

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ІВАНА ПУЛЮЯ  
ФАКУЛЬТЕТ КОМП'ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ І ПРОГРАМНОЇ  
ІНЖЕНЕРІЇ

**КОРЧУК ЛЕОНІД ВІКТОРОВИЧ**

УДК 004.4'2

**АНАЛІЗ ТЕХНОЛОГІЙ ОПИСУ РЕСУРСІВ З ВИКОРИСТАННЯМ RDF  
В СЕМАНТИЧНІЙ ПАВУТИНІ**

8.05010201 «Комп'ютерні системи та мережі»

**Автореферат**

дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль  
2017

Роботу виконано на кафедрі комп'ютерних систем та мереж Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України

**Керівник роботи:** доктор технічних наук, професор  
**Лупенко Сергій Анатолійович,**  
Тернопільський національний технічний університет імені Івана Пулюя,

**Рецензент:** кандидат технічних наук, доцент  
**Медвідь Володимир Романович,**  
Тернопільський національний технічний університет імені Івана Пулюя

Захист відбудеться 21 лютого 2017 р. о 9<sup>00</sup> годині на засіданні екзаменаційної комісії №35 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд.603.

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми роботи.** Сьогоднішні веб-ресурси розробляються здебільшого для використання людьми. Незважаючи на поступову появу в мережі даних, призначених для машинного сприйняття, ці дані в основному поширюються окремим файлом у певному форматі. Як наслідок, веб-браузери можуть забезпечити користувачів лише мінімальною підтримкою в аналізі та обробці мережових даних, адже браузері тільки представляють інформацію.

Завдяки семантичній павутині комп'ютерні програми можуть використовувати зростаючі обсяги структурованих даних, які розподілені і децентралізовано розсіяні по мережі в даний час. RDF представляє собою абстрактну модель, що забезпечує спосіб розбиття знань на дискретні частини.

Головна відмінність Semantic Web від Web полягає в тому, що кожна сторінка семантичної мережі містить інформацію на двох мовах: на звичайній, зрозумілій людині, яка відображується браузером, і на спеціальній, інформація на якій прихована від людських очей, але зрозуміла інтелектуальним програмам-агентам, роботам. Ця спеціальна мова описує представлений на сторінці матеріал за допомогою тегів і атрибутів, що зрозумілі машинам [47].

В основі семантичної мережі лежать три принципи: агрегація, безпека і логіка. Агрегація означає спільне використання даних.. В основу безпеки, що забезпечує довіру до семантичної мережі, покладені цифрові підписи, які можуть використовуватися агентами і комп'ютерами для перевірки того, що інформація отримана з достовірного джерела, наприклад від якогось публічного сервісу або персонального агента іншого довіреної користувача. Логіка – це набір правил опису інформаційної структури даних, протоколи і мова опису сторінок. Саме логіка дає семантичній мережі правила виводу для проведення міркувань і методики вибору тактик виконання операцій з даними, щоб отримати відповіді на питання. Ключовим аспектом технології семантичної павутини також є багаторівнева структура, частиною якої і є технологія RDF. Технологія RDFa, що являється її різновидом, дозволяє супроводжувати дані машинними підказками за допомогою набору атрибутів.

Робота присвячена дослідженню використання технології опису ресурсів RDFa в семантичній павутині та підвищенню рівня розуміння пошуковими системами змісту Інтернет сторінки.

**Мета роботи:** створення рішення, що дасть змогу виразити прості і складніші набори даних за допомогою RDFa на веб сторінці.

**Об'єкт, методи та джерела дослідження.** Стандарти та процеси аналізу інформації веб-сайтів пошуковими системами.

**Наукова новизна отриманих результатів:**

Для відображення попередньої видачі були використані розробки Google, яка називається Rich Snippets Testing Tool та веб-сервіс Structured Data linter.

**Практичне значення отриманих результатів.**

Дані експерименту показали, що пошук по властивостям сутностей RDFa, які були відображені в контенті є ефективнішим, ніж пошук по звичайним текстовим даним. Також вибрана технологія підвищує інформативність видачі при

пошуковому запиту. З використанням RDFa скорочується кількість слів, необхідних для пошуку відповідного сайту.

