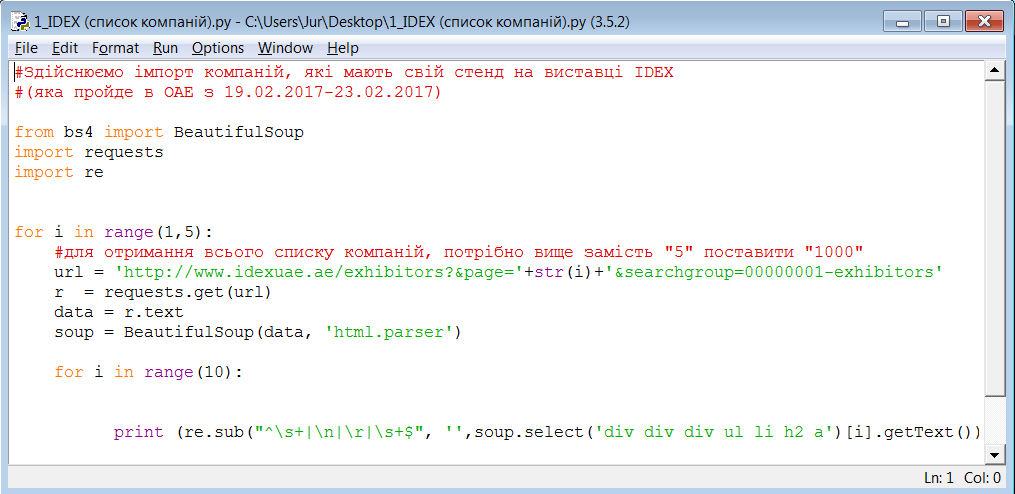
# 1. Парсинг інтернет сторінок на Python 3

Для підготовки до міжнародної виставки IDEX в Дубаї був написаний код, який робить парсинг виставкового сайту для того щоб отримати окремо всіх учасників, їх сферу діяльності та номер розташування стенду в разі наявності інтересу до цієї компанії.

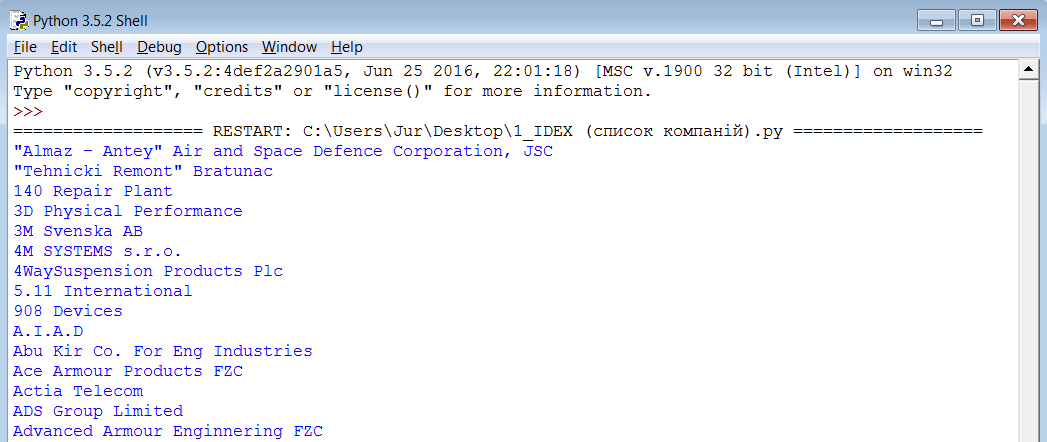
Код можна знайти за адресою:

### 1.1. Отримання списку всіх компаній з сайту виставки: [www.idexuae.ae](http://www.idexuae.ae)

**Код:**



**Результат:**

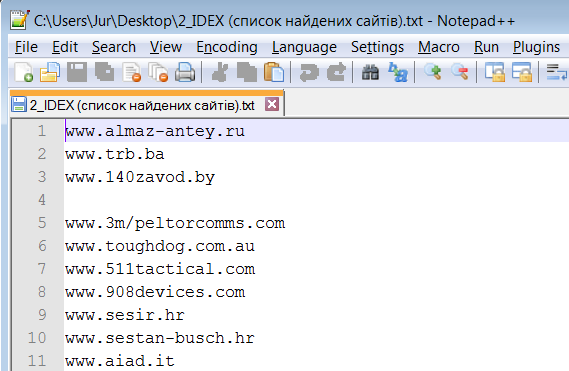


Так само можна отримати додаткову інформацію, а саме номера стендів, та інтернет сторінки компаній, частково змінюючи вищенаведений код.

