МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

### Кафедра САПР

ЗВІТ

до лабораторної роботи № 6

на тему:

ВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK, ДЛЯ ОПРАЦЮВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ.

ВИКОРИСТАННЯ РЕГУЛЯРНИХ ВИРАЗІВ ДЛЯ ОБРОБКИ ТЕКСТУ.

з дисципліни “Комп’ютерна лінгвістика”

Виконала:

Студентка групи ПРЛм-12

Гречух Ю.

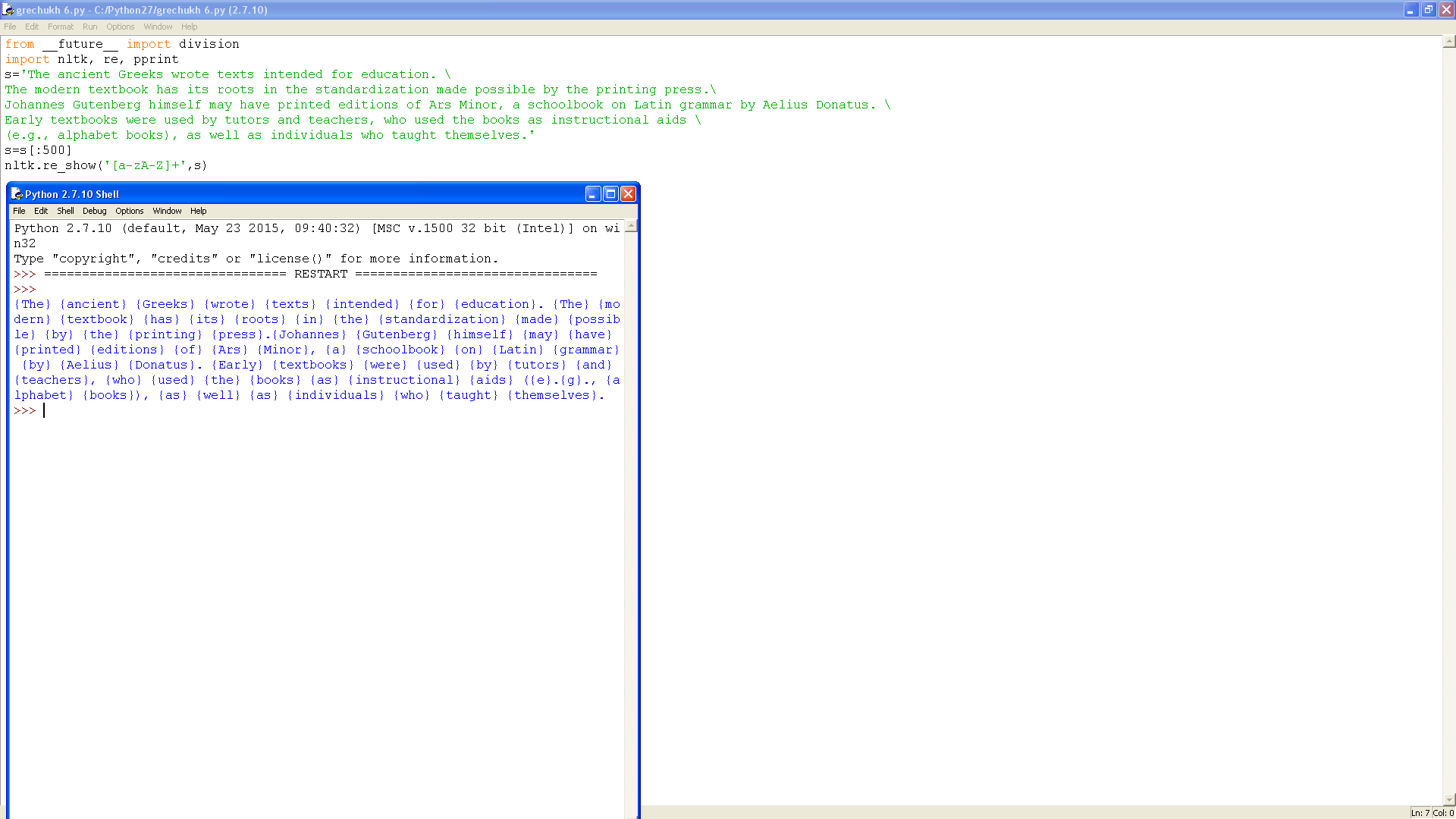
Перевірив:

