

ПЛОВДИВСКИ УНИВЕРСИТЕТ “ПАИСИЙ ХИЛЕНДАРСКИ”
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Седми Междунниверситетски Турнир по Програмиране
Пловдив, 08.05.2004г.

Задача F. КОМПЮТЪРНА МОРФОЛОГИЯ

Дадена е таблица с 10 словоизменителни типа на прилагателни имена в българския език. Всяка колона от таблицата определя парадигмата на думата и започва с номера на словоизменителния тип. Всеки ред съдържа словоформи с едни и същи граматични характеристики и започва с пореден номер на словоформа. Първият ред съдържа основната форма на всяка парадигма. Произволна основна форма извън таблицата принадлежи на този словоизменителен тип, който има най-голям брой различни представители в даден текст. При равенство се гледа общият брой представители. Ако и тогава има равенство, принадлежността е на типа с по-малък номер. Например **бесен** може да принадлежи на 1, 2, 4 или 8 тип, но следното примерно извлечение от текста на наредени двойки (словоформа, брой срещания на тази словоформа): (бесен,2); (бесени,1); (бесния,2); (бясна,1) дава преимущество 8-ми тип (бр. различни представители–3, общ брой – 5). Следва в класацията 4-ти тип (бр. различни представители–2, общ брой – 4), после 1-ви и накрая 2-ри тип (бр. различни представители–2, общ брой – 3). За последните 2 класацията е само по номер на тип. По този евристичен алгоритъм за даден текст произволна основна форма получава уникален номер на словоизменителен тип.

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
|----|--------|---------|-------------|---------|---------|--------|---------|---------|-----------|---------|
| 1. | нов | висш | български | силен | добър | бял | рядък | верен | строен | кози |
| 2. | нови | висши | български | силни | добри | бели | редки | верни | стройни | кози |
| 3. | нова | висша | българска | силна | добра | бяла | рядка | вярна | стройна | козя |
| 4. | ново | висше | българско | силно | добро | бяло | рядко | вярно | стройно | козе |
| 5. | новия | висшия | българския | силния | добрия | белия | редкия | верния | стройния | козия |
| 6. | новият | висшият | българският | силният | добрият | белият | редкият | верният | стройният | козият |
| 7. | новите | висшите | българските | силните | добрите | белите | редките | верните | стройните | козите |
| 8. | новата | висшата | българската | силната | добрата | бялата | рядката | вярната | стройната | козията |
| 9. | новото | висшето | българското | силното | доброто | бялото | рядкото | вярното | стройното | козето |

следва пример с по един представител в основна форма на всеки тип:

| | | | | | | | | | | |
|--|-------|------|-------|-------|-----|------|-------|-------|---------|-------|
| | ядрен | нисш | мъжки | течен | зъл | сляп | рязък | серен | стихиен | крави |
|--|-------|------|-------|-------|-----|------|-------|-------|---------|-------|

Да се създаде програма **F.EXE**, който прочита входната информация от стандартния вход със следния формат: на първия ред стои **K** - брой на тестове. На следващите редове – K на брой тестове, като във всеки тест се задава **име на текстов файл** (до 10 M, до 250 знака на ред) съдържащ текст с DOS-овска кирилица (кодове на букви на кирилица: A→128...Я→159, a→160...я→191) и **прилагателно** в основна форма (над същата азбука), което принадлежи на един от десетте типа.

Като резултат от всеки тест се извеждат на стандартния изход:

- номерът на словоизменителен тип на въведеното прилагателно или 0, ако не се среща в текста;
- всички срещащи се в текста словоформи от парадигмата на тази основна форма заедно с техния брой срещания, подредени по номер на словоформа;
- в лексикографски наредените думи от текста – първата дума с максимален брой срещания и последната дума с минимален брой срещания, заедно с техния брой срещания.

Забележка: Дума на кирилица се дефинира като последователност от букви на кирилица. Игнорира се разлика между малки и големи букви при броенето (ЗЪЛ, Зъл и зъл са една и съща дума)

ПРИМЕР:

Стандартен вход:

1
tekst1.txt
черен

Съдържание на tekst1.txt

Черната шатра беше разположена на върха на черната могила. Там пасяха два черни коня. Никой не знаеше кой е техният собственик. А това беше една черна магьосница.

Стандартен изход:

4
черни, 1
черна, 1
черната, 2
беше, 2
шатра, 1

Забележка: Не е задължително основната форма да се среща в текста. Наличието в текста на думите *черни*, *черна* и *черната* доказва, че словоизменителният тип на тази основна форма е 4.

| Лексикографски сортиран списък с думите от текста. | |
|---|---|
| а,1 беше,2 върха,1 два,1 е,1 една,1 знаеше,1 кой,1 коня,1 магьосница,1 могила,1 на,2 не,1 никой,1 пасяха,1 разположена,1 собственик,1 там,1 техният,1 това,1 черна,1 черната,2 черни,1 шатра,1 | Максимален брой срещания на дума: 2 Първа такава: <i>беше</i> . Минимален брой срещания на дума: 1 . Последна такава: <i>шатра</i> . |