

# **Рассредоточение реализации приложений в распределенной среде предметных посредников**

Вовченко Алексей

Институт Проблем Информатики РАН

Научный руководитель: Калиниченко Л.А.

Диссертационный семинар

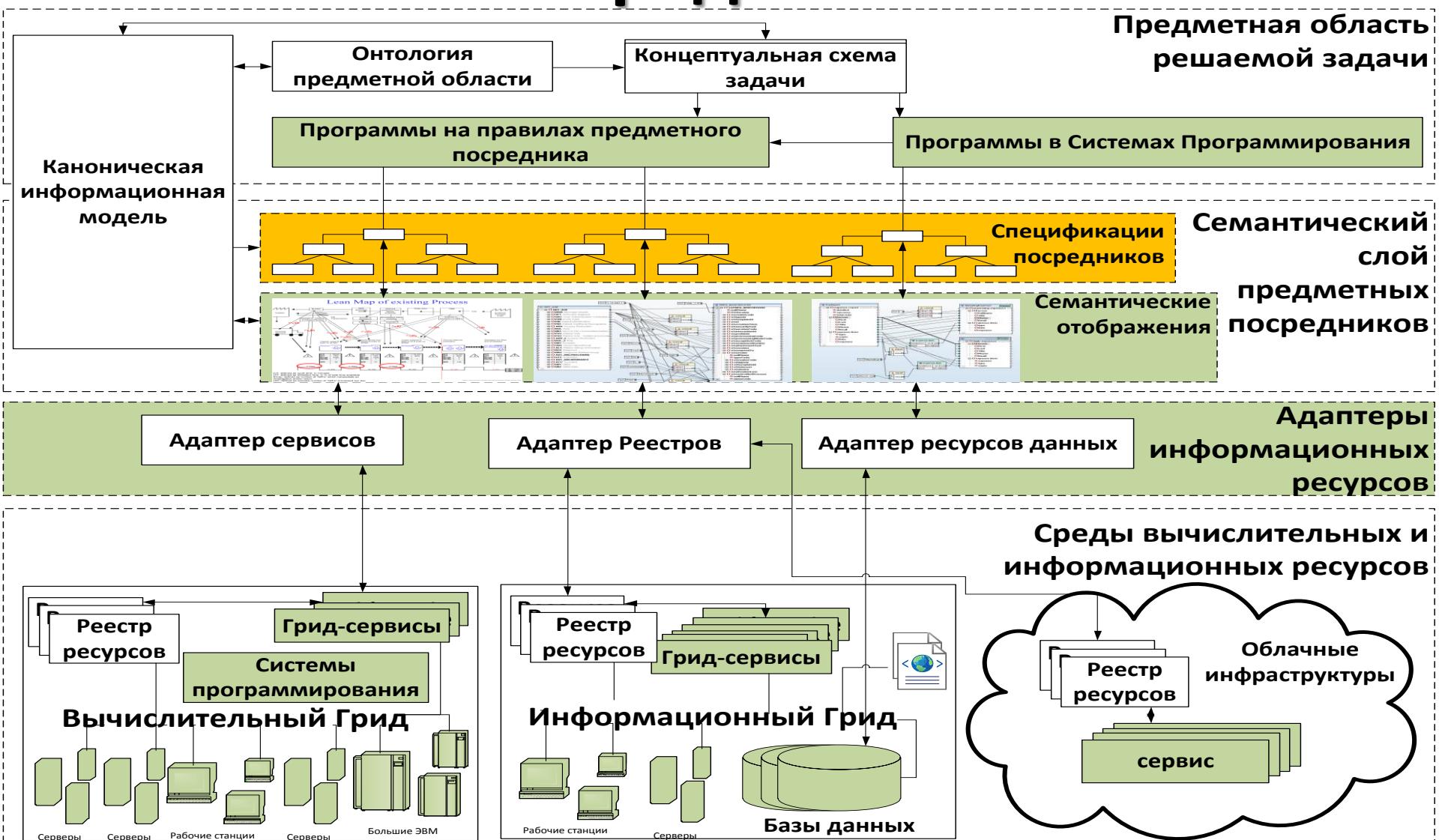
21.10.2011

RCDL'2011

# План

- Среда предметных посредников
- Постановка задачи рассредоточения
- Беглая характеристика других задач работы
- Подход к построению эффективного рассредоточения
- Выводы

# Инфраструктура предметных посредников



# Подход движимый приложениями

- *Подход движимый приложениями включает:* абстрактную спецификацию предметной области в виде онтологических концептов, структур данных, функций, процессов независимо от существующих спецификаций ресурсов
- Спецификация предметной области определяемая специалистами определяет предметный посредник
- Регистрация релевантных ресурсов в посреднике основана на семантическом отображении схемы ресурса в схему посредника
- Среда предметных посредников предоставляет:
  - 1) конструирование канонической информационной модели для унифицированного представления неоднородных онтологий, данных, сервисов и процессов;
  - 2) построение схемы предметного посредника как спецификации предметной области;
  - 3) идентификацию и семантическую интеграцию релевантных ресурсов в посреднике;
  - 4) спецификацию алгоритма решения задачи и ее выполнение.

