

Персонализация предметного посредника в области регуляции экспрессии генов*

© Д.О.Брюхов¹, В.Н.Захаров¹, Л.А.Калиниченко¹, С.А.Ступников¹, Н.Л.Подколотный^{2,3}

¹ИПИ РАН, г. Москва

²ИЦиГ СО РАН, г. Новосибирск

³ИВМиМГ СО РАН, г. Новосибирск

Аннотация

В данной статье описывается процесс создания персонализированных коллекций в области регуляции экспрессии генов. Он включает в себя определение как классы и типы посредника выражаются через классы и типы персонализированной коллекции.

Этот процесс основан на методе композиционного проектирования, при котором персонализированная коллекция является спецификацией требований, а федеративная схема посредника – спецификацией ресурса.

1 Введение

Для организации эффективных исследований в области биоинформатики необходимо обеспечить доступ и выполнение сложных запросов к большому числу распределенных информационных ресурсов, включающих разнообразную информацию о первичной и пространственной структуре макромолекул ДНК, РНК, белков и их комплексов, а также об особенностях их взаимодействия друг с другом.

Для обеспечения семантической интеграции многочисленных неоднородных и независимых информационных ресурсов требуется организовать глобальное унифицированное представление включаемых информационных ресурсов и предоставляемых услуг. Предлагается создать специальный промежуточный слой, состоящий из предметных посредников [3], обеспечивающих унифицированный интерфейс запросов к многочисленным источникам данных и освобождающих пользователя от необходимости находить подходящую коллекцию, осуществлять в ней поиск требуемой информации и вручную объединять информацию, полученную из различных коллекций.

Основными функциями предметного посредника являются:

- регистрация разнообразных электронных коллекций в данной предметной области [4];
- формирование и ведение интегрированной базы метаинформации для данной предметной области;
- поддержка персонализации информации для определенных групп пользователей.

Метаинформация предметного посредника разбивается на 3 уровня: локальный, федеративный и персонализированный. Локальный уровень содержит информацию, соответствующую каждой коллекции в форме канонической модели посредника. Федеративный уровень задает федеративные схемы, предназначенные для унифицированного доступа к множественным коллекциям с взаимосвязанными данными как к предметной области.

Персонализированный уровень [2] метаинформации требуется для представления персонализированных виртуальных электронных библиотек, создаваемых для конкретных пользователей или групп пользователей. Для формулирования требований к персонализированным электронным библиотекам требуются специальные средства моделирования. Эти средства основаны на канонической модели посредника.

Проектирование персонализированных электронных библиотек планируется как процесс поиска спецификаций федеративного уровня посредника и их фрагментов, композиция которых может служить уточнением (refinement) требований к персонализированной электронной библиотеке.

2 Описание посредника для регуляции экспрессии генов

Описание посредника включает описание онтологических понятий для области регуляции экспрессии генов и описание федеративной схемы. Онтологическая модель предметной области (регуляции транскрипции генов эукариот) включает онтологическое определение системы понятий, тезаурусов и словарей.

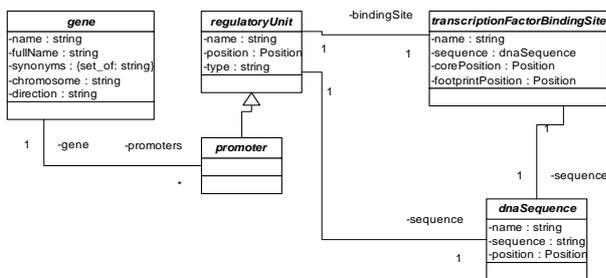


Рис.1 Фрагмент UML спецификации посредника

Пример онтологического понятия, представленного в библиотеке онтологий, и соответствующего атрибуту *name* типа *regulatoryUnit*:

Name "regulatoryUnit_name"

Definition "name of the extended DNA sequence (promoter, enhancer, silencer, etc.), that alter the levels of gene transcription and can determine spatial patterns of gene expression"

Схема посредника включает описание источников данных, типов экспериментов и порождаемых ими типов данных, методов анализа данных и построения теории предметной области и модели.

На Рис.1 представлен фрагмент спецификации схемы посредника для области регуляции экспрессии генов, содержащий спецификации типов, их атрибутов, ассоциаций между ними и ограничений, налагаемых на такие типы.

3 Пример описания персонализированной коллекции

Ресурсы персонализированной информации могут формироваться на основе различных подходов, например:

- выбор подмножества метаинформации федеративного уровня посредника, которая отражает интересы соответствующей группы пользователей;
- введение специальных словарей (онтологий), более точно соответствующих конкретным группам пользователей и обеспечивающих отображение словарей в тезаурус посредника;
- введение взглядов над федеративным уровнем, отражающих информационные потребности конкретных групп пользователей;
- использование существующих коллекций в качестве персонализированного представления над посредником.

В качестве примера рассмотрим случай, когда в качестве персонализированной коллекции используется существующая коллекция EPD, широко используемая в области молекулярной биологии. Использование существующих коллекций позволяет пользователям этой коллекции использовать возможности посредника в привычной для них форме и без дополнительного изучения структуры посредника.

