УДК 004.415

# ІНТЕГРАЦІЯ ДАНИХ У МЕРЕЖІ ІНТЕРНЕТ: ЗВ'ЯЗАНІ ДАНІ

*О.В. Новицький*

Інститут програмних систем НАН України,
03187, Київ, проспект Академіка Глушкова, 40, alex@zu.edu.ua

Розкривається підхід до інтеграції даних, у тому числі наукових результатів, у мережі Інтернет в рамках концепції зв'язаних даних. Зокрема демонструється методологія, за якою семантично описаний контент може автоматично публікуватися та втягуватися в єдину базу RDF посилань.

The paper reveals an approach to data integration, including research results on the Internet within the concept of linked data. In particular, demonstrate the methodology by which, semantically described content can automatically be published and to get involved and unified database RDF links.

## Вступ

В інтеграції інформації можна виділити наступні проблеми: інтеграція схеми [1], сховищ даних, інтеграція даних (також відома як інтеграція інформації підприємства, EII enterprise information integration) та інтеграція каталогу. Підхід до інтеграції даних з використанням онтологій називається інтеграція даних на основі онтологій. Загалом, є ряд кроків, які необхідно виконувати при інтеграції інформаційних систем з використанням онтологій. До них належать:

- інтерпретація запиту в термінологію загальної онтології;
- виявлення відповідності між семантично пов'язаними сутностями в локальній і загальній онтології;
- переклад відповідних даних з локальних інформаційних джерел (що беруть участь в обробці запиту користувача) в формалізм представлення знань системи інтеграції інформаційних систем;
- узгодження результатів отриманих з різних локальних інформаційних джерел, а саме, виявлення та усунення, наприклад, надлишку, дублювання та ін.

Семантична гетерогенність [2], як правило, відрізняється від синтаксичної та структурної гетерогенності в сімействі баз даних [3–6].

Синтаксична неоднорідність пов'язана з неоднорідністю форматів даних. Стандартизація форматів даних приймається як підхід до вирішення проблем синтаксичної неоднорідності. Наприклад, XML використовується як стандартний формат для всіх видів доступних Web-даних.

Структурні неоднорідності пов'язані з різними моделями даних, структур даних або схем, наприклад, реляційних і об'єктно-орієнтованих моделей бази даних. Прикладом вирішення проблеми структурної неоднорідності є використання RDF, який заснований на синтаксисі XML і забезпечує уніфікований спосіб структури джерел інформації.

Незважаючи на те, що в електронній бібліотеці інформація може бути представлена в різних видах, семантика цієї інформації видається за допомогою текстових метаданих, відповідно будемо зосереджуватися на інтеграції семантичних метаданих.

Якщо два інформаційні джерела, змодельовані в одному й тому ж форматі даних із застосуванням однієї й тієї ж моделі даних, можуть виникати проблеми семантичної неоднорідності [7]:

- семантичні конфлікти. Різні розробники моделей не відчувають точно такий набір об'єктів реального світу, але замість цього вони представляють набори, які перетинаються (включення або перекриття елементів набору). Наприклад, "Студент", об'єкт класу може виникнути в одній схемі, тоді як більш обмежений об'єкт класу "Студенти спеціальності інформатика" знаходиться в іншій схемі. При інтеграції двох схем клас "Студент спеціальності інформатика" буде інтегрований як підклас класу "Студент";
- описові конфлікти належать до конфліктів іменування внаслідок омонімів та синонімів.
- структурні конфлікти відрізняються від структурної неоднорідності. Навіть якщо два розробники моделей, використовують одну й ту ж модель даних, вони можуть вибирати різні конструктори для подання об'єктів реального світу. Наприклад, в об'єктно-орієнтованій моделі розробник описуючи компонент об'єкта типу *O* постає перед вибором створення нового типу об'єкта або додати атрибут до *O*.

