

# Автоматизированные методы пополнения знаний на основе онтологий

Научный руководитель:  
Пальчунов Д.Е.  
д.ф-м.н, в.н.с., ИМ СО РАН

Автор:  
Гуревич М.И.  
ФИТ, 12228

Новосибирск, 2014г.

# Предметная область

- ❑ Онтология – это совокупность множеств ключевых понятий предметной области и аналитических предложений, описывающих смысл этих понятий.
- ❑ Семантическая паутина – представляет информацию в удобном для машинной обработке виде (онтологии RDF).
- ❑ Прувер – автоматическая система доказательства теорем.

# Цели работы

- Осуществление автоматического пополнения известного множества предложений на основе знаний о предметной области
- Генерация логического вывода на основе пополненных данных

# Задачи

- ❑ Анализ предметной области
- ❑ Обзор существующих подходов хранения и обработки информации
- ❑ Разработка метода автоматического пополнения знаний
- ❑ Создание программной системы, реализующей пополнение знаний
- ❑ Интеграция пружера в созданную программную систему

# Используемые технологии

- **Онтологии:**
  - WordNet
    - «Русский WordNet»
    - «Тезаурус Рутез»
  - OWL
    - «Ontology for biomedical investigations»
- **Провер:**
  - Prover9.

# Задача автоматического пополнения знаний

Пусть дана сигнатура  $\zeta = \langle R, C \rangle$  и некоторая модели  $A$  данной сигнатуры, где  $R$  и  $C$  – кортежи предикатов и констант.

Пусть дано множество бескванторных предложений сигнатуры  $\zeta$ . Необходимо максимально дополнить это множество истинными предложениями этой же сигнатуры используя онтологии предметной области.

# Алгоритм пополнения

Пусть имеются:  $p(c_1)$  и  $c_n$ , такие что существует цепочка понятий  $c_1, \dots, c_n$  в онтологии связанные отношением *is-a-superclass*.

Выберем те, которые также являются константами в имеющейся сигнатуре  $(c_1, c_{i_1}, \dots, c_n)$ .

Таким образом, вместе с  $p(c_1)$  будут истинны и  $p(c_{i_1}), \dots, p(c_n)$ .

Аналогично и для большей арности.

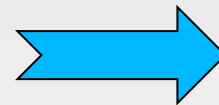
# Пример работы

File Info

Step 1.  
Input, data.

имеет{Джон, helicobacter\_pylori}

Continue



File Info

Step 3.  
View additional data.

имеет{джон, helicobacter\_pylori}  
имеет{джон, helicobacter}  
имеет{джон, helicobacteraceae}  
имеет{джон, campylobacterales}  
имеет{джон, epsilonproteobacteria}  
имеет{джон, proteobacteria}  
имеет{джон, bacteria}  
имеет{джон, грамотрицательная\_бактерия}  
имеет{джон, одноклеточная\_бактерия}  
имеет{джон, бактерия}  
имеет{джон, инфекция\_жкт}  
имеет{джон, инфекция}  
имеет{джон, заболевание}  
имеет{джон, болезнь}

Back Continue



# Интегрирование прuvera

- На вход прuverу подаются данные, полученные за счет автоматического пополнения.
- За счет использования онтологий предметной области увеличивается количество известных данных, что увеличивает возможность доказательства теоремы.

# Пример работы

File Info

Step 2.  
Input, ask.

