

Н.В. Шилов, Р.И. Идрисов, А.А. Акинин, А.В. Зубков

# РАЗРАБОТКА ПОРТАЛА ЗНАНИЙ О КЛАССИФИКАЦИИ КОМПЬЮТЕРНЫХ ЯЗЫКОВ

DICR'2010

Новосибирск, 30 ноября — 3 декабря 2010 г.



## Проект: «Портал знаний о классификации компьютерных языков»

*Компьютерный язык* –

искусственный язык, созданный для “обработки информации”, т. е. для представления, манипуляции (в том числе передачи), управления и обработки данными.

*Классификация* –

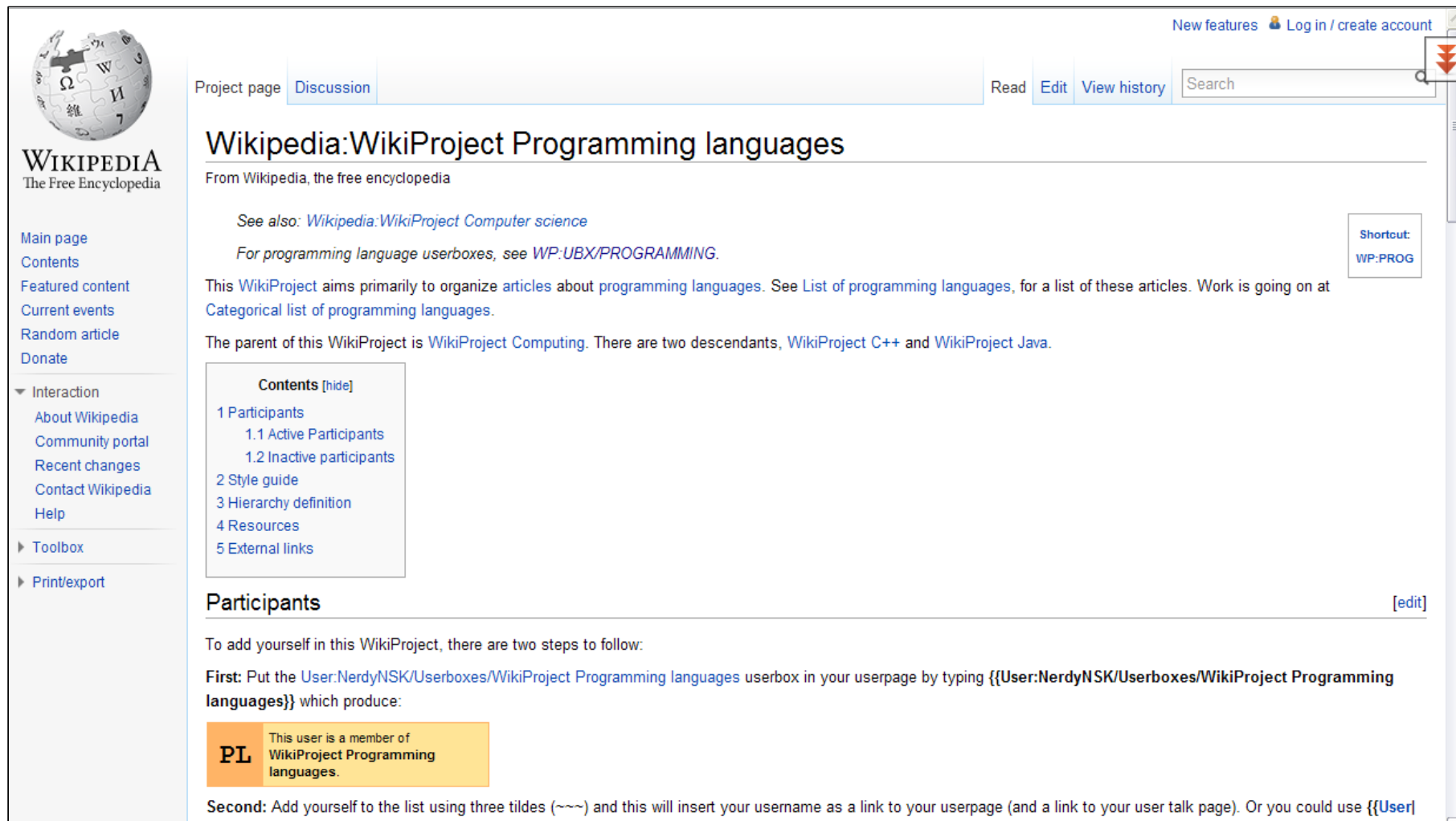
метод выделения, идентификации классов объектов и отношений между классами, а также навигации между классами (в том числе через отношения между классами).