Кожен домен використовує локальні онтології, які є результатом концептуалізації домену. Оскільки процес концептуалізації не є однозначним, це породжує гетерогенність джерел. Для того, щоб їх об'єднати, необхідно зробити більше, ніж простий механізм маркування відповідності об'єктів, класів або змісту. Насправді, часто виникає ситуація, коли поняття не зовсім збігаються, оскільки вони можуть мати відмінності за властивостями в видовій або родовій класифікації. У цілому можна виділити декілька видів співставлення онтологій:

© О.В. Новицький, 2010

• розширення – передбачає визначення онтології домену, пов'язуючи деякі поняття між двома вихідними онтологіями. Дві концептуальні моделі доповнюють одна одну, наприклад, концепти першої онтології уточнюються в другій через додаткові атрибути, які не зазначені в першій;

• гармонізація – припускає семантичну еквівалентність між доменом і прикладними онтологіями, що стосується одного й того ж онтологічного зобов'язання. У цьому випадку поточний домен можна розглядати як спеціалізацію в іншому домені, який є більш загальним або розташований на абстрактно-формальному рівні;

• вирівнювання – припускає узагальнення онтології домену через загальні поняття і аксіоми. Обидві моделі мають (багато / кілька) загальних спільних концептів.

Для електронних бібліотек вирішення проблеми семантичної гетерогенності можна вирішувати на двох рівнях – метаданих і контенту. В першому випадку маємо справу з описовими схемами даних, в другому – зі смисловими даними.

## Зв’язані дані

Для часткового розв’язання вищеописаних проблем зручним механізмом у рамках мережі Інтернет є сервіс DNS. Доменна система імен (англ. Domain Name System, DNS) – розподілена система перетворення імені хоста (комп'ютера або іншого мережевого пристрою) в IP-адресу. Кожен комп'ютер в Інтернет має свою власну унікальну IP-адресу – число, яке складається з чотирьох байт, система адресації є глобальною.

У даному випадку для нас є важливим та властивість, що ім’я хоста (доменне ім’я) є унікальним, і у випадку ідентифікації концепту з використанням DNS знімає проблему описових конфліктів. Моделлю даних, яка заснована на DNS є Linked Data.

Основні принципи Linked Data висвітлено в [http://www.w3.org/DesignIssues/LinkedData.html]. Перевага зв'язаних даних полягає в тому, що цінність і корисність даних збільшується, чим більше вони пов'язані з іншими даними. Основні принципи зв’язаних даних є.

1. Слід використовувати URIs як імена для сутностей;
2. Слід використовувати HTTP URIs, щоб люди могли побачити ці імена;
3. Якщо хтось шукає в URI, слід представляти корисну інформацію;
4. Містити посилання на інші URI для того щоб вони могли дізнатися більше сутності.

Природно постає питання розв’язання узгодженості описання ресурсів у рамках вищеописаних конфліктів.

Для розуміння природи проблеми необхідно визначити саме поняття Semantic Web, яке можна представити як бачення або ціль, де семантично багата анотація даних використовується машинними агентами для пошуку інформації. Ми перебуваємо на шляху до цієї мети або цілі, при цьому інтерпретуючи, що Semantic Web є більше процесом, ніж станом.

Саме поняття Semantic Web є багатогранним і чітко не визначеним. Його слід розуміти як здатність машин обробляти та розуміти дані, які розміщені в інформаційних ресурсах. Частим є запитання як відносяться Semantic Web та Linked Data. Фактично LD це перенесення технології гіперпосилань Веб-документів для зв’язування RDF трійок.

Деякі автори в своїх роботах [8–9] асоціюють це поняття з Semantic Web. Однак цей підхід повністю не відображає всіх аспектів Semantic Web, наприклад таких, як динамічність Semantic Web. Окрім того що інформація постійно змінюється, в середовищі Semantic Web функціонують і агенти, які цю інформацію оброблюють і в рамках концепції LD можуть вносити до неї певні зміни. Тому ми схильні думати, що LD – приклад частинної реалізації Semantic Web. Це дає відповідь як шукати документи в Semantic Web.

Застосування обох принципів призводить до створення в Веб спільних даних, які часто називають Веб-даними або Semantic Web.

Доступ до Веб-даних можна отримати при використанні LD браузерів, так само, як традиційний доступ до Веб-документів – за допомогою HTML-браузерів.

