computing@computingonline.net
www.computingonline.net

ISSN 1727-6209
International Journal of Computing

ПАРАДИГМА SEMANTIC WEB В КОНТЕКСТІ ЕЛЕКТРОННОЇ БІБЛІОТЕКИ: СЕРВІСИ ТА ІНТЕГРАЦІЯ ІНФОРМАЦІЇ

О.В. Новицький

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

Резюме: *В роботі представлено ряд теоретичних ідей та прикладні технології які можуть бути втілені при створенні семантичної електронної бібліотеки (ЕБ). Зокрема значна увага приділена, яким чином технологія Semantic Web використовується в різних аспектах ЕБ. Виділено основні рівні в структурі ЕБ які можуть бути носіями семантичного описання. Описані переваги такого підходу. Також висвітлено питання які постають при інтеграції класичних електронних бібліотек та місце Semantic Web в цих процесах. Зроблено короткий огляд провідних світових проектів з створення ЕБ з використанням технології Semantic Web.*

Ключові слова: *Semantic Web, електронна бібліотека, WSMO, семантична анотація.*

ВСТУП

Електронна бібліотека є складною інформаційною структурою. Саме поняття електронна бібліотека на даний момент конкретно не визначено. В роботах [1], [2], [3] було проведено аналіз стосовно цього поняття. Ми зупинимося на такому підході, що електронна бібліотека це об'єднання через мережу електронних текстів, документів, зображень, звуків, наукових даних та програмного забезпечення яке є ядром сьогоднішнього Інтернету, а в майбутньому через організацію доступу до електронних бібліотек буде утворюватися база знань людства. Цей підхід породжує так звану колективну пам'ять. Поняття колективної пам'яті саме для цифрових бібліотек виникло відносно недавно. Проте цей термін набув широкого поширення, зокрема комітет IEEE Technical Committee on Digital Libraries, трактує це поняття як сукупність електронних бібліотек, електронних музеїв, електронних архівів. Зазвичай основна інформація яка передавалася це була текстова інформація, про те на разі, передається інформації інших типів відео, звук, фотографії та ін.. [4].

Такий підхід є інноваційним оскільки дає цілісний доступ до інформації будь-яким користувачам будь-де. Розвиток сучасних інформаційних технологій дає можливість реалізувати це.

Оскільки сучасний науковий пошук пов'язаний з великою кількістю даних, тому важливо щоб цю інформацію могли використовувати всі науковці крім цього для наукових даних необхідним є прослідкування виникнення цих даних. Тим самим можна гарантувати їх достовірність, водночас зберігати унікальні екземпляри в бібліотеках, музеях та архівах.

Сьогодні в багатьох музеях та бібліотеках зберігається величезна кількість безцінної інформації, проблема в тому, що до неї є поки тільки фізичний доступ. Створення колективної пам'яті позитивно відобразиться на науці, такий підхід повинен стати серйозним поштовхом.

Відомо, що кожна наукова група збирає та поповнює свій інформаційний фонд, причому якщо кожна така група працює в одному напрямі то відповідно і інформаційні фонди будуть одного змісту, але можливо різної структури, що призводить до неефективної роботи, та різного тлумачення наукових понять.

Обмін інформацією дасть можливість аналізувати дані спостережень одночасно багатьом науковим групам навіть якщо вони будуть знаходитися на дуже великих відстанях. Таким чином утворюється спільний робочий простір, що дозволяє всім взаємовигідно працювати над проблемо.

Колективна пам'ять утворює так звані портали знань та контенту, що представляє собою мережу з розподіленими ресурсами.

Розвиток колективної пам'яті одночасно

потребує розвитку інших напрямів.

•Зберігання. Система Колективної пам'яті повинна та здатна зберігати великі об'єми інформації різнорідних форматів.

•Інтерфейс користувача. Один з найважливіших компонентів колективної пам'яті, який повинен представляти велику кількість сервісів, для взаємодії між користувачем та інформацією яку він шукає.

•Класифікація та індексація. Дає змогу групувати об'єкти. Однак виявлено, що на це сильно впливає індивідуальне сприйняття та великий обсяг інформації яку необхідно індексувати.

•Інформаційний пошук. В цій області існує багато методів пошуку, включаючи пошук мета даних та контенту. Визначити корисність результату пошуку може тільки сам користувач. Для покращення ефективності використовують додаткові метадані, які описують документ. Дослідники також зосереджуються на автоматизації створення і обслуговування параметрів користувача для використання їх в процесі пошуку.

•Адміністрування та збереження. Традиційні бібліотеки зберігають копію книги, музеї зберігають фізичний експонат. Система колективної пам'яті дозволяє зберігати декілька версій документа. Окрім того цифрова бібліотека може розмежовувати права доступу до авторських екземплярів тим самим зберігаючи авторське право. І всі перегляди будуть автоматично фіксуватися. Механізм захисту повинен бути надійним для виключення несанкціонованого доступу. Зміни технологій організації структури середовища зберігання інформації, доступу до неї та старіння засобів збереження становить серйозну проблему яка повинна також вирішуватися.

Окрім розглянутих вище напрямків необхідно також збільшувати степінь деталізації. Створення нових схем мета даних, зосередження на інформаційному вмісту а не на інформаційних об'єктах. [5].

Отже оперування такими обсягами інформації породжує певні проблеми. З погляду на вище сказане ці проблеми можна розділити на дві великі групи, перша група проблем пов'язана з технічними труднощами організації збереження інформації та доступу до неї, друга група проблем пов'язана з логічною організацію колективної пам'яті та забезпечення доступу до неї з подальшим аналізом змісту. Ця робота стосується вирішення проблем 2-гої групи.