*Цели проекта:*

- IT-образование и мировоззрение (онтология компьютерных языков),
- помощь в выборе языков для IT-проектов,
- направления разработки новых языков.



# Актуальность: WikiProject Programming Languages



The screenshot shows the Wikipedia page for the WikiProject Programming Languages. At the top left is the Wikipedia logo and the text "WIKIPEDIA The Free Encyclopedia". Below it is a navigation menu with links like "Main page", "Contents", "Featured content", "Current events", "Random article", "Donate", "Interaction", "About Wikipedia", "Community portal", "Recent changes", "Contact Wikipedia", "Help", "Toolbox", and "Print/export". At the top right, there are links for "New features", "Log in / create account", and a search bar. The main content area has a "Project page" tab set to "Discussion" and buttons for "Read", "Edit", and "View history". The title is "Wikipedia:WikiProject Programming languages". Below the title is the text "From Wikipedia, the free encyclopedia". There is a "See also" section with a link to "Wikipedia:WikiProject Computer science" and a note about programming language userboxes. The main text describes the project's aim to organize articles about programming languages and lists its parent project, WikiProject Computing, and its descendants, WikiProject C++ and WikiProject Java. A "Contents" box lists sections: 1 Participants (with sub-sections 1.1 Active Participants and 1.2 Inactive participants), 2 Style guide, 3 Hierarchy definition, 4 Resources, and 5 External links. Below the contents is a "Participants" section with an "[edit]" link. It explains how to add oneself to the list using a userbox and provides a code example: `{{User:NerdyNSK/Userboxes/WikiProject Programming languages}}`. A yellow box shows the resulting userbox: **PL** This user is a member of WikiProject Programming languages. The final instruction is to add oneself to the list using three tildes (~~~) to create a link to the user's page and talk page.

## Проект: «Портал знаний о классификации компьютерных языков»

- Специализированный инструмент (ориентированный на предметную область)
- Веб-приложение
- В основе — формальная онтология
- Свободное пополнение редактирование содержания
- Темпоральность
- Обработка пользовательских запросов (на основе проверки модели)
- Отчуждаемость данных (за счёт представления данных на языке OWL)



## Проект: «Портал знаний о классификации компьютерных языков»

Основные элементы:

- Объекты (компьютерные языки)
- Элементарные классы (базовые парадигмы)
- Определяемые классы (средствами логики описаний понятий)
- Элементарные отношения между объектами
- Производные отношения (средствами логики описаний ролей)
- Статические атрибуты (аннотации объектов и ролей)
- Темпоральные аннотации (времени правки)
- Авторство
- «Законы» предметной области (представленные Т-боксом)



## Поддержка аксиом и разрешения запросов

Синтаксис языка запросов, описания классов, отношений и аксиом:

$$C ::= \top \mid \perp \mid EC \mid \neg C \mid (C \sqcup C) \mid (C \sqcap C) \mid (\forall R. C) \mid (\exists R. C) \mid (C \uparrow R) \mid (C \downarrow R)$$
$$R ::= \top \mid \perp \mid \text{Id} \mid ER \mid \neg R \mid (R \sqcup R) \mid (R \sqcap R) \mid (R \circ R) \mid R^{-1} \mid R^+ \mid R^*$$
$$A ::= C \sqsubseteq C \mid C \equiv C$$

где  $C$  — класс,  $R$  — отношение,  $A$  — аксиома,

$EC$  — элементарный класс,  $ER$  — элементарное отношение.

Запросами к онтологии являются классы ( $C$ ), построенные в соответствии с указанным синтаксисом.



## Как совместить логический вывод из аксиом с выводом при помощи проверки модели?

Предложенное (от безысходности) решение:

**Никак.** Данные должны соответствовать аксиомам. В противном случае наполнение онтологии считается противоречивым.

Проверка на непротиворечивость «законов» и фактов

- при создании новых аксиом,
- при пополнении базы фактов.