Викладач кафедри САПР

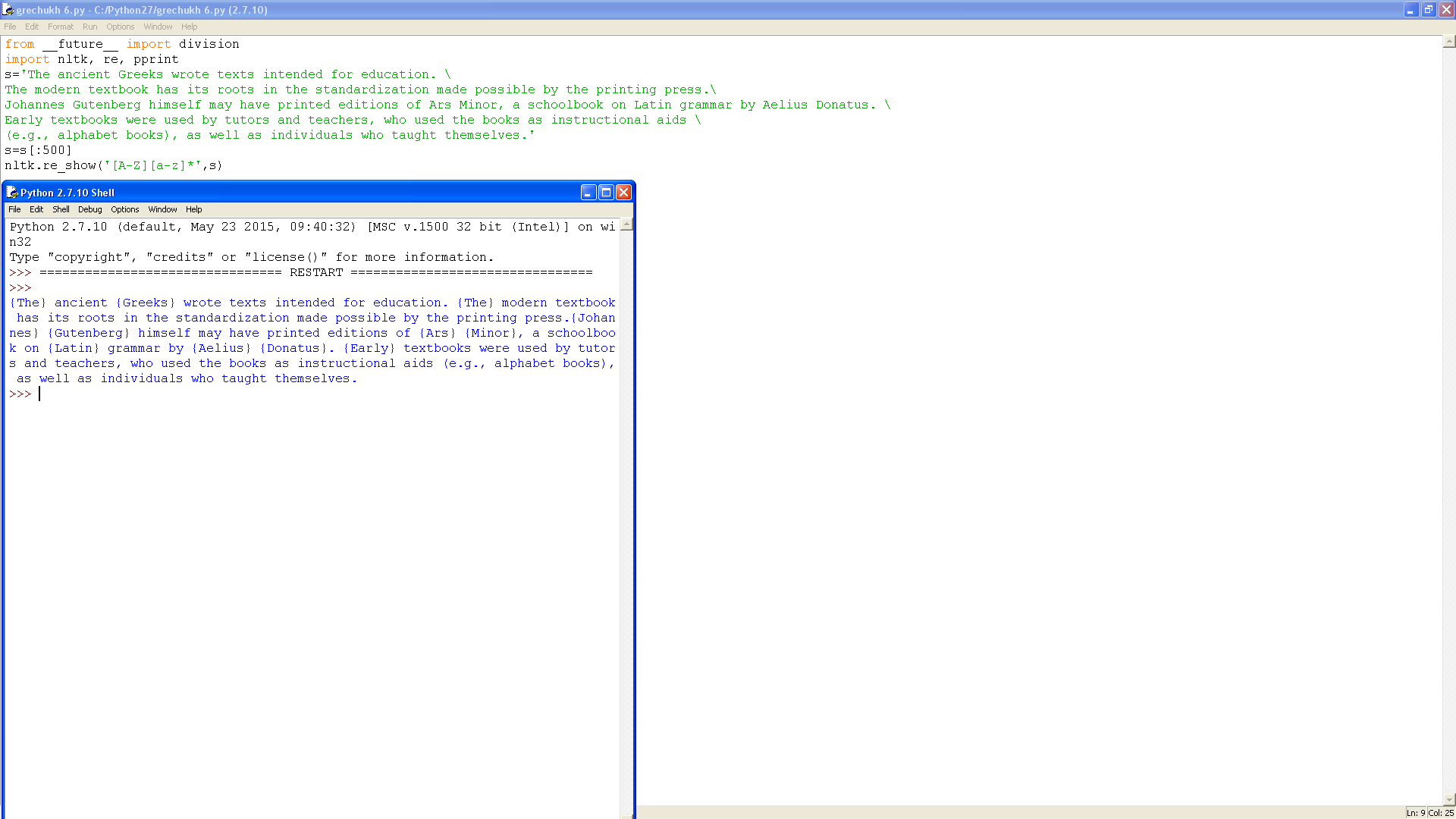
Дупак Б. П.

