Визначення та обґрунтування теми дослідження курсової роботи

Виконала Софія Геворгян

План:

1. Визначення та опис проблеми, якій буде присвячена курсова робота
2. Опис коротких функціональних можливостей АРІ
3. Приклад використання
4. Пропозиція по проведенню дослідження, вимоги на систему
5. Визначення та опис проблеми, якій буде присвячена курсова робота

Тематика: тема курсової роботи і, відповідно, область проблеми, вирішенню якої вона присвячена буде стосуватися загальнодоступної онлайн-енциклопедії «Вікіпедія» та роботи з її організаційною системою. За допомогою цього онлайн-ресурсу будь-який користувач мережі Інтернет може почерпнути для себе корисну і потрібну інформацію, а також за допомогою посилань і пов’язаних статей має змогу дізнатися про сторінки із подібною інформацією або ж дізнатися про тему більше.

Проблематика: з допомогою Вікіпедії безліч користувачів щодня дізнаються нову інформацію, тому часто виникає потреба отримати суттєві і корисні матеріали по всій тематиці, для прикладу: з допомогою Вікіпедії Ви дізнаєтеся про отруйні речовини ціаніди і для більшої обізнаності в темі Вас цікавлять речовини з подібними властивостями та характеристиками. Кожна сторінка Вікіпедії містить розділ «Див. також», який пропонує Вам статті на подібну тему.

Проблеми:

* Така розгалужена система пошуку із багатьма можливостями вибору може заплутати користувача, що призведе до неефективного netsurfing з можливістю пропуску важливої інформації (відкриття безлічі вікон, необхідність перевіряти чи усі посилання були переглянуті).
* При такому варіанті роботи часто користувач не має відчуття, що охопив весь матеріал теми, а тривалий і наполегливий пошук, який приводить до вже пройдених матеріалів, втомлює
* З іншого боку, Вікіпедія містить чудову організаційну структуру з допомогою якої можна знайти усю необхідну інформацію, просто систематично опрацьовуючи посилання за посиланням, що допомагає зберегти чітку організацію своєї роботи. Проте такий підхід інколи буває неефективний, коли ми, для прикладу, шукаємо лише деяку специфічну інформацію, для прикладу, колір речовини, на кожній сторінці, і тоді для використання організаційної структури необхідно виділити левову частку часу.

Суть роботи: створити можливість для користувачів опрацьовувати інформацію із певного розділу організаційної структури Вікіпедії, тобто на одній сторінці/в одному файлі подавати невелику кількість інформації із кожної із сторінок з обраного розділу (для прикладу, картинку, короткий опис та загальні характеристики об’єкту, один з яких може цікавити користувача – уся чітко структурована інформація, що відображається на сторінці результатів при введені пошукового запиту). Сторінки, кількість яких для ефективності роботи також можна буде вибрати, обиратимуться відповідно до кількості відвідувачів цієї сторінки за певний проміжок часу. Такий підхід дасть змогу також порівнювати отриману інформацію, для прикладу, отримати перелік найвідоміших творів епохи Ренесансу, що допоможе краще вивчити, яка назву і автора кожної картини і допоможе краще їх впізнавати.

Мета роботи: структурувати інформацію, узагальнити великі об’єми даних з метою кращого засвоєння/вивчення/опрацювання інформації користувачами Вікіпедії.

Цільова аудиторія: користувачі Вікіпедії

1. Опис коротких функціональних властивостей АРІ
2. Коротко про документацію АРІ

Для доступу до даних та функцій Вікіпедії використовується MediaWiki action API. Для цього використовується URL-адреса, формування якої було розглянуто на прикладі[[1]](#endnote-1).

<https://en.wikipedia.org/w/api.php?action=query&titles=Main%20Page&prop=revisions&rvprop=content&format=json&formatversion=2> – це простий приклад адреси для відправлення GET запиту.

<https://en.wikipedia.org/w/api.php> – частина, яка є базовим URL для англійської Вікіпедії.

action=query – параметр action вказує, на що саме ми просимо доступ, найчастіше це query - модуль, який використовується для отримання інформації(info) про сторінку (унікальну адресу, назву, останню зміну сторінки, кількість відвідувачів і тд), а також revisions – інформація про останню зміну сторінки[[2]](#endnote-2). Є також багато інших значень цього параметру.

titles=Main%20Page – параметр для дії query, який повідомляє назву сторінки, інформацію про яку ми хочемо отримати.

prop=revisions – або ж properties – параметр, який вказує, що саме ми можемо запитувати, у даному випадку – останні зміни сторінки. Має також значення info

rvprop=content – додатковий параметр, як уточнення для отримання останньої версії сторінки

format=json – параметр, який вказує на те, інформацію якого типу ми хочемо отримати, json у даному випадку.

