

Декларативная аналитика в мультидиалектной среде

Ковалев Д.Ю.

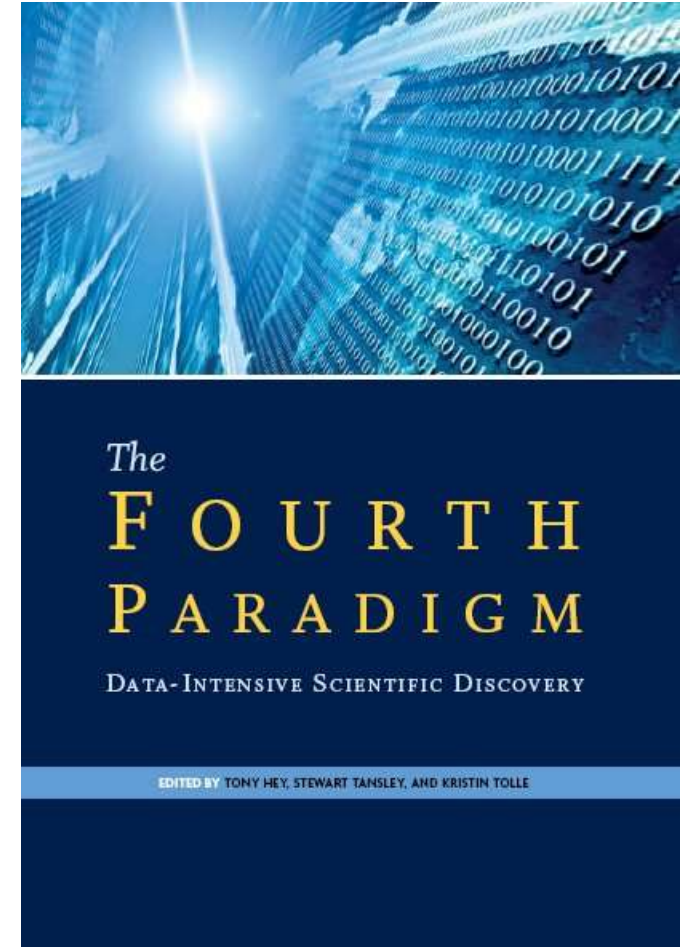
сотрудник лаборатории 23 ИПИ РАН,
аспирант ВМК МГУ

Науки с интенсивным использованием данных

Работа с данными - сбор,
хранение, обработка, анализ,
визуализация и т.д.

