**Методичні вказівки**

до виконання лабораторної роботи №4

"Використання засобів для створення онтологій"

з курсу

“Обробка надвеликих масивів даних”

Київ – 2018

**Ціль роботи:**

Навчитись створювати онтології. Навчитись працювати з застуванням Protege.

Зрозуміти взаємозв'язок онтологій з надвеликими масивами даних.

**Завдання:**

1. Продивитись одну з існуючих онтологій та описати її елементи у протоколі (наприклад, взяти одну зі списку: <https://www.bbc.co.uk/ontologies>)

2.Зареєструватись на <https://webprotege.stanford.edu>(також, можна використовувати офлайн-версію Protege)

3.Створити власну онтологію, з близько 10-ма класами у ній (можна взяти предметну область з бакалаврської/магістерської роботи)

4.Дати відповіді на контрольні запитання.

**Теоретичні відомості:**

Онтологія формально представляє знання як ієрархію понять в межах одного домену, використовуючи загальний словник для позначення типів, властивостей та взаємозв'язків цих понять.

Онтології є структурними фреймворками для організації інформації та використовуються в штучному інтелекті, Semantic Web, системній інженерії, програмній інженерії, біомедичній інформатиці, бібліотечній справі, закладках підприємств та інформаційній архітектурі, як форми подання знань про світ або деяку його частину. Створення доменних онтологій також є основою для визначення та використання структури корпоративної архітектури.

Що стосується тенденції надвеликих масивів даних:

Наприклад корпоративні застосунки Google. Замість того, щоб інтегрувати безліч різних корпоративних програм в організацію, щоб отримати, наприклад, 360-градусний погляд на клієнтів, онтологія дозволяє користувачам шукати схематичну модель усіх даних у програмах. Вони витягують відповідні дані з вихідної програми, такі як CRM-система, Big Data застосунки, файли, гарантійні документи тощо. Ці витягнуті семантики зв'язані в пошуковий граф замість схеми, щоб дати користувачам потрібні результати.

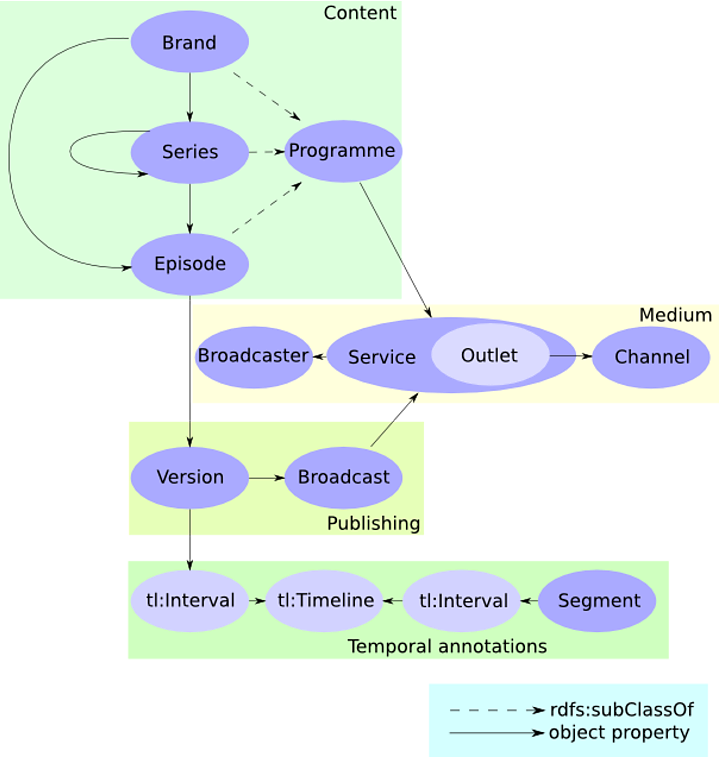
Онтологія надає користувачам інший підхід у застосуванні корпоративних програм, позбавляючи потреби інтегрувати різні програми. Це дозволяє користувачам шукати та об'єднувати додатки, бази даних, файли, електронні таблиці та ін. Продукт онтології дуже цікавий тим, що за минулі роки організації розробили і використовували величезну кількість корпоративних додатків для різних потреб і різних вимог. Інтеграція цих програм із метою інтегрованого перегляду в масштабах компанії є складною, дорогою та часто ризикованою.

