

Информационная среда
интеллектуальной системы типа ДСМ
для фармакологических задач

Ольшанский Дмитрий Леонидович

Аспирант ВИНТИ РАН

Метод ДСМ

ДСМ - метод автоматического порождения гипотез, предложенный В.К.Финном [1983]

Назван по инициалам английского философа Джона Стюарта Милля

Синтез процедур индукции, аналогии и абдукции

Успешно применяется в областях социологии, технической диагностики, фармакологии

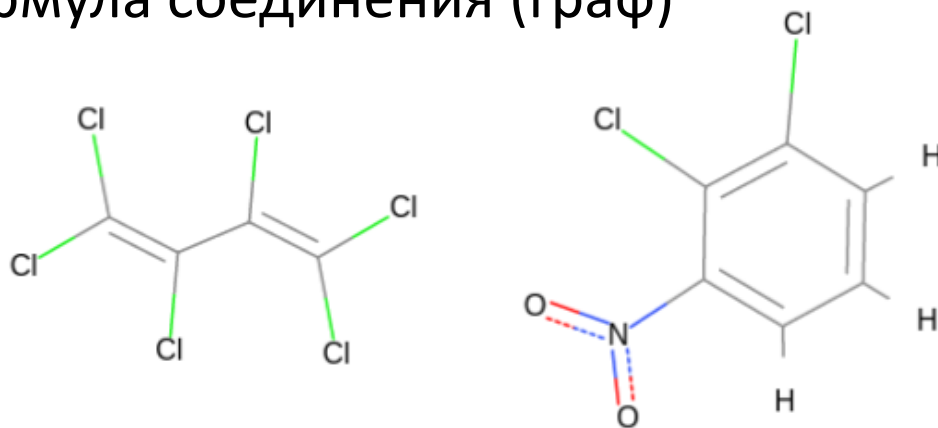
Фармакология - исторически первое применение интеллектуальной ДСМ-системы (связь структура – активность) [1997]



Джон Стюарт Милль

Сходство в химических структурах

Объект - структурная формула соединения (граф)



«Простое» решение

- сходство объектов на уровне элементов графа

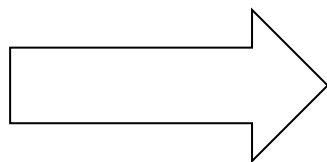
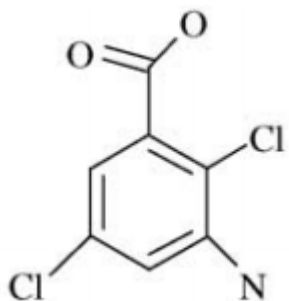
Проблемы:

- изоморфизм графа и максимальный общий подграф - NP-трудные задачи
- зашумление результата – не каждый фрагмент молекулы имеет одинаковый смысл (часть цикла и углеродная цепь)

Дескрипторные языки

Предметно-ориентированные языки описывают отдельные фрагменты соединений или функциональные группы

Цель – большая информативность и линейность записи



HC	CCOO
HN	NHHC
HO	OHC
CHCC	OC
CCCC	ClC
CCCN	
CCCCl	

Язык MNA/1 [A.Lagunin et al 2005]

Язык ФКСП

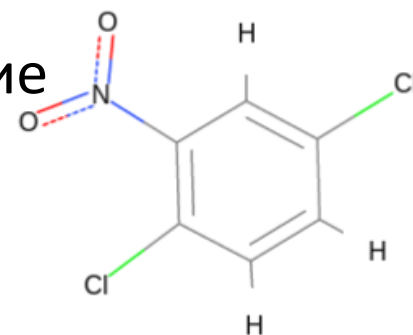
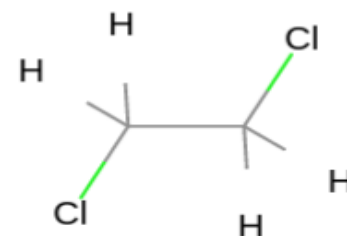
Фрагментарный код суперпозиций подструктур
[В.В. Авидон, 1972] для исследования
биологической активности соединений