1. ПАРАДИГМА SEMANTIC WEB В КОНТЕКСТІ ЕЛЕКТРОННОЇ БІБЛІОТЕКИ. ОГЛЯД СВІТОВИХ ПРОЕКТІВ

Використання семантичних технологій в електронних бібліотеках було приділено увагу в багатьох Європейських проектах:

В проекті SWHi [6] онтологія розроблена на основі електронної бібліотеки з точки зору, коли наші основні джерела даних в репозиторії описані метаданими. Це метадані відображається і зберігається в онтології, яка базується на онтології схеми. Крім того, буквені значення в метаданих, наприклад, заголовок, піддається аналізу в наслідок якого видобуваються імена сутностей, події та термінологія. Для збагачення онтології, також видобувають нову зв'язану інформацію з обраних веб-документів.

Пошук в цій системі реалізований у двох формах, простий та складний. Система використовує мову запитів RDF таку як SeRQL. Процесор генерування запитів SeRQL стикатися, принаймні з двома проблеми. По-перше, він не знає, в якому класі чи властивості можуть бути знайдені слова. Щоб уникнути цю проблему, прикладне програмне забезпечення Semantic Web, таке як OpenAcademia [13] вимагає від користувачів вводити ключові слова у відповідне поле (автор, назва або рік) в її розширений пошуковий інтерфейс. По-друге, існують деякі обмеження в підстроках відповідності SeRQL при використанні символу загальності '*'. Цю проблему можна вирішити за допомогою інформаційно-пошукових програм, таких як Lucene яка забезпечує потужний алгоритм, точного і ефективного пошуку.

Окрім самого пошуку, важливим також є питання представлення результатів пошуку. Одним із напрямків є візуалізація пошуку. У Semantic Web, візуалізація стає все більш важливою. Існують випадки складних взаємини між ресурсами, які не можуть бути представлені за допомогою простого списку. Крім того, як правило, відображається тільки невелику кількість результатів пошуку (в діапазоні 10-20 результатів на сторінці). До документів які знаходяться в хвості результату пошуку, швидше за все, ніколи не будуть зверталися.

Загальна архітектура системи SWHi показано на Рис. 1.

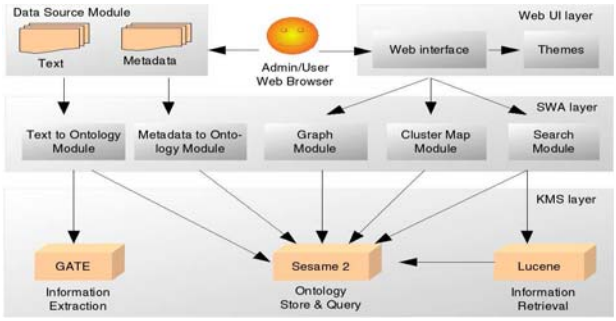

Рис. 1 – Архітектура SWHi

Для розвитку SWHi онтології, повторно використовуються наявні онтологічні ресурси для структурування і збереження історичної інформації, а саме: PROTON базова онтологія, таксономія предметної класифікації NewsBank/Readex, Дублінське Ядро та та словник FOAF Vocabulary. Це онтології зберігається з використанням Sesame2.

Проект eCulture є семантична пошукова система, яка дозволяє одночасно шукати в кількох колекціях установ культурної спадщини [7]. Це робиться шляхом перенесення цих колекцій в RDF шляхом зв'язування об'єктів колекцій як екземплярів класів через загальнодоступні словники, тим самим створюючи великий RDF граф.

Потім, в ході пошуку, цей граф трасується і деякі підграфи повертаються у вигляді результату.

Основним механізмом пошуку є використання Prolog.

Більшість Semantic Web додатків реалізовано за зразком реляційних баз даних, де прикладна логіка має доступу до бази даних на основі SQL [8]. Існує ряд еквівалентів SQL для Semantic Web, таких як SeRQL [9] і рекомендація W3C SPARQL [17]. Обидва дозволяють виразити граф, що складається з низки обов'язкових і необов'язкових ребер та вузлів, з розширеними умовами літеральних значень.

SeRQL відповідає вираженню графу на транзитивно закритому виразі використовуючи семантику RDFS. Стандарт SPARQL не визначає, чи виконується логічний наслідок, який виконаний механізмом судження СУБД.

Однак додатки eCulture, не використовують SeRQL або SPARQL. Замість цього, запити прикладної логіки виражаються як Prolog цілі на необроблених даних RDF та/або модулях судження RDFS/OWL.

Проект IPISAR (Image Preservation, Information Systems, Access and Research) досліджує розповсюдження, вивчення і раціональне використання культурної спадщини, та спроби представити вирішення загальних проблем в цих областях в рамках Semantic Web (SW) [11].

В рамках проекту розроблений додаток "Pescador", який буде зберігати каталогізовані дані в трійках які зберігатимуться в хранилищах (чиї функції будуть такі ж, що в реляційної бази даних в традиційних системах). Який як показала практика необхідно в подальшому вдосконалювати, зокрема забезпечення більш гнучких механізмів з подолання обмежень мови DSL яка була створена в рамках IPISAR.

Для досягнення цієї мети ми пропонуємо семантичну компоненту архітектуру (SCA), тобто, адаптація компонентів архітектури відповідно до принципів SW, в якому дані, структура та правила прикладної логіки, тісно пов'язані між собою.

SCA повинен координувати "компоненти", що підключаються, які б були б обгорнуті оболонкою яка б могла взаємодіяти з наступними типами: схемами; обмеженнями; правилами виводу; онтологіями; визначення шляхів; програмних кодом; специфікацією виводу; інформацією про конфігурацію Abox; посиланнями до зовнішніх джерел даних.