# Исполнительный слой предметных посредников

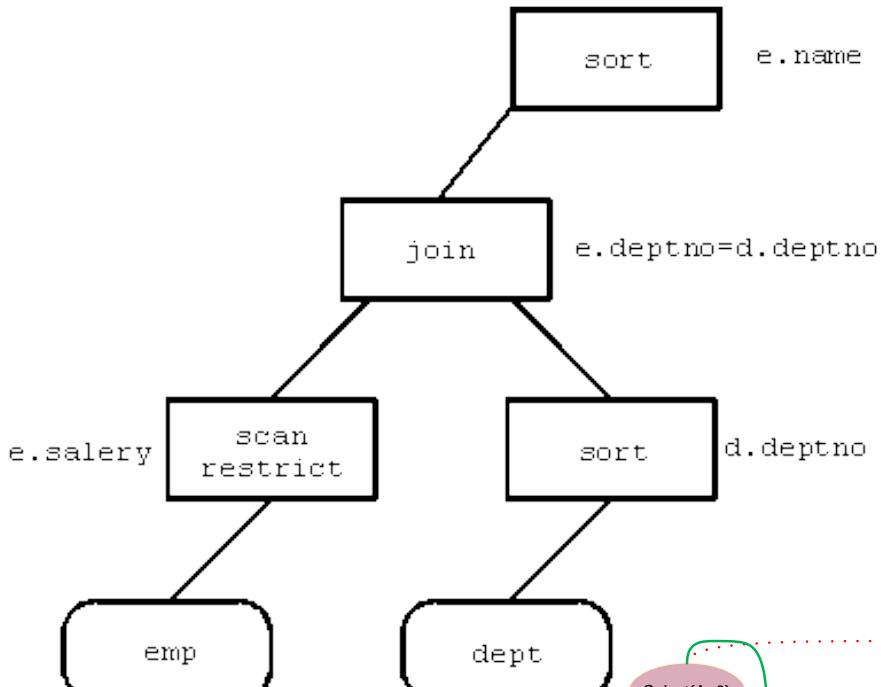


# Постановка задачи

- Разработка подхода к построению эффективного рассредоточения для реализации алгоритма решения задач в среде предметных посредников
  - Много языческая среда программирования (Java, Datalog, PL/SQL, T-SQL, SYNTHESIS) используется для спецификации алгоритма решения задачи
  - Каждый язык программирования(ЯП) может использоваться для реализации различных частей спецификации алгоритма решения задачи
  - Множество возможных вариантов образует модель рассредоточения, представленную как граф зависимостей операций

# Задача планирования

```
select * from emp e dept d
where e.salery>2000
and e.deptno=d.deptno
order by e.name
```



**Class1**  
A:integer  
B:integer  
C:string

Select(A>0)

GetResult(A,C,E,F)

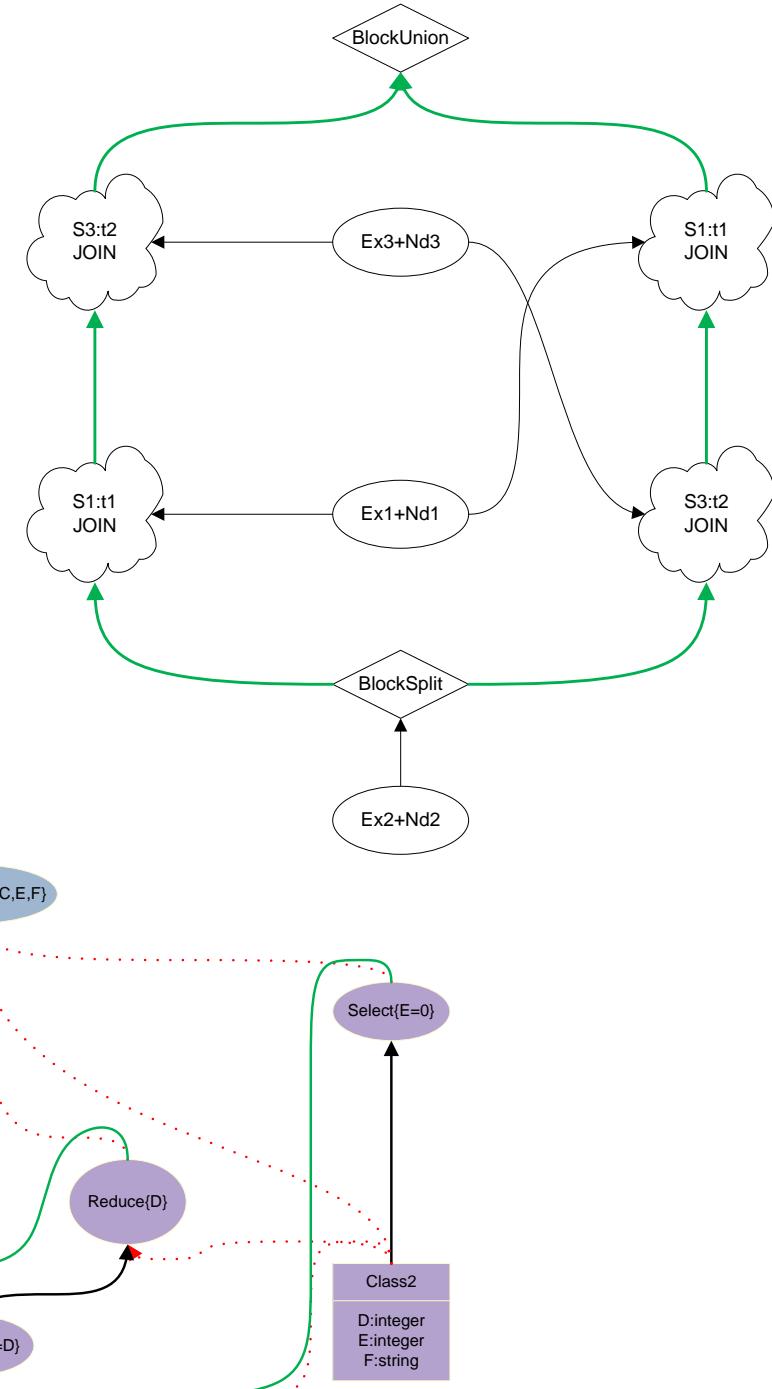
Reduce(B)

Select(B=D)