Однак замість того, щоб переміщатися між посиланням HTML-сторінок, LD браузери дозволяють користувачам переміщатися між різними джерелами даних, виконавши RDF посилання.

Це дозволяє почати з одного джерела даних, а потім пройти через потенційно нескінченну кількість джерел Web-даних, пов'язаних з RDF посиланням. При таких переходах внаслідок семантичних конфліктів та неоднорідності можуть виникати помилкові посилання, що будуть спотворювати зміст. Водночас неможливо повністю формалізувати всі взаємовідносини між концептами реального світу, а також відобразити асоціації, які властиві явищам та предметам реального світу.

Так як у традиційних Веб-документах можуть бути проскановані всі гіпертекстові посилання та переходи RDF-посиланнь Веб-даних. У даному випадку робота з такими даними має ряд переваг. Пошукові системи можуть надавати складні запити можливості, аналогічні тим, які передбачені звичайними реляційними базами даних. Оскільки результати запиту до структурованих даних є структуровані дані, а не лише посилання на HTML-сторінки, вони можуть бути негайно оброблені, дозволяючи, таким чином, новий клас програм, заснованих на Веб-даних.

Основні принципи LD тісно пов’язані з архітектурою Інтернет. Одним з основних понять в архітектурі є ресурс та представлення. Детальний зміст цих понять наведено в [10].

**Ресурс.** Для опублікування даних в Інтернет спочатку маємо ідентифікувати елементи, що представляють інтерес у нашому домені. Вони є сутності, чиї властивості й відносини ми хочемо описати в даних. За термінологією Веб-архітектури всі елементи, що представляють інтерес, називаються ресурсами.

У роботі [11] розрізняються два види ресурсів: інформаційні та неінформаційні ресурси (які також називаються "інші ресурси"). Ця різниця є дуже важливою у цьому контексті LD. Всі ресурси, які ми знаходимо на традиційних Веб-документах, наприклад, документи зображень та інші мультимедійні файли є інформаційними ресурсами.

Поняття «інформаційний ресурс» введено в [http://www.w3.org/TR/webarch/], тому що було відмічено корисність його використання для технологій мережі Інтернет.

Насправді Technical Architecture Group (TAG) не дає чіткої відповіді на питання різниці між інформаційними та неінформаційними ресурсами. Якщо взяти за основу підхід викладений в [12], тобто, якщо на GET запит при розіменуванні повертається результат з кодом 303, то це не інформаційний ресурс.

Багато сутностей, дані про які ми хочемо спільно використовувати, не є даними в прямому розумінні цього слова, наприклад: особи, фізичні об’єкти, місця, наукові концепції і т. д.

Як правило, всі "об'єкти реального світу", які існують поза Інтернет, – неінформаційні ресурси.

**Ресурсні Ідентифікатори.** Ресурси ідентифікуються за допомогою Uniform Resource Identifiers (Уніфіковані Ідентифікатори).

У контексті LD обмежуються використанням тільки HTTP URIs і не допускаються інші URI схеми, такі як URNs і DOIs.

HTTP URIs хороші з двох причин: вони забезпечують простий спосіб створення глобально унікальних імен без централізованого управління, а також URIs працюють не тільки як назва, але й як засіб доступу до інформації про ресурс через Інтернет внаслідок роботи служби DNS.

**Представлення.** Інформаційні ресурси можуть мати представлення. Представлення це потік байтів у певному форматі, наприклад, HTML, RDF / XML або JPEG. Наприклад, рахунок-фактура є інформаційним ресурсом Він може бути представлений як HTML сторінка, для друку – PDF документом, або RDF документом. Один інформаційний ресурс може мати різні представлення, наприклад, у різних форматах або на різних природних мовах.

**Розіменування HTTP URIs ідентифікаторів.** Розіменування URI це процес перетворення URI для отримання інформації про розташування ресурсу в мережі. У висновках проекту W3C TAG [12] представлено різницю виявлення інформаційних та неінформаційних ресурсів при розіменуванні URI.

