**Проектування додатків Semantic Web. Проблема цілісності даних.**

**Новицький Олександр**

**alex.googl@gmail.com**

**Інститут програмних систем НАН України УКРАЇНА, м. Київ, проспект Академіка Глушкова 40,**

# Вступ

Для мережі Інтернет, при активному розвитку інтелектуальних систем все більше набувають значення технології пов’язані з можливістю машинного аналізу інформації. Зазвичай машинна обробка даних, буде ефективною за умови наявності формалізованої моделі даних. Одним із способів формалізації інформації є RDF. RDF є загально прийнятою моделлю представлення семантичної структурованої інформації. Він надає засоби для публікації та обміну машиночитаних даних про ресурси в Інтернеті.

Проте RDF я тільки зовнішньою обгорткою даних. При роботі з ним, аналогічно до баз даних виникають певні проблеми зокрема проблеми цілісності даних. Ці проблеми є нерозв’язаними в рамках самого RDF, тому постають відповідні задачі розвинення засобів та методів перевірки цілісності даних в середовищі Semantic Web.

# Цілісність даних RDF

Перевірка цілісності не дуже добре підтримується в RDF. Це не так важливо в RDF, тому що кожна трійка (атомна одиниця інформації в графі RDF) розглядається як частина деякої загальної інформації. Таким чином, будь-якої комбінації трійок є в певному сенсі має свою інтерпретацію з погляду на відкритість світу.

RDF Schema, незважаючи на свою назву, також не підтримує перевірку цілісності даних.

Якщо розглянути OWL то у літературі, передумова OWA (Open World Assumption) була визначена як "найбільшим перешкодою до розуміння OWL". Існує також поширена помилка відносно тлумачень аксіом в OWL як аналогічних обмеженнь в реляційних базах даних. Тим не менш, аксіоми в онтології призначені для виведення нових знань, а не для валідації цілісності.

Перевірити цілісність RDF (OWL) можна лише шляхом закриття світу, тобто обмежити якимось чином можливі інтерпретації. Це можливо здійснити з використанням онтологій які будуть визначати додаткові обмеження над даними. Таки підхід використовується в Pellet [1], Eyeball який є частиною Apache Jena, а також Stardog ICV [2]. Семантика цілісності даних в таких підходах передбачає включення обмежень до онтологій. Водночас такі обмеження є не правильним для онтологій які є загально прийнятими. Наприклад нехай для конкретної інформаційної системи нам необхідно зафіксувати певний обмежений набір полів метаданих з онтології Дублінського Ядра для опису інформаційного ресурсу. Нехай деякий інформаційний ресурс повинен мати в описі тільки два предикати dc:title та dc:creator.

Формально це можливо записати наступним чином



Для накладення обмежень без розширення існуючих онтологій виражати такі обмеження можливо через RIF (*Rule Interchange Format*).

RIF є стандартом для обміну правилами між різними програмними системами. RIF зосереджується саме на обміні, тому на відміну від інших Семантичних веб-стандартів, таких як RDF, OWL і SPARQL, RIF не задовольняє потребам багатьох популярних парадигмам представлення знань та моделювання процесів.

В основу сімейств діалектів RIF покладається однорідність та розширюваність. Під однорідністю RIF очікується, що діалекти будуть мати максимально спільний синтаксичний і семантичний апарат. Розширюваність RIF [3] означає, можливість за потреби визначення нового діалекту RIF, як синтаксичне розширення існуючих діалектів RIF, з новими елементами, для реалізації додаткових функцій. Під перевіркою цілісності даних в RDF ми розуміємо тільки наявність повноти RDF трійок про деякий факт, то для перевірки таких обмежень використовувати весь апарат RIF-BLD чи RIF Core, є не доцільним. Тому для визначення правил обмеження цілісності даних, ми використали підмножину виразів та формул RIF Core та RIF-BLD, отриманий діалект RIF-CR. Алфавіт RIF-CR аналогічний до RIF-BLD. Проте набір визначення основних темів та формул, спрямований виключно на вирішення проблеми цілісності даних. Це дозволяє простіше практично реалізувати інтелектуальну систему підтримки цілісності даних в Semantic Web.

# Висновок

Реалізації підтримки RIF на стороні додатку для обробки семантичних даних, є можливим рішення проблем валідації цілісності даних. В рамках RIF розроблено власний діалект для застосування в специфічних предметних областях.

# Література

|  |  |
| --- | --- |
|  | 1. **Clark & Parsia, LLC**. Clark & Parsia: Thinking Clearly. [Online]. <http://clarkparsia.com/pellet/icv/> |
|  | 1. **Clark & Parsia, LLC**. (2013, Apr.) Validating RDF with OWL Integrity Constraints. [Online]. <http://stardog.com/docs/sdp/icv-specification.html> |
|  | 1. **Gary Hallmark, Michael Kifer, Adrian Paschke, Axel Polleres, Dave Reynolds Harold Boley**. (2013, Apr.) RIF Core Dialect (Second Edition). [Online]. <http://www.w3.org/TR/rif-core/> |