

Язык высокого уровня для алгебры нечетких запросов в неоднородной среде

Выполнил: Савельев Николай, 544 группа

Научный руководитель : д.ф.-м.н., проф. Б.А. Новиков

Рецензент: асс. Г.А. Чернышев

Введение

- Много разнородных данных
- Разнообразиие источников информации
- Централизованная обработка
- Выбор между точностью результата и временем работы



Введение

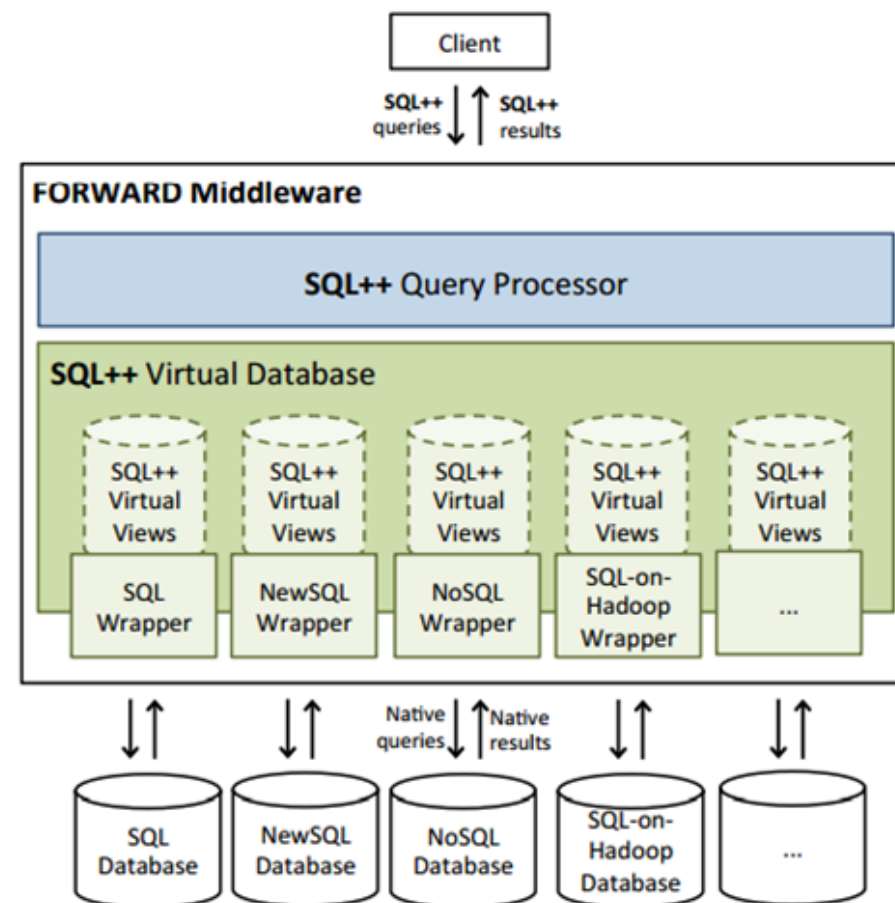
- Системы обработки неоднородных данных:
 - Выразительный язык запросов
 - Оптимизатор
 - Исполнитель запросов
 - Источники данных
- Высокие требования для языка запросов

Постановка задачи

- Цель: создать язык высокого уровня для алгебры нечетких запросов и разработать транслятор для него
- Задачи:
 - Реализовать представление модели алгебры нечетких запросов для последующего ее использования в процессе трансляции
 - Описать грамматику языка высокого уровня для алгебры нечетких запросов
 - Реализовать транслятор для языка высокого уровня в модель алгебры нечетких запросов
 - Провести апробацию реализованной функциональности

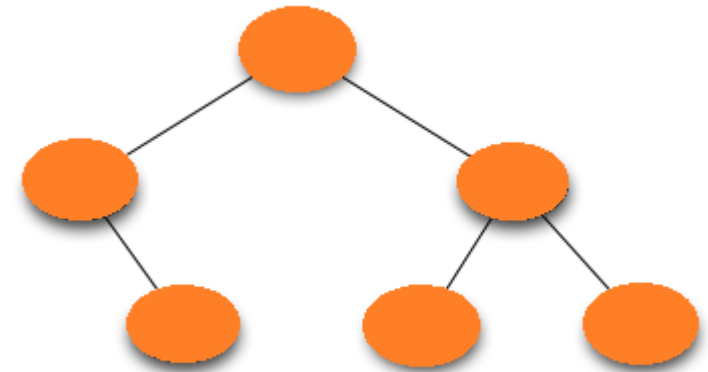
Существующие решения

- Системы хранения данных
- Интерфейсы – ADO, OLE DB, ODBC
- Forward
- Pentaho



Алгебра нечетких запросов

- Объект содержит релевантность запросу
- Операторы:
 - Фильтрация
 - Соединение
 - Группировка
 - Слияние
 - Nest / unnest
- Синтаксическая структура оператора:
 - Атрибуты объектов результата
 - Параметры
 - Аргументы