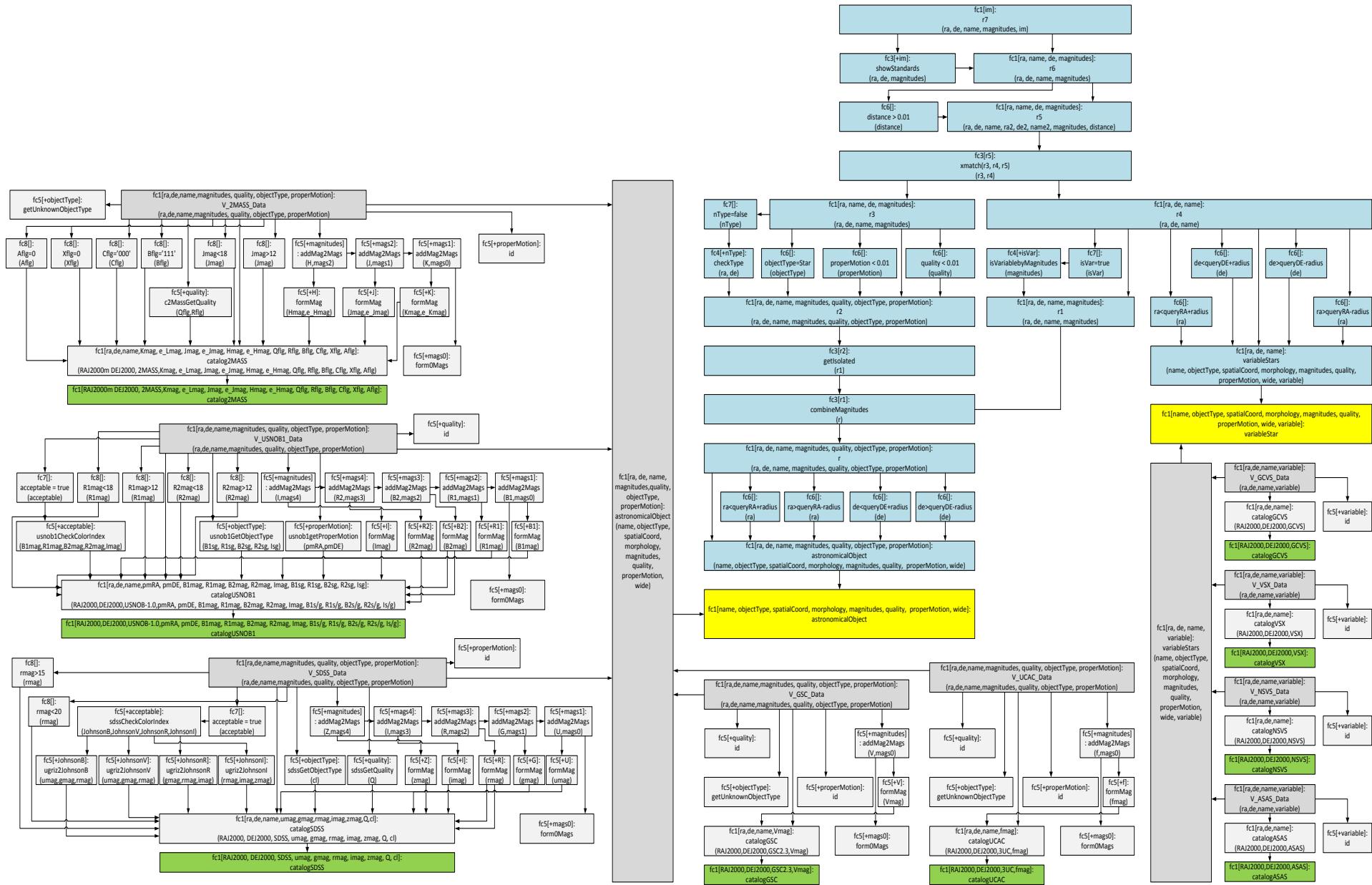
Reduce(D)

**Class2**  
D:integer  
E:integer  
F:string

Select(E=0)