Алгоритм вилучення інформації з графів часто потребує визначення шляхів між ресурсами. Ці шляхи в значній мірі відрізняються по довжині і складності, тому SCA повинна включати засоби визначення моделі шляху. Попередній огляд існуючих механізмів визначення шляхів показує, що SPARQLeR розширення SPARQL може бути кращим кандидатом для адаптації для SCA і Пескадор. Автори [12] вважають, що підтримка семантичного шляху запиту повинна бути невід'ємною складовою RDF мови запитів. SPARQLeR (SPARQL extended with Regular paths), розширення мови запитів SPARQL, який додає підтримку для семантичного шляху запиту. Пропоноване розширення вписується в загальний синтаксис і семантику SPARQL і дозволяє легко формулювати запити за участю широкого кола регулярних шляхів в моделях RDF графів.

Проект EPOCH та AMA [13] представляють ЕБ як великий індекс покликаний служити в якості довідника для пошуку і повернення цифрової інформації яка зберігається на веб-сайті в різних форматах і архівах.

Перша проблема полягає в тому, що кожний довідник має свою пошукову систему і використовує свою граматику метаданих для опису та індексації даних, зокрема, що вона ніколи не буде працювати на інших системах. Жодна з цих систем метаданих може проаналізувати всю інформацію на веб-сайті,

якщо ми не будемо робити їх доступними через машину зрозумілій формі з використанням RDF [14].

Друга проблема стосується безпосередньо інформації: величезна різноманітних форматів, що використовуються для індексування даних, є великою перешкодою на шляху до інтеграції, і повинні бути серйозно проаналізовані. Навіть якщо ми обмежуємо наші зусилля виключно для культурної спадщини, архіви (наприклад, бази даних музеїв і колекцій, археологічні розкопки звіти, доповіді та інші неструктуровані дані), ми змушені визнати, що інформація, також є гетерогенною.

Щоб створити єдиний концептуальний шар, семантична інформація повинна бути взята з бази даних, HTML-сторінок, описових текстів, метаданих і повинна буди представлена в стандартному форматі, з метою отримання концептуального змісту інформації створивши концептуальний мапінг.

Як тільки концептуальний шар для даних і метаданих готовий, семантична інформація буде зберігатися в контейнері засновані на RDF і онтології.

Це реальним місцем інтеграції, де об'єднані основні відомості з різних електронних архів можуть бути переглянуті та в яких можливо здійснити пошук як в єдиній цілій електронній бібліотеці. RDF мова надійною і достатньо гнучкою, щоб забезпечити сумісність і забезпечити загальну основу не тільки для цифрових бібліотек, але і з інших систем та послуг.

Мапінг є одним із самих важливих кроків при інтеграції даних. Для спрощення та доступності процесу мапінгу в проекті AMA було розроблене програмне забезпечення AMA Mapping Tool гнучкий інструмент який сприяє мапінгу різних археологічних та музейних колекції моделей даних (з різною структурою, а також неструктуровані дані, тобто текстовий опис) на загальний стандарт ґрунтується на CIDOC CRM-онтології.

Аналізуючи ці проекти можна стверджувати, що існують ряд проблем при створенні електронних бібліотек та їх інтеграції з використанням семантичних технологій. Наприклад така проблематика як створення підходів до семантичної анотації електронних об'єктів. Особливо це стосується семантичної анотації контенту електронних бібліотек у випадку, якщо контент представлений у різних форматах та у різних галузях знань людства.

Одним із способів вирішення цієї проблеми може бути узагальнена формальна модель анотації.

Іншою проблемою яка постає при оперуванні великої кількості гетерогенної інформації це забезпечення відповідних сервісів. Оскільки сервіси є специфічними для різних форматів документів та повинні враховувати особливості вимог користувачів для обробки цієї інформації.

2. ФОРМАЛЬНА МОДЕЛЬ АНОТАЦІЇ

Для того щоб показати переваги семантичного підходу до електронної бібліотеки на відміну від класичної необхідно виділити ті структурні елементи які раніше не мали семантичної моделі, і до яких ми пропонуємо цю модель побудувати.

Основними двома компонентами електронної бібліотеки є її контент та набір програмного забезпечення для роботи з цим контентом. Для початку розглянемо контент ЕБ. Інформація в електронних бібліотеках описується в термінах електронні об'єкти (Digital objects – DO), які являють собою мультимедійний контент і метадані [15]. Оскільки обсяги DO значні, то для спрощення пошуку та класифікації використовують анотування DO.

Формальна модель, яка запропонована в [16], виділяє два підходи до розуміння анотацій: анотації як метадані або анотації як контент.

У першому випадку ми маємо справу з різноманітними схемами метаданих (Dublin Core, MARC і ін.) які використовуються для опису інформаційних ресурсів. Ці анотації насамперед направлені на користувача.

У другому випадку анотації як представлення контенту призначені для автоматизованої машинної обробки. Ці анотації надають семантику документа. Семантична анотація – анотація написана формальною мовою з добре визначеною семантикою і заснована на онтологіях. Фактично ці анотації є формальною моделлю DO, з можливістю машинної обробки.

Семантика контенту в свою чергу, може визначатися на основі зовнішніх зв'язаних онтологій, що дозволить будувати семантичну модель документу, де зв'язок визначається між окремими сегментами DO, та на основі семантики зв'язків між структурними компонентами DO, де зв'язок визначається між логічними закінченими структурними компонентами Рис. 2.

Модель, яка представлятиме цифровий об'єкт повинна відображувати фактичний зміст даного об'єкту. Серед безлічі розроблених моделей, ми використовуємо модель, яка запропонована в [15], [16] зі змінами та уточненнями, які враховують використання онтологій.

Розташування кожного цифрового об'єкту

також як і анотації ідентифікується унікальним ідентифікатором – посиланням (link). Окрім цього посилання сполучає цифровий об'єкт і анотацію, і може відображувати відношення між об'єктами. Отже, можна виділити два типи посилань:

посилання анотації (Annotate link) – відображує відношення в середині цифрового об'єкту, який може бути як документом, так і анотацією;

посилання відношення (Relate-to link) – визначає відношення зовнішнього цифрового об'єкту до об'єкту, що анотується.