**Апробація.** Окремі результати роботи доповідались на V Міжнародній науково-технічній конференції молодих учених та студентів «Актуальні задачі сучасних технологій», Тернопіль, ТНТУ, 17-18 листопада 2016 року.

**Структура роботи.** Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 7 розділів, висновків, переліку посилань та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 118 арк. формату А4, графічна частина – 8 аркушів формату А1

## ОСНОВНИЙ ЗМІСТ РОБОТИ

**У вступі** наведено актуальність дослідження використання технології опису ресурсів RDFa в семантичній павутині та підвищенню рівня розуміння пошуковими системами змісту Інтернет сторінки.

**В першому розділі** проведено аналіз проблеми ефективного пошуку засобами технологій пошукових систем з використанням семантичних даних обґрунтовується вибір формату опису метаданих для вирішення завдань. Проаналізовано напрямки застосування розширеної для вирішення задачі покращення релевантності пошуку та видачі пошукової інформації.

Недоліком використання мікроформатів є відсутність єдиного центру розробки мікроформатів до того ж вони мають схожі властивості з RDFa.

**В другому розділі** розглянуті показники ефективності використання технологій опису метаданих ресурсів в семантичній павутині. Дослідження існуючих технологій реалізації семантичного пошуку на різних пошукових системах. Розглянуті підходи, моделі та алгоритми їх розв'язання. Обґрунтовано обраний метод, виконана практична частина роботи та проведено опис наявних елементів.

Для досліджень використані інші пошукові системи, що підтримують RDF (Yahoo!, SearchMonkey, Bing та інші). Це дає змогу проаналізувати актуальність формату для існуючих пошукових систем, інформативність видачі при пошуковому запиту та можливість розкрити семантичний зміст тексту.

Також використані інструменти та сервіси відображення, валідації, сервісу та відкладки сайтів з RDFa наповненням (Validator.ru, Structured Data Linter, Online parser/viewer and RSS generator ), завдяки ним оцінено витрати на підтримку і розширення закладеного інформаційного набору, необхідність використання особливої мови запитів, трудомісткість опису, релевантність інформації та можливість розкрити семантичний зміст тексту.

### **В третьому розділі**

За результатами проведених експериментів та їх статистичного аналізу, а також графічного представлення результатів роботи, можна зробити наступні висновки щодо ефективності технології опису ресурсів RDFa:

а) при використанні в поєднанні з пошуковою системою Google в певній мірі розкривається семантичний зміст тексту, так як при пошуку з включенням

властивостей RDFa пошуковий робот виділяє сутності поміж іншого контенту;

б) витрати на підтримку і розширення закладеного інформаційного набору скоріше за все не значні в порівнянні з користю яку може принести розглянутий формат, більш точну інформацію можна отримати лише вивчивши ресурсні затрати на розвиток формату;

в) при описанні сутностей та властивостей не відбувається дуплікація даних (немає необхідності окремо занотовувати контент в мові придатній для перегляду людиною та пошуковим роботом);

г) згідно з отриманих статистичних даних розміщення RDFa на сторінці певною мірою впливає на просування сайту, т.к. кількість слів для пошуку в скорочується приблизно в півтора рази, також позитивно на даний аспект впливає зміна графічної видачі, що виділяє сайт поміж інших, тим самим приваблюючи більше відвідувачів;

д) дещо відштовхує від обраної технології трудомісткість опису, особливо на початковому етапі, коли не знайома структура розмітки та словник сутностей і властивостей;

е) релевантність інформації при використанні технології підвищується, так як пошуковий робот визначає заанотовані сутності і властивості на сторінках ресурсів, що здебільшого є ключовими словами для пошуку, з статистичних даних також видно, що при використанні RDFa необхідна менша кількість слів для пошуку для видачі необхідної інформації;

ж) значно підвищується інформативність видачі при пошуковому запиті, а саме: до неї може додаватися графічне відображення та додаткова інформація про сутність;

з) при використанні технології в поєднанні з пошуковою системою

Google немає необхідності вдаватися до особливої мови запитів, що позитивно впливає на простоту пошуку, але з іншого боку спеціальна мова запитів могла б підвищити ефективність пошуку, а відповідно і використання самої технології;

и) технологія є актуальною для пошукової системи Google, щодо використання з іншими необхідно проводити додаткові дослідження.