Данные

- структура данных;
- объем
- скорость поступления



Концептуальное программирование

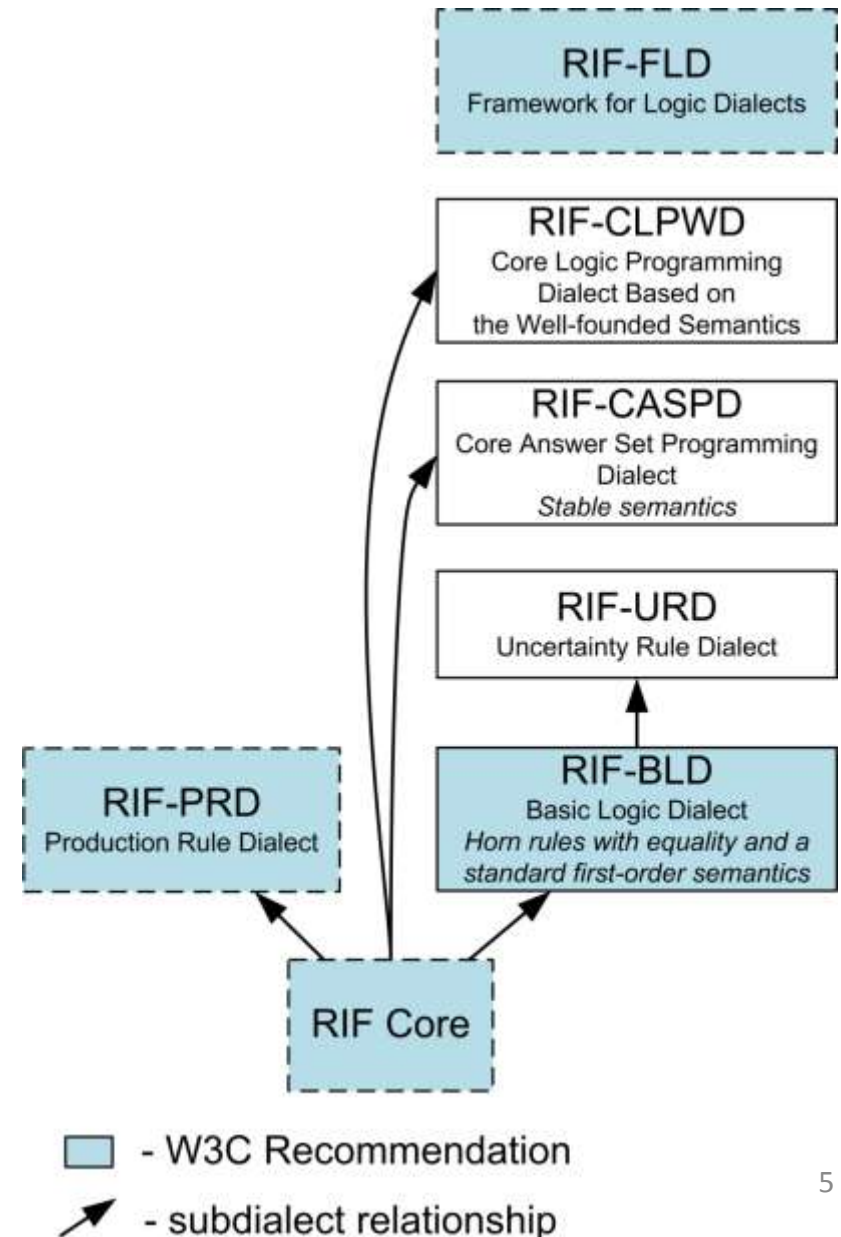
1. Концептуальное программирование, независимое от ресурсов
2. Идентификация релевантных ресурсов, построение отображений между концептуальной моделью и схемами ресурсов.
3. Спецификация задачи в терминах концептуальной модели

Декларативное программирование

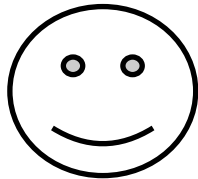
- Взрывной рост количества областей, где применяются декларативные высокоуровневые языки [Heller10]:
 - сетевые протоколы безопасности, анализ компиляторов, вероятностный вывод, управление роботами и т.д.
- Понятный для неспециалистов [Abit13]

Rule Interchange Format - RIF

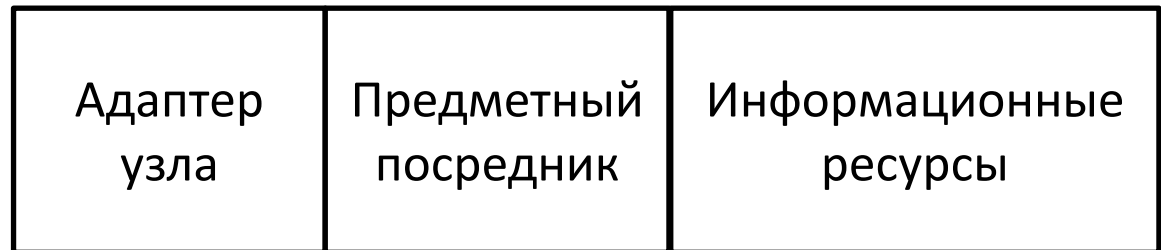
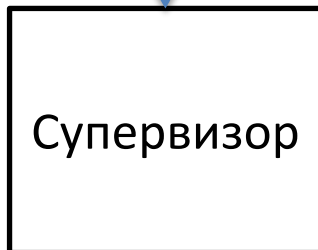
- Унифицированное семейство языков на правилах (диалектов)
 - каждый диалект обладает своей теоретико-модельной семантикой
- Отображение в диалект и обратно должно *сохранять семантику* (*отношение логического следования*)
- Возможность совместного использования с RDF и OWL