# Как реализовать модель открытого мира?

Предложенное (от безысходности) решение:

## Дуальные элементарные классы и отношения.

Элементарный класс:

- $C$  — множество объектов, про которые известно, что они входят в класс
- $\sim C$  — множество объектов, про которые известно, что они не входят в класс

Элементарное отношение:

- $R$  — пары объектов, про которые известно, что они связаны отношением
- $\sim R$  — пары объектов, про которые известно, что они не связаны отношением

Встроенная схема аксиом — дуальные классы и отношения не пересекаются:

$$\sim C \cap C \equiv \perp$$

$$\sim R \cap R \equiv \perp$$





## Какое выбрать внутреннее представление данных?

Онтология должна быть рассчитана на:

- Тысячи — десятки тысяч сущностей
- Десятки — тысячи пользователей
- Сохранение истории изменений и их авторов

Рассматриваемые варианты:

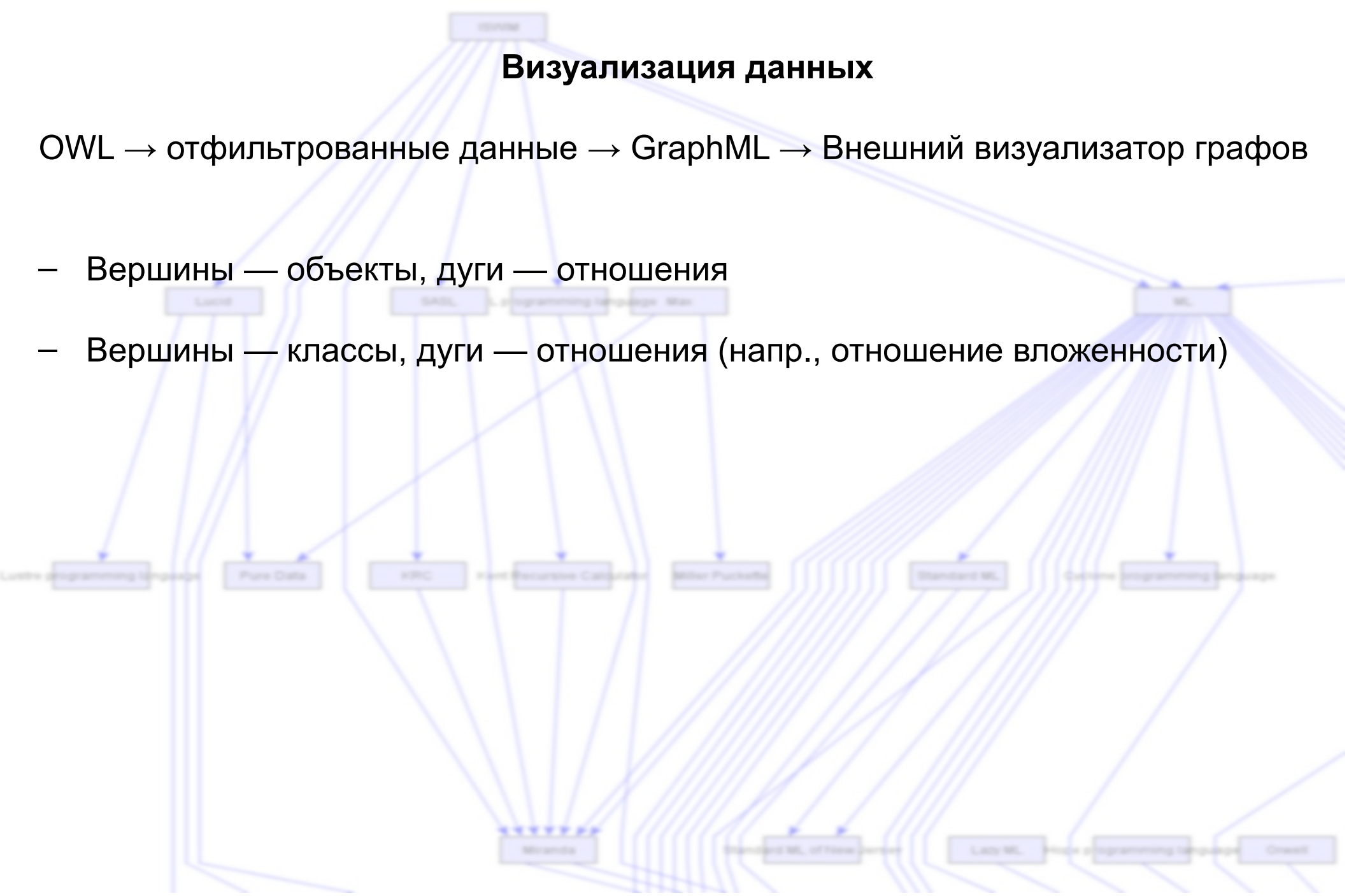
- OWL
- Реляционная база данных



## Визуализация данных

OWL → отфильтрованные данные → GraphML → Внешний визуализатор графов

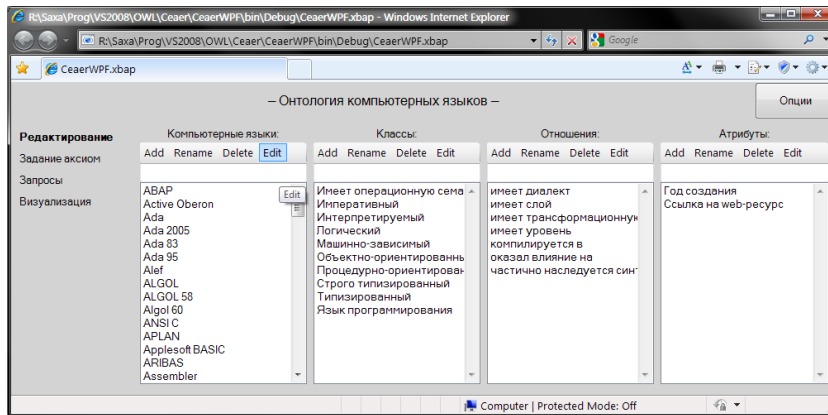
- Вершины — объекты, дуги — отношения
- Вершины — классы, дуги — отношения (напр., отношение вложенности)



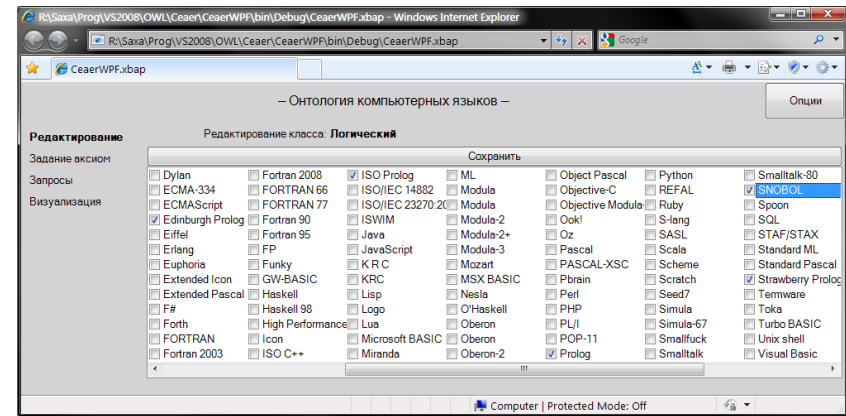
# Реализация интерфейса

## Веб-приложение Microsoft Windows Presentation Foundation (WPF Web Application)

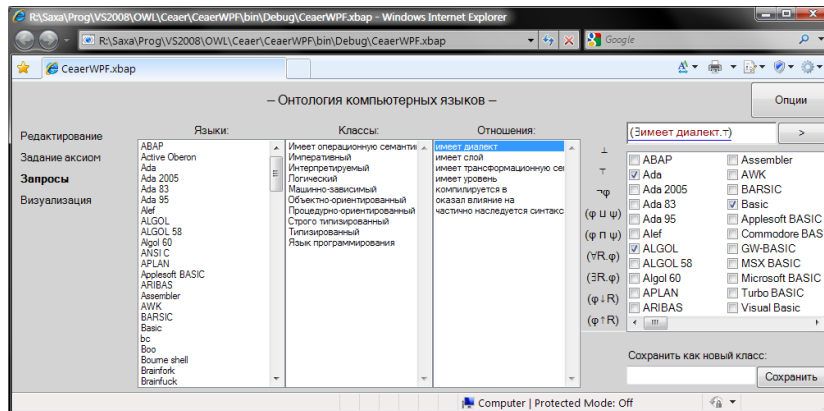
Редактирование списков основных объектов онтологии:



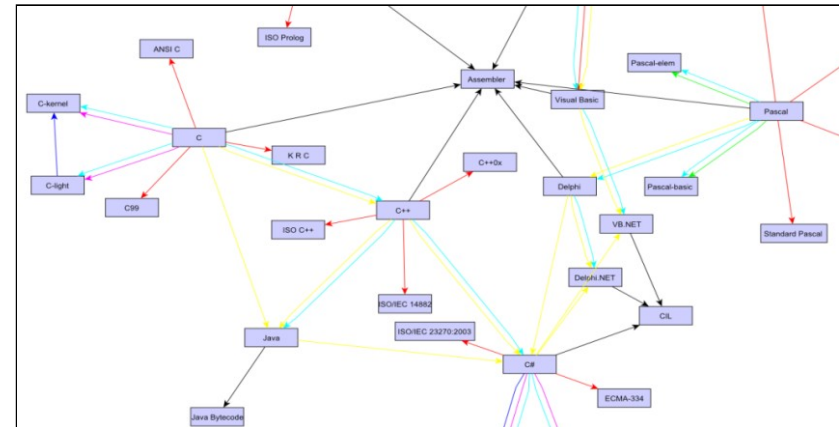
Редактирование элементарного класса:



Обработка запросов:



Визуализация фрагмента онтологии в виде графа:



# Источники данных

## Прогopedia (alpha) — энциклопедия компьютерных языков

<http://progopedia.ru>

Progopedia *alpha*

- Главная
- О сайте
- Парадигмы
- Типизации
- Новые статьи
- Свежие правки
- Блог
- Google-группа

Топ 10 пользователей:

- kit
- WellRus
- Nickolas
- helions8
- octagram
- kemiisto
- euginekosenko
- Antoni
- 218
- cosmicdustman

Регистрация

Google® Пользовательский поиск

Поиск

Progopedia in English

В закладки: ■ 🌟 🇺🇸 🇺🇦

Touch Pro 2 Accessories

### Энциклопедия языков программирования

Добро пожаловать!

Progopedia — **свободная** энциклопедия языков программирования: все тексты доступны на условиях лицензии **GNU Free Documentation License**, а редактором может стать каждый желающий.

На данный момент в энциклопедии языков программирования **Progopedia.ru** описано:

- языков программирования: 70,
- диалектов: 66,
- реализаций: 207,
- версий: 545.

### Примеры программ

- Hello, World! (примеров: 40)
- Квадратное уравнение (примеров: 11)
- Факториал (примеров: 56)
- Числа Фибоначчи (примеров: 40)

### Список всех языков, диалектов, реализаций и версий

Свернуть все | Развернуть все

- ABAP
- Ada (диалекты: Ada 2005, Ada 83, Ada 95)
- Alef
- ALGOL (диалекты: ALGOL 58, Algol 60)
- APLAN
- ARIBAS
- Assembler
- AWK
- BARSIC
- Basic (диалекты: Applesoft BASIC, Commodore BASIC, GW-BASIC, MSX BASIC, Microsoft BASIC, Turbo BASIC, Visual Basic)

## Freebase — the definitive open database of people, places, and things

<http://freebase.com>

Freebase

Explore Apps Developers Sign In

Search 11 million intercor

Start typing to get some suggestions

- Explore
- Apps
- Developers
- Help

[Sign In or Sign Up](#)

## Programming language

### Embed this Topic



A programming language is an artificial language designed to express computations that can be performed by a machine, particularly a computer. Programming languages can be used to create programs that control the behavior of a machine, to express algorithms precisely, or as a mode of human communication. Many programming languages have some form of written specification of their syntax (form) and semantics (meaning). Some languages are defined by... [More](#)

[Read article at Wikipedia](#)