Львів 2015

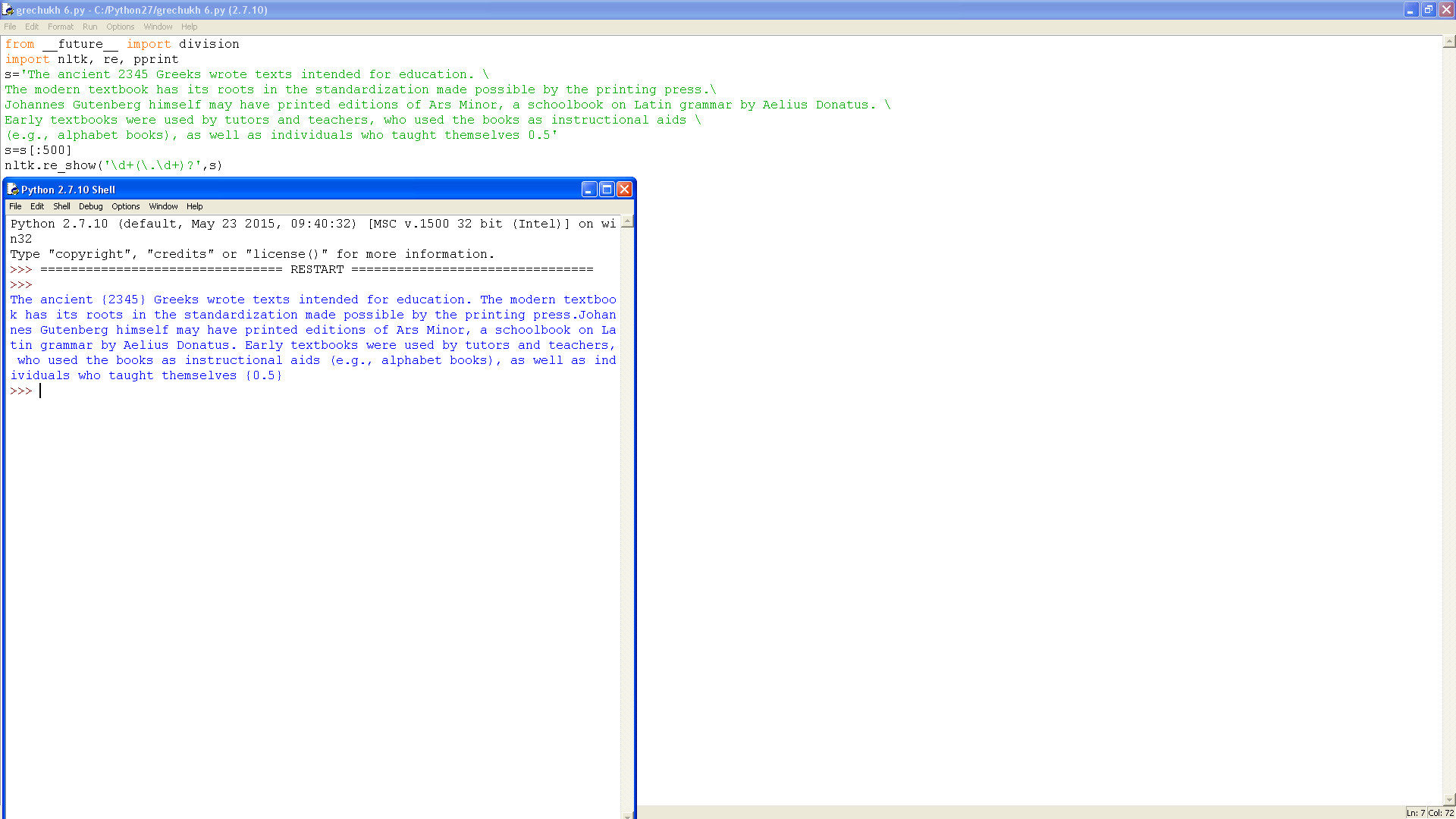
**Завдання 1.** Описати, які класи стрічок відповідають наступному регулярному виразу. [a-zA-Z]+. Результати перевірити використовуючи nltk.re\_show()