Мультидиалектная среда



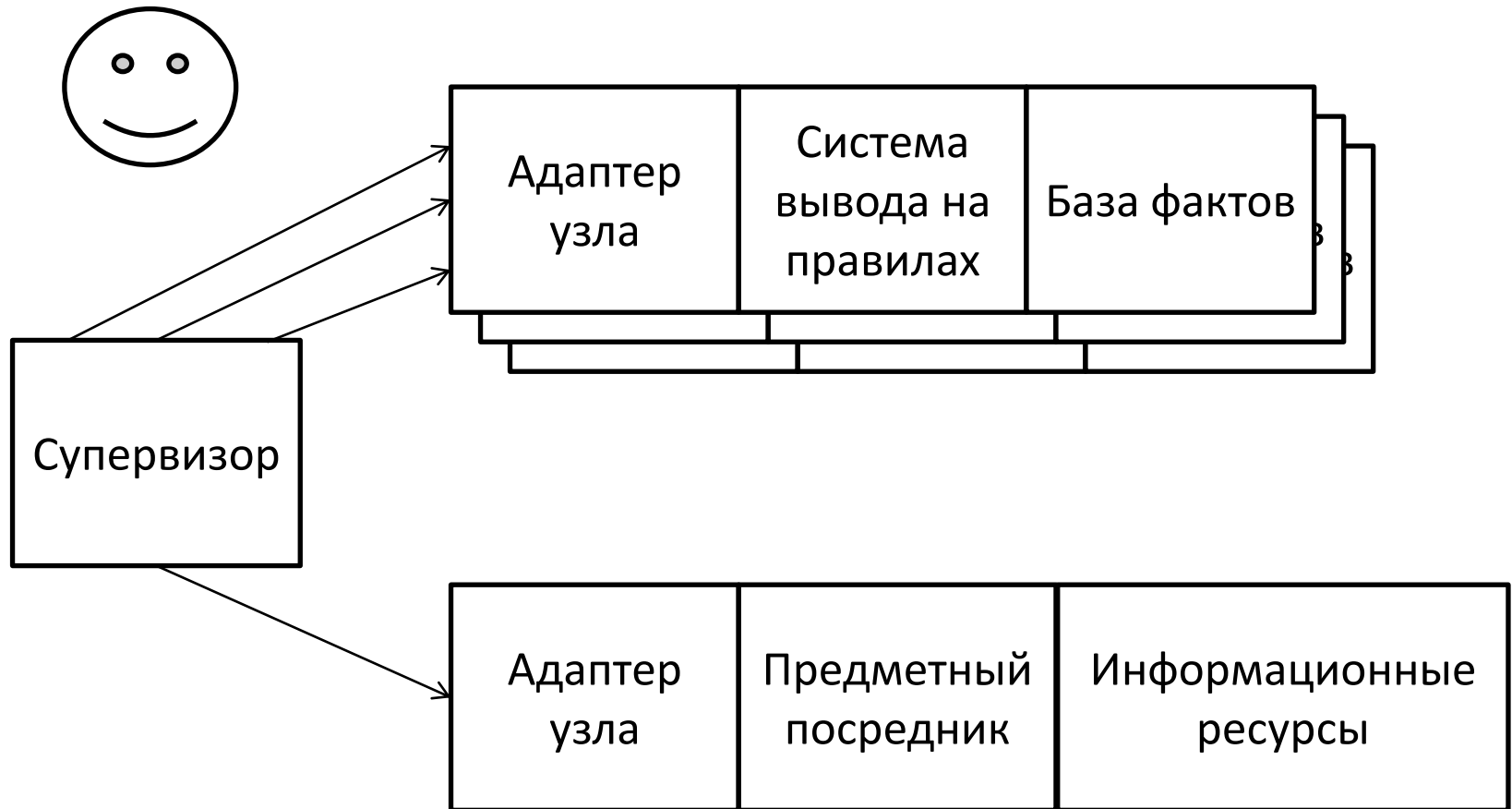
концептуальная
программы



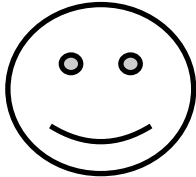
- Узел состоит из

- адаптер
 - переписывает программы в диалекты и обратно
 - реализует делегирование правил
- Система вывода или посредник
- Программа на правилах и коллекция фактов

