Оскільки, написана мною програма може працювати у два різні шляхи, то я опишу ці два шляхи.Результати виконання цих “двох” програм однаковий і це txt файли в папці з програмою, який виглядає приблизно, так:

1.Learn Programming with Python in 100 Steps - www.udemy.com/python-tutorial-for-beginners/

2.Basic Python for Beginners - www.udemy.com/basic-python-for-beginners/

3.Create a 3-Tier Application Using Azure Virtual Machines - www.udemy.com/free-azure/

4.Machine Learning A-Z™: Hands-On Python & R In Data Science - www.udemy.com/machinelearning/

5.Essential Training Photoshop CC 2018 - Basics (with subs) - www.udemy.com/essential-training-photoshop-cc-2018-basics-with-subs/

6.How To Make A Wordpress Website 2018 -Elementor Page Builder - www.udemy.com/how-to-make-a-wordpress-website-with-elementor/

7.Learning how to learn: brain hacks that actually work. - www.udemy.com/learning-how-to-learn-brain-hacks-that-actually-work/

8.The Web Developer Bootcamp - www.udemy.com/the-web-developer-bootcamp/

9.Learn Blockchain By Building Your Own In JavaScript - www.udemy.com/build-a-blockchain-in-javascript/

10.Complete Python Bootcamp: Go from zero to hero in Python 3 - [www.udemy.com/complete-python-bootcamp/](http://www.udemy.com/complete-python-bootcamp/)

Також сама програма може видати такі дані:

Program get wrong input data so program change name of 785 course!

Program get wrong input data so program change name of 826 course!

Program get wrong input data so program change name of 875 course!

Отже, ми бачемо , деякі стрічки з txt файлу і з самої програми, але що вони означають? Останнє, найлегше – ці три стрічки означають, що при записі у файл, виникла помилка, з вибраним кодуванням, при записі у файл, і тому програма змінила назву курсу, підкреслюю тільки назву курсу, url вона нечіпає. Тому через те що в назві курсу був поганий символ вона його просто видалила, крім того вона вказує порядковий номер курсу, назву якого вона змінила. З цим ми розібралися, проте що ж записує в файл програма. По-перше представлені вище рядки – це не єдине, що може видавати програма, ще може бути таке повідомлення:

Sorry, but program didnt search anything.

Дана повідомлення, означає, що в процесі обробки даних користувач ввів неправильні дані, або за введеними користувачем даними, програма не знайшла жодного курсу, таке теж можливо, якщо наприклад ввести категорію курсу з одної теми, наприклад ввести музику, а в підкатегорію ввести, наприклад IT & Software, тобто зрозуміло, що з такими даними програма не зможе знайти жодного курсу.

Що ж до перших даних, якщо програма виконала все правильно, тоді вона запускає функцію write\_list\_with\_courses або special\_write, перша функція записує дані, отримані напряму з courses-list, а друга з curriculum list, оскільки при роботі з першими даними, вони подаються у вигляді списку словників з даними про самі курси і в цих словниках перші є тільки один title і url, а при роботі з другим запитом отримується список списків словників, тобто в першому списку записані всі курси, в другому дані про курс. І в цих даних немає даних з ключами title і url, є дані з title, але вони не стосуються назви курсу, а стосуються лекцій та уроків. Тому при роботі з останніми прийшлося вручну добавити url та title про курс. Це були описані дані через розлоблені мною методи, при роботі напряму з udemy api все набагато простіше, в словниках зразу присутні всі необхідні ссилки та назви для курсів, проте немає більших даних про них.

Це були скорочені дані, виведені для користувача, та які ж повні отримані дані. Якщо працювати з udemy api, то можна отримати такі дані:

* Дата створення курсу;
* Імена авторів курсу;
* Посилання на курси;
* Ціна;
* Дані про матеріали для курсів.

Обсяг цих даних, залежить тільки від часу і кількості ключів для API, тобто для одного ключа можна отримати повні дані для десятьох курсів, тобто посилати запит, для кожного курсу окремо і таким чином отримати більш детальну інформацію про курс, або посилати запит для отримання інформації про десять курсів, проте ця інформація буде більш загальна. Таким чином за 50 секунд можна отримати дані про 10 курсів або про тисячу, оскільки максимальна допустима кількість запитів за 40 скунд становить 10, але якщо мати декілька ключів, то можна збільшити їхню кількість. Також обсяг даних обмежується розміром отриманої сторінки. Що ж таке цей розмір? Все дуже просто це кількість курсів за один запит при використанні courses-list, максимально можливо встановити 10000, проте реально повертаються тільки 100 курсів, решта що не “влізли” буде втрачено.

Для пошуку максимально подібних назв тем та підтем, було написано функцію short\_Jacquard\_algorithm – яка обробляє два тексти і повертає коефіцієнт подібності, при бажанні можна реалізувати його довгу версію, яка між ними відмінність? Перший видаляє пробіли, робить всі букви маленькими , а довгий видаляє з’єднувальні слова, закінчення і решту непотрібних символів, проте для цього треба мати файли з коренями слів та файл з з’єднувальних слів. А так алгоритм повертає коефіцієнт подібності двох текстів. На базі цього алгоритму працюють три інші функції. Як отримати більш доствірні дані? Чим більше коефіцієнт ближчий до 1 тим тексти подібніші. За допомогою цього одна функція, шукає подібні категорії до введених користувача і якщо такі знайшлись повертає знайдену категорію, якщо коефіцієнт менший за 0.5 вона повертає None, що означає що такої категорії не існує. За допомогою цього отримані результати будуть найбільш достовірні. Проте це не відміняє того, що знайдені дані і введені користувачем дані будуть суттєво відрізнятись, оскільки можна ввести althhe, а виведеться health.

Якість отриманих даних після обчислювальних операцій?

Загалом отримані вихідні дані є надзвичайно якісними, оскільки є велика кількість перевірки введеної мови, введеної категорії і т.д., після чого відсоток того, що отримані дані будуть невірними падає, оскільки якщо в переліку можливих категорій в файлах не знайдеться введеної категорії, програма попросить ввести ще раз категорію, або пропустити цей крок.

При обробці великих текстів збільшується шанс отримання недостовірних даних, оскільки при порівнянні велиеих текстів може бути присутня кількість непотрібних слів, які зменшать коефіцієнт і програма порахує що вони різні.

Взагалі виведені у файл дані не єдине що можна вивести, можна вивести наприклад які лекції воно порахувало за однакові, можна зробити сортування за середньою оцінкою користувачів.