****

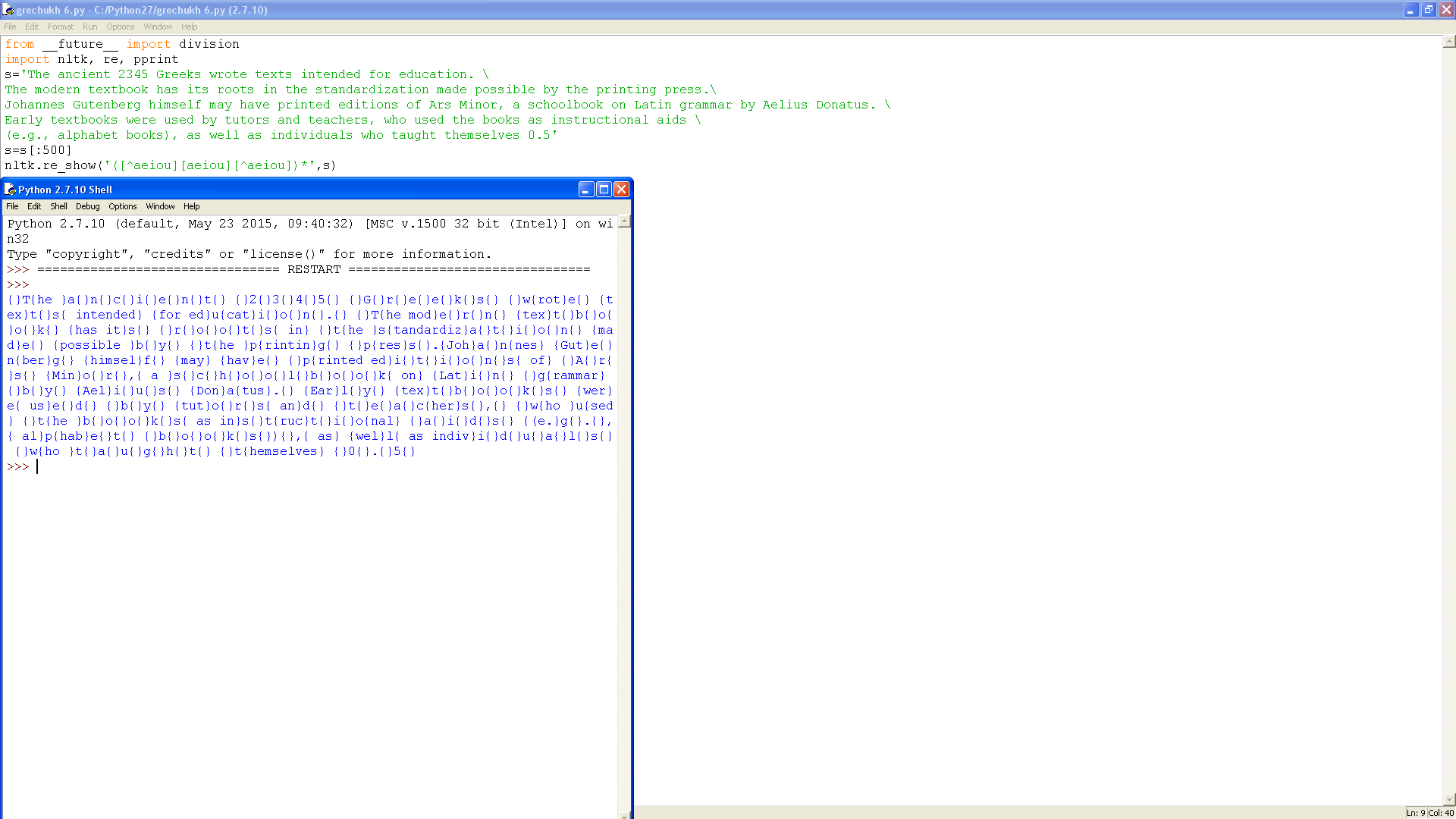
**Завдання 2.** Описати, які класи стрічок відповідають наступному регулярному виразу. [A-Z][a-z]\*. Результати перевірити використовуючи nltk.re\_show()