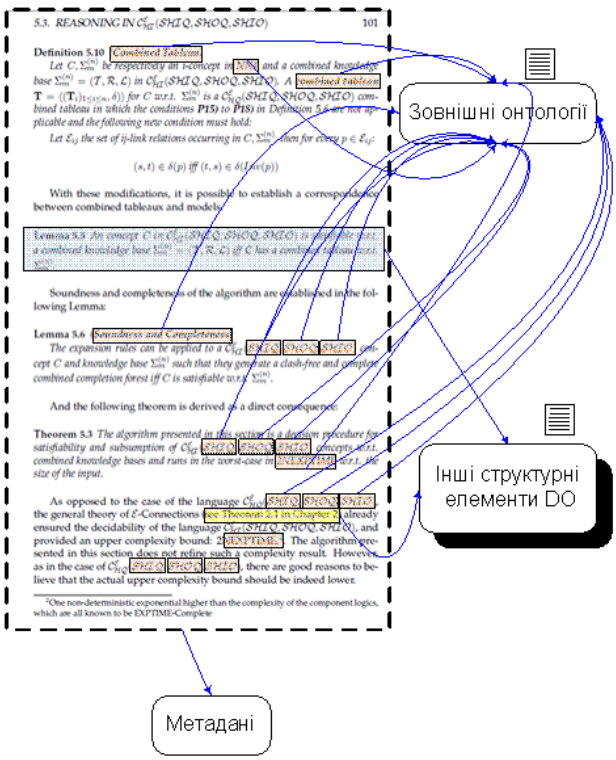

Рис. 2 – Анотування документа шляхом визначення відношень між термінами, лемами та зовнішніми отологіями а також визначення зв'язків між логічно закінченими елементами з іншими частинами DO. Окрім цього DO описується метаданими.

Нехай LT множина типів посилань, тоді LT містить наступні типи посилань $LT = \{AnnotateLink, Relate\text{-}toLink\}$.

Посилання надаються у вигляді ідентифікаторів. Загальноприйнятими ідентифікаторами можуть виступати URI, DOI, OPENURL, Persistent URL (PURL), PURL-based Object Identifier. $H(k)$ множина ідентифікаторів цифрових об'єктів у момент часу k.

У цій моделі цифровий об'єкт надається у вигляді потоку. Потік sm це кінцева послідовність:

$$DO(\Sigma) \to sm : I = \{1,2,...,n\},\ n \in \mathbb{N}$$

де e $= (j_1, j_2, ..., j_n)$ – алфавіт символів.

Якщо ми маємо потік sm :

$$DO(\Sigma) \to sm : I = \{1,2,...,n\},\ n \in \mathbb{N},$$

то в цьому потоці ми можемо виділити неперервний сегмент st_{sm} послідовності чисел a, b так що:

$$st_{sm} = [a,b],\ 1 \le a \le b \le n,\ n \in \mathbb{N}.$$

Безліч сегментів ми позначаємо ST , так що $\forall st_{sm_i} \in ST,\ i = 1,2,...,n.$

Якщо цифровий об'єкт DO має безліч унікальних ідентифікаторів H то функція hsm відображує унікальний ідентифікатор до сегментів які містяться в DO :

$$h \xrightarrow{hsm} st_{sm_i},\ n \in \mathbb{N}$$

Потік вимагає, щоб функція не мала властивостей сюрьїктивності і інєктивності. Кожен цифровий об'єкт може мати принаймні один потік.

SM визначає множину потоків, так що $\forall sm_i \in SM,\ i = 1,2,...,n$.

Анотацію можна розглядати як процес розширення онтології O_i. Розглянемо найпростіший випадок, коли розширення відбувається шляхом додавання нових екземплярів онтології O_i. Кожний клас онтології O_i будемо позначати через kl_i, а множину класів через KL.

Анотація $a \in A(k)$ це кортеж:

$$a = \begin{pmatrix} h_a \in H(k), \\ A_a \subseteq KL(k) \times LT \times ST(k) \times \\ \times SM(k-1) \times H(k-1) \end{pmatrix}$$

де h_a – унікальний власний ідентифікатор анотації a, тобто $h(h_a) = a$;

A_a множина n -арних відношень анотації a і визначається як добуток множин KL, LT, ST, SM та H.

У випадку анотування веб-документів, формальна модель зміниться. Нехай A множина всіх анотацій a, а D множина документів, відповідно, $DO = D \cup A$, причому підмножина множини DO позначатимемо do, тобто

$do \in DO$.

Анотацією веб-документа називатимемо мічений граф:

$$G := \left(\left(DO, E_{da} \subseteq A \times DO\right)\right),$$

де $DO = D \cup A$ вершини графа;

$$E_{da} = \left\{ \begin{array}{l} \left(a, do\right) \in A \times DO \mid \exists \alpha \in A_{\alpha}, \\ \alpha = \left(kl_i, sm_i, st_{sm_i}, hsm^{-1}, LT\right) \end{array} \right\}$$

– сторони графа.

3. СЕРВІСНА АРХІТЕКТУРА ЕБ

Як було сказано раніше при об'єднанні електронних бібліотек виникають проблеми з відмінностями в архітектурі ПЗ. Ми вважаємо, що використання сервіс-орієнтованої архітектури є ключовим елементом досягнення інтероперабельності. У такому середовищі складне програмне забезпечення може бути спроектоване на основі сервісів, що є у розпорядженні і які доступні не лише локально, але і через Інтернет. У такому контексті для побудови електронних бібліотек ми пропонуємо використовувати архітектуру сервіс-орієнтованої електронної бібліотеки (Service-Oriented Digital Library architecture – SODL) [17], це забезпечить зручний спосіб для досягнення побудови колективної пам'яті.