имеет{Джон, болезнь}

Back Continue



File Info

Calculation success

TRUE

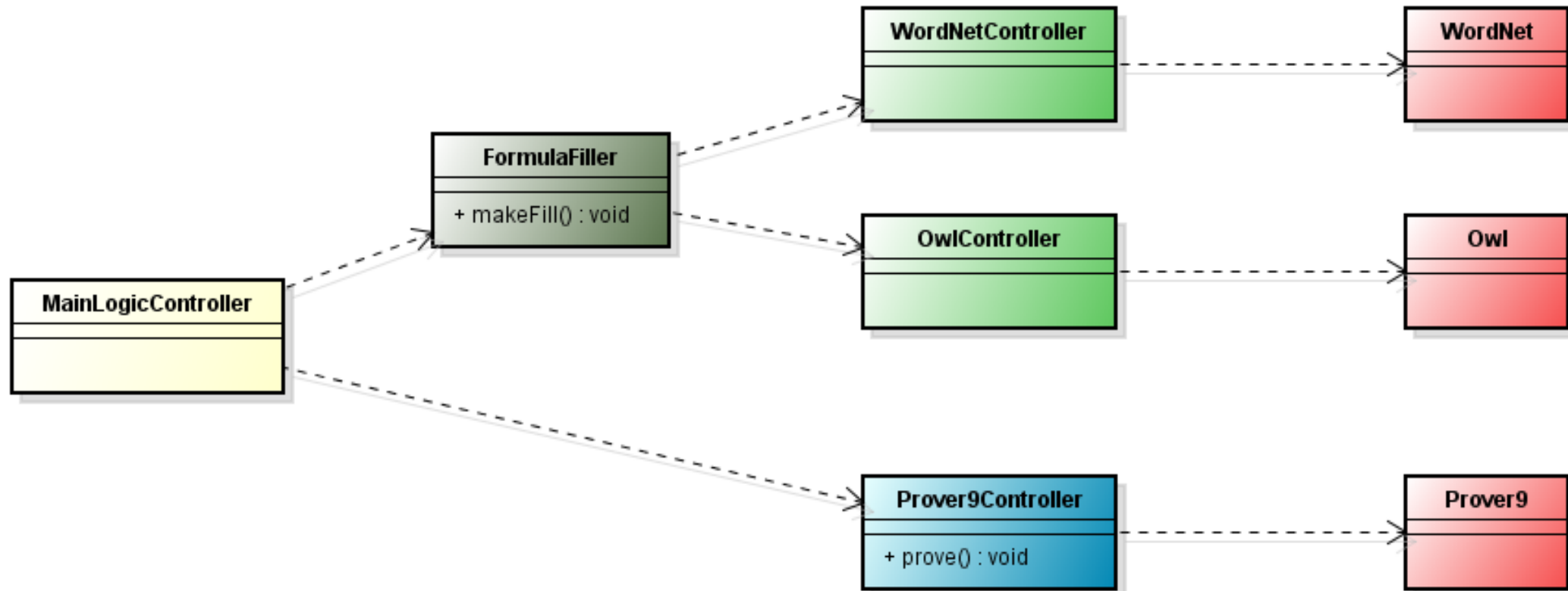
Back

# Используемые паттерны

Для удобства реализации и простоты расширения использовались:

- ❑ MVC – связь интерфейса и логики
- ❑ Builder – задание модулей пополнения
- ❑ Abstract Factory – внутренняя структура в модуле пополнения
- ❑ EventBus – связь контроллеров между собой

# Архитектура контроллера



# Результаты

- Обоснована возможность автоматического пополнения знаний на основе данных о предметной области в виде онтологий;
- Показаны условия и сфера применимости разработанного алгоритма пополнения;
- Создана программная реализация пополнения с использованием онтологий WordNet и OWL с возможностью автоматического логического вывода с Prover9.
- Реализованы различные методы ввода данных: консольный, графический, импорт из файла, из программы Махасоевой О.Г.

# Дальнейшее развитие

- Использование синонимии в названиях предикатов;
- Использование контекстного пополнения на основе дополнительных данных из структуры предложений на естественном языке;
- Пополнение благодаря другим связям в онтологии;
- Оптимизация внутренних алгоритмов;

# Публикации и конференции

- Гуревич М. И. Автоматизация построения нечеткого значения истинности на неполных данных // 51-я Международная научная студенческая конференция "Студент и научно-технический прогресс". - Новосибирск: НГУ, 2013. - С. 224.
- Гуревич М. И. Автоматизация логического вывода из имеющихся знаний на основе онтологий // 52-я Международная научная студенческая конференция "Студент и научно-технический прогресс". - Новосибирск: НГУ, 2014. - С. 238.
- Гуревич М. И. Разработка метода автоматизации пополнения множества бескванторных предложений новыми на основе данных онтологии предметной области // Альманах современной науки и образования. Тамбов: "Грамота", 2014.

**Спасибо за внимание!**



