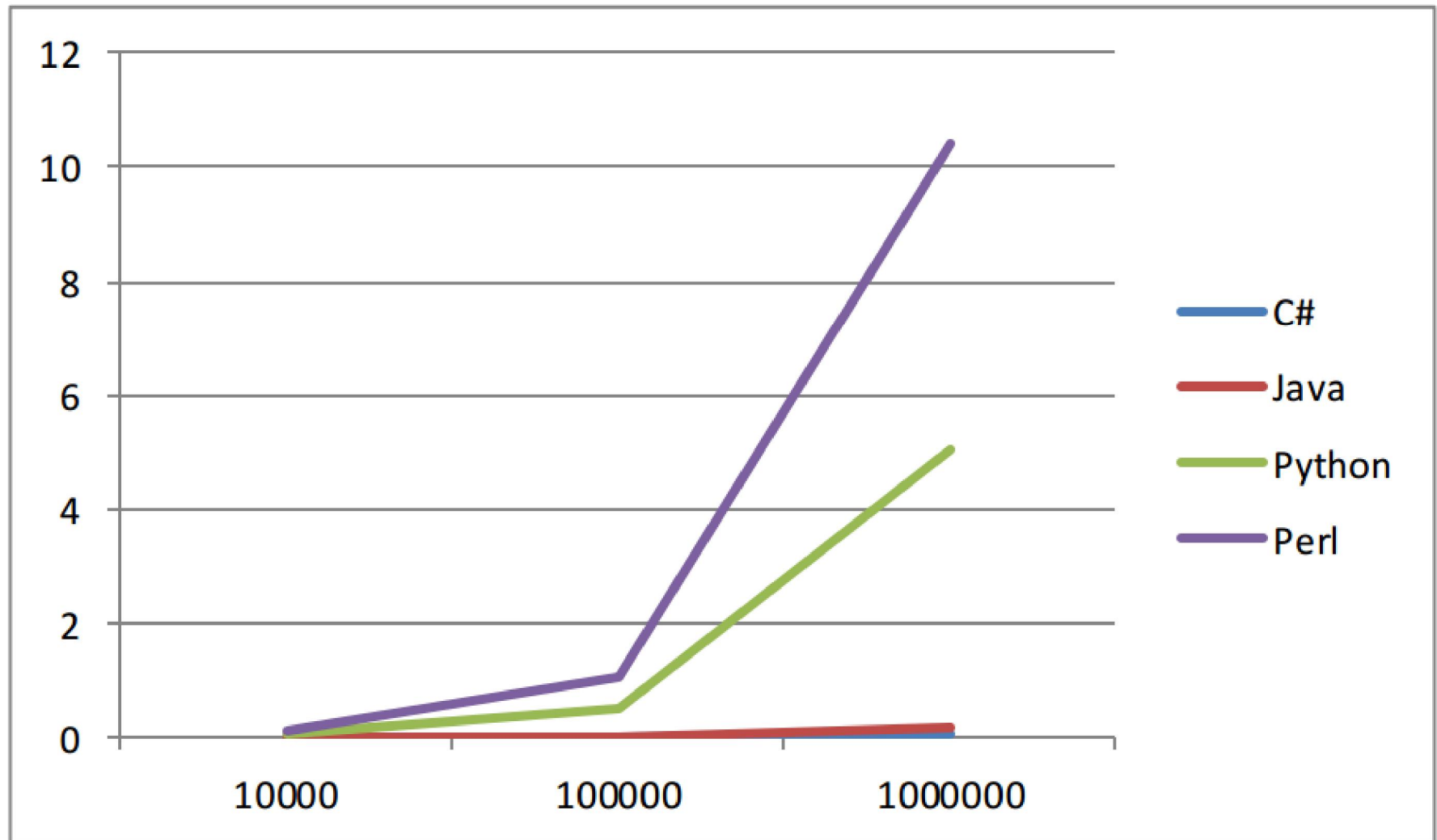
- 4 različite implementacije u 4 programska jezika
 - C#, Java, Perl, Python
- Korišteni primjeri od 100, 1000, 10 000, 100 000 te milijun znakova
- Testirano na operacijskom sustavu Biolinux 7
 - Virtualni stroj
 - 1GB RAM, 2.53 GHz procesor