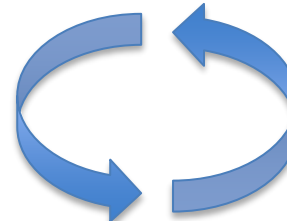
Мультидиалектная среда



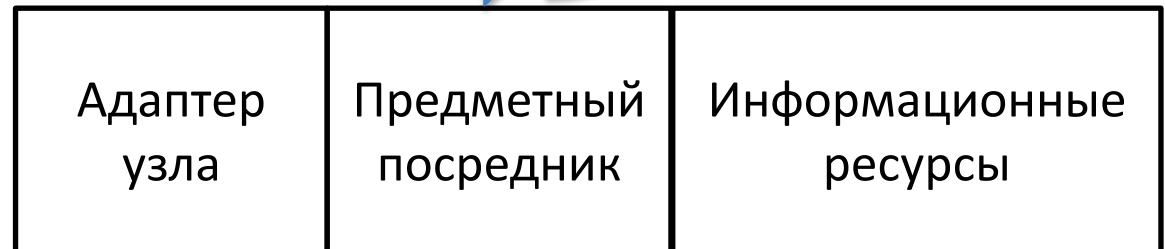
Мультидиалектная среда



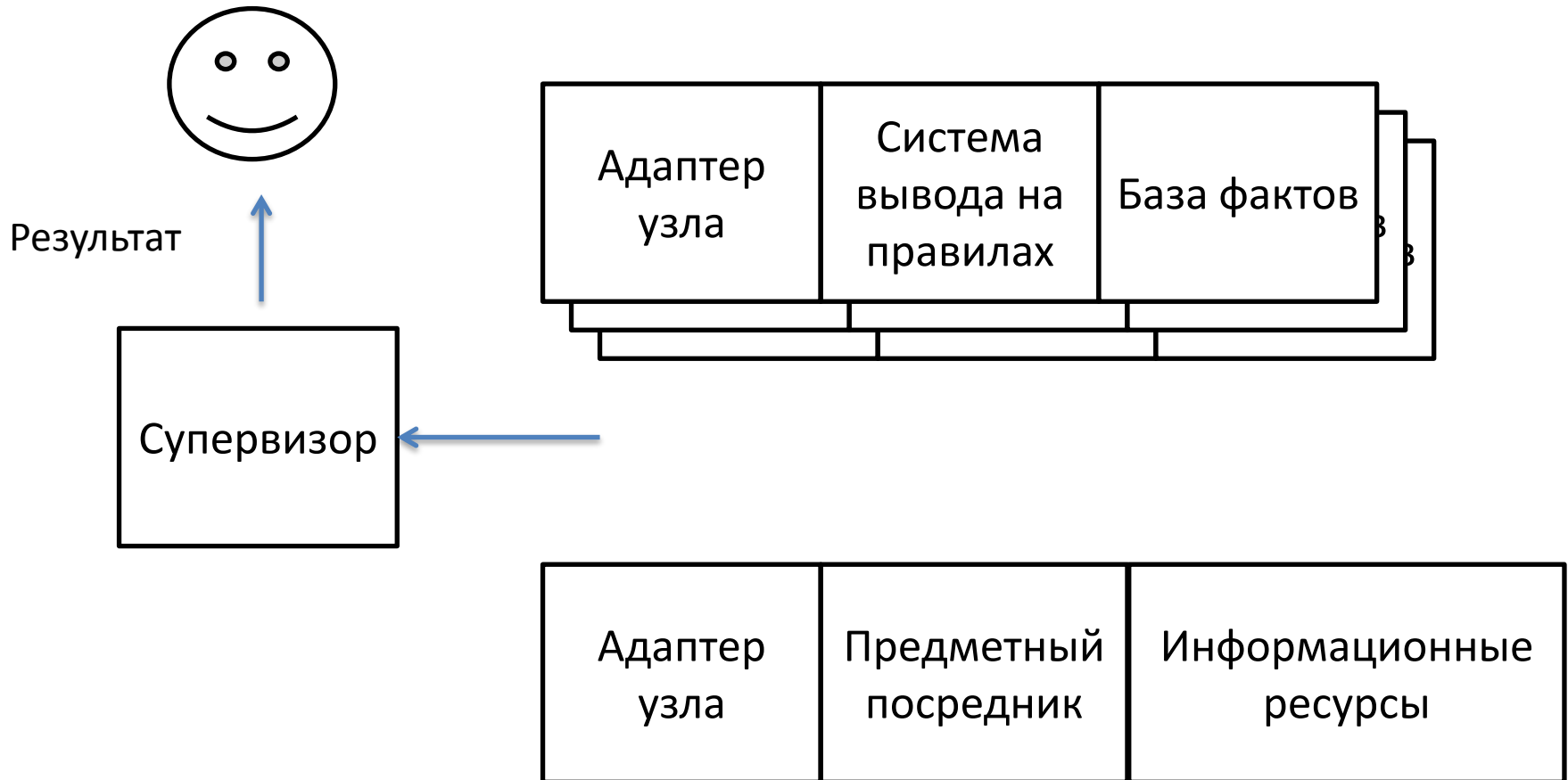
Супервизор



Узлы обмениваются фактами и правилами



Мультидиалектная среда



Цель и задачи

- Цель
 - исследование и развитие программной среды для сопровождения научных исследований, в которой концептуально и декларативно используются разнообразные инструменты – языки и системы на правилах, средства интеграции данных, аналитические методы
- Задачи
 - Пример задачи, требующий аналитики
 - Концептуальная - > распределенная программа
 - Теоретико-модельная семантика программы
 - Доказательство сохранения семантики отображением

Пример: Построение диверсифицированного инвестиционного портфеля

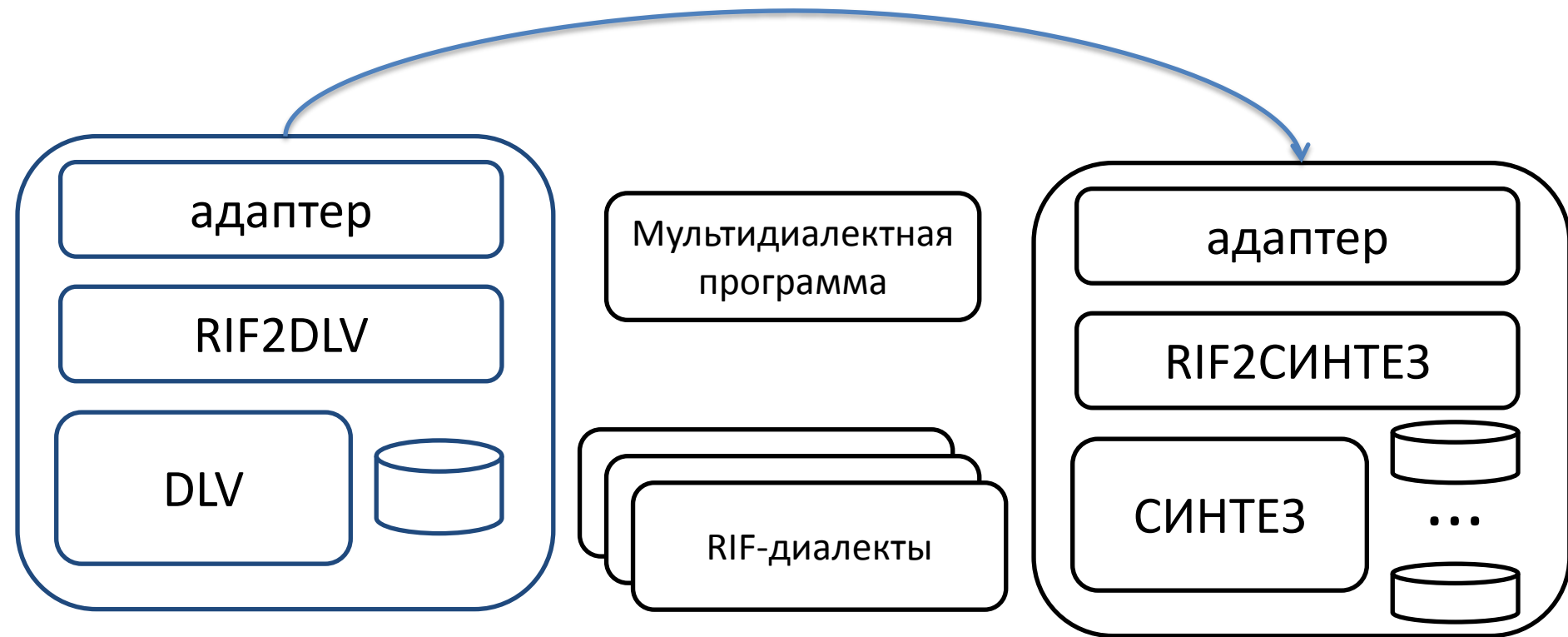
- *Портфель* – множество ценных бумаг
- *Проблема* – построить портфель максимального размера
 - колебание стоимости ценных бумаг не зависит от других бумаг
 - риск значительного падения стоимости портфеля существенно уменьшается
- Задачи
 1. вычисление попарных корреляций для выбранного временного периода
 2. вычисление максимального подмножества бумаг, удовлетворяющих условию
- Формализация задач
 1. Интегрировать исторические данные о стоимости ценных бумаг (*Google Finance, Yahoo! Finance*) используя предметные посредники
 2. Свести задачу поиска подмножества к задаче о поиске *максимальной клики* в неориентированном графе (NP-полная) и решить ее с помощью специально адаптированной системы на правилах

□ Детали: http://synthesis.ipi.ac.ru/synthesis/projects/RuleInt/Conceptual_Use_Case_Example.htm

Что сделано?