*Інформаційні ресурси:* коли ідентифікаційний URI інформаційного ресурсу є розіменованим, сервер, власник URI, як правило, породжує нове представлення або новий екземпляр інформаційного ресурсу в нинішньому стані та відправляє його назад клієнту, використовуючи HTTP код відповіді 200 OK.

*Неінформаційні ресурси* не можуть бути розіменовані напрямки. Тому Веб-архітектура використовує спеціальний прийом, щоб URIs могли ідентифікувати неінформаційні ресурси, які будуть розіменовані: Замість того щоб представити ресурс, сервер відправляє клієнту URI інформаційного ресурсу, який описує, неінформаційний ресурс з використанням HTTP коду відповіді 303. Це називається 303 переадресацією. Другий крок – клієнт розіменовує цей новий URI й отримує представлення ресурсу з описом неінформаційного ресурсу.

Серед великого різноманіття технологій, що пов’язані з Semantic Web, важливо встановити співвідношення в якому перебувають LD до цих технологій.

Розглянемо технологію RDFa та її місце щодо LD. RDFa призначений для створення семантичної розмітки контенту. Семантична розмітка або анотування – явний опис семантики контенту ресурсу за допомогою понять семантичний моделі (онтології або словника). Явний опис семантики виконується з зазначенням чіткої відповідності між певною частиною контенту ресурсу та його семантикою, описаною в семантичний моделі. Анотація це визначення семантики формальним способом. На даний момент в основу анотації покладають модель даних RDF. Сьогоднішні Web-ресурси розробляються здебільшого для використання людьми. Незважаючи на поступову появу в мережі даних, призначених для машинного сприйняття, вони, в основному, подаються окремим файлом у певному форматі.

При цьому відповідність машинної версії людському представленню досить обмежена. Як наслідок, Web-браузери можуть забезпечити користувачів лише мінімальною підтримкою в аналізі та обробці мережевих даних. Адже браузери тільки представляють інформацію. Технологія RDFa [13], дозволяє супроводжувати графічні дані машиночитаючими підказками з допомогою набору XHTML-атрибутів. RDFa – це спосіб вираження RDF-даних в XHTML, в рамках якого дані, призначені для повторного використання людиною.

Зв’язані дані дають можливість використання Інтернет для підключення відповідних даних, які раніше не були пов'язані між собою, або використовуючи Web знизити бар'єри для зв’язування даних які в даний час пов'язані з використанням інших методів. Або більш конкретно, LD – це термін, який використовується для опису рекомендованих найкращих методів для виявлення, спільного використання та підключення частин даних, інформації та знань в Semantic Web, використовуючи URIs і RDF [14].

**Вибір словників для представлення інформації.** Для здійснення анотування, яке в подальшому може бути легко оброблене програмними додатками, необхідно повторно використовувати терміни (де це можливо) з відомих словників. Нові терміни мають визначатися тільки тоді, якщо не знайдено необхідних термінів в існуючих словниках. Це є одним із можливих рішень проблематики семантичних конфліктів. Найбільш популярними словниками є:

Friend-of-a-Friend (FOAF) – словниковий запас для опису людей.

Dublin Core (DC) визначає загальні атрибути метаданих.

Semantically-Interlinked Online Communities (SIOC) словник для представлення онлайнових співтовариств.

Description of a Project (DOAP) – словниковий запас для опису проектів.

Simple Knowledge Organization System (SKOS) – словник для представлення таксономії і слабко структурованих знань.

Music Ontology забезпечує терміни для опису виконавців, альбомів та треків.

Review Vocabulary – лексика для подання відгуків.

Creative Commons (CC) – словниковий запас для опису умови ліцензії.

Більш великий список відомих словників ведеться в ESW Wiki [15].

Загальноприйнятою є практика змішування термінів з різних словників. Особливо рекомендується використовувати rdfs:label та foaf:depiction властивостей (якщо це можливо), оскільки ці терміни добре підтримуються клієнтськими додатками.

