



1 Diena 4 Uždavinys

Kalbos

Jums reikia parašyti interaktyvią programą, kuri pagal duotą Vikipedijos ištraukų seką (žr. pavyzdį žemiau) spėtų paeiliui kiekvienos ištraukos kalbą. Po kiekvieno spėjimo programai pateikiamas teisingas atsakymas, taigi programa gali mokytis ir ilgiau praktikuodamasi geriau spėti.

Kiekviena kalba koduojama numeriu L nuo 0 iki 55. Bet kurią ištrauką sudaro lygiai 100 simbolių, kurie vaizduojami sveikųjų skaičių nuo 1 iki 65535 masyvu E . Šie skaičiai nuo 1 iki 65535 buvo parinkti atsitiktinai ir neatitinka jokios standartinės koduotės.

Realizuokite procedūrą **excerpt(E)**, kur E yra 100 skaičių masyvas, vaizduojantis Vikipedijos ištrauką aukščiau aprašytu būdu. Procedūra turi vieną kartą iškviešti procedūrą **language(L)**, kur L yra Vikipedijos spėjamos ištraukos E kalba. Vertinimo serveris realizuoja funkciją **language(L)**, kuri įvertina jūsų spėjimą ir gražina teisingą kalbą. Spėjimas teisingas, jei **language(L) = L**.

Vertinimo serveris iškvies **excerpt(E)** 10 000 kartų, po vieną kartą su kiekviena ištrauka duomenų faile. Jūsų sprendimo tikslumas yra lygus daliai ištraukų, kurioms **excerpt(E)** teisingai atspėjo kalbą.

Uždaviniui spręsti galite naudoti bet kokį metodą. Roccio metodo tikslumas yra apie 0,4. Jo esmė yra nustatyti ištraukos panašumą su kiekviena jau matyta kalba L ir parinkti labiausiai panašią kalbą. Panašumas apibrėžiamas kaip skirtingų E simbolių, kurių kiekvieną galima rasti bent viename ankstesnių kalbos L ištraukų, kiekis.

Atkreipiame dėmesį, kad pradiniai duomenys buvo parsiję iš tikrų Vikipedijos straipsnių, ir juose gali būti keletas klaidingų simbolių ar teksto fragmentų. Taip ir turi būti, ir tai sudaro dalį uždavinio.


Pavyzdys.

Kaip pavyzdį pateikiame 56 kalbų Vikipedijų ištraukų tekstus.

0. Yshokkie word meestal in Kanada , die noorde van die VSA en in Europa gespeel. Dit is bekend as 'n b
1. وهو المنتج الذي يجعل المنظم لا يكسب ربحا ولا يخسر ويحصل على دخل يكف ، Marginal Producer المنتج الحدي
2. "BAKIL" Futbol Klubu 1995-ci ildə Misir Səttar oğlu Əbilov tərəfindən yaradılmış və həvəskar futbol
3. Квинт Фулвий Флак (Quintus Fulvius Flaccus; † 205 пр.н.е.) е политик и генерал на Римската републик
4. ইন্ডিয়ান ইনস্টিটিউট অফ সোশ্যাল ওয়েলফেয়ার অ্যান্ড বিজনেস ম্যানেজমেন্ট (সংক্ষেপে আইআইএসডব্লিউবিএম)
5. 5. juni (lipanj) (5.6.) je 156. dan godine po gregorijanskom kalendaru (157. u prestupnoj godini
6. La Caunette és un municipi francès , situat al departament de l' Erau i a la regió de Llenguadoc-Ros
7. Praha je malé městečko v Texasu , které leží cca 85 km na jihozápad od Austinu . Bylo založeno
8. Graeme Allen Brown (født 9. april 1979 i Darwin , Northern Territory , Australien) er en australsk
9. Der Plattiger Habach (3.214 m ü. A. , nach anderen Angaben nur 3.207 m [1])
10. Το Νησί Γκρέιτ Μπάρριερ (Αγγλικά : Great Barrier Island , Μαορί : Motu Aotea) είναι νησί στα βόρεια
11. Sid Bernstein Presents... is a 2010 feature-length documentary film by directors Jason Ressler and E
12. El término latino lex loci celebrationis aplicado al derecho internacional privado quiere decir: "le
13. Apollo 5 oli kosmoselaev , mis sooritas Apollo programmi teise mehitamata lennu. Lennu käigus testit
14. هزار و سیصد و پنجاهمین سیارک کشف شده اس (TA نامگذاری : 1934 ، سیارک ۱۳۵۰ (به انگلیسی : 1350
15. V. I. Beretti (myös Vikenty Ivanovitš Beretti , alk. Vincent Beretti ; 1781 Milano Italia – 18. elok
16. Le 5 e bataillon de parachutistes vietnamiens (ou 5 e BPVN ou encore 5 e Bawouan) est une unité par
17. Amina Sarauniyar Zazzau., wadda ta rayu daga shekarar 1533 zuwa 1610, daya ce daga cikin 'ya'ya biyu
18. ה שערות רימן היא השערה שהציע בשנת 1859 ה מתמטיקאי ברנרד רימן , מגדולי המתמטיקאים של אותה ע
19. Sudski proces Doe protiv Boltona je sudski proces iz 1973 . godine kojim je američki Vrhovni sud uki