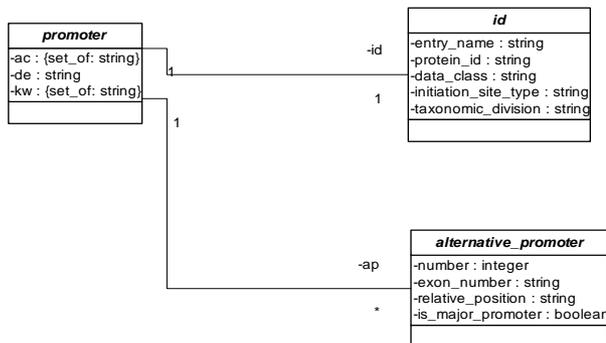


Рис.2 Фрагмент UML спецификации коллекции EPD

На Рис. 2 приведен фрагмент UML спецификации этой коллекции. Также, эта коллекция содержит список онтологических понятий не связанных с онтологическими понятиями посредника. Пример онтологического понятия, соответствующего атрибуту *entry_name* типа *Id*:

Name "entry_name"

Definition "unique entry identifier which obeys rigorous naming conventions"

4 Конструирование персонализированных коллекций

Процесс создания персонализированных коллекций включает в себя установление соответствия между онтологическими понятиями персонализированной коллекции и посредника, и определение как классы и типы посредника выражаются через классы и типы персонализированной коллекции. Классы посредника моделируются как множество экземпляров (объектов) типа экземпляра классов, а их описание в терминах персонализированной коллекции определяет ограничения, которым должны удовлетворять экземпляры данного класса. Этот процесс основан на методе композиционного проектирования [1,2], при котором персонализированная коллекция является спецификацией требований, а федеративная схема посредника – спецификацией ресурса.

Общая схема процесса персонализации выглядит следующим образом:

1. идентификация релевантных классов федеративной схемы

Для каждого класса персонализированной коллекции формируется список возможно релевантных классов федеративной схемы посредника, на основании интеграции онтологических понятий, соответствующих данным классам. Т.е. если между онтологическими понятиями, соответствующими классу коллекции и классу посредника, установлена позитивная связь или связь понятие-подпонятие, то данный класс посредника добавляется в список возможно релевантных классов. Аналогично формируются список релевантных типов для типов экземпляров

классов коллекций, список релевантных атрибутов (функций) для каждого атрибута (функции) типа.

2. конструирование наиболее общих редуктов

Для типа экземпляра каждого найденного релевантного класса посредника выполняется конструирование наиболее общего редукта между ним и типом экземпляров заданного класса персонализированной коллекции. В наиболее общий редукт попадают как общие атрибуты, так и атрибуты, соответствующие атрибутам типа посредника, которые могут быть получены из атрибутов типа коллекции. Для этого осуществляется поиск и разрешение структурных конфликтов между спецификациями коллекции и схемы посредника.

3. конструирование частичных взглядов

Для каждого релевантного класса персонализированной коллекции конструируется частичный взгляд, задающий ограничения в терминах классов посредника, которые должны удовлетворяться значениями соответствующих наиболее общих редуктов типов экземпляров класса коллекции.

4. композиция частичных взглядов

Для данного класса персонализированной коллекции конструируется взгляд как композиция частичных взглядов над его релевантными классами посредника. При этом конструируется композиция наиболее общих редуктов типов коллекции, полученных для типов экземпляров всех релевантных классов посредника. Полученный взгляд является материализованным взглядом, задающим формулу выражения класса персонализированной коллекции через классы посредника.

Примером результата процесса персонализации для коллекции EPD над посредником в области регуляции экспрессии генов являются следующий взгляд:

```
{v_promoter_promoter;  
  in: class;  
  class_section: {  
    lav: invariant, {  
      subseteq (v_promoter_promoter(p),  
        promoter(p/R_Promoter_Promoter)) }  
    };  
  instance_section: CR_Promoter_Promoter  
}
```

На основе полученных взглядов работает процедура переписывания запросов в терминах персонализированной коллекции (EPD) в запрос в терминах предметного посредника.

Литература

- [1] Брюхов Д.О., Калинин Л.А., Скворцов Н.А. Композиционное проектирование информационных систем: методы и средства. *Системы и средства информатики. Вып. 10*, М.: Наука, 2000, с. 128-147
- [2] Briukhov D.O., Kalinichenko L.A., Skvortsov N.A. Personalization through Specification Refinement

and Composition. In *Proc. of the Joint DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries*, Ireland, June 2001

- [3] Kalinichenko L.A., Briukhov D.O., Skvortsov N.A., Zakharov V.N. Infrastructure of the subject mediating environment aiming at semantic interoperability of heterogeneous digital library collections. *Proc. of the Second Russian National Conference on "Digital Libraries: Advanced Methods and Technologies, Digital Collections*, Sep. 26-28, 2000, Protvino.
- [4] Л.А.Калининченко, Д.О.Брюхов, В.Н.Захаров, О.А.Подколотная, Н.Л.Подколотный Проблемы создания предметного посредника для регуляции экспрессии генов. В *Трудах Четвертой Всероссийской конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции"*, Дубна, 15-17 октября 2002 г.

Personalization of subject mediator in the gene expression regulation domain

D.O. Briukhov, V.N. Zakharov, L.A. Kalinichenko, S.A. Stupnikov, N.L. Podkolodny

This paper describes the process of development of personalized digital collections in the gene expression regulation domain. It consists in definition how classes and types of the mediator are expressed in classes and types of the personalized collection.

This process is based on the compositional design method where the personalized collection is a specification of requirements and the mediator is a specification of resource.

* Работа частично поддержана РФФИ (грант 01-07-90084)