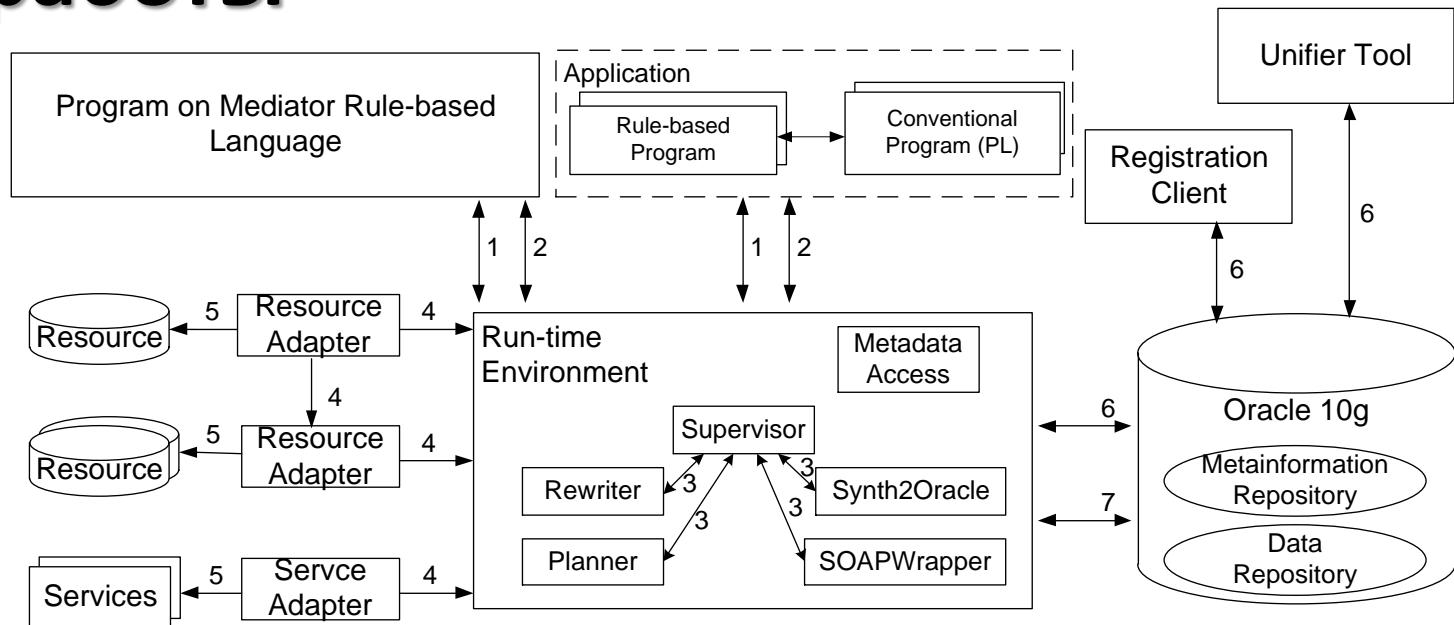


# Пример графа зависимостей операций



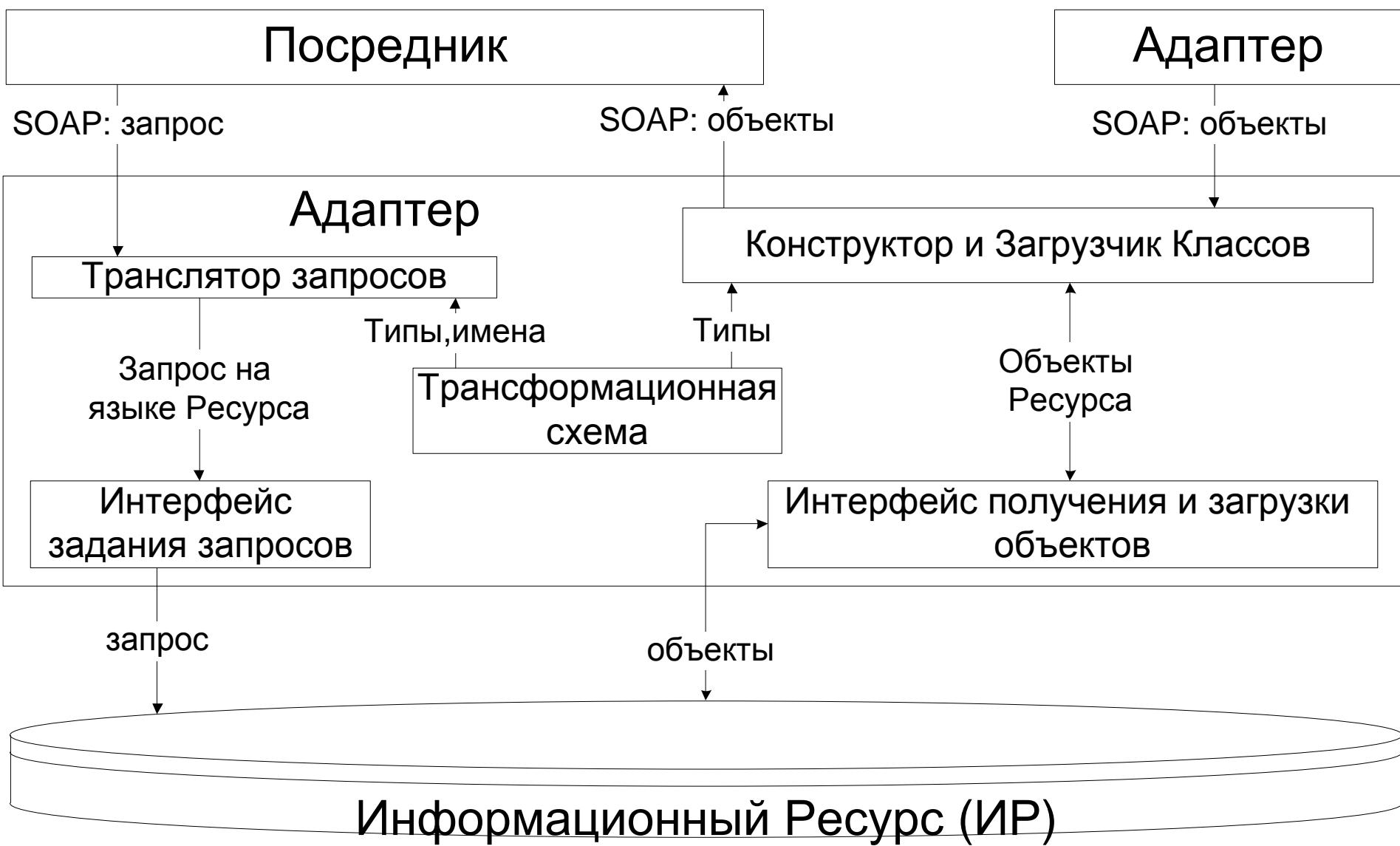
# Обзор Диссертационной работы

Работа может быть разделена на три независимые задачи



- Задача построения рассредоточения
- Задача построения адаптеров информационных ресурсов
- Задача связывания декларативного языка предметных посредников и процедурных языков программирования

# Построение адаптеров (1)



# Построение адаптеров (2)

- Все методы адаптер должны допускать одновременное использование из нескольких потоков (thread safe)
- Поддержка сессий
- Поддержка возможности повторного использования результата запроса
- Поддержка возможности управления загрузкой данных в адаптер, а также возможности материализации запросов в адаптерах
- Поддержка Capabilities адаптеров
- Поддержка оценочных запросов
- Поддержка возможности задания запроса над ограниченной выборкой
- Поддержка статистики
- Программируемые адаптера

# Связывание языка правил с ЯП (1)

	Синтакс	Типы	Связывание	Имена	Коллекции	Долговечность	Запросы	Ссылки	Рефакторинг
Тип-в-Тип		✓	✓*		✓*				
Тип-в-Объект									
ТипЯП-в-Тип			✓*		✓*				
Запрос-в-ЯП	✓		✓*	✓			✓		✓
Запрос-строка									
Запрос-объект	✓								✓
ЯМД-строка								✓	
ЯМД-в-ЯП								✓	
Д-Объекты						✓		✓	
Т-Объекты									
У-Коллекции					✓*				

# Связывание языка правил с ЯП (2)

- предложена система характеристик (features), которыми может характеризоваться и оцениваться подход к сопряжению
- для решения всех проблем несоответствия импеданса подход к сопряжению должен обладать следующим набором характеристик: **Тип-в-Тип, Запрос-в-ЯП, Долговечные Объекты, Универсальные Коллекции** (статический подход)
- ни один из существующих подходов (ODMG 3.0, JDO, JDBC, SQLJ, OCCI, LINQ) полностью не реализует статический подход
- Предлагается одновременно реализовать как статический подход, преодолевающий несоответствие импеданса, но накладывающий ряд ограничений на возможности посредников, так и динамический подход, никак не ограничивающий возможности посредников

# Определения: Граф зависимостей

**Определение 2.2.** Функциональной операцией будем называть конструкцию программы (спецификации), которой можно манипулировать в процессе построения рассредоточения как атомарной единицей.

**Определение 2.3.** Если функциональная операция  $op_1$  среди входных параметров имеет выходные параметры функциональной операции  $op_2$ , то говорят, что операция  $op_1$  зависит от операции  $op_2$ .