Якщо потрібне URI посилання на географічні місця, напрямки досліджень, загальні теми, книги тощо, необхідно використовувати Уніфіковані Ідентифікатори з джерел даних в рамках проекту W3C SWEO Linking Open Data [16], наприклад GeoNames, DBpedia, MusicBrainz, dbtune або RDF Book Mashup. Дві основних переваги використання Уніфікованих Ідентифікаторів з цих джерел даних.

• Уніфіковані Ідентифікатори розіменовуються, це означає, що опис цієї концепції може бути отриманий з Інтернет. Наприклад, за допомогою URI DBpedia http://dbpedia.org/page/Doom можна визначити значну інформацію про комп'ютерну гру Doom, у тому числі опис на різних мовах і різних класифікацій.

• URI вже пов'язані з URI з інших джерел даних. Наприклад, можливо переходити від даних URI DBpedia http://dbpedia.org/resource/Berlin до даних представлених на GeoNames і EuroStat. Використовуючи концепцію URI, ці дані з'єднуються з багатьма іншими даними, утворюючи мережу зв'язаних даних.

## Семантична анотація, як інтеграція даних

Зв'язані дані надають можливість інтегрувати між собою дані, які розміщені в мережі Інтернет.

У випадку структурованих коротких описових метаданих (наприклад, у рамках DC) цей процес можна автоматизувати. Але для автоматичного аналізу змісту документа таких анотацій явно недостатньо. Тому останнім часом велика увага приділяється більш докладному розкриттю сенсу контенту через його анотації.

На даний момент існує підгрунтя для онтологічного підходу інтеграції інформації. По-перше,нині вже створено достатню кількість онтологій у різних предметних областях, наприклад, Basic Formal Ontology [http://www.ifomis.org/bfo/], CIDOC Conceptual Reference Model [http://cidoc.ics.forth.gr/], Open Biomedical Ontologies [http://www.obofoundry.org/] і т. д. По-друге, розроблено ряд програм, які сприяють практичному впровадженню Semantic Web.

Важливим етапом на шляху інтеграції інформації в Semantic Web є прийняття рекомендації мови запитів SPARQL (W3C Recommendation, January 15, 2008) та рекомендації з повторного використання RDF-даних у XHTML RDFa (W3C Recommendation, October 18, 2008).

Семантична розмітка або анотування – це явний опис семантики контенту ресурсу за допомогою понять семантичної моделі (онтології або словника). Явний опис семантики виконується за значенням чіткої відповідності між певною частиною контенту ресурсу та його семантикою, описаної у семантичній моделі. Анотація при цьому базується на RDF.

Нинішні Web-ресурси розробляються здебільшого для використання їх людьми. Незважаючи на поступову появу в мережі даних, призначених для машинного сприйняття, ці дані в основному подаються окремим файлом у певному форматі. При цьому відповідність машинної версії людському поданню досить обмежена. Як наслідок, Web-браузери можуть забезпечити користувачів лише мінімальною підтримкою в аналізі та обробці мережевих даних, адже браузери тільки представляють інформацію. Технологія RDFa [17] дозволяє супроводжувати графічні дані машиночитаними підказками за допомогою набору XHTML-атрибутів. RDFa – це спосіб вираження RDF-даних в XHTML, в рамках якого дані, призначені для людини, використовуються повторно.

Прикладом використання RDFa може слугувати закладання фрагмента коду, що описує назву та автора статті, яка розташована в електронній бібліотеці. При описі використовується схема метаданих Дублінського Ядра (xmlns: dc = http://purl.org/dc/elements/1.1/).