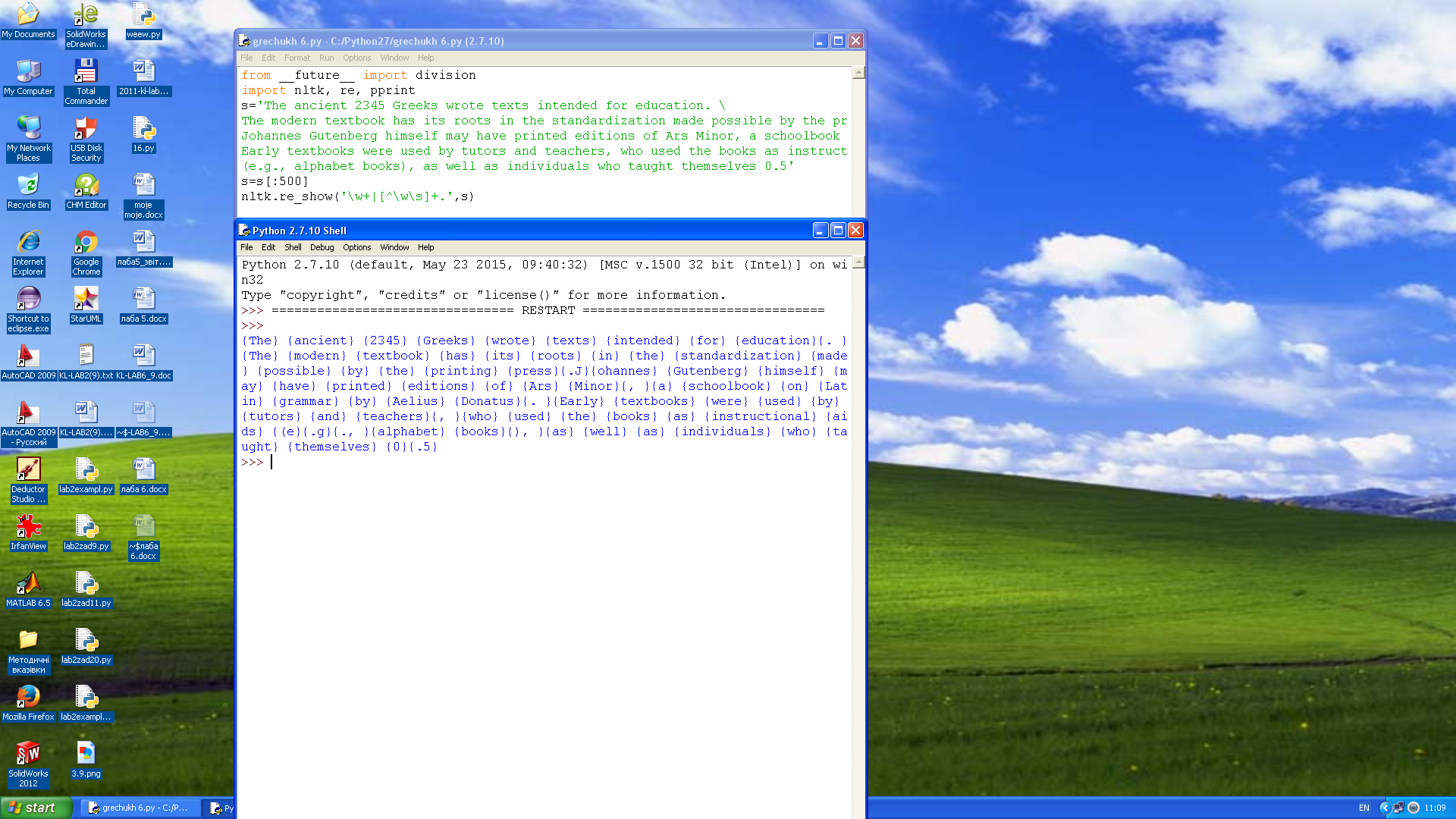
**Завдання 3.** Описати, які класи стрічок відповідають наступному регулярному виразу. \d+(\.\d+)?. Результати перевірити використовуючи nltk.re\_show()



