

„Przetwarzanie tekstu i modele języków naturalnych”

Kornel Jankowski

1 Zanim zaczniesz

Można wybrać dwa warianty oddawania zadań: można oddać zadania: A, B1, C, albo zadania A, B2, C. Odpowiedzi na zadania A i B[12] mogą być w treści maila, albo w załączonym PDFie. Skrypty z zadania C proszę dołączyć do maila. Każde z zadań warte jest 5 punktów.

Rozwiązania proszę wysłać, koniecznie z tagiem [WWW-NLP] w temacie na adres k3rnelj <at> gmail.com. Proszę nie zapomnieć się podpisać ;)

2 Zadanie A

W tym zadaniu założmy, że zamiast standardowego mamy bardzo prosty trzyliterowy alfabet: {a, b, c}. Patrząc na alfabet jako na zbiór, *słowem nad danym alfabetem* nazywamy dowolny ciąg znaków pochodzących z tego zbioru. Możliwymi słowami nad naszym alfabetem są np. aab, baba, caba, abc, baca, a, b, c.

Oblicz, ile jest słów nad takim alfabetem, które:

1. mają długość dokładnie pięciu znaków
2. mają dokładnie n znaków, ale wśród nich dokładnie dwie litery b
3. są palindromami długości 10?
4. są palindromami długości co najwyżej 10?

3 Zadanie B1

Przeczytaj artykuł o wyrażeniach regularnych pod tym adresem: <https://goo.gl/m3YK4r>. Bardzo proszę uważać z czytaniem artykułów w innych miejscach, zawierają bowiem sporo niebezpiecznych niedopowiedzeń, a w szczególności unikać proszę rozdziału o wyrażeniach regularnych Perla w polskim Wikibooks, który zawiera tyle wierutnych bzdur, że trzeba by go całościowo przepisać.

Spróbuj napisać, zgodnie z opisem z artykułu wyrażenia pasujące do:

1. Słów: ada, ala, ela, adela, hela. (wyrażenie ma pasować do wszystkich podanych słów *i żadnych innych!*)
2. Daty w formacie DD-MM. Dla ułatwienia uznajmy, że każdy miesiąc ma 30 dni.
3. Zapisów dziesiętnych liczb naturalnych od 0 do 127.

4 Zadanie B2

Napisz wyrażenie regularne opisujące prawidłowe zapisy adresów protokołu IP wersji 4 (np. 192.168.4.42).

5 Zadanie C

Napisz skrypty basha:

1. Skrypt `concat.sh`, który na wyjście wypisuje konkatencję swoich dwóch pierwszych argumentów, tj. np. dla wywołania `./concat.sh Tr alala` wypisuje `Tralala`, a dla wywołania `./concat.sh 12 34` wypisuje `1234`.
2. Skrypt `add.sh`, który na wyjście wypisuje arytmetyczną sumę swoich dwóch pierwszych argumentów (załóżmy, że całkowitych), tj. np. dla wywołania `./concat.sh 12 34` wypisuje `46`.
3. Skrypt `parrot.sh`, który w pierwszym argumencie przyjmuje ciąg znaków, a w drugim liczbę całkowitą `k`. Skrypt ma wypisać `k` razy na wyjście ciąg z pierwszego argumentu (w kolejnych liniach), tj. np. dla wywołania `./parrot.sh 'Litwo, ojczyzno moja!' 2` skrypt powinien wypisać

```
Litwo, ojczyzno moja!  
Litwo, ojczyzno moja!
```

Skrypt powinien po wypisaniu każdej linii przeczekać sekundę, zanim wypisze następną.

Każdy ze skryptów, jeśli zostanie wywołany nieprawidłową liczbą argumentów, powinien wypisać `EKHEM, EKHEM!` i zakończyć pracę z kodem wyjścia `1`. Poza ich liczbą, można nie sprawdzać poprawności argumentów.

Jeśli potrzebujesz, skorzystaj z materiałów dotyczących pisania skryptów w bashu podlinkowanych w opisie warsztatów i pytaj wujka google'a.