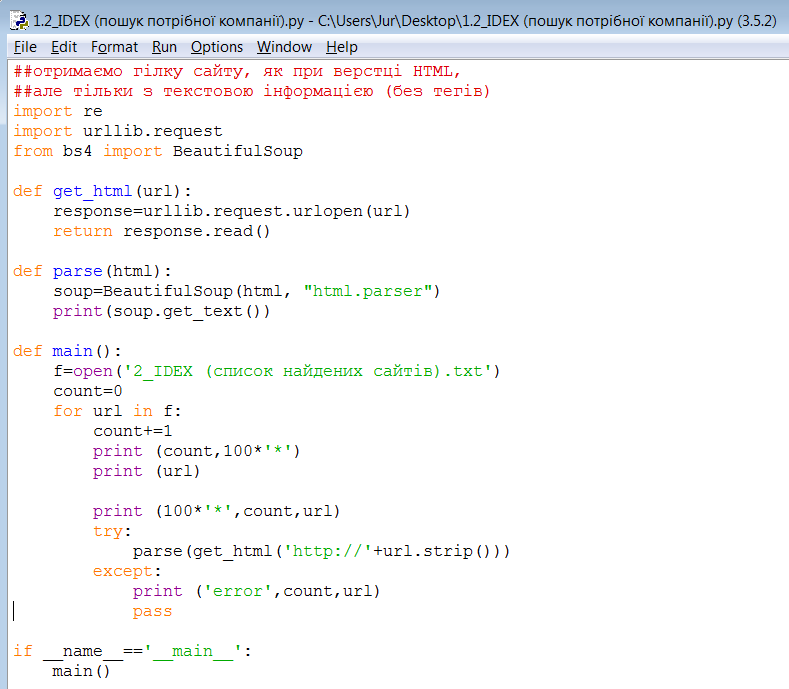
### 1.2. Отримання детальної інформації з першої інтернет сторінки всіх учасників виставки.

Система відкриває інтернет сайти, адреси яких знаходяться в текстовому файлі нижче, та відображає датальную інформацію.

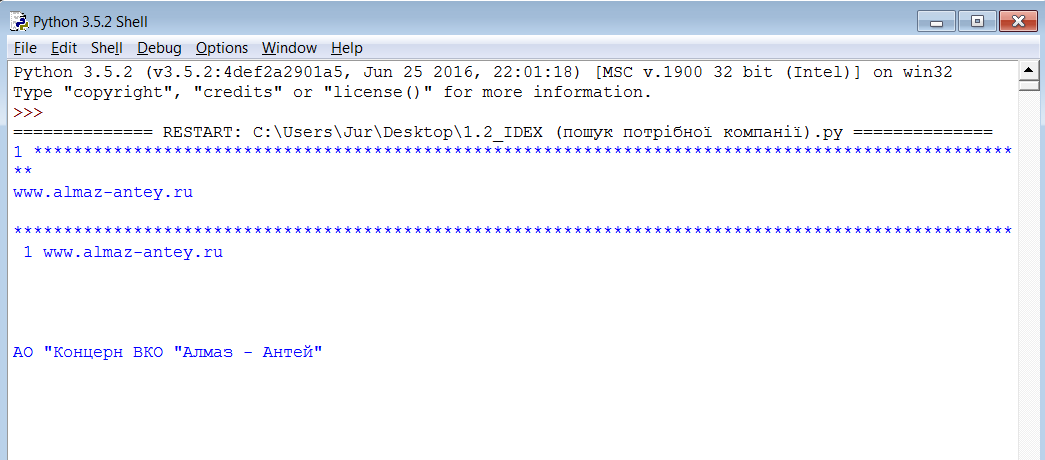
**Обробка даного текстового файлу:**



**Код:**



**Результат:**



Отриманий результат, це не що інше як DOM структура першої сторінки компанії, але всі теги виключені (залишена тільки текстова інформація з DOM структури).

Потім копіюємо цю інформацію в текстовий файл, та шукаємо за допомогою Ctrl+F ключові слова. Якщо потрібне слово знайдене, тоді можна заходити на інтернет сайт компанії, на сайті якої було знайдене ключове слово, та аналізувати більш детально.

**Висновки:** Всі вищенаведені маніпуляції дали змогу автоматизувати процес підготовки до виставки. Тобто, виконавши код, ми змогли відмовитися від ручної обробки сайту [www.idexuae.ae](http://www.idexuae.ae), де представлені більше ніж 800 компаній та автоматизували обробку всіх компаній за ключовими словами.