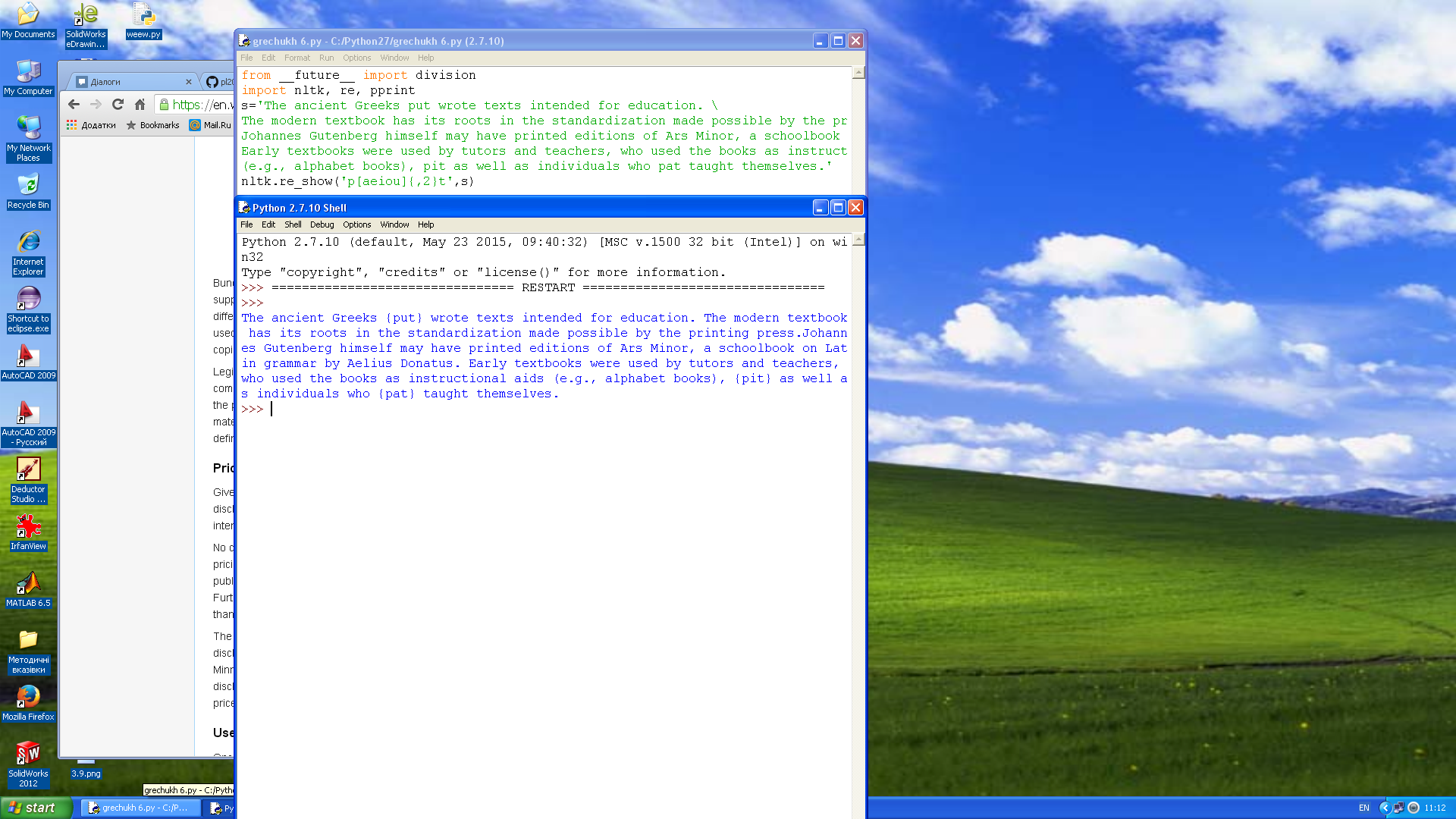
**Завдання 4**. Описати, які класи стрічок відповідають наступному регулярному виразу. ([^aeiou][aeiou][^aeiou])\*. Результати перевірити використовуючи nltk.re\_show()