Для забезпечення вільного доступу служб до бібліотеки вона має бути реалізована на відкритій архітектурі. Тобто взаємодія служб має бути описана за допомогою стандартизованих правил і середовища, що дасть можливість вільно вводити новий сервісний елемент, не перебудовуючи систему наново.

Для того, щоб забезпечити динамічне налаштування електронної бібліотеки, необхідно забезпечити механізми, які виконуватимуть цю функцію. Таким механізмом є введення семантики в середовище функціонування веб-сервісів. Цей підхід отримав назву Semantic Web, і стосується не тільки семантичного опису сервісів а також і сематичного опису інформаційних ресурсів Інтернету (веб-сторінок, документів тощо.)

Під сервісом ми розуміємо деякою послугу мета якої задовольнити запити користувача. Сервіси реалізуються за допомогою веб-сервісів. Веб-сервіс це програмне застосування. Основна відмінність між сервісом і веб-сервісом полягає в тому, що той самий сервіс одночасно можна реалізувати різними веб-сервісами. У свою чергу веб-сервиси можуть між собою об'єднуватися, створюючи новий веб-сервіс, цей процес називають композиція веб-сервісів.

Сервіси електронної бібліотеки можна класифікувати за різними аспектами. Зокрема результатом виконання деякого сервісу на вимогу запитуючої сторони буде результат у виді структурованої інформації про електронні об'єкти які містяться в ЕБ. Тому якщо розглядати ЕБ як замкнене середовище, очевидно, що інформаційне наповнення ЕБ не зазнало змін, сервіси з такою властивістю будемо називати сенсорні сервіси.

Поняття сенсорних сервісів для ЕБ ми ввели тому, що ці сервіси аналогічно до сенсорів фільтрують інформаційне середовище, фактично не впливаючи на зміст цього середовища.

В рамках підходу Semantic Sensor Web [18] ми додатково анотуємо результати виконання сервісів. Тому на відміну від класичної побудови семантичного середовища для веб-сервісів, де результат не анотується, ми передбачаємо анотування метаданими результатів виконання веб-сервісів. Характер метаданих повинен передбачати історію, спосіб, локалізацію та час отримання результатів.

Для побудови електронної бібліотеки на основі сервіс-орієнтованого підходу необхідно виділити сервіси, на основі яких виконуватиметься композиція [19]. Це обумовлено тим, що завдання композиції простіше вирішувати і, крім того, вимоги до електронної бібліотеки є такими, що в першу чергу необхідно вирішувати композиційні завдання. Базові сервіси виділено в роботі [20].

Нехай ми маємо запит r, у якого є початкові параметри r_{in} і бажані параметри виходу r_{out}, необхідно знайти веб-сервіс w такий, щоб виконував r причому виконувалися умови:

$r_{in} \supseteq w_{in}$

$r_{out} \supseteq w_{out}$

Проблема пошуку веб-сервісу, який може самостійно виконати запит r носить назву дослідження веб-сервісів (Web Service Discovery – WSD). Коли за допомогою одного веб-сервісу неможливо досягнути виконання запиту r, потрібно утворити композицію кількох веб-сервісів $\left\{w^1, w^2, w^3, ..., w^n\right\}$ послідовним чи паралельним способом, таким чином, щоб для всіх $w^i \in \left(w^1, w^2, w^3, ..., w^n\right)$, причому w^i_{in} може бути заснований на w^i_{out}, та виконувалось відношення $\left(r_{in} \cup w^1_{out} \cup ... \cup w^n_{out}\right) \supseteq r_{out}$. Ця проблема носить назву композиції веб-сервісів (Web Service Composition – WSC).

Більш складною проблемою є проблема

композиції веб-сервісів.

В процесі функціонування сервіс-орієнтованої електронної бібліотеки середовище в якому функціонують веб-сервіси постійно змінюється. Зміна середовища породжується двома основними факторами:

по-перше проблеми, які породжують розподілені системи часові затримки і ненадійність транспортного протоколу, недостача пам'яті спільного використання між частинами розподіленої системи, проблеми відмови доступу та паралельних запитів, а також проблеми пов'язані з програмною несумісністю в наслідок оновлення частини розподіленої системи;

по-друге людський фактор, в процесі обслуговування користувача, можуть змінитися вимоги користувача, а отже це певним чином впливатиме на результат. Пересічні користувачі цільової аудиторії не мають чіткого уявлення про архітектуру бібліотеки, а отже не можуть на перед чітко визначитися з цілями які повинні задовольняти сервіси.

Ми будемо розглядати композицію на основі ціле-орієнтованої парадигми, тобто виходячи початкових умов та наявної множини сервісів здійснити композицію. Причому, оскільки веб-сервіси розміщені в семантичному середовищі, то вибирати такі плани композиції, які можуть бути корисними для кінцевого користувача. Тобто на відміну від класичної постановки задачі, від специфікації цілі до пошуку сервісів, які зможуть цю ціль досягти, виходити з того припущення, що наявна множина сервісів може досягнути деякі наперед невідомі цілі, які можуть бути обрані користувачем.

В такій складній системі цілі, які постають перед композицією можуть змінюватися. Ми також виходимо з того, що знання про оточуючий світ є не повними а отже і цілі є неповними. Для таких конфліктних цілей Horst Rittel і Melvin Webber ввели поняття вікід (wicked) задачі. Тому класичні методи планування виявилися неефективними. В середовищі де цілі не є чітко визначеними і є динамічний процес планування приймає інше смислове значення, а саме, план виконання дій не є однозначним алгоритмом досягнення цілі, функція планів можна сформулювати наступним чином [21]: перевірка ресурсів; починають координаційні процеси або допомагають спростити координаційні процеси на початку; встановлюють відповідальність та індетифікацію; трековий прогрес і що більш важливо генерація намірів; впізнають та управляють ризиками; підтримують імпровізацію, і що саме важливо формалізують представлення людини (користувача) про проблему яку план намагається вирішити.