**Определение 2.4.** Граф зависимостей функциональных операций или граф зависимостей – это ориентированный, вообще говоря, несвязный граф, без циклов, в вершинах которого расположены функциональные операции.

Вершины в графе именуются также как и операции. Дуги в графе выражают зависимости операций, так что если операция  $op_1$  зависит от операции  $op_2$ , то в графе зависимостей существует дуга, направленная от вершины  $op_1$  к вершине  $op_2$ .

# Определения: Модель рассредоточения

**Определение 2.6.** Моделью рассредоточения будем называть граф зависимостей функциональных операций, для каждой операции которого определены назначения. Также в модели рассредоточения определены возможные назначения для каждой из операций.

**Определение 2.11.** Оценкой эффективности рассредоточения будем называть функционал ET (Execution Time), существенно зависящий от рассредоточения, такой что  $ET = T_{\text{Plan}} + TE$ .

**Определение 2.12.** Минимальным рассредоточением называется такое рассредоточение, при котором оценка эффективности рассредоточения ET будет минимальной среди всех возможных рассредоточений.

# Построение эффективного рассредоточения

- Перестановка операций в модели рассредоточения
- Execution Time (TE) представляет собой **максимум**, если все ресурсы независимы и возможно параллельное выполнение
- В противном случае execution Time (TE) представляет **сумму**, а выполнение последовательно

$$T_E = \max_{n=1..k} (T_n^R + T_n^T) + T^M + T^{PL} \quad T_E = \sum_{n=1}^k (T_n^R + T_n^T) + T^M + T^{PL}$$

- k – число информационных ресурсов,  $T_n^R$  – время выполнения запроса на n-ом ресурсе,  $T_n^T$  – время передачи данных от n-ого ресурса в посредник или другой ресурс,  $T^M$  – время выполнения запроса в посреднике,  $T^{PL}$  – время выполнения программы в языке программирования.
- Эффективное рассредоточение минимизирует ET (определение 2.11)

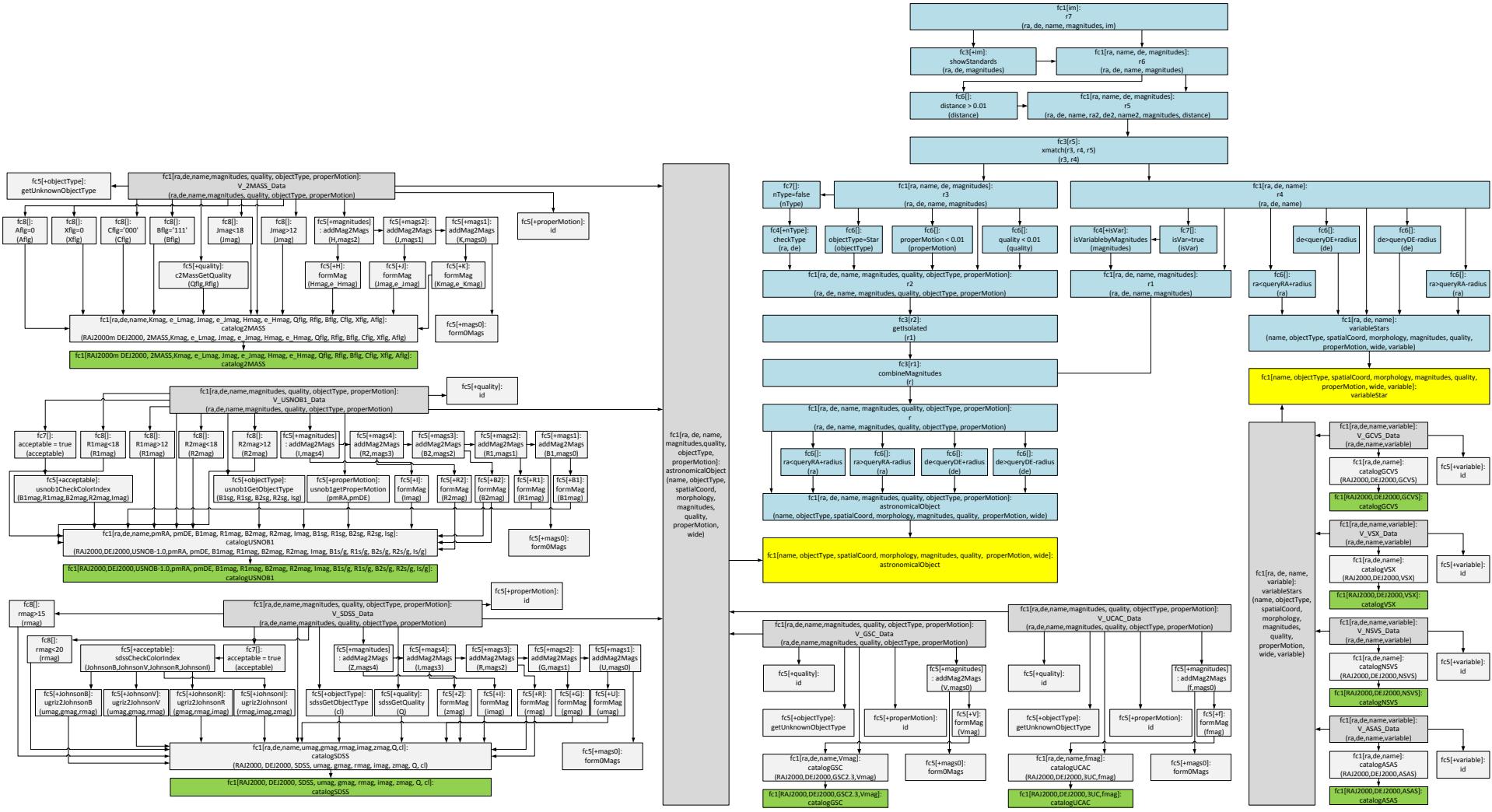
# Подход построения эффективного рассредоточения (общий случай)

