

# Реферат

магистерской аттестационной работы

на тему:

“Семантический Грид в молекулярных исследованиях”

Османовой Татьяны Мунировны

## **Цель работы**

Целью данной работы является исследование и анализ существующих моделей семантического Грид в молекулярных исследованиях.

## **Актуальность проведенных исследований**

Ученые многих дисциплин используют молекулярное моделирование для получения информации о свойствах молекул. Исследования свойств молекулярных структур это одна из самых главных задач современной науки. Их характеристики и свойства широко используют в физике, химии, биологии, генетике, медицине, палеонтологии, инженерии и других науках.

В данный момент известно более 30 миллионов сложных химических структур, которые очень важны для медицины, биологических наук и создания новых материалов. Чрезвычайно важным является понимание свойств этих структур. Для их исследования определяют поведение молекул, используя известные физические законы, временами эти характеристики могут быть измерены, но в большинстве случаев они должны быть рассчитаны. Большинство необходимых характеристик может быть рассчитаны с помощью уравнения Шрёдингера, но количество вычислений при этом слишком велико. При этом вычисление именно уравнения не является основной сложностью, намного сложнее анализировать полученные результаты, поскольку данные и коды неоднородные и не содержат упорядоченную структуру. По этому, в последнее время научно – исследовательские инициативы широко применяют Грид – вычисления для простоты и ускорения вычислений. Но при этом остается проблема

сохранения данных, поскольку каждая научно – исследовательская лаборатория сохраняет свои данные локально, что сдерживает поиск необходимой информации в несколько раз.

Так как в химическом сообществе основной источник данных это индивидуальные научные работы в традиционных химических лабораториях, дополненные на сегодняшний день приходом автоматизированных, высокопродуктивных, синтезированных и аналитических технологий (особенно в промышленных секторах). Данные генерируются в лабораториях и распределяются по всему сообществу, каждая лаборатория имеет большой вклад в научные открытия, которые консолидированы через мировую литературу. Необходимо создать систему, которая смогла бы собрать мировые изобретения в молекулярных исследованиях, тем самым облегчить доступ к ним и поиск. Таким инструментом и может стать разработка семантической Грид – инфраструктуры.

Пытаясь решить проблемы медицинского обслуживания, здравоохранение все больше обращается к информационным технологиям, в которых видит возможность управления ресурсами, уменьшения очередей, исключение ошибок и обеспечения современного уровня лечения населения отдаленных регионов. Однако прогресс в модернизации здравоохранения тормозится рядом таких факторов, как: прохождение и понимание записи истории болезни пациента между организациями внутри страны и между странами; неуверенность в том, что информация защищена, и доступ к ней регулируется; выявление надежных источников информации для сравнения; управления большими объемами данных, особенно в области медицинской генетики; использования традиционных информационных сетей и технологий в здравоохранении. Однако Грид технологии позволят клиницистам преодолеть многие из перечисленных трудностей.

### **Решаемые в работе задачи**

В работе представлены теоретические сведения о существующих моделях семантического Грид для молекулярных исследований, рассмотрены

разные типы онтологий, с помощью которых описываются метаданные химических структур в этих моделях, рассмотрены способы обнаружения ресурсов в семантических Грид, а именно поиску ресурсов, что удовлетворяют запросам пользователей.

Проведен анализ существующих систем семантического Грид для молекулярных исследований, учитывая назначение каждой и перечень заданий, которые они используют с учетом современных требований к семантическому Грид. Эти требования включают в себя совместимость, масштабируемость, децентрализованность и динамизм.

### **Достигнутые результаты**

Результатом проведенных исследований является определение требований к семантическим Грид в области вычислительной химии, с учетом молекулярных исследований в Украине и соответствия его к мировым стандартам существующих моделей, таким как: Molecular GRID (World Wide Molecular Matrix (WWMM)), проект CombeChem та проект Collaboratory for Multi-scale Chemical Science(CMCS).

### **Научная новизна**

Научная новизна работы заключается в анализе и изложении основных требований в контексте применения семантического Грид для молекулярных исследований. На сегодняшний момент в мире очень широко употребляется и развивается технология семантического Грид во многих областях наук, но к сожалению нет такого стремительного применения в области молекулярных наук. По этому, данная работа может служить отправной точкой в исследовании разработки семантического Грид для молекулярных исследований в Украине.

### **Практическая ценность**

Практическая ценность работы заключается в получении систематической, теоретической и информационной основы для построения инфраструктуры семантического Грид для работы в молекулярных

исследованиях. Результаты работы могут быть использованы как рекомендации и предложения для построения украинской Грид – инфраструктуры в молекулярных науках, а так же как руководство по использованию существующих.

### **Выводы и рекомендации**

В данной работе были рассмотрены основные требования, которые ставятся при построении инфраструктуры семантического Грид. Оценено состояние молекулярных исследований в мире. Показаны основные проблемы, которые возникают во время молекулярных исследований и тем самым доказана целесообразность и актуальность использования данной инфраструктуры в молекулярных науках.

В работе приведено детальное описание моделей представления семантического Грид в мире, а именно Molecular GRID (World Wide Molecular Matrix (WWMM)), проект CombeChem та проект Collaboratory for Multi-scale Chemical Science(CMCS).

На основе подробного анализа этих систем обнаружены некие недостатки, а именно все системы нацелены на то, что бы решать вопросы поиска и сохранения данных, ни в одной работе не реализован агентный подход поиска. С точки зрения совместимости, данные модели охватывают очень узкую прикладную область, такую как описание молекул и их свойств, но в проекте CombeChem из-за применения очень детальной онтологии, есть возможность расширить эту область.

Даная работа не может считаться целостной для разработки и применения семантического Грид в молекулярных исследованиях, поскольку каждая отдельная группа ученых занимается определенным кругом исследований и проанализированные модели являются только пример применения технологии семантического Грид для определенного круга задач и проблем, которые возникают во время молекулярных исследований.

Так же в работе проведен подробный анализ научных исследований в Украине и показано, что большая часть научных институтов использует в

своих исследованиях молекулярные структуры и их свойства. Именно по - этому в работе предложены пути использования семантического Грид в национальной Грид – инфраструктуре. Есть два варианта для этого: создать собственную модель или же использовать существующую. И для решения каждого из них в работе разработаны определенные рекомендации.

Работа на 112 страницах содержит 22 иллюстрации. При подготовке работы использовалась литература из 27 разных источников.

Перечень ключевых слов:

*семантический Грид, семантический Веб, онтология, молекулярная динамика, метаданные, , CML, XML- схема, Protein Data Bank, World Wide Molecular Matrix (WWMM)), CombeChem, Collaboratory for Multi-scale Chemical Science(CMCS), Dublin Core, InChI, молекулярне структуры, фолдинг белков.*