- Выбраны два диалекта RIF-BLD и RIF-CASPD
- Построено отображение из DLV в RIF-CASP
 - требуется доказать, что отображение сохраняет отношение логического следования
- Показана эквивалентность проблемы задаче поиска максимальной клики в графе
- Специфицирована концептуальная программа
- Реализовано *делегирование* правил и супервизор, реализующий исполнение программы

Делегирование правил

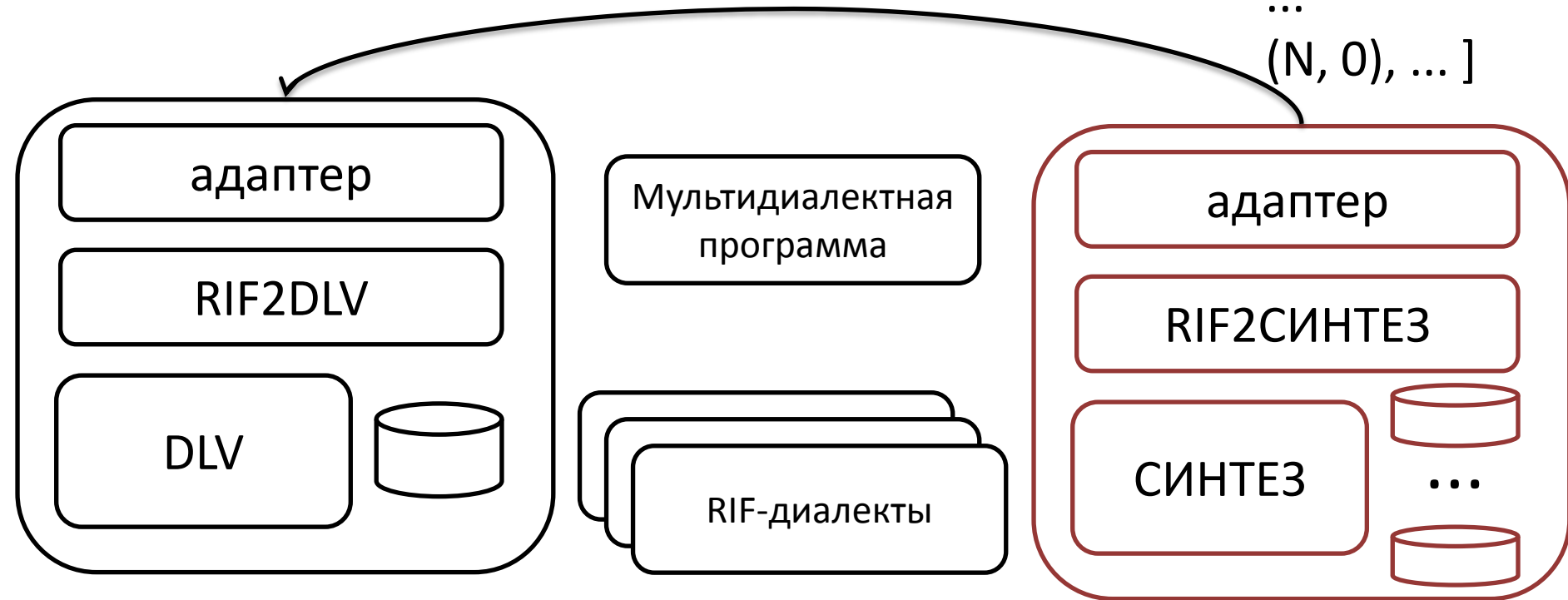


$\text{uedge}(X, Y) \quad :- \quad \text{edge@synthesis}(X, Y), \quad X < Y.$

Делегирование правил

$(X, Y) = [(0, 1), (0,$

$\dots$
 $(N, 0), \dots]$



$\text{uedge}(X, Y) \text{ :- edge@synthesis}(X, Y), X < Y.$

Поиск максимальной клики на узле DLV

`uedge(X,Y) :- edge@synthesis(X,Y), X < Y.`

`uedge(Y,X) :- edge@synthesis(X,Y), Y < X.`

Предикаты из
узла СИНТЕЗ

% вершина либо в клике, либо нет

`clique(X) v nonClique(X) :- node@synthesis(X).`

% две вершины не могут быть в клике, если между ними нет ребра

`:- clique(X), clique(Y), not uedge(X,Y), X < Y.`