Онтологія використовує унікальну комбінацію спадкової семантичної моделі на основі графів та семантичного пошуку, що дозволяє поступово зменшувати часові показники та вартість складних інтеграційних завдань. Онтологія переосмислює збір даних, співвідношення даних та проекти міграції даних у світі після Google.

**План виконання роботи:**

Описати одну з існуючих онтологій. Для прикладу візьмемо [онтологію програм](онтологію%20програм) з сайту <https://www.bbc.co.uk/ontologies>

Ця онтологія спрямована на надання простого словника для опису програм. Вона охоплює бренди, серіали (сезони), епізоди, трансляції подій, мовлення та ін.



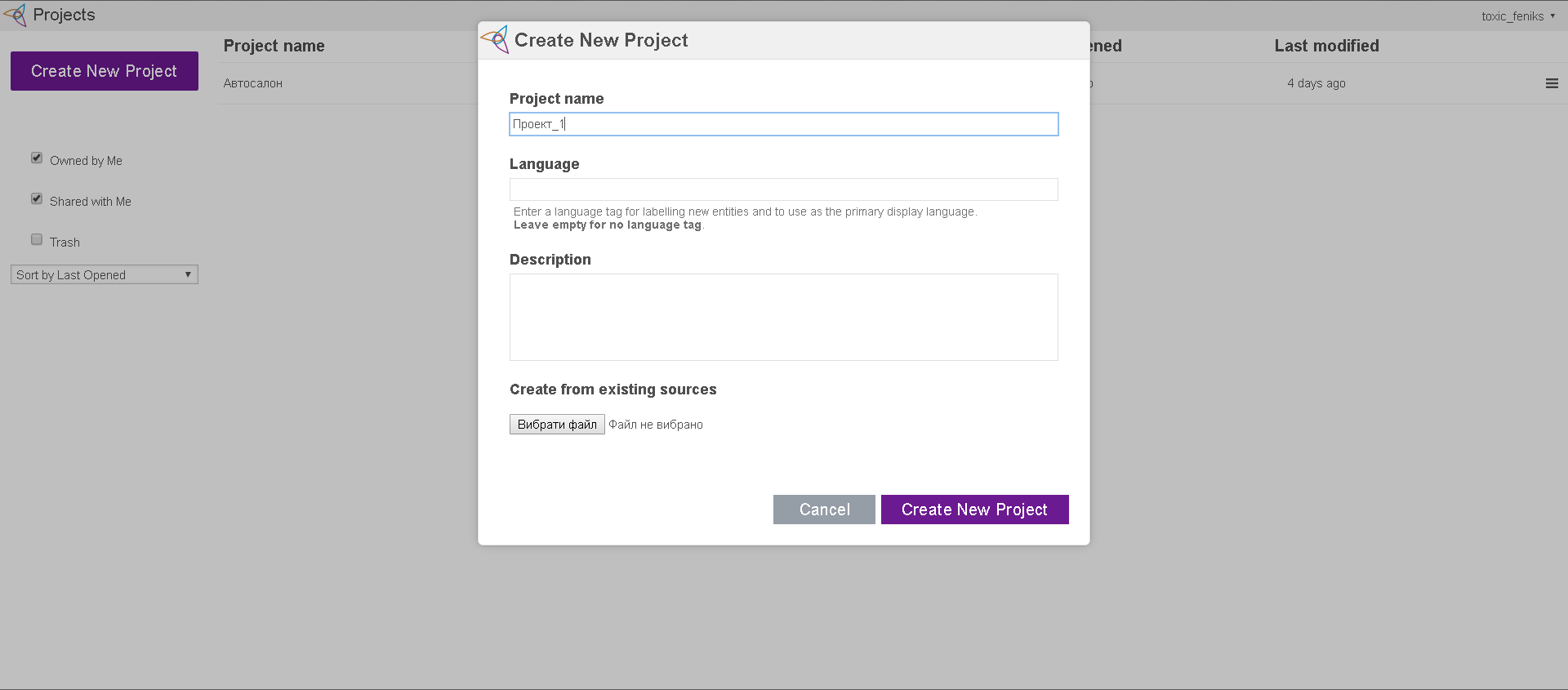
По даній діаграмі бачимо, що в онтології є такі класи:

* Контент
  + Бренд (підклас – Програма, властивості: Сезон(Серії), Епізод)
  + Серії (підклас – Програма, властивості: Сезон(Серії), Епізод)
  + Епізод (підклас – Програма, властивість – Версія)
  + Програма (властивість – Сервіс)
* Подача (Medium)
  + Транслятор
  + Сервіс (властивості – Транслятор, Канал)
    - Випуск (властивість − Канал)
  + Канал
* Публікація
  + Версія (властивості – Передача, Інтервал)
  + Передача (властивість – Сервіс)
* Тимчасові анотації
  + Інтервал (властивість – Хронологія)
  + Хронологія
  + Сегмент (властивість – Інтервал)

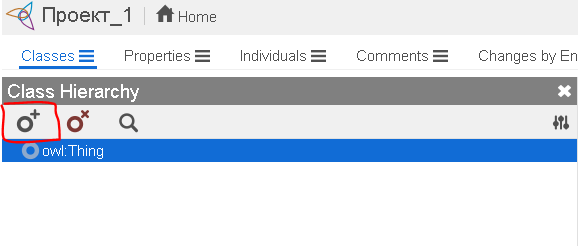
Усі класи пов’язані між собою за двома принципами: властивість (стрілка) або підклас (пунктир).

Зайти на <https://webprotege.stanford.edu> та виконати просту реєстрацію або завантажити оффлайн версію з офіційної сторінки <https://protege.stanford.edu/>.

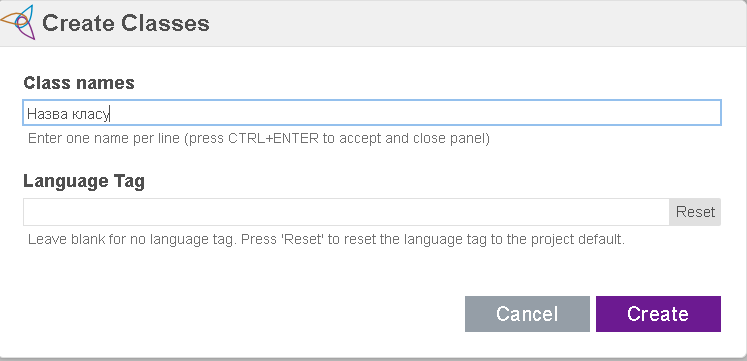
Після реєстрації створимо проект



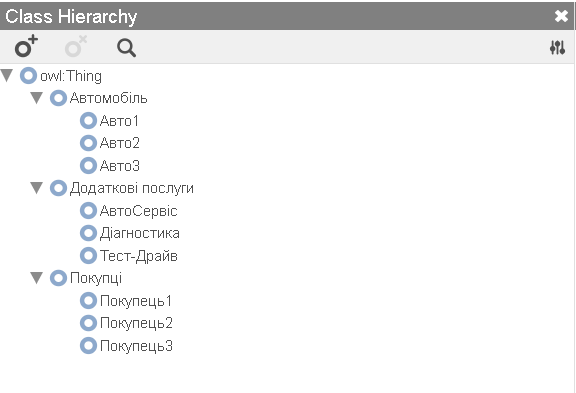
Створимо декілька класів онтології, елементи яких будуть взаємодіяти між собою. Для цього потрібно натиснути на кнопку Create в меню Ієрархія класів.



Задати назву класу.