- Для каждого ЯП определяются его функциональные возможности в виде функциональных операций выражимых на нем
- По множествам функциональных возможностей определяются классы функциональных операций (универсальные конструкты) для группы языков
- Каждая операция в модели рассредоточения, характеризуется своим функциональным классом, а также назначением для реализации
- Множество зависимостей операций в модели рассредоточения представлено дугами
- Для модели рассредоточения и каждого ЯП реализуются трансляторы прямого и обратного отображения

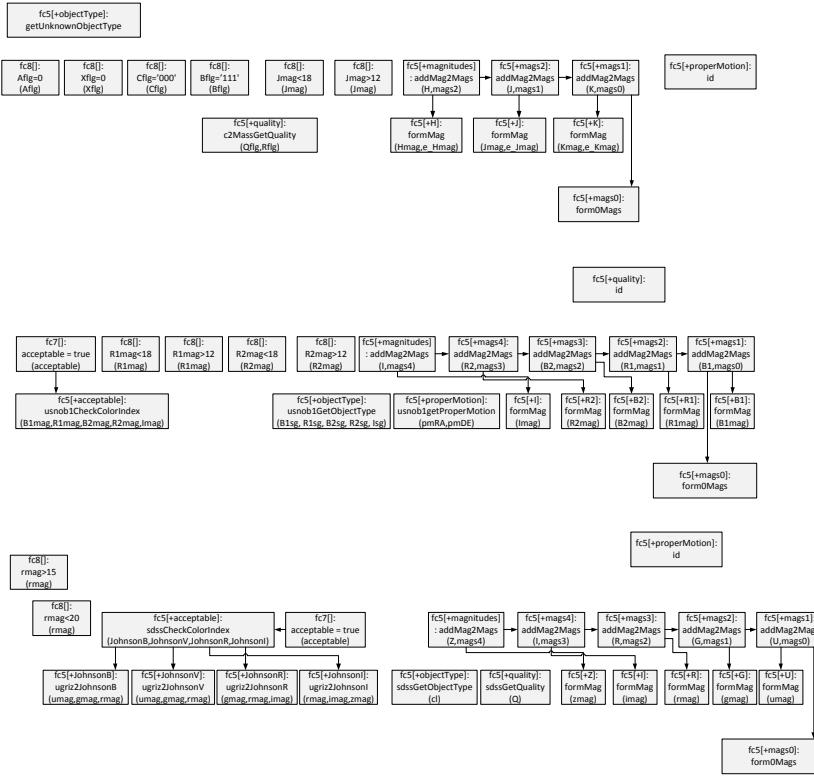
# Подход построения эффективного рассредоточения (общий случай)

- Исходная спецификация алгоритма решения задачи представлена в многоязычной среде отображается в модель рассредоточения
- Автоматически (применение эвристических правил) или полуавтоматически (с участием эксперта) производится перестановка операций в модели рассредоточения
- Для оценки эффективности модель рассредоточения транслируется в спецификацию алгоритма решения задачи в многоязычной среде, после чего и происходит интерпретация (выполнение)
- В случае полного перебора необходимо проверить все варианты рассредоточений
- Для ускорения оценки эффективности, используется выполнение на ограниченной выборке

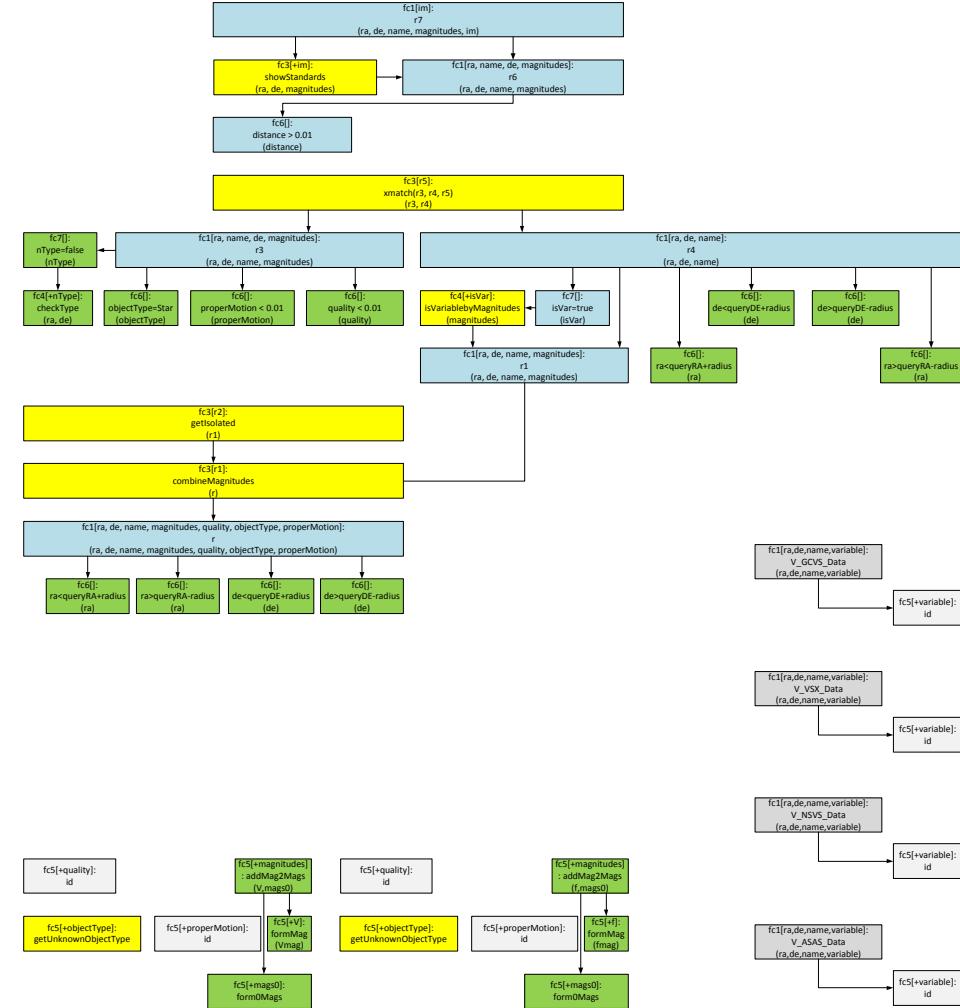
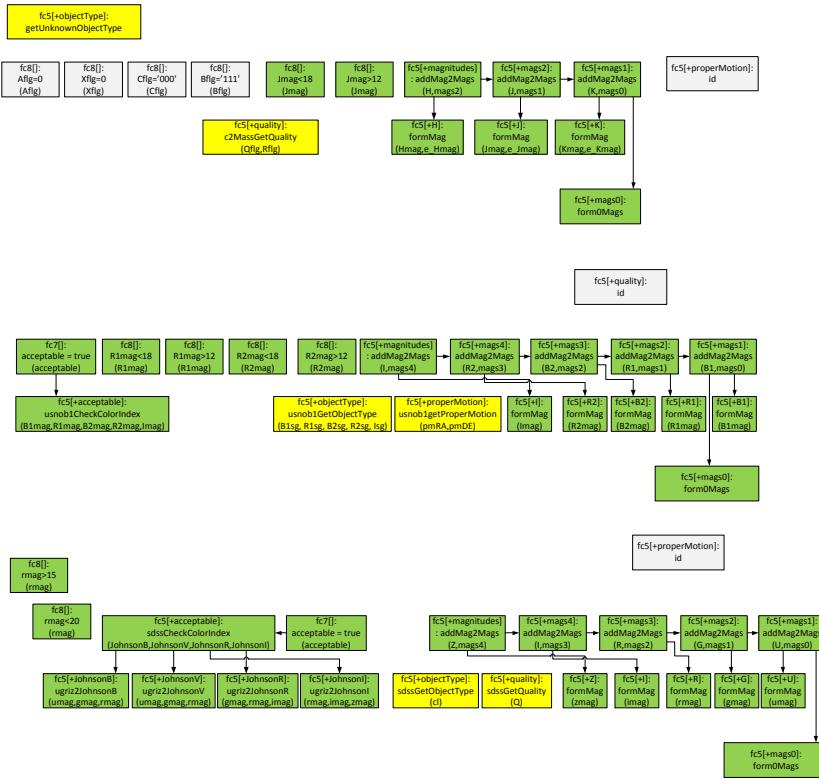
# Пример