Така комбінація функціональних властивостей призводить до того, що плани необхідно додатково досліджувати і проводити репланування. Тобто змінювати плани під час виконання цих планів.

4. ПРАКТИЧНА РЕАЛІЗАЦІЯ

Реалізація викладених положень є надзвичайно складною, оскільки постає ряд задач прикладного характеру. На даний момент нема довершеного фреймворку для моделювання автоматичної композиції веб-сервісів. Однак на наш погляд найбільш перспективним є WSMT [22] та WSMX [23], які базується на WSMO.

WSMO[1] (WSMO – Web Service Modeling Ontology) підхід в якому відображено всі аспекти які пов'язані з семантичними сервісами, які доступні через веб інтерфейс. Кінцевою ціллю цього підходу являється представлення інформації для автоматичного машинного вирішення задач дослідження, вибору, композиції, співставлення, виконання та моніторингу.

WSMO оперує 4 головними поняттями:

Онтології представляють термінологію яку використовують інші компоненти WSMO.

Веб-сервіси представляють обчислювальні об'єкти, таким чином, що в деякій предметній області ці об'єкти являють собою сервіси.

Цілі представляють побажання користувача, відповідно для задоволення яких можна відшукати веб-сервіс.

Посередник – описує ті елементи, які відповідають за вирішення питань функціональної сумісності між іншими елементам WSMO.

Синтаксис WSMO визначається мовою WSML – Web Service Modeling Language.

Сервіс підключається до середовища **WSMT** за допомогою WSDL інтерфейсу. Тому нами було розроблено два сервіси для існуючого вільного ПЗ управління електронними бібліотеками Eprints[2]. Сервіс getEprint який повертає метадані для запису з вказаним на вході id. Другий сервіс є набагато потужнішим. Цей сервіс має назву searchEprint, він дозволяє виконувати повнотекстовий пошук та пошук в метаданих електронної бібліотеки. На сонові останнього сервісу можна моделювати ряд інших сервісів, таких наприклад як сенсорні сервіси класифікації. Кожен сервіс має інтерфейс який

[1] http://www.wsmo.org/

[2] http://www.eprints.org/

відповідає специфікації WSDL 1.1. Дані розробки викладені на офіційному сайті Eprints а також обидва сервіси впроваджено на сайті http://eprints.zu.edu.ua/.

5. ВИСНОВОК

В даній роботі коротко викладено теоретично-практичні основи створення ЕБ з використанням семантичних технологій. Зроблено огляд провідних Європейських проектів з інтеграції технології Semantic Web в ЕБ. В статті також вказано на принципи застосування Semantic Web до сервісів ЕБ.

Проте проблематика створення таких ЕБ потребує подальшого вивчення, зокрема в роботі не формалізовано поняття онтології. Водночас також не вказано і типи зв'язків які можуть бути між екземплярами та онтологією. Проте вже зараз можна стверджувати, що необхідно буде вирішувати проблему вирівнювання між онтологіями, яка виникне в наслідок інтеграції двох або більше семантичних ЕБ.

6. СПИСОК ЛІТЕРАТУРИ

[1] Licklider, J. C. R. Libraries of the Future. Cambridge : MIT Press, 1965.

[2] *Going digital: a look at assumptions underlying digital libraries.* Marshall, D. M. Levy and C. C. 8, 1995 p., Communications of the ACM, T. 38, Pp. 77-84.

[3] *What are digital libraries? competing visions.* Borgman, C. L. 3, 1999 p., Information Processing and Management, T. 35, Pp. 227-243.

[4] IEEE Technical Committee on Digital Libraries. http://www.ieee-tcdl.org.

[5] Neuhold, Erich J. Position Statement Past Chairman. http://www.ieee-tcdl.org/posstatement.html.

[6] Ismail Fahmi, Junte Zhang, Henk Ellermann, Gosse Bouma. SWHi System Description: A Case Study in Information Retrieval, Inference, and Visualization in the Semantic Web. *The Semantic Web: Research and Applications, 4th European Semantic Web Conference.* Innsbruck, Austria : Springer, 2007, Pp. 769-778.

[7] *Porting Cultural Repositories to the Semantic Web.* Omelayenko, B. Tenerife, Spain : 2008. Proceedings of the First Workshop on Semantic Interoperability in the European Digital Library (SIEDL-2008). Pp. 14-25.

[8] Wielemaker, J., Hildebrand, M., Ossenbruggen, J.R. Van. Using Prolog as the fundament for applications on the semantic web (2008). *Proceedings of the 2nd Workshop on Applicatiions of Logic Programming and to the web, Semantic Web and Semantic Web Services.* Porto, Portugal : 2007.

[9] *SeRQL: A Second Generation RDF Query Language.* Broekstra J, Kampman A. Proc SWAD-Europe Workshop on Semantic Web Storage and Retrieval 2003.

[10] Eric Prud'hommeaux Andy Seaborne. SPARQL Query Language for RDF. *W3C.* 2008 p. http://www.w3.org/TR/rdf-sparql-query.

[11] *Solutions for a Semantic Web-Based Digital Library Application.* Martínez, Andrew Russell Green and José Antonio Villarreal. 2008. First Workshop on Semantic Interoperability in the European Digital Library.

[12] *SPARQLeR: Extended Sparql for Semantic Association Discovery.* Krys Kochut, Maciej Janik. 2007. 4th European Semantic Web Conference (ESWC2007). http://www.eswc2007.org/pdf/eswc07-kochut.pdf.

[13] *Semantic Maps and Digital Islands: Semantic Web technologies for the future of Cultural Heritage Digital Libraries.* A. Felicetti, H. Mara. Tenerife, Spain : 2008. SIEDL 2008: Semantic Interoperability in the European Digital Library. Pp. 51-62.