Rezultati

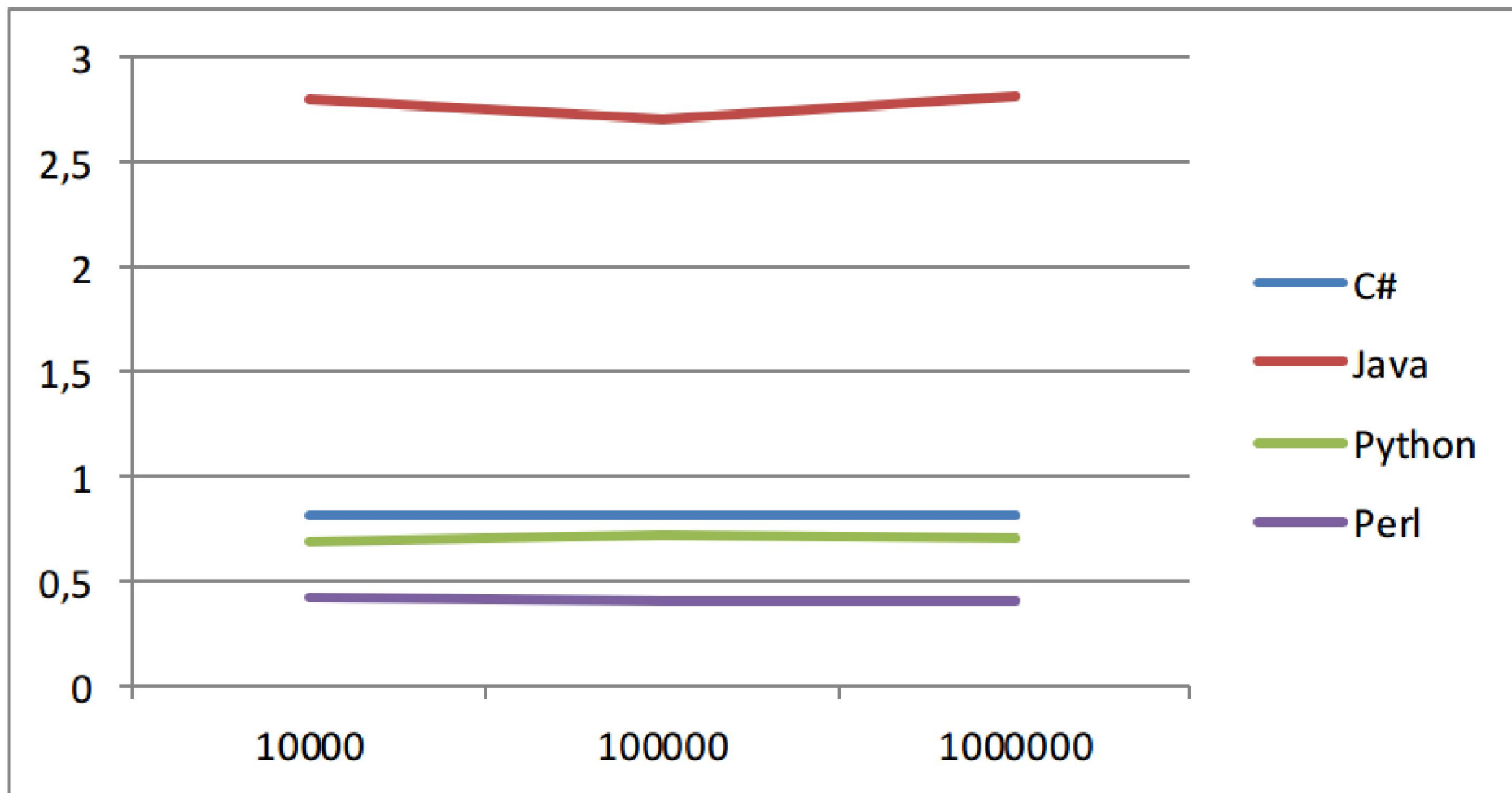
Jezik		C#	Java	Python	Perl
Duljina prve sekvence	Duljina druge sekvence				
10^2	10^2	0.01842 s	0.06656 s	0.06025 s	0.09196 s
10^3	10^2	0.02149 s	0.10451 s	0.14958 s	0.43023 s
10^3	10^3	0.09987 s	0.32992 s	3.54217 s	9.48722 s
10^4	10^3	0.36812 s	0.63270 s	17.39963 s	46.447793
10^4	10^4	6.14174 s	12.75127 s	427.21399 s	1002.60669 s
10^5	10^5	630.68986 s	1327.45527 s	832 min 42 s	1731 min 19 s
10^6	10^6	1157 min 12 s	2555 min 27 s	-	-

Jezik		C#	Java	Python	Perl
Duljina prve sekvence	Duljina druge sekvence				
10^2	10^2	6632 kB	18856 kB	5908 kB	3104 kB
10^3	10^2	6680 kB	19992 kB	5908 kB	3104 kB
10^3	10^3	6840 kB	21880 kB	5908 kB	3632 kB
10^4	10^3	7146 kB	22508 kB	5908 kB	3632 kB
10^4	10^4	8096 kB	27984 kB	6960 kB	4164 kB
10^5	10^5	13320 kB	33068 kB	14296 kB	13932 kB
10^6	10^6	67540 kB	78644 kB	86680* kB	113128* kB

Usporedba vremenske složenosti



Usporedba memorijske složenosti



Zahvaljujem na pažnji!