**В четвертому розділі** розглянуто нестандартні способи просування сайтів в пошукових системах.

**В п'ятому розділі** розраховано основні техніко-економічні показники науково-дослідної роботи.

**В шостому розділі** розглянуто питання планування робіт з охорони праці. Описано небезпечні і шкідливі фактори середовища проживання. Джерела забруднення навколишнього середовища. Наслідки негативних дій небезпечних і шкідливих факторів середовища, їх вплив на людину. Мінімізація радіоактивного ураження та втрат серед промислово-виробничого персоналу у випадку застосування наземних ядерних вибухів під час надзвичайних ситуацій воєнного часу.

**В сьомому розділі** розглянуто роль матеріало- та ресурсозбереження у вирішенні екологічних проблем і комплексне оцінювання екологічності виробництва.

**У загальних висновках щодо дипломної роботи** описано прийняті в роботі технічні рішення.

В графічній частині приведено загальну структуру Semantic WEB, семантичні мікроформати, показники ефективності використання технології опису метаданих семантичних ресурсів, сторінки веб-сайту, відображення та представлення властивостей сутностей у форматі RDFa, результати експериментального пошуку.

## **ВИСНОВКИ**

У результаті виконання дипломної роботи був розроблений веб-сайт на мові XHTML, який включає в себе п'ять сторінок. На основі документації по використанню технології опису ресурсів RDFa та словників сутностей і властивостей була занотована частина контент сайту.

Сайт був включений до пошукової системи Google. Для відображення попередньої видачі були використані розробка Google, яка називається Rich Snippets Testing Tool та веб-сервіс Structured Data linter. Також було проведено тестування сайту в пошуковій системі, а вихідні дані були статистично опрацьовані та представлені у вигляді таблиць. На їх основі було побудовані гістограми за допомогою програмного пакету Excel.

Дані експерименту показали, що пошук по властивостям сутностей RDFa, які були відображені в контенті є ефективнішим, ніж пошук по звичайним текстовим даним. Також вибрана технологія підвищує інформативність видачі при пошуковому запиті. З використанням RDFa скорочується кількість слів, необхідних для пошуку відповідного сайту.

Отже, технологія опису ресурсів RDFa має наступні переваги:

- а) підтримується найпопулярнішою пошуковою системою Google, а також рекомендована робочою групою W3C;
- б) не потребує дуплікації даних;
- в) використовується з розширюваною мовою розмітки XHTML;
- г) при пошуковій видачі надається додаткова графічна та текстова інформація;
- д) словники сутностей та властивостей, що використовуються технологією, постійно поповнюються;
- е) позитивно впливає на просування сайту в пошуковій системі.

## **СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ**

1. Л.В. Корчук. Сумісні стандарти та мови розмітки RDF / Корчук Л.В. – Тези доповіді на V Міжнародній науково-технічній конференції молодих учених та студентів «Актуальні задачі сучасних технологій». Том II, Тернопіль, 17-18 листопада 2016 року.– Тернопіль, ТНТУ, 2016. – с. 56-57.

## **АНОТАЦІЯ**

**Корчук Л.В. Аналіз технологій опису ресурсів з використанням RDF в семантичній павутині.**

Дипломна робота на здобуття освітнього ступеня магістра 8.05010201 - Комп'ютерні системи та мережі. - Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль 2017.

В роботі проаналізовано засоби вирішення проблеми тобто сучасні технології опису метаданих ресурсів. Після проведеного попереднього аналізу сучасних мов описання в семантичній павутині можна зробити висновок що на сьогоднішній день слід розглядати лише технології RDFa, оскільки вони є найбільш оптимальним варіантом анотування та використовуються при пошуку в найвідомішій у світі пошуковій системі Google, до того ж має свою базу атрибутів і сутностей та покращує якість видачі інформації по запиту.