# Пример



# Пример



# Пример ВО взглядах

```
// GAV for SDSS
v_SDSS_Data(x/[ra, de, name, magnitudes, quality, objectType, properMotion])
:- CATALOG_SDSS.catalogSDSS(x/[ra:RAJ2000, de:DEJ2000, name:SDSS, umag, gmag, rmag,
imag, zmag, Q, cl])
& ProgramableA.ugriz2JohnsonB(umag, gmag, rmag, JohnsonB)
& ProgramableA.ugriz2JohnsonV(umag, gmag, rmag, JohnsonV)
& ProgramableA.ugriz2JohnsonR(gmag, rmag, imag, JohnsonR)
& ProgramableA.ugriz2JohnsonI(rmag, imag, zmag, JohnsonI)
& ProgramableA.formMag(umag, 0, 'UGRIZ_U', U)
& ProgramableA.formMag(gmag, 0, 'UGRIZ_G', G)
& ProgramableA.formMag(rmag, 0, 'UGRIZ_R', R)
& ProgramableA.formMag(imag, 0, 'UGRIZ_I', I)
& ProgramableA.formMag(zmag, 0, 'UGRIZ_Z', Z)
& ProgramableA.form0Mags(mags0)
& ProgramableA.addMag2Mags(U, mags0, mags1)
& ProgramableA.addMag2Mags(G, mags1, mags2)
& ProgramableA.addMag2Mags(R, mags2, mags3)
& ProgramableA.addMag2Mags(I, mags3, mags4)
& ProgramableA.addMag2Mags(Z, mags4, magnitudes)
& ProgramableA.sdssGetQuality(Q, quality)
& ProgramableA.sdssGetObject Type(cl, objectType)
& ProgramableA.sdssCheckColorIndex(JohnsonB, JohnsonV, JohnsonR, JohnsonI,
acceptable)
& acceptable = true
& id(0, properMotion)
```

# Пример ВО взглядах

```
// GAV for SDSS
  v_SDSS_Data(x/[ra, de, name, magnitudes, quality, objectType, properMotion])
:- CATALOG_SDSS.catalogSDSS(x/[ra:RAJ2000, de:DEJ2000, name:SDSS, umag, gmag, rmag,
imag, zmag, Q, cl])
  & CATALOG_SDSS.getMagnitudes(umag, gmag, rmag, imag, zmag, magnitudes)
  & CATALOG_SDSS.sdssGetQuality(Q, quality)
  & CATALOG_SDSS.sdssGetObject_type(cl, objectType)
  & CATALOG_SDSS.sdssCheckColorIndex(umag, gmag, rmag, imag, zmag, acceptable)
  & acceptable = true
  & id(0, properMotion)
```

# Пример в правилах и ЯП (1)

```
1      r(x/[ra, de, name, magnitudes, objectType, properMotion,  
quality])  
:- astronomicalObject(x1/[ra: spatialCoord.ra, de:  
spatialCoord.de, objectType,  
properMotion, quality, magnitudes])  
& ra < queryRA + radius & ra > queryRA - radius  
& de < queryDE + radius & de > queryDE - radius;  
  
2      combineMagnitudes (r/AstronomicalObject, r1);  
  
3      getIsolated(r1, r2);  
  
4      r3(x/[ra, de, name, magnitudes])  
:- r2(x1/[ra, de, name, objectType, properMotion, quality,  
magnitudes])  
& checkType(ra, de, 'G', nType) & nType = false  
& objectType = Star  
& properMotion < 0.01  
& quality < 0.01;
```

# Пример в правилах и ЯП (2)

```
5      r4(x/[ra, de, name])
:- r1(x1/[ra, de, name, magnitudes])
& isVariablebyMagnitudes(ra, de, iVar) & iVar = true;

6      r4(x/[ra, de, name])
:- variableStar(x1/[ra: spatialCoord.ra, de: spatialCoord.de,
name]);
& ra < queryRA + radius & ra > queryRA - radius
& de < queryDE + radius & de > queryDE - radius;

7      xmatch(r3, r4, r5);

8      r6(x/[ra, de, name magnitudes])
:- r5(x1/[ra, de, name, magnitudes, distance])
& distance > 0.01;

9      r7(im/Image)
:- r6(x/ra, de, name, magnitudes])
& showStandards(ra, de, radius, magnitudes, im);
```