1 Diena 4 Uždavimas

Kalbos

20. Owen Cunningham Wilson (1968 . november 18. , Dallas , Texas , Egyesült Államok) amerikai színész
21. Հայ Մարտիրիկե Եկեղեցին պատկանում է Արևելյան Մարտիրիկե Եկեղեցիների և այսպիսով ենթարկվում է Հռոմի Պապի ա
22. Dionysios dari Halicarnassus (Bahasa Yunani : Διονύσιος Ἀλεξάνδρου Ἀλικαρνασσεύς , Dionysios putra
23. Nnamdi "Zik" Azikiwe , bu onye isi-ala izizi Nijiria nwere. Ochichi ya bidolu na afo 1954 welu ruo n
24. La Riserva naturale orientata Serre della Pizzuta è un'area protetta del dipartimento Regionale di S
25. 石橋和義 (いしばし かずよし/まさよし、生没年不詳)は、詳。石橋氏 初代当主。初名氏義。尾張 三郎を通称とし、官途は、左近将監 → 三河守 → 左衛門佐。足利直義
26. კობინ ბლოიუ (ინგლ. Corbin Bleu ; დ. 21 თებერვალი , 1989 , დაბადების ადგილი ბრუკლინი , ნიუ-იორკი , ა
27. Тárья Кáарина Хáлонен (Tarja Kaarina Halonen)); 24 желтоқсан , 1943 , Каллио , Хельсинки , Финлан
28. 딜롱 (Dilong)은 중국 랴오닝(Liaoning) 지방의 익시안층(Yixian Formation)에서 온전한 4구의 화석으로 발견되었다. 이 공룡은 가장 원시적인 초기의 티
29. Сүймөнкул Чокморов - советтик актёр. Жетинин айынын 9 (ноябрь) 1939-жылы, Фрунзе шаарын жанындагы Чо
30. D' Mirjam van Abellin war eng Nonn a Mystikerin , och " Maria vum gekräizegte Jesus " genannt. Si as
31. Panopea abrupta (angl. Geoduck) - jūrinīų dvigeldžių moliuskų rūšis, priklausanti Hiattellidae šeim
32. "Dzimis Latvijā" ir Liepājas dueta Fomins & Kleins 2004 . gada 23. februārī izdotais otrais albu
33. I Ludwik Lejzer Zamenhof dia dokotera mpijery maso nipetraka any Polonia . Fantantsika izy ankehitri
34. Седумстотини милиони малечки алвеоли во белите дробови , всушност се шупливи чаури - алвеоли прекрие
35. Энэхүү шувуу нь Бутан , Хятад , Гонконг , Энэтхэг , Пакистан , Иран , Япон , Казакстан , Солонгос ,
36. भारतीय म्हाराष्ट्र राज्याच्या नागपूर पासुन २१६ कि.मी. दूर असलेले एक गाव. ते वैनगंगा नदीच्या काठावर
37. De Slotervaart was oorspronkelijk de waterweg die sinds de Middeleeuwen het dorp Sloten verbond met
38. Macierz S (macierz rozpraszania , od ang. scattering matrix) jest centralnym elementem w mechanice k
39. A Hora do Rush 3 (Rush Hour 3 , no original) é o terceiro filme da franquia Rush Hour . Dirigido po
40. Coordonate : 51°34'0"N 12°3'0"E / 51.56667 , 12.05 Brachstedt este o comună din landul Saxonia-A
41. Гробнициы императоров династии Мин и Цин — памятник Всемирного наследия ЮНЕСКО , состоящий из несколько
42. Kovalentni radijus atoma - ponekad se naziva i valentni radijus. Kovalentni radijus je srednje rasto
43. Konięcpol je mesto v Pofsku v Sliezskom vojvodstve v okrese Powiat częstochowski v rovnomennej gmine
44. Hoxhë Vokrrri vije nga Shqipëria ishte një klerik shqiptar i cili luftonte për Çështjen Kombëtare . A
45. Гурдијеље је насеље у општини Тутин у Рашком округу . Према попису из 2002. било је 93 становника (п
46. Underhållsstöd betalas ut av Försäkringskassan (FK) till en förälder som är vårdnadshavare och bor e
47. இந்தியாவின் தேசிய நெடுஞ்சாலைகள் நடுவண் அரசின் தேசிய நெடுஞ்சாலைத் துறையால் பராமரிக்கப்படுகின்றன. பெரு
48. Дар он зиндаги .маишат ,фаолияти мехнати,муборизаи ичтимои, русуму омор, хислат ва эхсосоти халк ифо
49. ไฟฟางเหาะา ฝืนเฟสพันส (ฝงภษ : Phytophthora infestans) คือเชื้อราโอบโอมซึ่งท หรือ ราน้ำ ที่เป็นสาเหตุ
50. ABUL FAWARIS BERRANY - 11. asyrda Orta Aziýadaky oguz taýpalarynyň berrany dinastýasynyň wekili. OI
51. Egemenlik ya da hâkimiyet , bir toprak parçası ya da mekan üzerindeki kural koyma gücü ve hukuk yara
52. Темне фентезі (від англ. Dark Fantasy - темне, похмуре фентезі) - піджанр літератури, який включає
53. Paris By Night 84: In Atlanta - Passport to Music & Fashion (Âm nhạc và Thời trang) là chương tr
54. ISO 3166-2:GU ni akoole ninu ISO 3166-2 , apa opagun ISO 3166 ti International Organization for Stan
55. 下卡姆斯克 俄文 Нижнека́мск 韃靼語 Түбән Кама/Түбән Кама) 是 俄羅斯 韃靼斯坦共和國 東北部的一個城市，位於 卡馬河 南岸。2002年 人口22