[14] RDF Core Working Group. Resource Description Framework (RDF). *Resource Description Framework.* W3C. http://www.w3.org/RDF/.

[15] *Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries.* Marcos André Gonçalves, Edward A. Fox, Layne T. Watson, Neill A. Kipp. 2, ACM New York, NY, USA, 2004 p., ACM Transactions on Information Systems (TOIS), T. 22. ISSN:1046-8188.

[16] *A formal model of annotations of digital content.* Maristella Agosti, Nicola Ferro. 1, 2007 : ACM New York, NY, USA, T. 26.

[17] *A service-oriented architecture for digital libraries.* Yves Petinot, C. Lee Giles, Vivek Bhatnagar, Pradeep B. Teregowda, Hui Han, Isaac Councill. ACM New York, NY, USA, 2004. Proceedings of the 2nd international conference on Service oriented computing. ISBN:1-58113-871-7.

[18] *Web, Semantic Sensor.* Sheth, A., Henson, C. та Sahoo, S.S. 4, 2008 p., Internet Computing, IEEE, T. 12. 10.1109/MIC.2008.87.

[19] *A service-oriented architecture for digital libraries.* Yves Petinot, C. Lee Giles, Vivek Bhatnagar, Pradeep B. Teregowda, Hui Han, Isaac Councill. ACM New York, NY, USA, 2004. Proceedings of the 2nd international

conference on Service oriented computing.
[20] Інноваційні підходи до створення колективної пам'яті на основі парадигми Semantic Web. О.В., Новицький. Львів : ПП Вежа і Ко, 2008. Computer Science and Information Technologies 2008.
[21] Flexecution as a Paradigm for Replanning, Part 1. Klein, Gary. 5, Los Alamitos, CA, USA : 2007 p., IEEE Intelligent Systems, T. 22, Pp. 79-83. 1541-1672.
[22] Web Service Modeling Toolkit (WSMT). http://sourceforge.net/projects/wsmt/.
[23] WSMX (Web Service Modelling eXecution environment). http://www.wsmx.org/.
[24] OpenAcademia. www.openacademia.org.

Олександр Новицький, *молодший науковий співробітник Інституту програмних систем НАН України.*

Наукові інтереси: технологія Semantic Web та її прикладне застосування в електронних бібліотеках.

computing@computingonline.net
www.computingonline.net

ISSN 1727-6209
International Journal of Computing

PARADIGM SEMANTIC WEB IN THE CONTEXT OF DIGITAL LIBRARY: SERVICES AND INFORMATION INTEGRATION

O. Novytskyi

Institute of Software System NAS of Ukraine 40 Academician Glushkov Ave., 03187, Kyiv, Ukraine
alex@zu.edu.ua

Abstract: *This paper presents a several of theoretical ideas and applied technology that can be embodied in the creation of Semantic Digital Library (SDL). In particular, considerable attention paid to how the Semantic Web technology is used in various aspects of DL. A basic level in the DL that may be carriers of semantic description. Described the advantages of this approach. It also highlights issues that arise during the integration of classical electronic libraries and Semantic Web a place in these processes. Made a brief overview of the world's leading projects to create DL using Semantic Web.*

Keywords: *Semantic Web, digital library, WSMO, semantic annotation.*

INTRODUCTION

Electronic Library is understood from position the collective memory and a complex structure. Committee of IEEE Technical Committee on Digital Libraries, interprets this concept as a set of Digital libraries, digital museums, digital archives [1].

Collective memory forms the so-called knowledge portals and content, with a network of distributed resources.

The development of a collective memory requires the development of other areas such as: storage, user interface, classification, information search, management and conservation.

In addition to above discussed areas should also increase the degree of detail descriptive information. [2].

1. SEMANTIC WEB ISSUES IN THE CONTEXT OF DL

Using semantic technologies in digital libraries has been given attention in many projects such as: SWHi [3], eCulture [4], IPISAR [5], EPOCH та AMA [6].

The first problem is that each directory has its own search engine and uses a grammar to describe the metadata and data, including that it will never work on other systems. Out of this situation will be the presentation of information in machine understandable form, using RDF [7].

The second problem concerns directly to information: a huge variety of formats that are used for indexing data, is a major obstacle to integration.

To create a single conceptual layer, semantic information must be taken from the database, HTML-page, descriptive text, and metadata to be presented in a standard format in order to obtain the conceptual content of information created conceptual mapping.

One way to address this problem can be generalized formal model of annotations.

Another problem that arises when handling large amounts of heterogeneous information is to ensure appropriate services.

2. FORMAL MODEL OF ANNOTATIONS

The main two components of the electronic library is its content, and set the software to work with this content. To start, consider the content of EB. Information in libraries is described in terms of electronic facilities (Digital objects DO), which are multimedia content and metadata [8]. Formal model that suggested in [9], identifies two approaches to understanding the annotation: annotations as metadata or annotation as content.

Among the many models, we use a model that suggested in [8], [9] with some changes and refinements, which include the use of ontology. Let LT the set of types of links.

$H(k)$ set of identifiers of digital objects in time k.

Set of segments ST, we indicate that

$\forall st_{sm_i} \in ST,\ i = 1,2,...,n.$

SM defines the set of streams, so that $\forall sm_i \in SM,\ i = 1,2,...,n.$

Annotation can be viewed as the process enlargement of ontology O_i.

Each class of ontology O_i shall designate kl_i, a set of classes KL.

Annotation $a \in A(k)$ is tuple:

$$a = \begin{pmatrix} h_a \in H(k), \\ A_a \subseteq KL(k) \times LT \times ST(k) \times \\ \times SM(k-1) \times H(k-1) \end{pmatrix}$$

where h_a – own unique identifier annotations a, that is $h(h_a) = a$;

A_a set n-arity relations annotation a and is defined as the product sets KL, LT, ST, SM and H.

