

Zadanie: PIS

Pisarze



XXVII OI, etap I. Plik źródłowy pis.* Dostępna pamięć: 256 MB.

21.10 – 18.11.2019

Uwaga: W tym zadaniu limit na długość kodu źródłowego wynosi 10 kB (a nie, jak zazwyczaj, 100 kB).

Bajtek natrafił na fragment tekstu w języku polskim. Wie tyle, że fragment pochodzi z jednej z trzech książek:

- „Pan Tadeusz” Adama Mickiewicza,
- „Quo Vadis” Henryka Sienkiewicza,
- „Lalka” Bolesława Prusa.

Napisz program, który pomoże mu ustalić, z której z tych książek pochodzi ów fragment.

Wejście

W pierwszym wierszu wejścia znajduje się liczba T ($1 \leq T \leq 1000$) oznaczająca liczbę przypadków testowych. Każdy z kolejnych T wierszy zawiera fragment jednej z trzech podanych wyżej książek. Fragment pozbawiony jest polskich znaków, zawiera od 10 do 2000 słów oraz zaczyna się i kończy pełnym słowem (lub znakiem interpunkcyjnym).

Łączna długość wszystkich fragmentów na wejściu nie przekroczy dwóch milionów znaków.

Wyjście

Na wyjściu powinno znaleźć się T wierszy; i -ty z nich powinien zawierać jedno słowo Mickiewicz, Prus lub Sienkiewicz, oznaczające autora i -tego fragmentu z wejścia.

Przykład

Dla danych wejściowych:

```
3
Petroniusz obudzil sie zaledwie kolo...
Litwo! Ojczyzna moja! ty jestes jak...
W poczatkach roku 1878, kiedy swiat...
```

poprawnym wynikiem jest:

```
Sienkiewicz
Mickiewicz
Prus
```

Wyjaśnienie przykładu: Fragmenty z wejścia są początkami odpowiednich dzieł. Przykładowe wejście widoczne w treści zadania zostało skrócone w celu utrzymania struktury dokumentu.

Ocenianie

Uwaga: Jest to zadanie z **analizy danych**. Nie musi istnieć idealne rozwiązanie poprawnie rozpoznające wszystkie fragmenty. Ocenianie zostanie przeprowadzone następująco:

- Niech T będzie liczbą fragmentów na wejściu.
- Niech P będzie liczbą poprawnych odpowiedzi udzielonych przez program uczestnika.
- Jeśli $P \geq 0,9 \cdot T$, uczestnik otrzymuje 100% punktów za test.
- Jeśli $P \leq \frac{T}{3}$, uczestnik otrzymuje 0% punktów za test.
- W przeciwnym wypadku uczestnik otrzymuje $100 \cdot \frac{P - \frac{T}{3}}{0,9 \cdot T - \frac{T}{3}}$ % punktów.

Fragmenty dzieł pochodzą z plików dostarczonych razem z treścią zadania. Razem z nimi uczestnikom udostępniony został program `pistestgen.py`, który pozwala wygenerować losowy test za pomocą komendy `python3 pistestgen.py grupa nazwa katalog [ziarno]`, przy czym:

- `grupa` jest numerem grupy testów i tym samym numerem podzadania (liczbą całkowitą od 1 do 4),
- `nazwa` jest nazwą testu (program stworzy pliki `nazwa.in` oraz `nazwa.out`),
- `katalog` jest ścieżką do katalogu zawierającego teksty dzieł,

- *ziarno* jest opcjonalnym parametrem (ciągiem znaków) pozwalającym powtarzać generowanie identycznego testu (podanie dwa razy takiego samego ciągu znaków wygeneruje dwa razy taki sam test).

Program ten został również użyty do wygenerowania (wszystkich) przypadków testowych, na których sprawdzane będą rozwiązania uczestników.

Podzadanie	Warunki	Liczba punktów
1	$T \leq 100$, fragmenty zawierają od 500 do 2000 słów	20
2	$T \leq 1000$, fragmenty składają się ze zdań	20
3	$T \leq 1000$, fragmenty zawierają od 30 do 80 słów	30
4	$T \leq 1000$	30