### Software In Genre:

- Smalltalk
- Perl
- Ruby
- PHP

## Источники данных

Basic data extraction instruments was created for Progopedia and Freebase

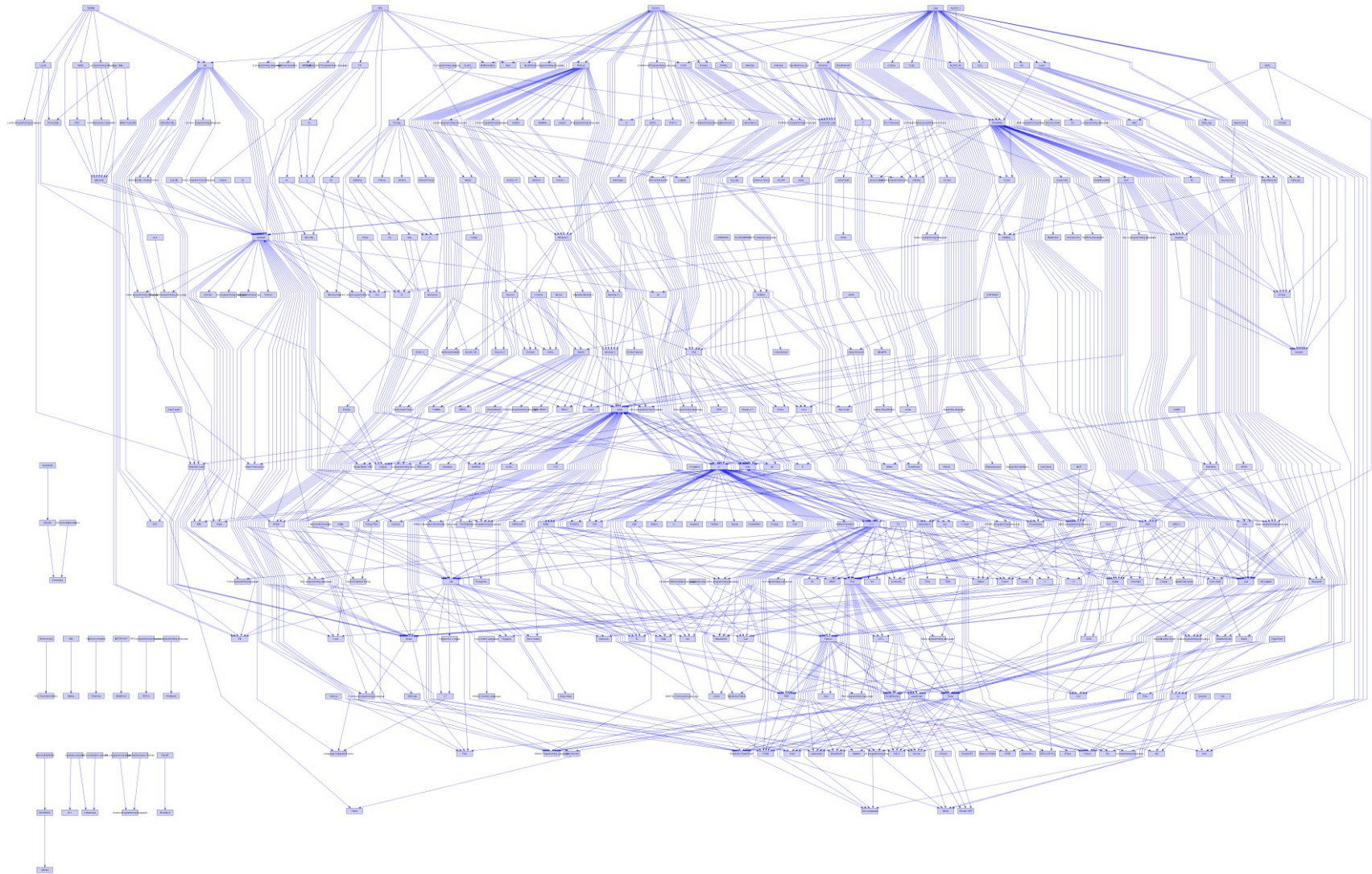
May 2010 statistics:

	Progopedia	Freebase	Progopedia and Freebase
– языков	131	1181	1238
– классов	35	46	72
– отношений	2	2	2
– атрибутов	3	1	3
– «законов»	—	—	—