Грамматика языка запросов

- SQL
- Только операция выборки
- Каждое ключевое слово – свой оператор алгебры
- Блоки параметров
- Поддержка нечетких запросов на уровне функций

```
SELECT *  
FROM table1  
    FUZZY JOIN (algorithm      = 'merge_join',  
                precise_bound = 0.8)  
    table2  
ON table1.a = table2.b
```

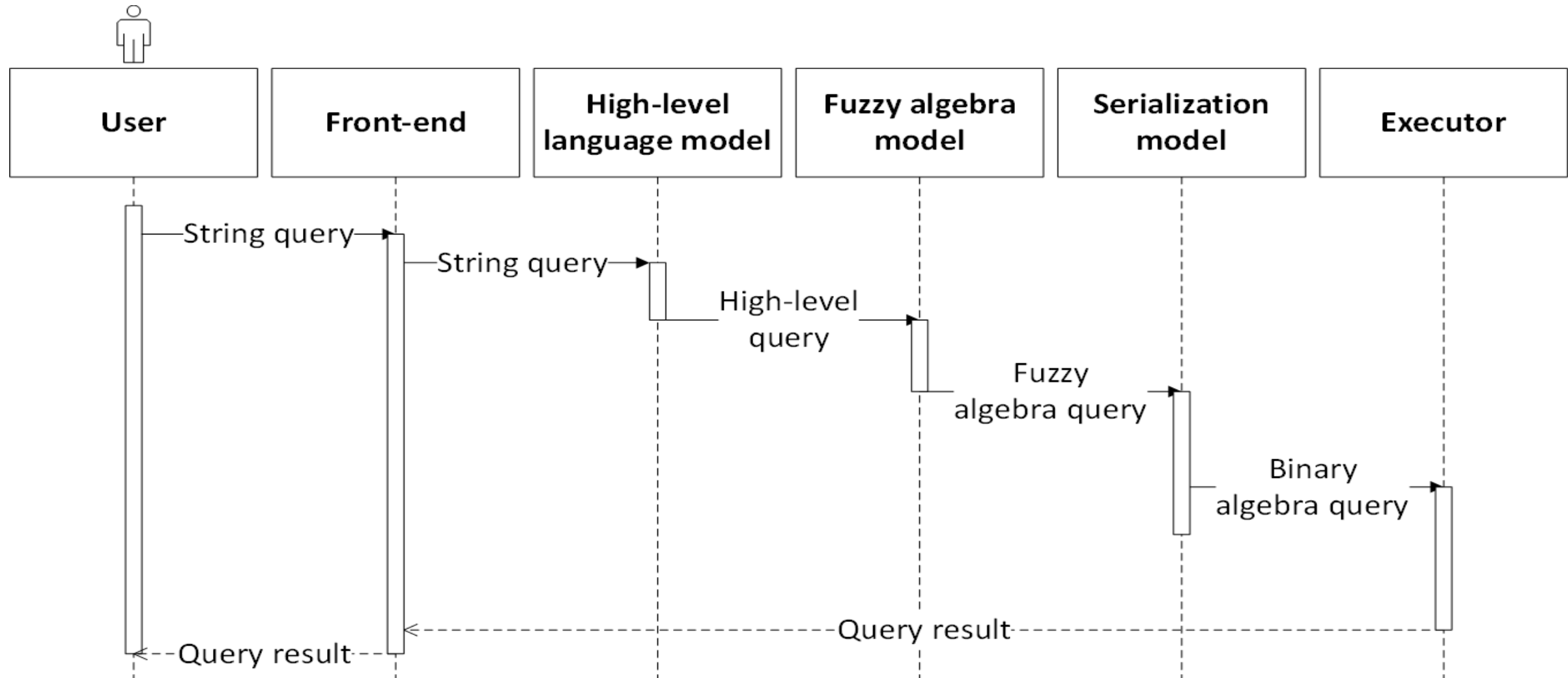
Реализация транслятора

- Java
- Грамматика в расширенной форме Бэкуса-Наура
- Генерация парсера по правилам грамматики
- ANTLR
- Промежуточная модель языка

Реализация транслятора(2)

- Трансформация в операторы алгебры
- Удобство разделения трансляции на несколько моделей
- Сериализация
- Apache Thrift
- XStream