% максимизация размера клики

`:~ nonClique(X).`

Поиск максимальной клики на узле DLV

Вывод

```
Cost ([Weight:Level]): <[71:1]>
```

```
Best model: {clique(0), clique(12), clique(14),  
clique(20), clique(29), clique(61)}
```

```
Cost ([Weight:Level]): <[71:1]>
```

```
Best model: {clique(0), clique(12), clique(17),  
clique(23), clique(29), clique(68)}
```

...

Будущие исследования

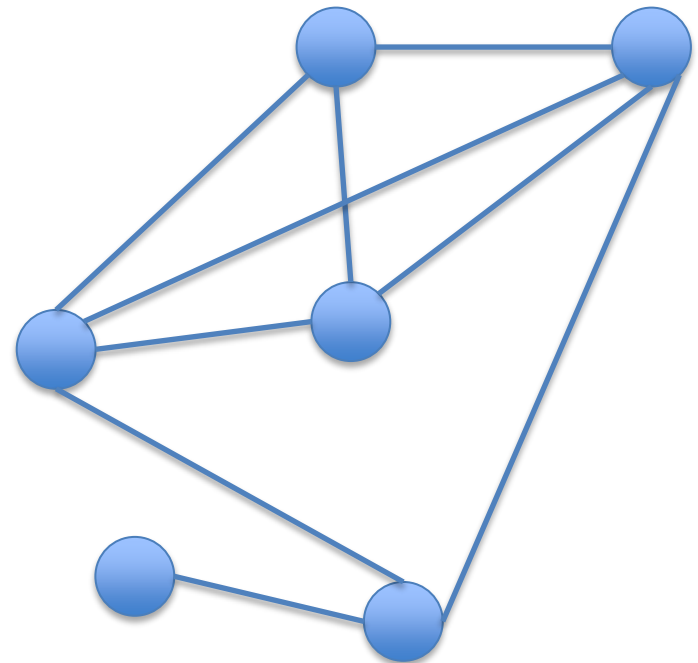
- Доказательство сохранения отношения выполнимости
- Описание теоретико-модельной семантики концептуальной и распределенной программ
- Улучшение делегирования и спецификация распределенного алгоритма
- Другие примеры задач

Спасибо за внимание!

Back up

Задача о максимальной клике

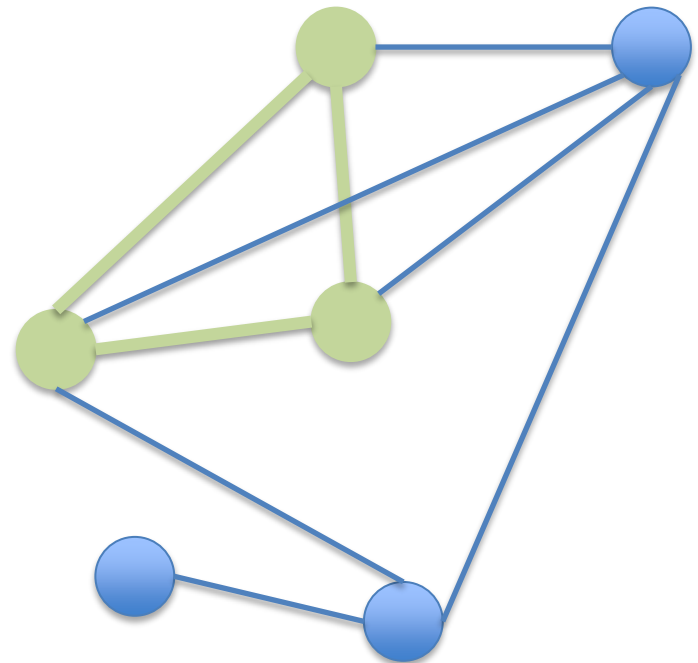
- Клик в неориентированном графе называется подмножество вершин, каждые две из которых соединены ребром графа.
- Размер клики - число вершин в ней.
- Требуется найти клику максимального размера