Описывает соединения как набор из 3 видов
активных подструктур

Линейный дескриптор – пары активных центров,
связанных углеродной цепочкой

Циклический дескриптор – циклы и циклические
системы в молекуле

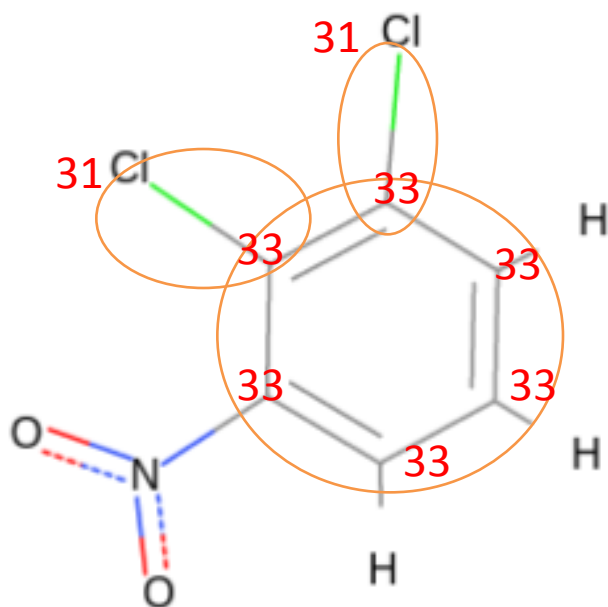
Дескриптор замещения – взаимное расположение
активных центров в ароматических системах



Язык ФКСП

Пример кодирования

Соединение



КОД ФКСП

6,06 3100331 3100331 3162311

31 – активный центр (атом Cl)

33 – активный центр (атом ароматического цикла)

3100311 – линейный дескриптор

6,06 – циклический дескриптор

3162311 – дескриптор замещения для пары 31 центров в цикле

Комплекс WebChem-JSM

Задачи

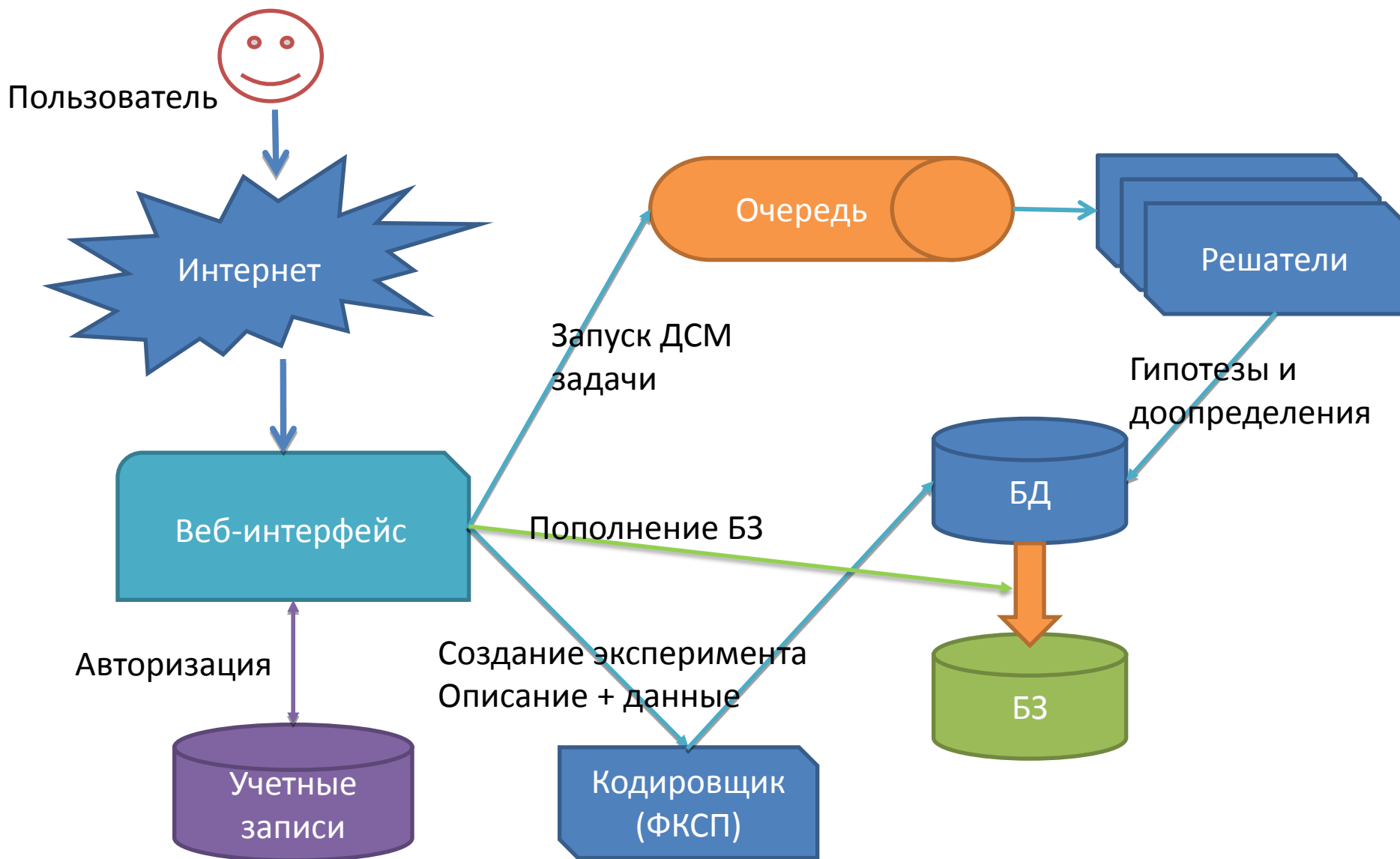
- создать информационную среду для проведения ДСМ экспериментов в области фармакологии
- вести базу данных (БД) по проводимым ДСМ экспериментам
- с участием экспертов формировать базу знаний (БЗ) по результатам экспериментов
- предоставлять визуализацию данных, гипотез и доопределений