In the case of annotation of web documents, the formal model of change. Let A set all annotation a, and D set DO, accordingly, $DO = D \cup A$, a subset of the set DO to mark do, that is $do \in DO.$

Annotation Web-document called tagging graph:

$$G := \left(\left(DO, E_{da} \subseteq A \times DO\right)\right),$$

where $DO = D \cup A$ vertex graph;

$$E_{da} = \begin{Bmatrix} (a, do) \in A \times DO \mid \exists \alpha \in A_a, \\ \alpha = \left(kl_i, sm_i, st_{sm_i}, hsm^{-1}, LT\right) \end{Bmatrix}$$

– side of the graph.

3. SERVICE-ORIENTED DIGITAL LIBRARY ARCHITECTURE

We believe that the use of service-oriented architecture is a key element in achieving interoperabelnosti. To build digital libraries, we propose to use the architecture of service-oriented e-Library (Service-Oriented Digital Library architecture – SODL) [10], it provides a convenient way to achieve the construction of collective memory

To ensure a dynamic setting digital library must provide mechanisms that perform this function. This mechanism is the introduction of semantics in the operation environment of web services.

Digital library services can be classified according to various aspects. Services are in the process of change is not content DL, is called sensory services.

As part of the approach Semantic Sensor Web [11] we additionally anotuyemo results of services. Therefore, unlike classical building environment for semantic web services, which result not anotuyetsya, we anticipate the results of the metadata annotation of web services.

Basic services allocated in the [12].

We will consider the composition based on goal-oriented paradigm is based initial conditions and the existing set of services to make the composition. And, because Web services are located in the semantic environment, then choose the plan of composition that may be useful for the end user. That is, unlike the classic statement of the problem of specification aims to find services that can achieve this goal, go with the assumption that the available set of services can reach some pre unknown targets that may be selected by user.

This combination of functional properties has meant that the plans need further study and conduct replanning. That is changing plans during the implementation of these plans.

4. CONCLUSIONS

However, in our opinion the most promising compositions for framework Web-services is WSMT [13] and WSMX [14], which is based on WSMO.

We developed two services for the existing free software management digital libraries Eprints1: service getEprint and searchEprint. Each services the interface that meets the specification WSDL 1.1. What allows these services to connect WSMT.

This paper briefly outlined the theoretical and practical bases of DL using semantic technologies. An overview of European projects to integrate technology in the Semantic Web DL. The article also stated on the application of principles of Semantic Web services to the DL.

However, creating such problems DL needs further study, particularly in the no formal notion of ontology. At the same time also has both types of connections that can be between instances and ontologies. But now we can say that it will be necessary to solve the problem of alignment between ontologies, which arise as a consequence of the integration of two or more semantic DL.

5. REFERENCES

[1] IEEE Technical Committee on Digital Libraries. [Online] http://www.ieee-

[1] http://www.eprints.org/, http://files.eprints.org/401/

tcdl.org.
[2] Neuhold, Erich J. Position Statement Past Chairman. [Online] http://www.ieee-tcdl.org/posstatement.html.
[3] Ismail Fahmi, Junte Zhang, Henk Ellermann, Gosse Bouma. SWHi System Description: A Case Study in Information Retrieval, Inference, and Visualization in the Semantic Web. *The Semantic Web: Research and Applications, 4th European Semantic Web Conference.* Innsbruck, Austria : Springer, 2007, p. 769-778.
[4] *Porting Cultural Repositories to the Semantic Web.* Omelayenko, B. Tenerife, Spain : s.n., 2008. Proceedings of the First Workshop on Semantic Interoperability in the European Digital Library (SIEDL-2008). p. 14-25.
[5] *Solutions for a Semantic Web-Based Digital Library Application.* Martínez, Andrew Russell Green and José Antonio Villarreal. 2008. First Workshop on Semantic Interoperability in the European Digital Library.
[6] *Semantic Maps and Digital Islands: Semantic Web technologies for the future of Cultural Heritage Digital Libraries.* A. Felicetti, H. Mara. Tenerife, Spain : s.n., 2008. SIEDL 2008: Semantic Interoperability in the European Digital Library. p. 51-62.
[7] RDF Core Working Group. Resource Description Framework (RDF). *Resource Description Framework.* [Online] W3C. http://www.w3.org/RDF/.
[8] *Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries.* Marcos André Gonçalves, Edward A. Fox, Layne T. Watson, Neill A. Kipp. 2, s.l. : ACM New York, NY, USA, 2004, ACM Transactions on Information Systems (TOIS), Vol. 22. ISSN:1046-8188.
[9] *A formal model of annotations of digital content.* Maristella Agosti, Nicola Ferro. 1, 2007 : ACM New York, NY, USA, Vol. 26.
[10] *A service-oriented architecture for digital libraries.* Yves Petinot, C. Lee Giles, Vivek Bhatnagar, Pradeep B. Teregowda, Hui Han, Isaac Councill. s.l. : ACM New York, NY, USA, 2004. Proceedings of the 2nd international conference on Service oriented computing. ISBN:1-58113-871-7.
[11] *Web, Semantic Sensor.* Sheth, A., Henson, C. e Sahoo, S.S. 4, 2008, Internet Computing, IEEE, Vol. 12. 10.1109/MIC.2008.87.
[12] *Innovation Approaches to design of collective memory on the paradigm base of Semantic Web.* O.V. Novitsky. Lviv : PP Vezha and Co, 2008. Computer Science and Information Technologies 2008.
[13] 13. Web Service Modeling Toolkit (WSMT). [Online] http://sourceforge.net/projects/wsmt/.
[14] WSMX (Web Service Modelling eXecution environment). [Online] http://www.wsmx.org/.