Для організації проведення досліджень за визначеними в першому розділі показниками ефективності в першу чергу використані засоби Google:

а) пошукова система Google – підтримує відображення інформації представленої за допомогою RDFa, дає змогу оцінити наскільки RDFa надає можливість розкрити семантичний зміст тексту, впливає на просування сайту, видачу релевантної та додаткової інформації.

б) Rich Snippets Testing Tool – дає змогу оцінити які дані, занотовані за допомогою RDFa розуміє пошуковий робот Google

в) Google Webmaster Tools – програма для оперативної роботи в сегменті SEO, і отримання статистичних оперативних даних. Користувачам цієї системи набагато простіше підвищувати параметри сайту. Набір можливостей достатній і необхідний для ведення комплексної статистики сайту, а крім цього система досить інтелектуальна і може проаналізувати і вказати на допущені при розробці сайту помилки, які вебмастери не помітили самі.

Для досліджень також використані інші пошукові системи, що підтримують RDFa (Yahoo!, SearchMonkey, Bing та інші). Це дасть змогу проаналізувати актуальність формату для існуючих пошукових систем, інформативність видачі при пошуковому запиті та можливість розкрити семантичний зміст тексту.

Також використані деякі інструменти та сервіси відображення, валідації, сервісу та відкладки сайтів з RDFa наповненням (Validator.ru, Structured Data Linter, Online parser/viewer and RSS generator ), що дасть змогу оцінити витрати на підтримку і розширення закладеного інформаційного набору, необхідність використання особливої мови запитів, трудомісткість опису, релевантність інформації та можливість розкрити семантичний зміст тексту. Деякі з них дають можливість графічного відображення аналізу.

Проведено експериментальні дослідження.

**Ключові слова:** СЕМАНТИЧНА ПАВУТИНА, RDF, XHTML, РЕСУРС, ВЕБ-СТОРІНКА, РОЗМІТКА, ФОРМАТ, АТРИБУТ, МЕТАДАНИ.

## ANNOTATION

**Korchuk L. V. Analysis of resource description technologies using Semantic Web RDF**

The diploma paper for obtaining the Master's degree 8.05010201 - Computer systems and networks - Ivan Puluj Ternopil National Technical University, Ternopil 2017.

The analysis tools that address the problem of modern technology metadata describing resources. After a preliminary analysis of modern languages in the description of semantic web can conclude that today should only technology RDFa, as they are the best option and annotation used to search the world's most famous search engine Google, moreover, has its own database and attributes entities and improves delivery of information upon request.

For the organization of research by their first section performing primarily used tools Google:

a) the search engine Google - can display the information provided via RDFa, allows you to evaluate how RDFa allows you to uncover the semantic meaning of the text affects the promotion, issuing relevant and more information.

b) Rich Snippets Testing Tool - allows you to evaluate that data, tracked using RDFa aware of Google search

c) Google Webmaster Tools - a program for operational segment SEA and obtain statistical operational data. The users of the system much easier to increase the site settings. Set of capabilities sufficient and necessary for maintaining comprehensive statistics website and in this system is quite intelligent and can analyze and point out the development site made mistakes that webmasters did not notice themselves.

For studies also used other search engines that support RDFa (Yahoo !, SearchMonkey, Bing and others). This will allow to analyze the relevance format for existing search engines, information in issuing the search request and to discover semantic content of the text.

We also used some tools and services mapping, validation, service and vidkladky sites with RDFa content (Validator.ru, Structured Data Linter, Online parser / viewer and RSS generator), which will allow to estimate the costs of maintaining and expanding the set of embedded information, the need for special query language, the complexity of the description and relevance of information to discover semantic content of the text. Some of them make it possible to graphically display analysis.

In the third section conducted experimental studies.

**Key words:** SEMANTIC WEB, RDF, XHTML, RESOURCE, WEB PAGES, LAYOUT, FORMAT, ATTRIBUTES, METADATA.