Комплекс WebChem-JSM

Ключевые свойства

- множество одновременных пользователей
- удобство применения: без установки, настройки, привязки к ОС и т.п.
- высокая эффективность, масштабируемость
- гибкость: для внесения изменений и обновлений
- централизованное хранение для БД и БЗ
- контроль и ограничение доступа
- простота поддержки системы

Архитектура WebChem-JSM



Интерфейс WebChem-JSM

Новый эксперимент

Название

Эксперимент-2

Дата

27.05.2015

Активность

Токсичность

Уровень активности

2

Метод

Прямой

Кодировщик

FCSP

Комментарий

Первый эксперимент.

Загрузить MOL файлы

Импортировать активность

Создать эксперимент

Название	Активность		
10.MOL	+ -	+ - ?	🗑
11.MOL	+ -	+ - ?	🗑
12.MOL	++	+ - ?	🗑
13.MOL	++	+ - ?	🗑

Интерфейс WebChem-JSM

Список экспериментов

admin

Эксперимент "Эксперимент-1"

Данный эксперимент был выполнен. Перейти к [гипотезам](#).

Входные данные

Выберите пример из списка, чтобы получить детальную информацию.

#	MOL	Активность
1	10.MOL	+ -
2	11.MOL	+ -
3	12.MOL	++
4	13.MOL	++
5	14.MOL	++
6	15.MOL	++
7	16.MOL	--
8	17.MOL	--

Схема молекула и коды



6,06 0200331 0200331 0263020

Интерфейс WebChem-JSM

Список экспериментов

admin

Гипотезы

'+' гипотезы

'-' гипотезы

Предсказания

Гипотезы

Коды

Родители

Молекулы

1 (+0)

0200331

11.МОЦ(1)

2 (+0)

0200331

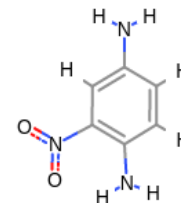
12.МОЦ(2)

3 (+0)

6,06

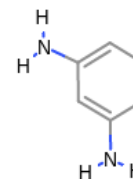
4 (++)

Родитель



Гипотеза - выберите фрагмент

1 2



Решатель

Для большей эффективности впервые применяются параллельные алгоритмы для шага индукции

В качестве основы – семейство алгоритмов «Замыкай по одному»

Close By One [Kuznetsov, 2002]

В том числе модификации

FCbO [Krajca et al, 2004]

In-Close2/InClose-3 [Andrews, 2012-2015]

Критерий отбора алгоритмов – минимальный объем состояния в памяти, без центральной структуры данных

Заключение

Созданная информационная среда реализует

- веб-интерфейс пользователя
- автоматическое кодирование и визуализацию данных предметной области
- высокоэффективный решатель с параллельным шагом индукции
- базу данных и базу знаний

Модульная архитектура позволяет добавить альтернативные реализации или заменить отдельные компоненты