# Пример в правилах и ЯП (3)

```
1      # Synchronized Parallel
Aladin.getImage(queryRA, queryDE, radius);

2      # Synchronized Parallel
Set<String> galaxies = GetNedGalaxies(queryRA, queryDE, radius);

3      # Synchronized Parallel
#SYNTHESIS
Class r(x/[ra, de, name, magnitudes])
:- astronomicalObject(x1/[ra: spatialCoord.ra, de:
spatialCoord.de, name, objectType,
properMotion, quality, magnitudes])
& ra < queryRA + radius & ra > queryRA - radius
& de < queryDE + radius & de > queryDE - radius;
& objectType = Star
& properMotion < 0.01
& quality < 0.01;
#SYNTHESIS
```

# Пример в правилах и ЯП (4)

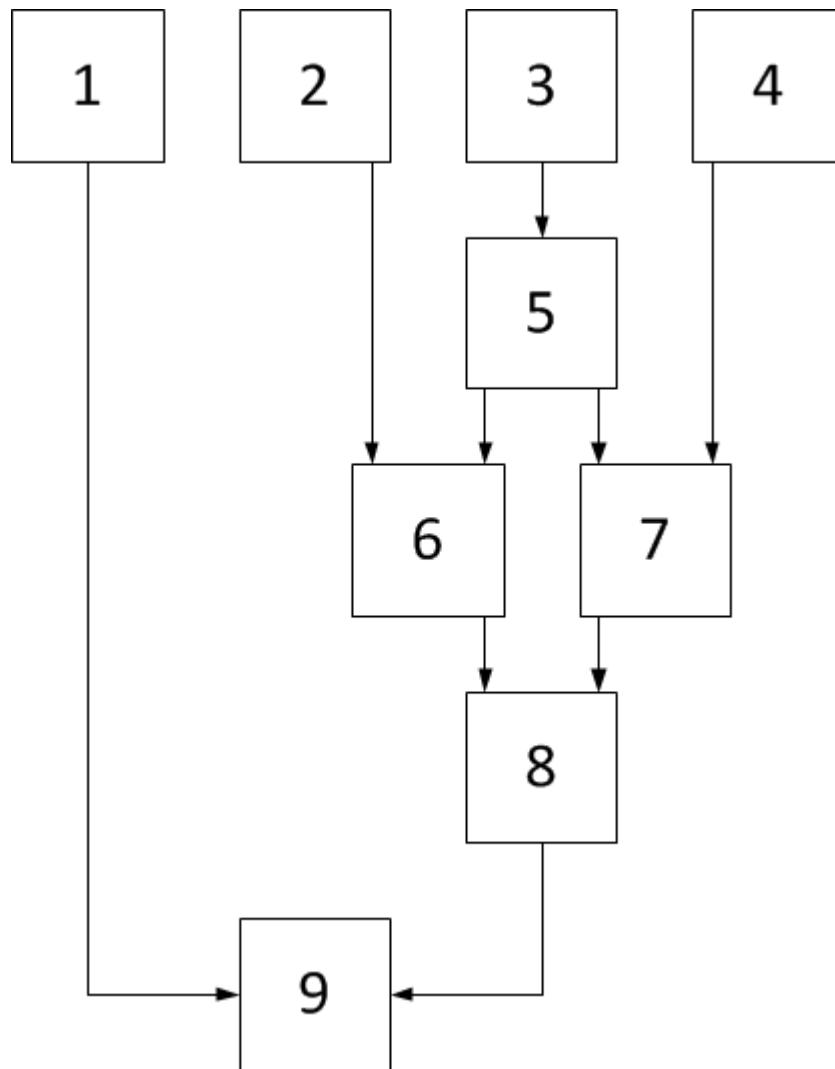
```
4      # Synchronized Parallel  
#SYNTHESIS  
Class r4(x/[ra, de, name])  
:- variableStar(x1/[ra: spatialCoord.ra, de: spatialCoord.de,  
name]);  
& ra < queryRA + radius & ra > queryRA - radius  
& de < queryDE + radius & de > queryDE - radius;  
      #SYNTHESIS  
  
5      Set<Class_r> r1 = combineMagnitudes (r);
```

# Пример в правилах и ЯП (5)

```
6      # Synchronized Parallel  
Set<Class_r> r2 = getIsolated(r1);  
      Set<Class_r> r3 = getNotGalaxie(r2, galaxies);  
  
7      # Synchronized Parallel  
Set<Class_r4> r5 = getVariablebyMagnitudes(r1);  
      r5.add(r4);  
  
8      Set<Class_r> r6 = xmatch(r3, r5, 0.01);  
  
9      Aladin.putStandardsOnImage(r6);
```

**подобное изменение программы позволило улучшить среднее время выполнения почти в 10 раз**

# Пример в правилах и ЯП (6)



# Выводы

- Представлена общая постановка задачи рассредоточения
- Даны понятия, необходимые для постановки задачи рассредоточения:
  - модель рассредоточения,
  - рассредоточенная реализации или рассредоточение,
  - минимальное рассредоточение.
- Для обобщенной среды предметных посредников определены классы функциональных операций
- На основе анализа возможности перестановок операций, разработан набор экспертных правил для оптимизированного построения эффективного рассредоточения
- Разработан метод полного перебора для автоматического построения эффективного рассредоточения
- Разработан метод полуавтоматического построения эффективного рассредоточения

# Научная новизна диссертационной работы

- подход к построению квази-минимального рассредоточения реализации алгоритма решения задач, позволяющий генерировать варианты рассредоточений, а также оценивать их эффективность;
- предложен подход сопряжения систем программирования с предметными посредниками, включающий реализаций как статического подхода, ориентированного на решения максимального числа проблем не соответствия импеданса, так и динамического подхода, ориентированного на предоставление максимальных возможностей для разработчиков;
- разработана архитектура программируемых адаптеров информационных ресурсов, обеспечивающая эффективное рассредоточение реализации программ посредников среди ресурсов и адаптеров, а также подход к полуавтоматическому созданию адаптеров.

Спасибо за внимание!  
Вопросы?