Pradinių duomenų faile grader.in.1 įrašyta 10 000 tokių pavyzdžių 56 kalbomis, kurios buvo nurodytos kaip "gimtoji kalba" IOI 2010 registracijos anketose. Kiekvienam fragmentui kalba yra atsitiktinai parinkta iš tų 56 kalbų, o pati ištrauka paimta iš atsitiktinai parinkto Vikipedijos straipsnio ta kalba pirmosios pastraipos. Kiekvienoje failo eilutėje įrašyta:

- Kalbos ISO kodas, sudarytas iš dviejų simbolių;
- 100 skaičių nuo 1 iki 65 535, atitinkančių pasirinktojo straipsnio pirmosios pastraipos pirmuosius 100 simbolių.
- UTF-8 koduote pavaizduoti tie patys 100 simbolių, kuriuos galite matyti tekstiniam

1 Diena 4 Uždavinys

Kalbos

redaktoriuje arba Firefox naršyklėje. Šitai pateikta tiktai jūsų patogumui ir tai nėra pradiniai duomenys jūsų programai.

Oficialusis vertintojas naudos 10 000 skirtingų pavyzdžių, atrinktų aukščiau aprašytu būdu. Tačiau vertintojas skirtingoms kalboms priskirs skirtingus numerius nuo 0 iki 55 ir kiekvienam simboliui priskirs skirtingus numerius nuo 1 iki 65 535.

1 Uždutis [30 taškų]. Jūsų programa vertinimo serveryje turi pasiekti 0,3 arba geresnį tikslumą.

2 Uždutis [iki 80 taškų]. Jūsų rezultatas bus lygus $114(\alpha - 0,3)$, suapvalinta iki artimiausio sveiką skaičiaus, kur α yra jūsų sprendimo tikslumas.

Realizacija.

- naudokite RunC programavimo ir derinimo aplinką
- Dalyvio katalogas: `/home/ioi2010-contestant/language/` (prototipas: `language.zip`)
- Reikia realizuoti: `lang.c` arba `lang.cpp` arba `lang.pas`
- Dalyvio sąsaja: `lang.h` arba `lang.pas`
- Vertintojo sąsaja: `grader.h` arba `graderlib.pas`
- Pavyzdinis vertintojas: `grader.c` arba `grader.cpp` arba `grader.pas` ir `graderlib.pas`
- Pavyzdinio vertintojo duomenų failas: `grader.in.1`
Pastaba: Kiekvienoje duomenų failo eilutėje įrašyta: kalbos kodas sudarytas iš dviejų simbolių; fragmentas sudarytas iš 100 tarpais atskirtų skaičių; ištrauka tekstiniu formatu.
- Pavyzdinio vertintojo pradinių duomenų failo rezultatas: jei procedūra iškviečia **language** aprašytu būdu kiekvienai iš 10 000 ištraukų, pavyzdinis vertintojas įrašys OK `alpha`, kur `alpha` yra tikslumas.
- Kompiliavimas ir vykdymas (komandinė eilutė): `runc grader.c` arba `runc grader.cpp` arba `runc grader.pas`
- Kompiliavimas ir vykdymas (gedit įskiepis): *Control-R*, redaguojant uždavinį sprendžiančią procedūrą.
- Pateikimas (komandinė eilutė): `submit grader.c` arba `submit grader.cpp` arba `submit grader.pas`
- Pateikimas (gedit įskiepis): *Control-J*, redaguojant bet kurį procedūros arba vertintojo failą.