**Фрагмент XHTML коду ЕБ з розміткою RDFa**

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:dc="http://purl.org/dc/elements/1.1/">
<head profile="http://www.w3.org/2003/g/data-view">
<title> Доповідь про http://oai.org.ua</title>
</head>
<body>
  <h1>Ресурс http://oai.org.ua</h1>
  <dl about="http://eprints.zu.edu.ua/2648/">
  <dt>Назва доповіді</dt>
  <dd property="dc:title">Інтеграція наукових електронних бібліотек України: всеукраїнський портал збору та пошуку метаданих http://oai.org.ua</dd>
   <dt>Автор</dt>
<dd property="dc:creator">Новицький, О.В.</dd>
  </dl>
 </body>
</html>
```

Сам по собі механізм RDFa мало цікавий, хоч і визначає семантику контенту. Необхідною умовою є можливість вилучення зі сторінок семантичної анотації. Такий механізм розроблений та має назву Gleaning Resource Descriptions from Dialects of Languages GRDDL (http://www.w3.org/TR/grddl/).

За допомогою GRDDL можливо витягати мікроформатний контент. Специфікація GRDDL визначає розмітку на основі існуючих стандартів для оголошення про те, що XML документ містить у собі дані сумісні з RDF, а також посилання на алгоритм (як правило, представлений в XSLT), для отримання даних з документа.

Розмітки містять визначення простору імен загального призначення для XML-документів, а також посилання на профіль відносин для використання в валідних XHTML документах.

Далі представлений фрагмент XHTML коду відповідно до GRDDL.

**Фрагмент XHTML коду з розміткою RDFa та GRDDL**

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/">
<head profile="http://www.w3.org/2003/g/data-view">
<link rel="transformation" href="RDFaRDF.xsl"/>
<title> Доповідь про http://oai.org.ua</title>
</head>
<body>
  <h1>Ресурс http://oai.org.ua</h1>
  <dl about="http://eprints.zu.edu.ua/2648/">
  <dt>Назва доповіді</dt>
  <dd property="dc:title">Інтеграція наукових електронних бібліотек України: всеукраїнський портал збору та пошуку метаданих http://oai.org.ua</dd>
  <dt>Автор</dt>
  <dd property="dc:creator">Новицький, О.В.</dd>
  </dl>
 </body>
</html>
```

При обробці даного фрагмента засобами XSLT буде отримана модель даних і представлена в RDF за допомогою XML.

**RDF представлений за допомогою XML**

```
<rdf:RDF xmlns:h="http://www.w3.org/1999/xhtml" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about="">
<transformation xmlns="http://www.w3.org/1999/xhtml" rdf:resource="RDFa2RDFXML.xsl"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://eprints.zu.edu.ua/2648/">
  <dc:title xmlns:dc="http://purl.org/dc/elements/1.1/">Інтеграція наукових електронних бібліотек України: всеукраїнський портал збору та пошуку метаданих http://oai.org.ua</dc:title>
  </rdf:Description>
  <rdf:Description rdf:about="http://eprints.zu.edu.ua/2648/">
  <dc:creator xmlns:dc="http://purl.org/dc/elements/1.1/">Новицький, О.В.</dc:creator>
  </rdf:Description>
</rdf:RDF>
```

Варто звернути увагу на можливість GRDDL перетворення розмітки RDFa (для якої, наприклад, використовується схема даних Дублінського Ядра) безпосередньо в інші схеми метаданих, такі як CIDOC-CRM.

Такий підхід застосовується для Веб-документів, але в майбутньому можливе застосування даної технології до мультимедіа форматів.

Ще одним застосуванням даного підходу може бути процес запропонований у [18].

Приклад автоматичного внесення документів (з можливістю розподіленості) та побудови індексів. Ідея полягає в GRDDL обробці джерел документів та витягування вбудованого RDFa для підключення в сховища RDF. Такими сховищами RDF є проект W3C SWEO Linking Open Data [19], який об'єднує понад 142 млн. RDF посилань (рис. 1).