Простий модуль url\_for\_api.py демонструє просте використання URL адреси для отримання інформацій про сторінку.

Детальніше було розглянуто модуль Query, адже саме він зосереджений на читанні або ж отриманні інформації. Модуль Query має багато підмодулів, один із яких – properties (prop), значення якого дозволяють отримувати ту чи іншу інформацію (модуль url\_for\_api.py за допомогою значення prop=info, inprop=watchers отримує інформацію про кількість користувачів). Як бачимо, інколи підмодулі можуть мати свої підмодулі, які уточнюють тип інформації, на яку надіслано запит.

1. Оболонки для АРІ

wikipedia – це бібліотека Python (одна з оболонок для MediaWiki API), що сфокусована на доступі і опрацюванні даних з Вікіпедії, таким чином ми фокусуємося на тому, щоб опрацьовувати дані, а не отримати їх[[3]](#endnote-3).

wikipedia.search(query) – функція, повертає список із назвами сторінок Вікі, які містять запит query.

wikipedia.suggest(query) – функція, повертає можливу назву сторінки Вікі або ж None

wikipedia.summery(query) – повертає короткий зміст статті з можливістю вказувати кількість речень/символів. Також можливе виникнення помилок через існування кількох сторінок або ж відсутністю такої сторінки у Вікі.

wikipedia.page(query) – дозволяє присвоїти змінній тип WikipediaPage і надає змогу отримувати дані, такі як назва сторінки, url-адресу, фото, зміст і т.п.

wikipedia.set\_lang(“fr”) – змінює мову, якою надаються відповіді

class wikipedia.WikipediaPage(title) – містить дані з вікі сторінки, такі як

categories – список категорій сторінки

content – загальний контент сторінки без зображень і таблиць

coordinates – якщо є інформація, то повертає ширину і довготу

html() – повертає повну html сторінку

images – список посилань зображень на сторінці

1. Приклади використання

Один із прикладів проектів, який уже реалізовано на основі інформацій із Вікіпедії - <http://stars.chromeexperiments.com/>

Проект, базуючись на інформації із Вікісторінок на астрономічну тематику, будує модель галактики, таким чином узагальнюючи інформацію про зірки і планети і даючи змогу користувачу в зручному і оперативному режимі структурувати інформацію.

У модулях get\_wiki\_table.py, simple\_functions.py та url\_for\_api.py наведено приклади використання деяких функцій бібліотеки wikipedia.

Зокрема розглянуто, як саме доступатися до конкретних даних на сторінці вікі, як отримувати такі дані, як зображення, короткий зміст.

Надзвичайно важливу для проекту інформацію можна почерпнути за допомогою формування url-адреси, такі як кількість переглядів сторінки, дані про останні зміни на сторінці та багато іншого.

1. Пропозиція по проведенню дослідження, вимоги на систему

* Спонсор проекту:

Ініціатор: Геворгян Софія

* Бізнес-потреба:

Підвищити і розширити організаційні структури Вікіпедій, надати можливість користувачу краще засвоювати інформацію, підвищити функціональну здатність організаційної структури, надати можливість візуально вивчати інформацію про кілька творів мистецтва за фотографіями, які знаходяться на одній сторінці.

* Бізнес-вимоги

надати користувачу можливість отримувати узагальнену інформацію з певної категорії вікісторінок. надати можливість за вказаною назвою Вікісторінки шукати подібні сторінки, із певної категорії сторінок вибирати ті, які відвідуються найчастіше і надавати коротку інформацію про кожну з них

* Питання та обмеження:

Граничний термін розробки – 25 травня

1. https://www.mediawiki.org/wiki/API:Main\_page [↑](#endnote-ref-1)
2. http://2015.compjour.org/tutorials/exploring-wikipedia-api-intro/ [↑](#endnote-ref-2)
3. https://pypi.python.org/pypi/wikipedia [↑](#endnote-ref-3)