Последовательность работы



Апробация

- AsterixDB
- Файловый адаптер
- Исполнитель запросов

Апробация (2)

```
SELECT  
asterix.name, file.name  
FROM file JOIN asterix  
ON asterix.dep_id =  
file.id  
WHERE asterix.age > 3
```

Апробация (2)

SELECT

asterix.name, file.name

FROM file **JOIN** asterix

ON asterix.dep_id =

file.id

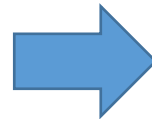
WHERE asterix.age > 3



```
<query id="file" type="PRIMARY">
  <parameters>
    <parameter name="sourceName">
      <expression type="CONSTANT" value="file"/>
    </parameter>
  </parameters>
</query>
```

Апробация (2)

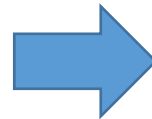
```
SELECT  
asterix.name, file.name  
FROM file JOIN asterix  
ON asterix.dep_id =  
file.id  
WHERE asterix.age > 3
```



```
<query id="asterix" type="FILTER">  
  <parameters>  
    <parameter name="filterExpression">  
      <expression type="GT">  
        <arguments>  
          <expression id="age"  
            type="ATTRIBUTE"  
            value="asterix.age"/>  
          <expression type="CONSTANT" value="3"/>  
        </arguments>  
      </expression>  
    </parameter>  
  </parameters>  
<arguments>  
  <query id="asterix" type="PRIMARY">  
    <parameters>  
      <parameter name="sourceName">  
        <expression  
          type="CONSTANT"  
          value="asterix"/>  
        </parameter>  
      </parameters>  
    </query>  
  </arguments>  
</query>
```

Апробация (2)

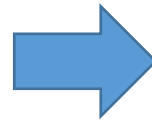
```
SELECT
asterix.name, file.name
FROM file JOIN asterix
ON asterix.dep_id =
file.id
WHERE asterix.age > 3
```



```
<query id="asterix" type="FILTER">
  <parameters>
    <parameter name="filterExpression">
      <expression type="GT">
        <arguments>
          <expression id="age"
            type="ATTRIBUTE"
            value="asterix.age"/>
          <expression type="CONSTANT" value="3"/>
        </arguments>
      </expression>
    </parameter>
  </parameters>
  <arguments>
    <query id="asterix" type="PRIMARY">
      <parameters>
        <parameter name="sourceName">
          <expression
            type="CONSTANT"
            value="asterix"/>
        </parameter>
      </parameters>
    </query>
  </arguments>
</query>
```

Апробация (2)

```
SELECT
asterix.name, file.name
FROM file JOIN asterix
ON asterix.dep_id =
file.id
WHERE asterix.age > 3
```



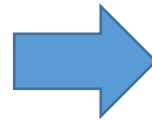
```
<query id="asterix" type="FILTER">
  <parameters>
    <parameter name="filterExpression">
      <expression type="GT">
        <arguments>
          <expression id="age"
            type="ATTRIBUTE"
            value="asterix.age"/>
          <expression type="CONSTANT" value="3"/>
        </arguments>
      </expression>
    </parameter>
  </parameters>
  <arguments>
    <query id="asterix" type="PRIMARY">
      <parameters>
        <parameter name="sourceName">
          <expression
            type="CONSTANT"
            value="asterix"/>
        </parameter>
      </parameters>
    </query>
  </arguments>
</query>
```


Апробация (2)

```
SELECT
asterix.name, file.name
FROM file JOIN asterix
ON asterix.dep_id =
file.id
WHERE asterix.age > 3
```



```
<query id="file" type="PRIMARY">
  <parameters>
    <parameter name="sourceName">
      <expression type="CONSTANT" value="file"/>
    </parameter>
  </parameters>
</query>
```



```
<query id="asterix" type="FILTER">
  <parameters>
    <parameter name="filterExpression">
      <expression type="GT">
        <arguments>
          <expression id="age"
            type="ATTRIBUTE"
            value="asterix.age"/>
          <expression type="CONSTANT" value="3"/>
        </arguments>
      </expression>
    </parameter>
  </parameters>
  <arguments>
    <query id="asterix" type="PRIMARY">
      <parameters>
        <parameter name="sourceName">
          <expression
            type="CONSTANT"
            value="asterix"/>
        </parameter>
      </parameters>
    </query>
  </arguments>
</query>
```

Результаты

- Реализовано представление модели алгебры нечетких запросов
- Создан язык высокого уровня для алгебры нечетких запросов
- Реализован транслятор для языка высокого уровня в модель алгебры нечетких запросов
- Выполнена апробация реализованной функциональности
- Результаты данной работы представлены на ISSAT2015