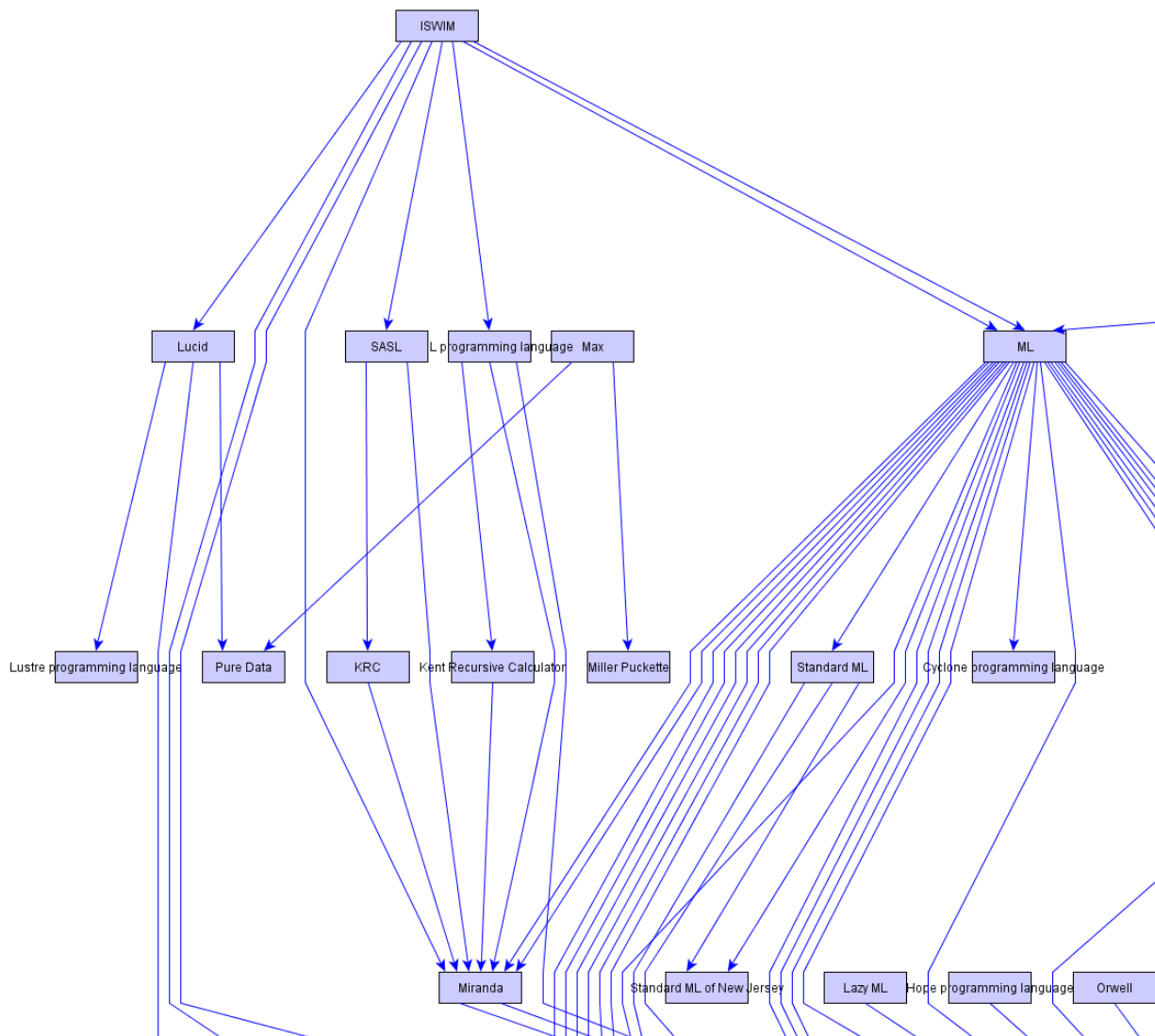
# Источники данных: визуализация

## Граф влияния компьютерных языков



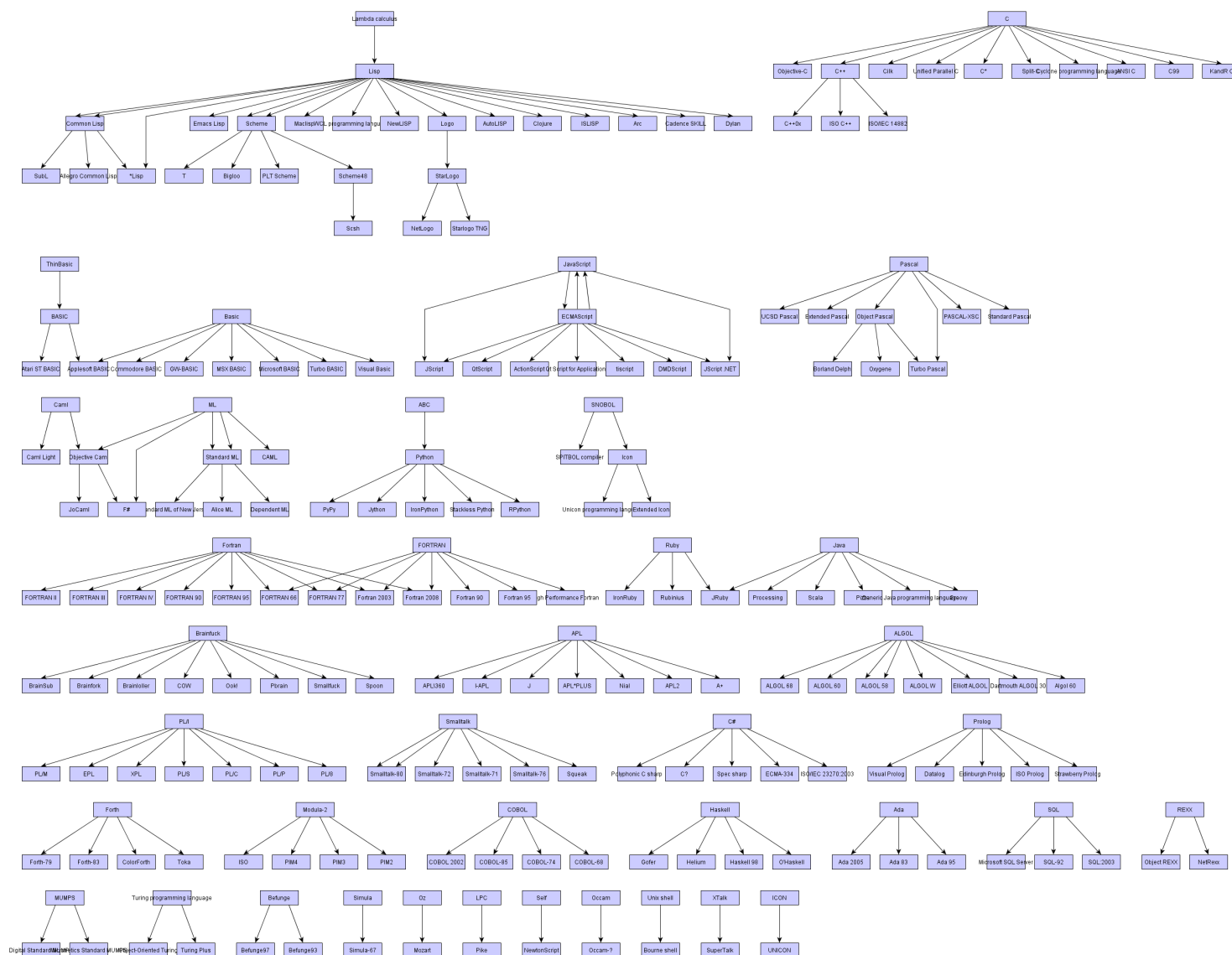
# Источники данных: визуализация

Граф влияния компьютерных языков (фрагмент)



# Источники данных: визуализация

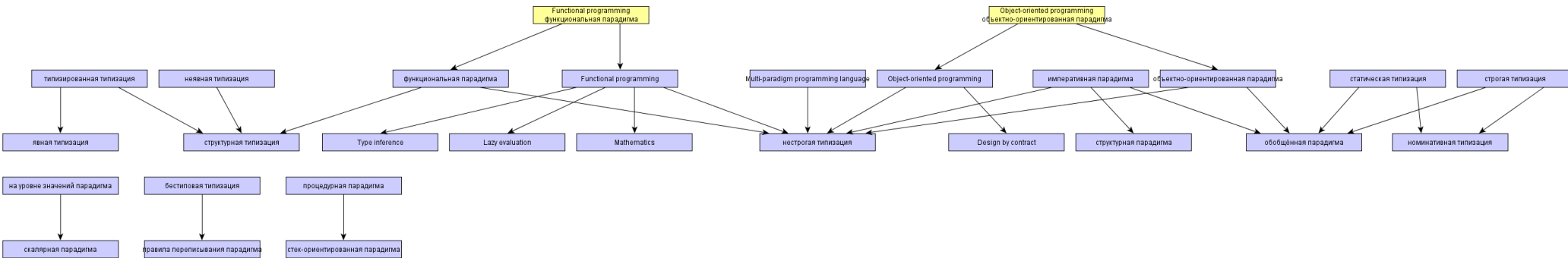
## Граф диалектов компьютерных языков





# Источники данных: визуализация

Граф вложенности классов (только классы, содержащие более 3 элементов)



## Источники данных: результаты

- Были найдены противоречия:

Атрибут	Язык	Значение во “Freebase”	Значение в “Прогпедии”
Дата создания	C	1972	1969
Дата создания	Forth	1970	1970s
Дата создания	Icon	1977	1970
Дата создания	PHP	1995	1994
Дата создания	Python	1991	1990

- Граф вложенности классов может быть использован для создания аксиом
- Граф влияния языков содержит циклы. Чтобы избежать противоречия, необходим темпоральный подход



## Что дальше?

- Открыть доступ через Интернет.
- Привлечь «экспертов» к пополнению.
- Автоматические средства пополнения из открытых источников.
- Использование в образовательном процессе.