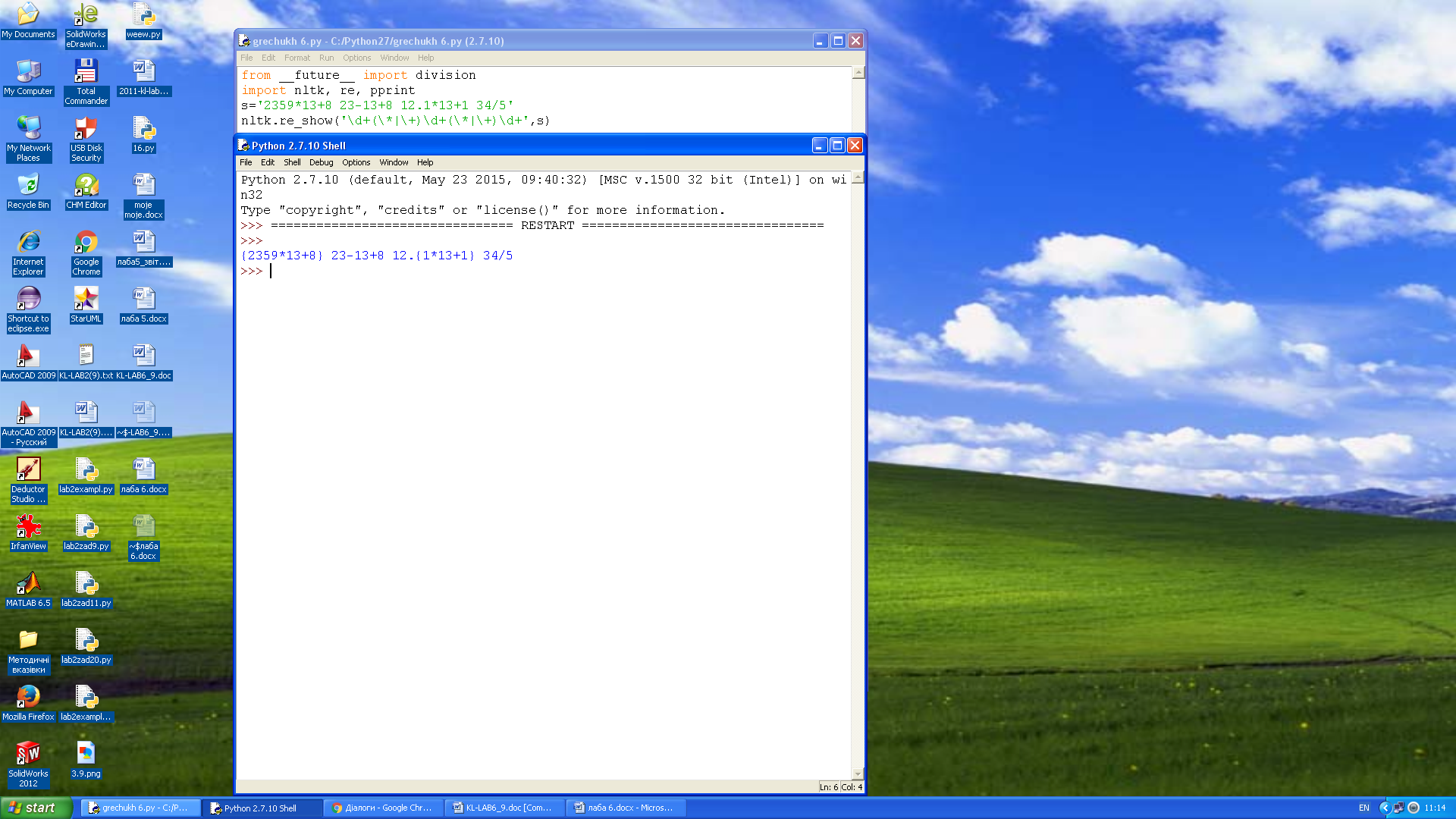
**Завдання 5.** Описати, які класи стрічок відповідають наступному регулярному виразу. \w+|[^\w\s]+.. Результати перевірити використовуючи nltk.re\_show()