Задача о максимальной клике

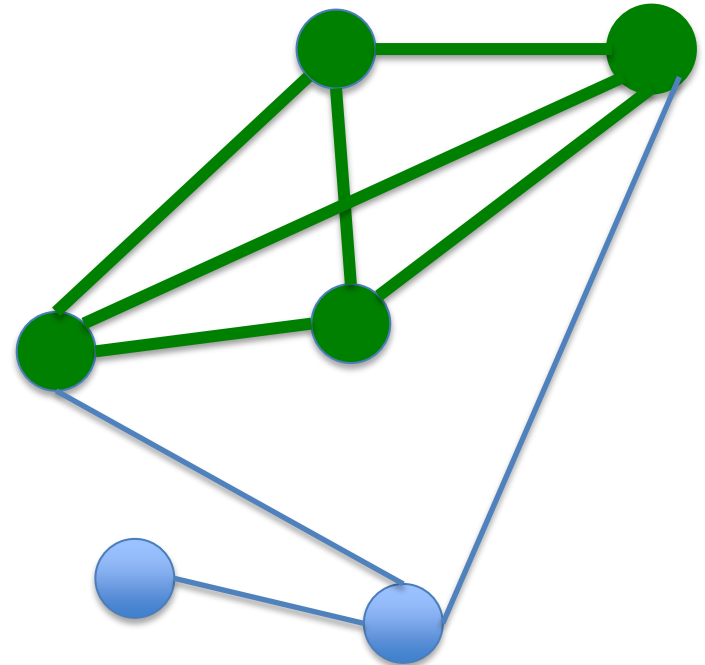
- Клик в неориентированном графе называется подмножество вершин, каждые две из которых соединены ребром графа.
- Размер клики - число вершин в ней.
- Требуется найти клику максимального размера

клика с 3 вершинами

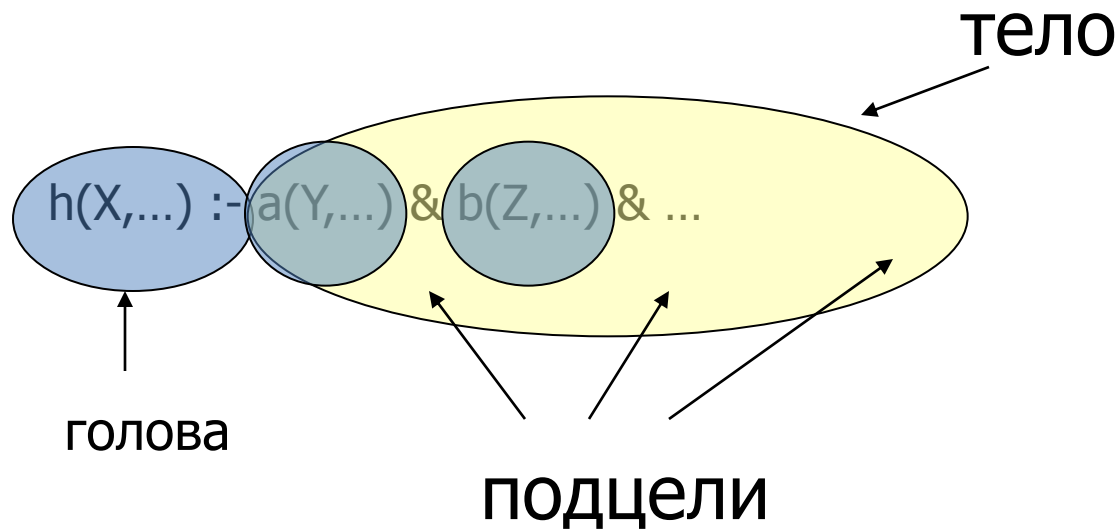


Задача о максимальной клике

- Клик в неориентированном графе называется подмножество вершин, каждые две из которых соединены ребром графа.
- Размер клики - число вершин в ней.
- Требуется найти клику максимального размера
- **NP-полная задача!**



Логические правила



“Голова правила является истиной, если все подцели истины.”

Virtual Integration in Subject Mediators

- A problem is formulated in terms of a mediator's schema and mapping of the resource schemas into it
- Is transformed into a set of tasks (queries) to resources registered at the mediator
- Tasks are executed in the resources, results are transferred to the mediator
- Results are combined and transferred to the user

