

Семантичний ГРІД в молекулярних дослідженнях

Османова Т. М., ННК «ІПСА» НТУУ «КПІ», Лунченко О.О, ННК
«ІПСА» НТУУ «КПІ»

На даний час відомо більше ніж 30 мільйонів складних хімічних структур, що є важливими для медицини, біологічних наук та створення нових матеріалів. Надзвичайно важливим є розуміння властивостей цих структур. Для їх дослідження визначають поведінку молекул використовуючи відомі фізичні закони, часом ці характеристики можуть бути виміряні, проте у більшості випадків вони повинні бути розраховані. Більшість необхідних характеристик може бути розраховано використовуючи рівняння Шрьодінгера, але кількість обчислень при цьому надзвичайно велика. При цьому обчислення власне рівняння не є основною складністю, набагато складнішим є аналіз отриманих результатів: дані та коди молекул – дуже неоднорідні без впорядкованої структури.

Саме така задача є однією з найважливіших для розробки семантичного ГРІДу. Основна семантика – надзвичайно стійка, програмні коди також є дуже надійними з явною вираженою семантикою. Більше того, необхідність у обчислювальній хімії є значною у багатьох науках, що не належать до хімічних – матеріалознавство, безпека, біологічні науки, науки про Землю та нанотехнології. Ці користувачі вимагають підходу «чорної скрині», що пропонує отримання результатів на вимогу, без розуміння того, як обчислюється результат.

Враховуючи вимоги щодо зберігання та обробки даних в молекулярних дослідженнях, було розглянуто модель семантичного ГРІДу, що передбачає мінімальні зусилля з боку користувачів для роботи з даними. Всі компоненти містять відкритий вихідний код та відкриті дані, що дозволяє будь – кому завантажувати або працювати з даними. Модель передбачає вільне обчислювання молекулярних властивостей у «чорній коробці» використовуючи методи квантової механіки. Користувач завантажує дані у відомому та зручному йому текстовому редакторі (наприклад Microsoft Word), далі дані в залежності від структури перетворюються в CML (Chemical Markup Language) або PML (Polymer Markup Language), потім відбувається перетворення у XML (Extensible Markup Language) і на виході дані зберігаються в XML – репозиторії.

Найбільшу увагу у роботі приділено опису мов CML та PML, що були розроблені для опису семантики даних, що використовуються в молекулярних дослідженнях. Було проаналізовано основні переваги та недоліки кожної мови та сфери їх застосування. В роботі запропоновано декілька можливих покращень, що дозволять розглядати дані мови не лише для опису конкретних фізичних проблем (наприклад, розрахунку структури полімерних ланцюжків), а для усього спектру задач розв'язуваних за допомогою семантичного ГРІДу в молекулярних науках.

Список літератури

1. P. Murray-Rust and H. S. Rzepa, J. Chem. CML Schema, – Inf. Comp. Sci., 2003. – 43p.

2. P. Murray-Rust and H. S. Rzepa. STMML. A markup language for scientific, technical and medical publishing, –Data Science, 2002, –1-65p.

3. P Murray-Rust, RC Glen, HS Rzepa, JJP Stewart, JA Townsend, EL Willighahen, Y. Zhang. A semantic GRID for molecular science [Электронный ресурс]/ P Murray-Rust. .–Режим доступа: <http://www.dspace.cam.ac.uk/bitstream/1810/197085/1/34E7446C.pdf>