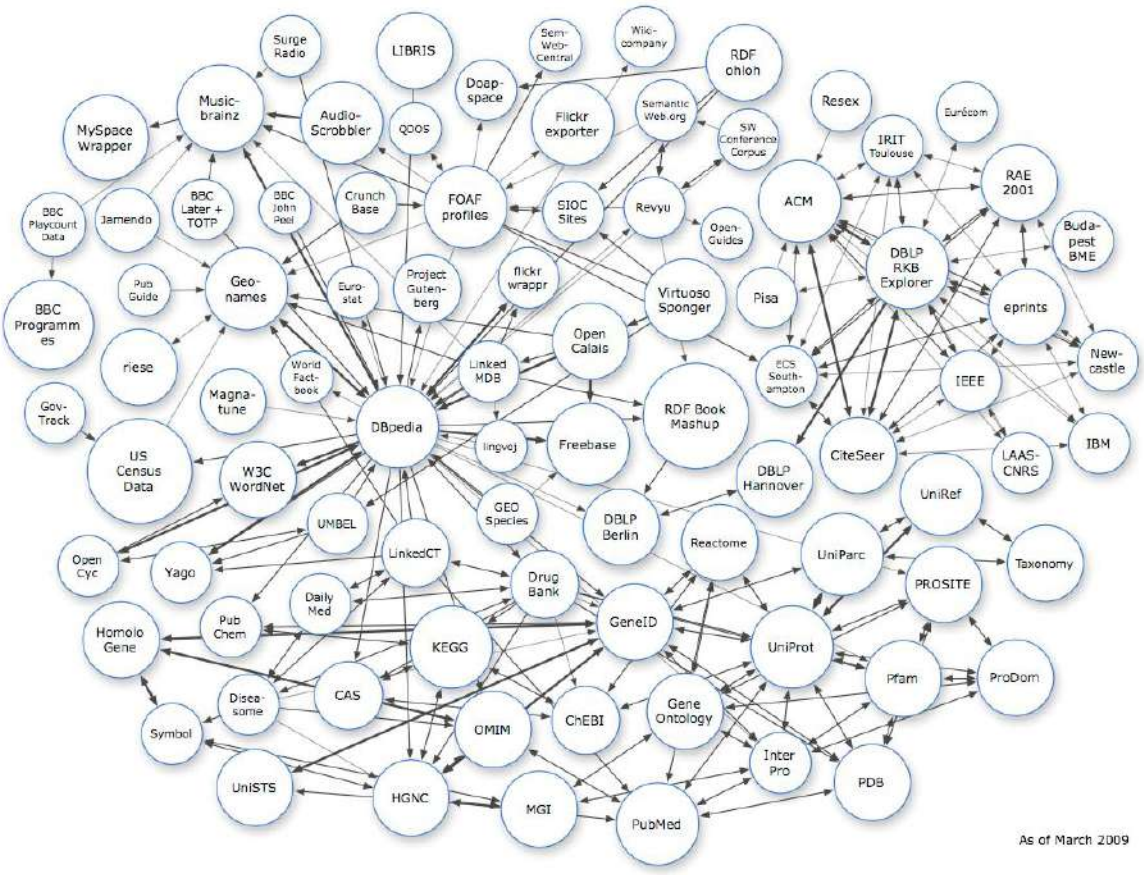


Рис. 1. Хмара зв’язаних даних Linking Open Data

Далі SPARQL запити вибирали б із цього сховища відповідні результати, які були б представлені у вигляді Веб-сторінки, що автоматично генерується (рис. 2).