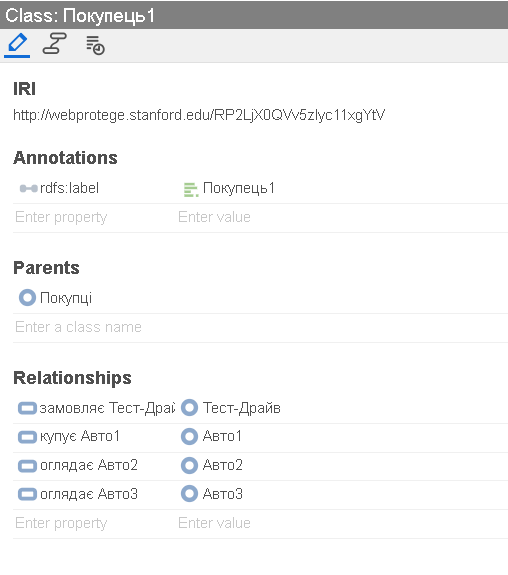


Для прикладу ми створили таку ієрархію класів.



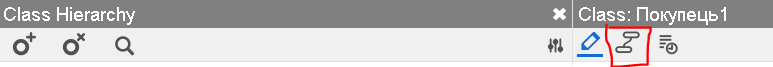
У нас є клас Автомобіль, що містить декілька автомобілів, Додаткові послуги, а також Покупці.

Задамо певні властивості для покупців. Наприклад покупець 1 оглядає авто, замовляє Тест-Драйв та купує Авто1. Для інших аналогічні властивості.

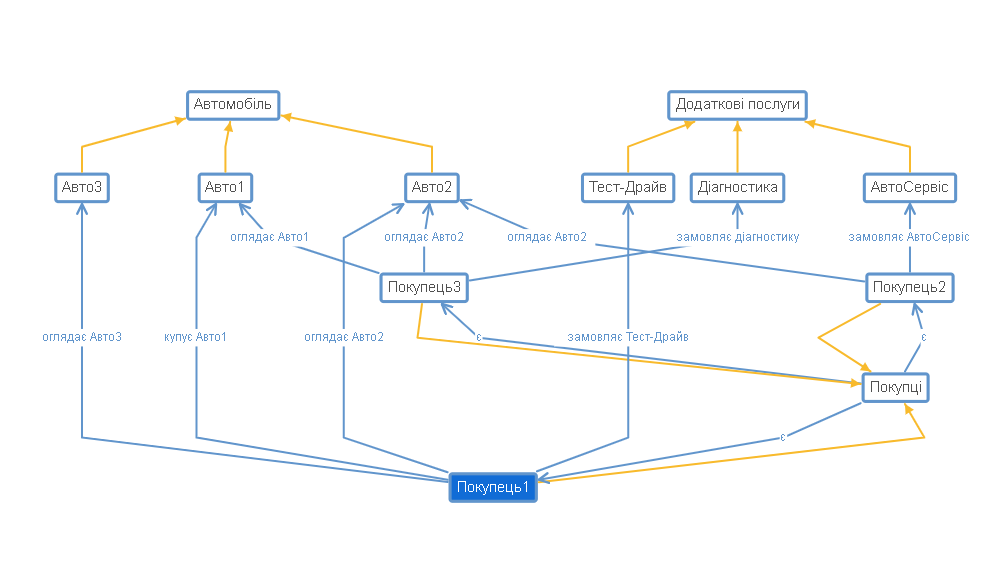


Коли задали всі властивості, що нам потрібні, відкриємо представлення онтології у вигляді графу.

Для цього потрібно натиснути на кнопку Entity Graph.



В результаті отримали такий граф онтології.



Детальну російськомовну інструкцію по оффлайн версії ви можете знайти за даним посиланням <http://profi.net.ua/component/content/article/123.html>.

**Контрольні запитання:**

1) Що таке онтологія? Які сфери їх застосування?

2) Які складові частини онтологій?

3) Що таке мова OWL?

4) Яку роль відіграють онтології у науці про великі дані?

5) Що таке Protege?

**Список посилань:**

1. Стаття про відношення онтологій до Big Data − <https://datafloq.com/read/role-ontology-plays-big-data/749>
2. Інструкція до веб-інтерфейсу WebProtege −

<https://protegewiki.stanford.edu/wiki/WebProtegeUsersGuide->

1. Російськомовна інструкція до оффлайн застосування Protege −

<http://profi.net.ua/component/content/article/123.html>

1. Що таке онтологія в сфері IT − <https://is.gd/eEblWp>