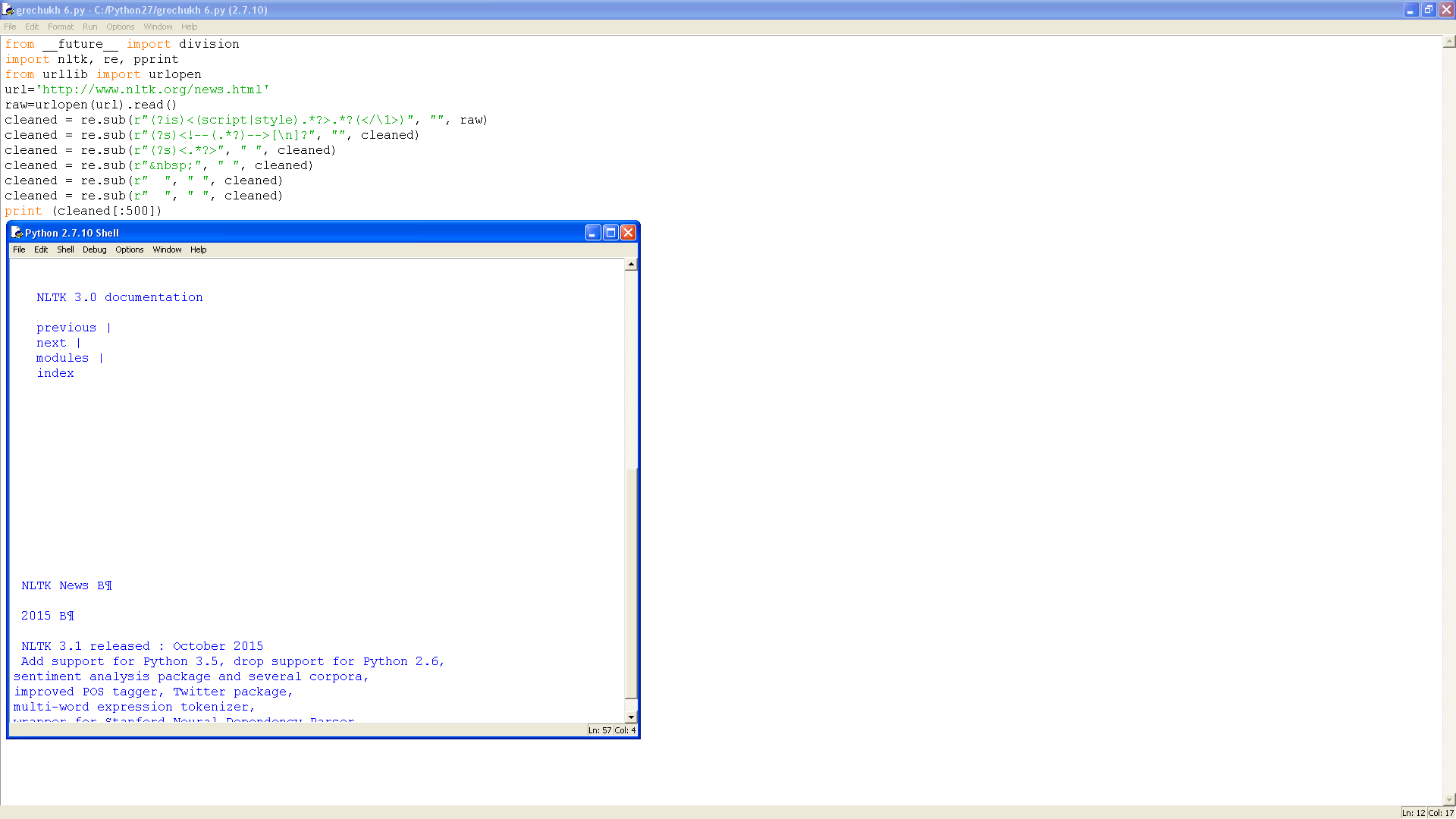
**Завдання 6.** Описати, які класи стрічок відповідають наступному регулярному виразу. p[aeiou]{,2}t Результати перевірити використовуючи nltk.re\_show()

****

**Завдання 8.** Написати регулярний вираз, який встановлює відповідністьнаступному класу стрічок:арифметичний вираз з цілими значеннями і, який містить операції множення та додавання *(2\*3+8*).

****

**Завдання 14.** Прочитати файл допомоги про функцію re.sub() використовуючи help(re.sub) . Використовуючи re.sub напишіть програму видалення HTML розмітки замінивши її на пробіли.

****

**Завдання 15.** Прочитати Додаток А. Дослідити явища описані у Додатку А використовуючи корпуси текстів та метод findall()для пошуку в токенізованому тексті.