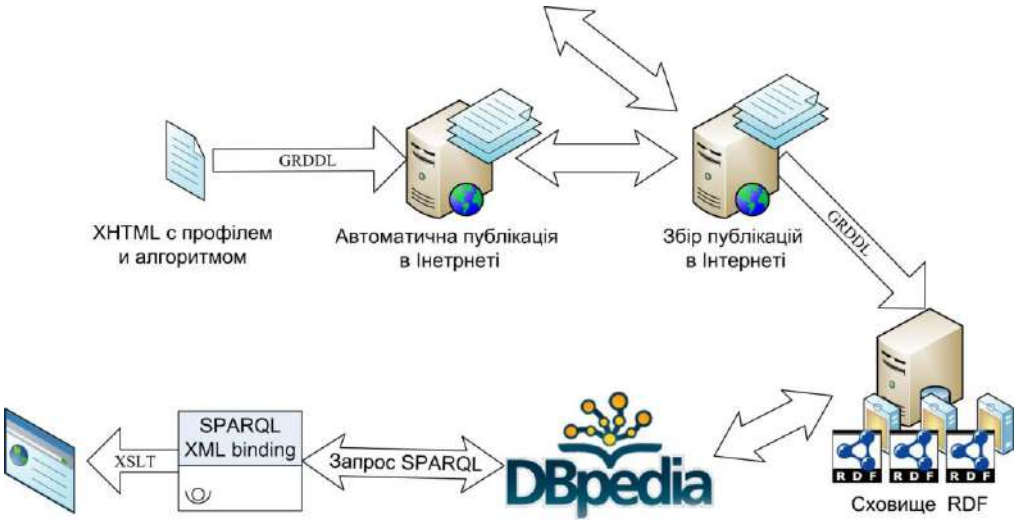


Рис. 2. Приклад RDFa та GRDDL

1. *Pavel Shvaiko.* "Iterative schema-based semantic matching," Informatica e Telecomunicazioni, Trento, Technical Report DIT-06-102, 2006.
2. *Yun Lin.* Semantic Annotation for Process Models: Facilitating Process Knowledge Management via Semantic Interoperability.: Department of Computer and Information Science Norwegian University of Science and Technology.
3. *Ioana Manolescu and Donald Kossmann Daniela Florescu.* "Answering XML queries over heterogeneous data sources," in 27th International Conference on Very Large Data Bases (VLDB 2001), P. 241–250.
4. *Maurizio Lenzerini.* "Data integration: a theoretical perspective," in 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2002), New York, 2002.
5. *Yuan Wang, Shawn R. Jeffery and David J. DeWitt. Leonidas Galanis.* "Locating data sources in large distributed systems.," in 29th International Conference on Very Large Data Bases (VLDB 2003), P. 874–885.
6. *Alon Y. Levy.* "Combining artificial intelligence and database for data integration," in In Artificial Intelligence Today: Recent Trends and Developments, P. 249–268.
7. *Christine Parent and Yann Dupont Stefano Spaccapietra.* "Model independent assertions for integration of heterogeneous schemas // LDB J. – 1992. – Vol. 1, N. 1. – P. 81–126.
8. *Tom Heath's.* Tom Heath's Displacement Activities. (2009) [Online]. http://tomheath.com/blog/2009/03/linked-data-web-of-data-semantic-web-wtf/
9. *Richard Cyganiak, Tom Heath Chris Bizer.* Welcome to WWW4, the research application server of the Lehrstuhl für Wirtschaftsinformatik. (2007, July) [Online]. http://www4.wiwiss.fu–berlin.de/bizer/pub/LinkedDataTutorial/
10. *Tim Bray, Dan Connolly, Paul Cotton, Roy Fielding, Mario Jeckle, Chris Lilley, Noah Mendelsohn, David Orchard.* Norman Walsh, and Stuart Williams Tim Berners–Lee. (2004, Nov.) World Wide Web Consortium (W3C). [Online]. http://www.w3.org/TR/webarch/
11. *World* Wide Web Consortium (W3C). (2007, Oct.) [Online]. http://www.w3.org/2001/tag/doc/httpRange–14/2007–05-31/HttpRange-14
12. *Roy T. Fielding.* [httpRange-14] Resolved. [Online]. http://lists.w3.org/Archives/Public/www-tag/2005Jun/0039.html
13. *Mark Birbeck, Shane McCarron, Steven Pemberton, Ben Adida.* The World Wide Web Consortium (W3C). (2008, Sep.) [Online]. http://www.w3.org/TR/rdfa-syntax/
14. *Linked* Data community. Linked Data - Connect Distributed Data across the Web. (2009) [Online]. http://linkeddata.org/
15. *ESW* Wiki. (2009) [Online]. http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies
16. *ESW* Wiki. (2009) [Online]. http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
17. *Mark Birbeck, Ben Adida.* RDFa Primer. (2009, Sep.) [Online]. http://www.w3.org/TR/xhtml-rdfa-primer/
18. *Fabien Gandon.* Institut National de Recherche en Informatique et en Automatique. (2009) [Online]. http://www-sop.inria.fr/acacia/personnel/Fabien.Gandon/tmp/grddl/rdfaprimer/PrimerRDFaSection.html
19. *Linking* Open Data community. ESW Wiki. (2010) [Online]. http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData