

Машинное обучение для формирования контекстной подсказки (контент-ассист) при разработке на платформе 1С:Предприятие

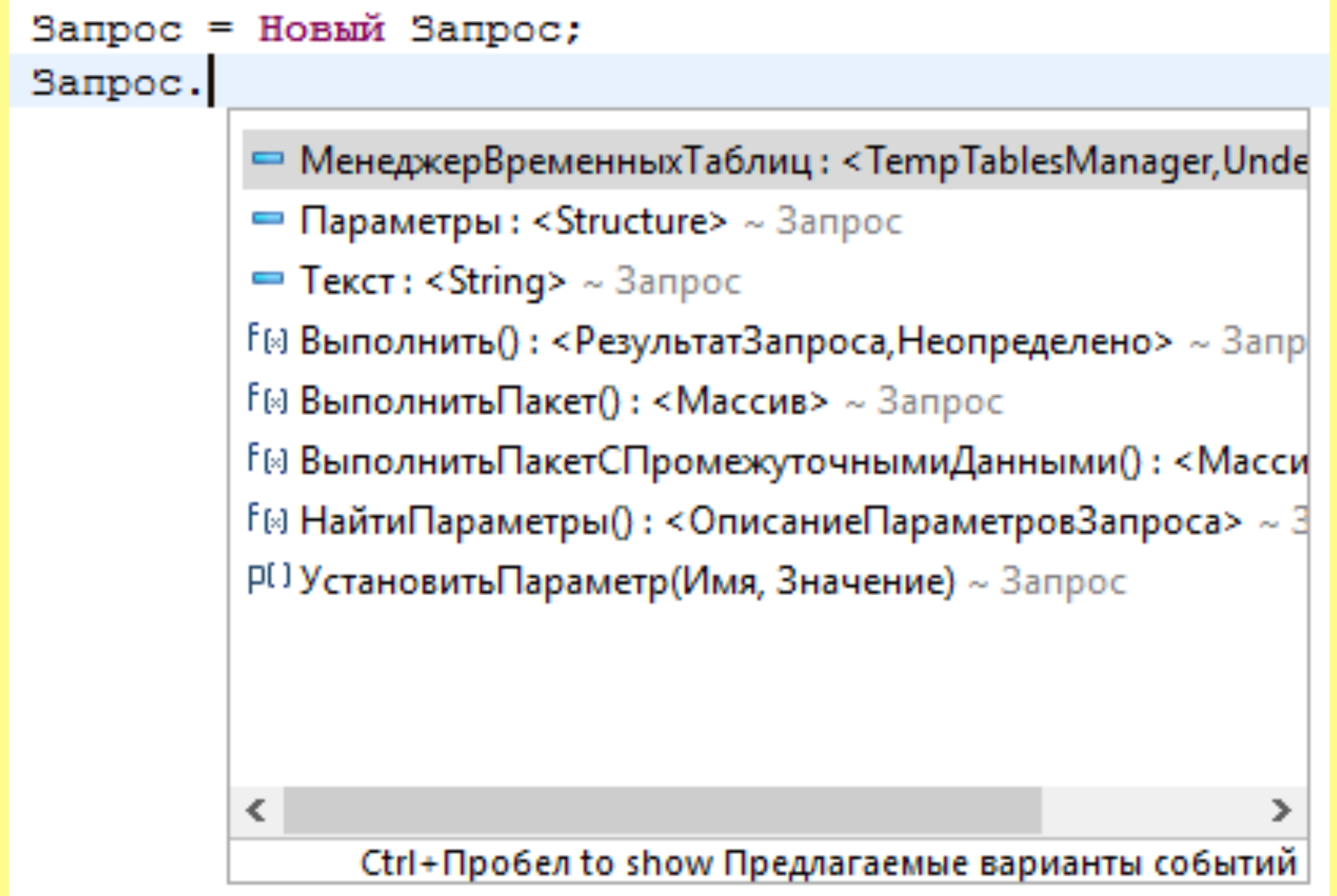
Студент 2 курса магистратуры, М05-912е гр.:
Каразеев А. А.

Научный руководитель:
Дзюба М. В.

МФТИ, 2021

Введение

Контекстная подсказка — это инструмент, который помогает нам писать и редактировать код программы. С его помощью мы быстрее вводим текст и допускаем меньше ошибок и опечаток.



Цели и задачи работы

- Провести обзор имеющихся решений для построения контекстной подсказки;
- Исследовать предметную область и реализовать механизм переупорядочивания контекстной подсказки с применением машинного обучения для встроенного языка платформы “1С: Предприятие 8”;
- Провести анализ полученных результатов;
- Описать процесс встраивания полученного алгоритма в качестве плагина для среды разработки “1С Enterprise Development Tools” (1С:EDT).

Существующие решения

- Eclipse Code Recommenders
 - Проект архивирован, более не поддерживается;
 - Статистические алгоритмы;
- Eclipse Orion
 - Заточен под веб-разработку;
 - Разработан с учетом особенностей языка JavaScript;
- Jedi
 - Разработан для языка Python;
 - Статический анализатор на основе синтаксического дерева;
 - 120 соавторов, 4.9k звезд на GitHub.

Существующие решения

- Их объединяет следующее:
 - Основаны на знаниях о языке и завязаны на конкретный язык, что делает невозможным их переиспользование с другими языками;
 - А также не учитывают контекст.

Предполагаемый алгоритм

- Алгоритм будет переупорядочивать элементы контекстной подсказки, которые наиболее релевантны в месте вызова подсказки.

Предполагаемый алгоритм

- Алгоритм будет переупорядочивать элементы контекстной подсказки, которые наиболее релевантны в месте вызова подсказки.

Например:

```
запрос = Новый Запрос;  
  
запрос.<|>
```

Таблица 1 — Пример выдачи контекстной подсказки (слева) и желаемая последовательность элементов (справа).

1. МенеджерВременныхТаблиц	1. Текст
2. Параметры	2. МенеджерВременныхТаблиц
3. Текст	3. УстановитьПараметр()
...	...
8. УстановитьПараметр()	8. Параметры

Под знаком “<|>” понимается место вызова контекстной подсказки.

Имеющиеся алгоритмы

- **Синтаксический анализатор** (не учитывает контекст, может предлагать редко используемые элементы);
- **Статистический подход** - предлагает элементы на основании аналитических данных множества проектов (не учитывает контекст);
- Все они завязаны на конкретный язык и базу знаний о нём, нет возможности переноса опыта на другие языки;

Имеющиеся алгоритмы

- **Синтаксический анализатор** (не учитывает контекст, может предлагать редко используемые элементы);
- **Статистический подход** - предлагает элементы на основании аналитических данных множества проектов (не учитывает контекст);
- Все они завязаны на конкретный язык и базу знаний о нём, нет возможности переноса опыта на другие языки;
- **Машинное обучение** - может предсказывать элементы в зависимости от контекста вызова.

Машинное обучение

- Проанализировав имеющиеся алгоритмы, было принято решение использовать подход **машинного обучения**;
- **Условные Случайные Поля (CRF)** - сложный процесс обучения, ручной подбор признаков;
- **Graph NNs / Gated Graph NNs** - требуют большой объём вычислительных ресурсов, ручной подбор признаков;
- **RNN / LSTM**

Машинное обучение

- Проанализировав имеющиеся алгоритмы, было принято решение использовать подход **машинного обучения**;
- **Условные Случайные Поля (CRF)** - сложный процесс обучения, ручной подбор признаков;
- **Graph NNs / Gated Graph NNs** - требуют большой объём вычислительных ресурсов, ручной подбор признаков;
- **RNN / LSTM:**
 - word2vec - теряет часть информации об отношениях между переменными, классами и методами;
 - code2seq - не требует ручного подбора признаков, применимо для различных языков программирования, а также может быть использовано для решения других задач и не требует семантического анализа.

Архитектура решения - АСД



- Было решено использовать **Абстрактное Синтаксическое Дерево** в качестве описания текста программ;
- АСД наиболее полно характеризует место, из которого была вызвана контекстная подсказка;
- А также хорошо характеризует контекст, в котором был произведён вызов, и “знает” о программе до и после точки вызова;
- Может быть расширено дополнительными параметрами.

Архитектура решения - АСД

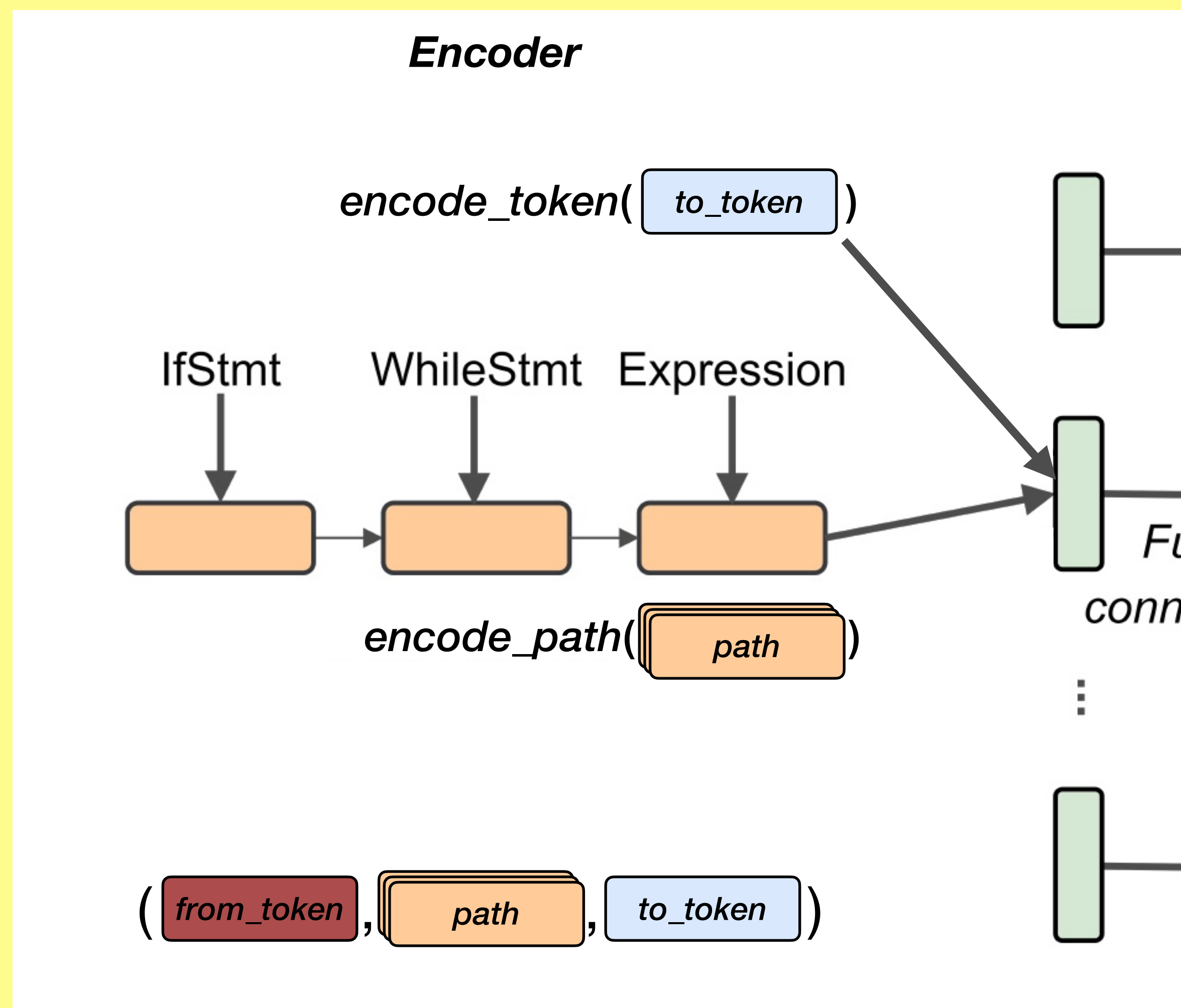


- По тексту программы строится Дерево Разбора, а затем по нему - Абстрактное Синтаксическое Дерево:
- **Внутренние вершины** сопоставлены с операторами языка программирования (“Пока”, “+”, “=”, “>” и т.п.)
- **Листья** - операнды (“х”, “42” и т.п.)
- Полученная информация представляет из себя граф, над которым можно производить преобразования и обучать модели.

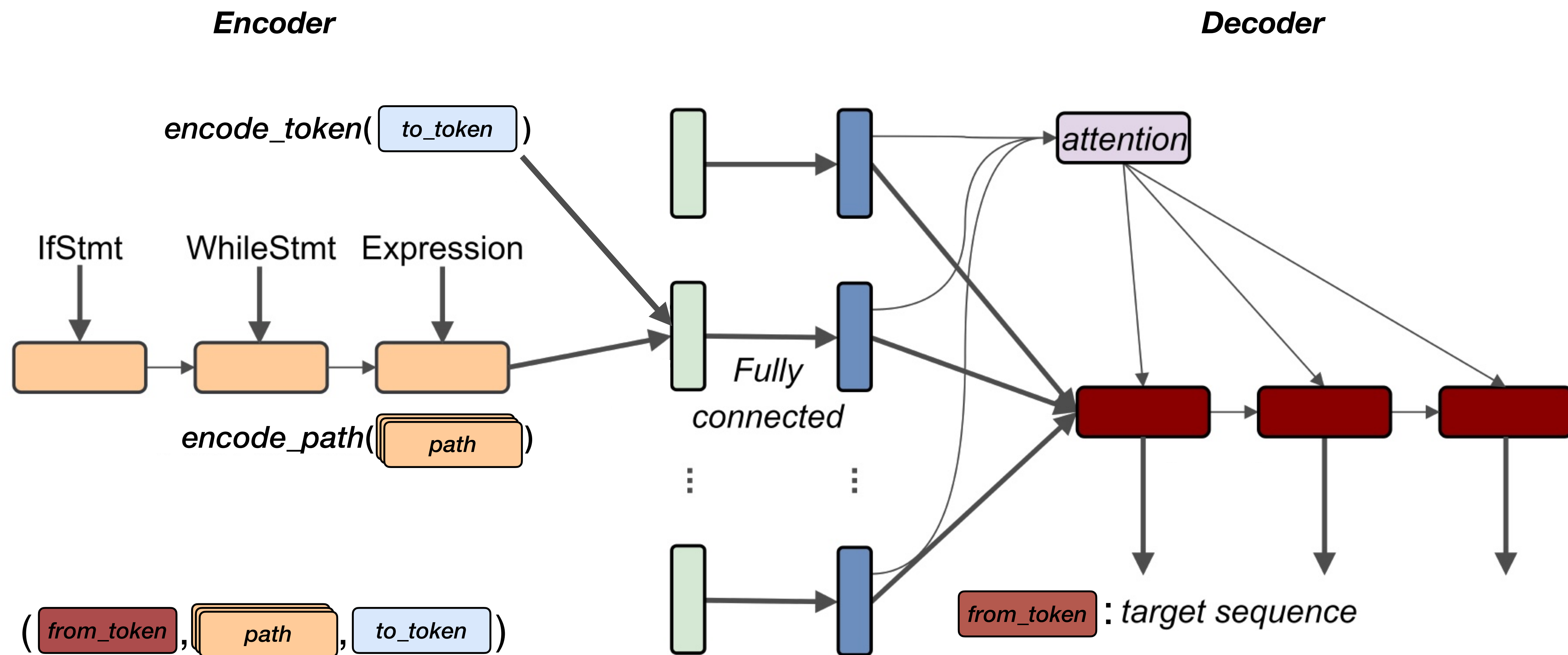
Построение модели - code2seq

- Проанализировав варианты, было решено взять за основу модель code2seq, с помощью которой можно строить векторные представления текста программ;
- Датасет для обучения состоит из набора контекстных цепочек между листьями АСД:
 - `(from_token, path, to_token)`, где
 - `from_token` - начало цепочки (операнд)
 - `path` - последовательность внутренних вершин (операторы)
 - `to_token` - конец цепочки (операнд)

Построение модели - code2seq



Построение модели - code2seq



Обучение модели - code2seq

- Предсказывается `from_token` на основании `path` и `to_token`;
- Модель представляет из себя Энкодер + Декодер;

$$h_2, \dots, h_l = LSTM(E_{v_2}^{nodes}, \dots, E_{v_l}^{nodes})$$

$$\text{encode_path}(v_2, \dots, v_l) = [h_l^{\rightarrow}; h_2^{\leftarrow}]$$

$$\text{encode_token}(w) = \sum_{s \in \text{split}(w)} E_s^{\text{subtokens}}$$

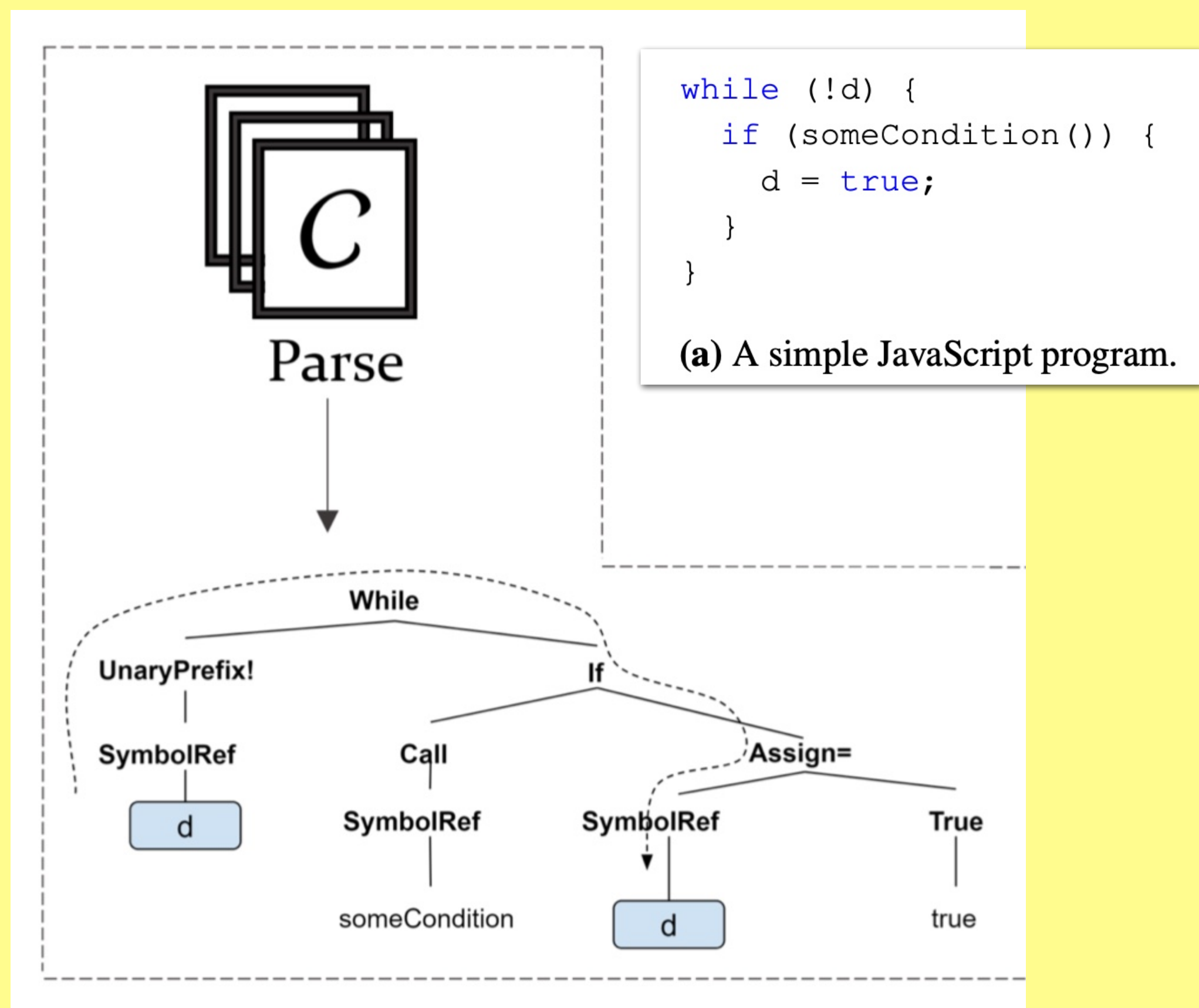
$$z = \tanh(W_{in}[\text{encode_path}(v_2 \dots v_l); \text{encode_token}(\text{value}(v_l))])$$

Пример работы

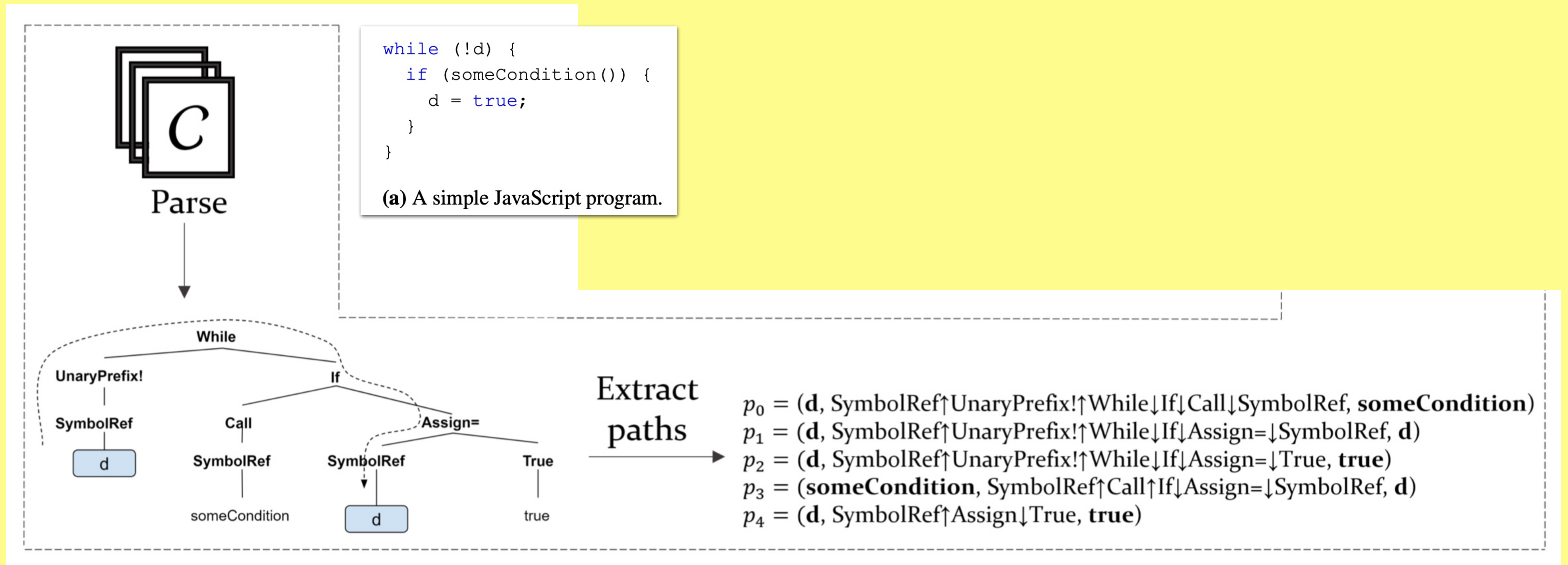
```
while (!d) {  
  if (someCondition()) {  
    d = true;  
  }  
}
```

(a) A simple JavaScript program.

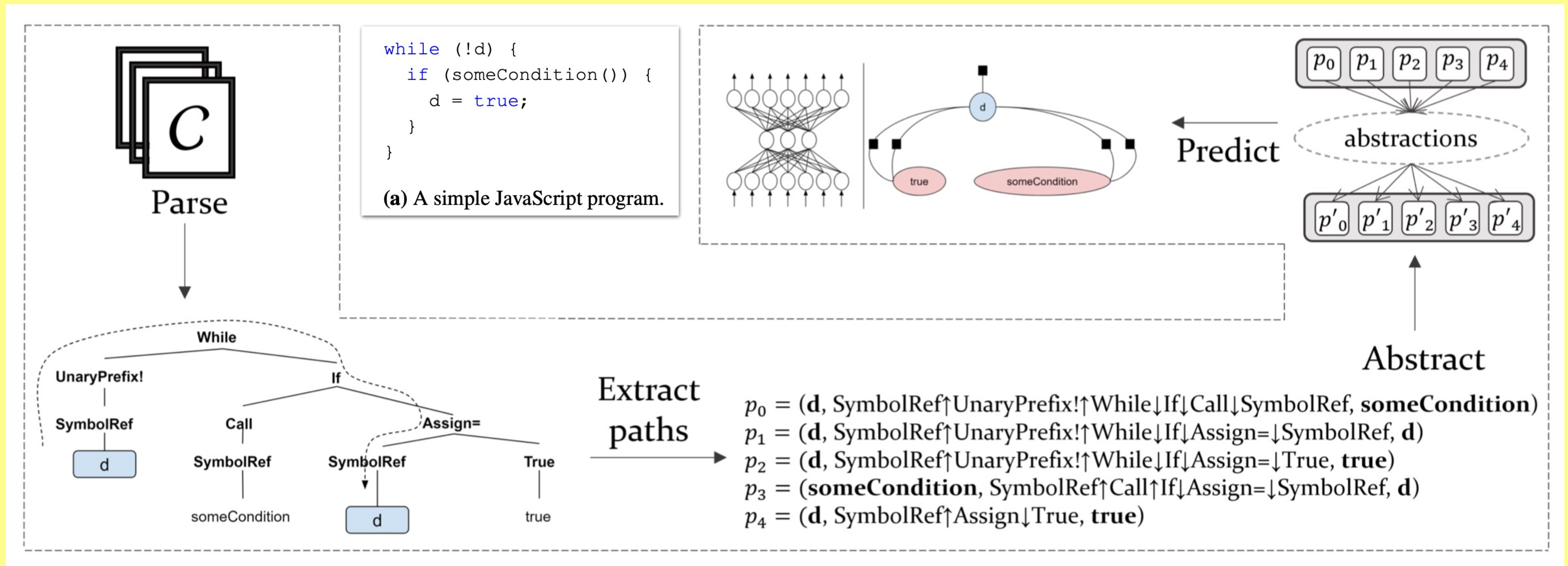
Пример работы



Пример работы



Пример работы



Подготовка данных - АСД

Document

Documents → Document → Object module

Процедура

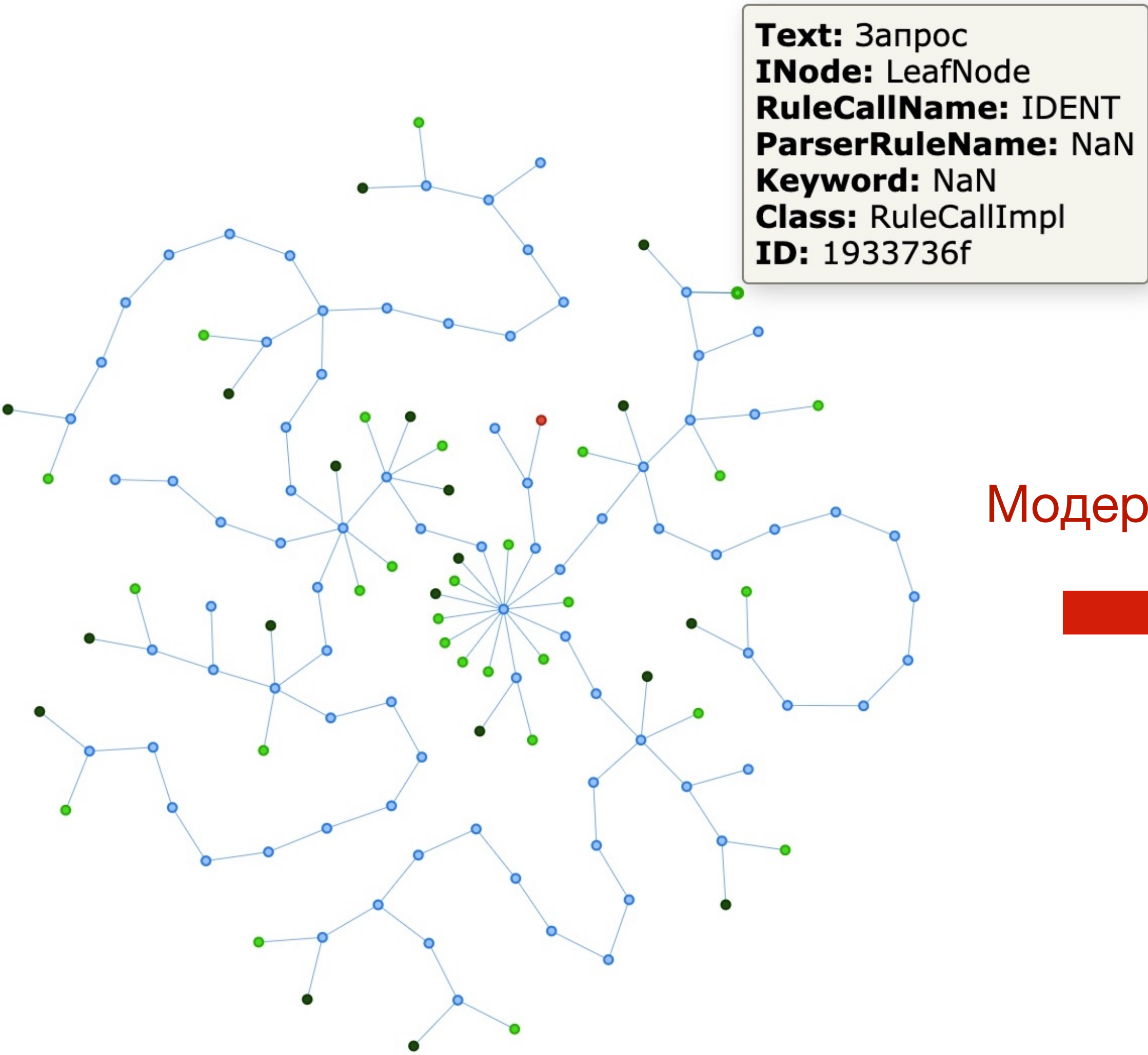
 имяПроцедуры(Параметр)

Экспорт

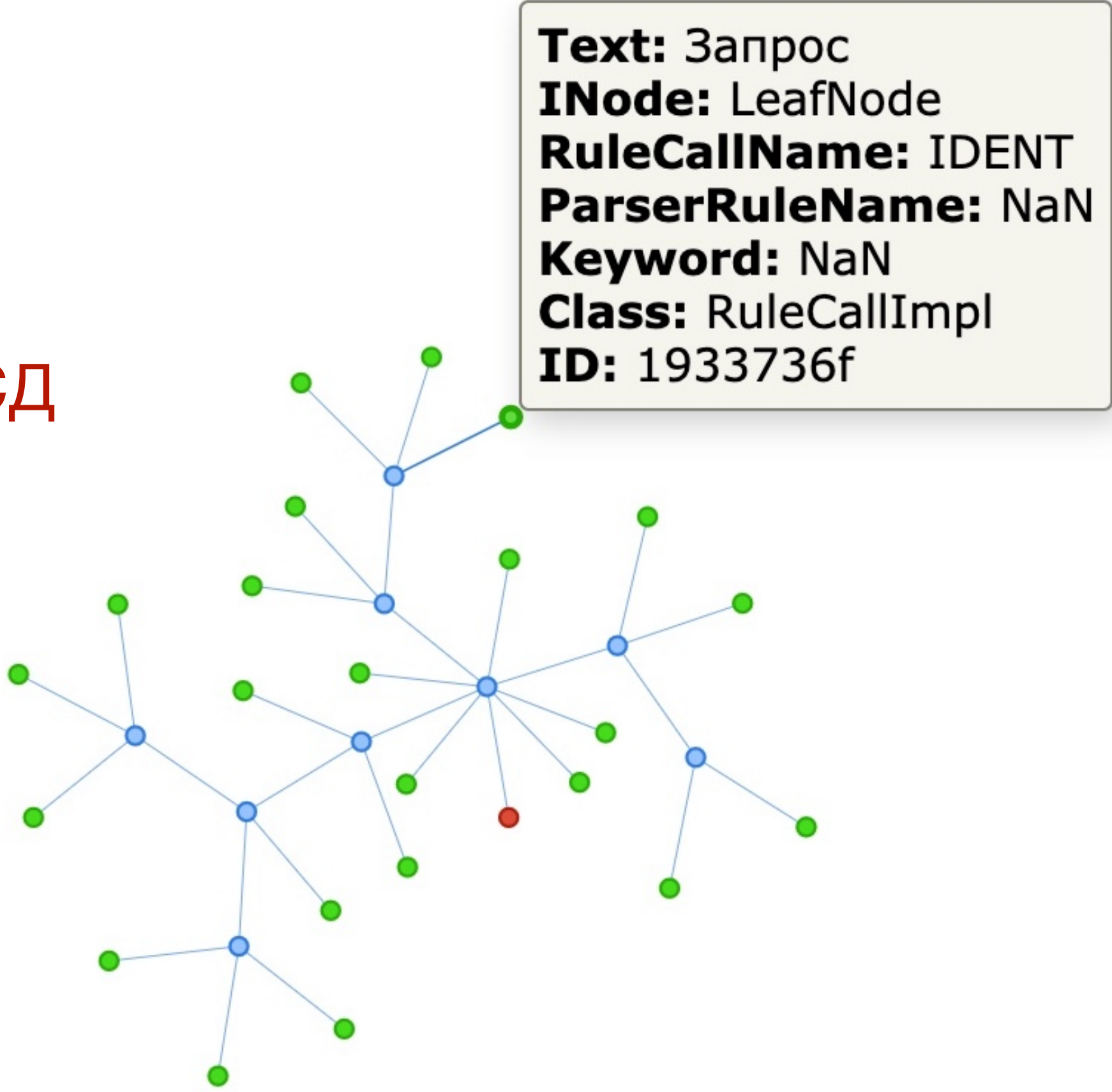
Запрос = **Новый** Запрос;

Если Параметр = **Истина** Тогда
Запрос.Текст = "текст запроса";
КонецЕсли;

КонецПроцедуры



Модернизация АСД



Модернизация АСД

- Удалены листовые вершины с типом HiddenLeafNode, а также вершины, отвечающие за отступы и пробелы;
- Сокращены пути от оставшихся листовых вершин к корню за счёт удаления промежуточных вершин;
- Максимальная допустимая длина пути ограничена сверху величиной `max_length` (проведены эксперименты со значениями 5 и 9).

Встраивание в “1C:EDT”

- Плагин встраивается в процесс build проекта через точку расширения Eclipse и обрабатывает каждый модуль встроенного языка;
- Собирает информацию о контекстах (Деревья Разбора);
- Происходит модернизация деревьев и на выходе имеем АСД, по которым уже строятся контекстные цепочки;
- Цепочки подаются на вход модели, которая предсказывает релевантность элементов контекстной подсказки.

Результаты

- Полученные результаты были сделаны только на основании АСД;
- Это не конечный результат, АСД можно расширить признаками: учитывать типизацию переменных, типы модулей (обработчик, класс, метод и т. п.);
- Поставленные задачи были достигнуты и предложены способы улучшения алгоритма.

Catalogs → ТиповыеОперации → Forms → ФормаВариантаЗаполнения → Module

```
34      ЭтоЗначение = 1;
35      НеИзменять = 0;
36      Элементы.Параметр.Доступность = Ложь;
37  КонецЕсли;
38
39  // Установить связь параметров выбора
40  Если СвязьПоВладельцу = "Параметр" Тогда
41      ЭтоПараметр = 1;
42      ЭтоЗначение = 0;
43      НеИзменять = 0;
44      Элементы.ЭтоЗначение.Доступность = Ложь;
45      Элементы.Значение.Доступность = Ложь;
46      Элементы.Параметр.Доступность = Истина;
47  ИначеЕсли ЗначениеЗаполнено(СвязьПоВладельцу) Тогда
48      МассивСвязейПараметровВыбора = Новый Массив;
49      // НоваяСвязь = Новый СвязьПараметраВыбора("Отбор.Владелец", "Св
50      // МассивСвязейПараметровВыбора.Добавить(НоваяСвязь);
51      МассивСвязейПараметровВыбора.
52      Элементы.Значение.СвязиПараме
53  КонецЕсли;
54
55  Элементы.Значение.ОграничениеТипа
56
57  КонецПроцедуры
58
59  #КонецОбласти
60
61
62
63
```

ВГраница() : <Число> ~ Массив

Вставить(Индекс) ~ Массив

Вставить(Индекс, Значение) ~ Массив

Добавить() ~ Массив

Добавить(Значение) ~ Массив

Количество() : <Число> ~ Массив

Найти(Значение) : <Неопределено,Число> ~ Массив

Очистить() ~ Массив

Получить(Индекс) : <Произвольный> ~ Массив

Удалить(Индекс) ~ Массив

Установить(Индекс, Значение) ~ Массив

Space to show Event proposals

from_token → to_token:
Добавить → МассивСвязейПараметровВыбора

- Топ-5:
1. Вставить
 2. **Добавить**
 3. Удалить
 4. Очистить
 5. ВГраница

Catalogs → ВариантыОтчетов → Object module

```
73  КонецЕсли;
74
75  // Удаление из табличной части подсистем, помеченных на удаление.
76  МассивУдаляемыхСтрок = Новый Массив;
77  Для Каждого СтрокаРазмещения Из Размещение Цикл
78      Если СтрокаРазмещения.Подсистема.ПометкаУдаления = Истина Тог
79          // МассивУдаляемыхСтрок.Добавить(СтрокаРазмещения);
80      МассивУдаляемыхСтрок.
81  КонецЕсли;
82  КонецЦикла;
83  Для Каждого СтрокаРазмещения
84      Размещение.Удалить(Строка
85  КонецЦикла;
86
87  // Заполнение реквизитов "Наи
88  ПроиндексироватьНастройки();
89  КонецПроцедуры
90
91  #КонецОбласти
92
```

ВГраница() : <Число> ~ Массив

Вставить(Индекс) ~ Массив

Вставить(Индекс, Значение) ~ Массив

Добавить() ~ Массив

Добавить(Значение) ~ Массив

Количество() : <Число> ~ Массив

Найти(Значение) : <Неопределено,Число> ~ Массив

Очистить() ~ Массив

Получить(Индекс) : <Произвольный> ~ Массив

Удалить(Индекс) ~ Массив

Установить(Индекс, Значение) ~ Массив

Space to show Event proposals

from_token → to_token:
Добавить → МассивУдаляемыхСтрок

- Топ-5:
1. Вставить
 2. Найти
 3. Установить
 4. **Добавить**
 5. Получить

```
123
124 &НаСервере
125 Процедура УстановитьУсловноеОформление( )
126
127     Список.УсловноеОформление.Элементы.Очистить( ) ;
128     Элемент = Список.УсловноеОформление.Элементы.Добавить
129
130     ГруппаОтбора = Элемент.Отбор.Элементы.Добавить(Тип("ГруппаОтбора.ТипГруппы = ТипГруппыЭлементовОтбораКомп
131     ГруппаОтбора.ТипГруппы = ТипГруппыЭлементовОтбораКомп
132     ОтборЭлемента = ГруппаОтбора.Элементы.Добавить(Тип("Э
133     ОтборЭлемента.ЛевоеЗначение = Новый ПолеКомпоновкиДан
134     ОтборЭлемента.ВидСравнения = ВидСравненияКомпоновкиДан
135     // ОтборЭлемента = ГруппаОтбора.Элементы.Добавить(Тип
136     ОтборЭлемента = ГруппаОтбора.
137     ОтборЭлемента.ЛевоеЗначение =
138     ОтборЭлемента.ВидСравнения =
139     ОтборЭлемента.ПравоеЗначение
140
141     Элемент.Оформление.Установить
142
143 КонецПроцедуры
```

- ВидСравнения : <ВидСравненияКомпоновкиДанных> ~ Эл
- ИдентификаторПользовательскойНастройки : <Строка> ~
- Использование : <Булево> ~ ЭлементОтбораКомпоновки
- ЛевоеЗначение : <ПолеКомпоновкиДанных> ~ ЭлементО
- ПравоеЗначение ~ ЭлементОтбораКомпоновкиДанных
- Представление : <Строка> ~ ЭлементОтбораКомпоновки
- ПредставлениеПользовательскойНастройки : <Строка> ~
- Применение : <ТипПримененияОтбораКомпоновкиДанн
- РежимОтображения : <РежимОтображенияЭлементаНаст
- Родитель : <ГруппаЭлементовОтбораКомпоновкиДанных,
- ТипГруппы : <ТипГруппыЭлементовОтбораКомпоновкиДа
- Элементы : <КоллекцияЭлементовОтбораКомпоновкиЛан

from_token → to_token:
Элементы → ГруппаОтбора

- Топ-5:
1. Применение
 2. ТипГруппы
 3. Использование
 4. **Элементы**
 5. РежимОтображение

```
139
140     КорневойЭлемент = Справочники.ОбщероссийскийКлассификаторОсновныхФондов.
141     Если КорневойЭлемент.Пустая() Тогда
142         ЭлементОбъект = Справочники.ОбщероссийскийКлассификаторОсновныхФондо
143         ЭлементОбъект.Код = "ОК 013-94";
144         ЭлементОбъект.Наименование = НСтр("ru='Утвержден Постановлением Госс
145         ЭлементОбъект.НаименованиеГруппировки =
146             НСтр("ru='Утвержден Постановлением Госстандарта РФ от 26 декабря
147
148         ОбновлениеИнформационнойБазы.ЗаписатьОбъект(ЭлементОбъект);
149
150         // КорневойЭлемент = ЭлементОбъект.Ссылка;
151         КорневойЭлемент = ЭлементОбъект.
152     КонецЕсли;
153
154     Пока Выборка.Следующий() Цикл
155         ЭлементОбъект = Выборка.Ссылка.П
156         ЭлементОбъект.Родитель = Корнево
157
158         ОбновлениеИнформационнойБазы.Зап
159     КонецЦикла;
160
161 КонецПроцедуры
162
```

- КонтрольноеЧисло : <Строка> ~ СправочникОбъект.Обще
- НаименованиеГруппировки : <Строка> ~ СправочникОбъе
- ЭтотОбъект : <СправочникОбъект.ОбщероссийскийКласс
- ВерсияДанных : <Строка> ~ СправочникОбъект.Общерос
- Владелец : <Неопределено> ~ СправочникОбъект.Общере
- ДополнительныеСвойства : <Структура> ~ СправочникОб
- ЗаписьИсторииДанных : <ПараметрыЗаписиИсторииДанн
- ИмяПредопределенныхДанных : <Строка> ~ СправочникС
- Код : <Строка> ~ СправочникОбъект.ОбщероссийскийКла
- Наименование : <Строка> ~ СправочникОбъект.Общерос
- ОбменДанными : <ПараметрыОбменаДанными> ~ Справ
- Помет

from_token → to_token:
Ссылка → ЭлементОбъект

- Топ-5:
1. ЭтотОбъект
 2. Владелец
 3. Код
 4. **Ссылка**
 5. Наименование

```
10
11 Если ЗначениеЗаполнено(ЗначениеШтри
12 Запрос = Новый Запрос;
13 Запрос.УстановитьПараметр("Знач
14 // Запрос.Текст ="ВЫБРАТЬ ПЕРВЫ
15 // | ШтрихкодыУпаковокТоваро
16 // |ИЗ
17 // | Справочник.ШтрихкодыУпа
18 // |ГДЕ
19 // | ШтрихкодыУпаковокТоваро
20 Запрос.
21
22 Результ
23 Если НЕ
24 Тек
25 Общ
26 КонецЕс
27 КонецЕсли;
```

МенеджерВременныхТаблиц : <МенеджерВременныхТаблиц>
Параметры : <Структура> ~ Запрос
Текст : <Строка> ~ Запрос
ТребуемаяАктуальностьДанных : <ТребуемаяАктуальностьДанных>
ТребуемоеВремяАктуальностиДанных : <Дата,СтандартнаяАктуальность>
Выполнить() : <Неопределено,РезультатЗапроса> ~ Запрос
ВыполнитьПакет() : <Массив> ~ Запрос
ВыполнитьПакетСПромежуточнымиДанными() : <Массив>
НайтиПараметры() : <ОписаниеПараметровЗапроса> ~ Запрос
УстановитьПараметр(Имя, Значение) ~ Запрос

Space to show Event proposals

from_token → to_token:

Текст → Запрос

Top-5:

1. УстановитьПараметр
2. **Текст**
3. МенеджерВременныхТаблиц
4. Выполнить
5. Параметры

```
147
148 &НаКлиенте
149 Процедура СписокПередНачаломДобавления(Элемент, Отказ, Копирование, Родитель, Группа, П
150
151 Отказ = Истина;
152
153 ПараметрыФормы = Новый Структура;
154 Если Копирование Тогда
155 // Если Не Элементы.Список.ТекущиеДанные.Свойство("Ссылка") Тогда
156 Если Не Элементы.Список.ТекущиеДанные.
157 Возврат.;
158 КонецЕсли;
159 ПараметрыФормы.Вставить("ЗначениеКопир
160 КонецЕсли;
161 ПараметрыФормы.Вставить("ЗначенияЗаполнени
162 ПараметрыФормы.ЗначенияЗаполнения.Вставить
163
164 ОткрытьФорму("Справочник.БанковскиеСчета.Ф
165
166 КонецПроцедуры
167
168
```

Основной : <Булево> ~ ДанныеФормыСтруктура
ПометкаУдаления : <Булево> ~ ДанныеФормыСтруктура
Свойство(Ключ) : <Булево> ~ ДанныеФормыСтруктура
Свойство(Ключ, НайденноеЗначение) : <Булево> ~ ДанныеФормыСтруктура
Банк : <СправочникСсылка.Банки> ~ ДанныеФормыСтруктура
ВалютаДенежныхСредств : <СправочникСсылка.Валюты>
ВидСчета : <Строка> ~ ДанныеФормыСтруктура
Владелец : <СправочникСсылка.Контрагенты,Справочник
ДатаЗакрытия : <Дата> ~ ДанныеФормыСтруктура
Код : <Строка> ~ ДанныеФормыСтруктура
Наименование : <Строка> ~ ДанныеФормыСтруктура
НоменСчета : <Строка> ~ ДанныеФормыСтруктура

Space to show Event proposals

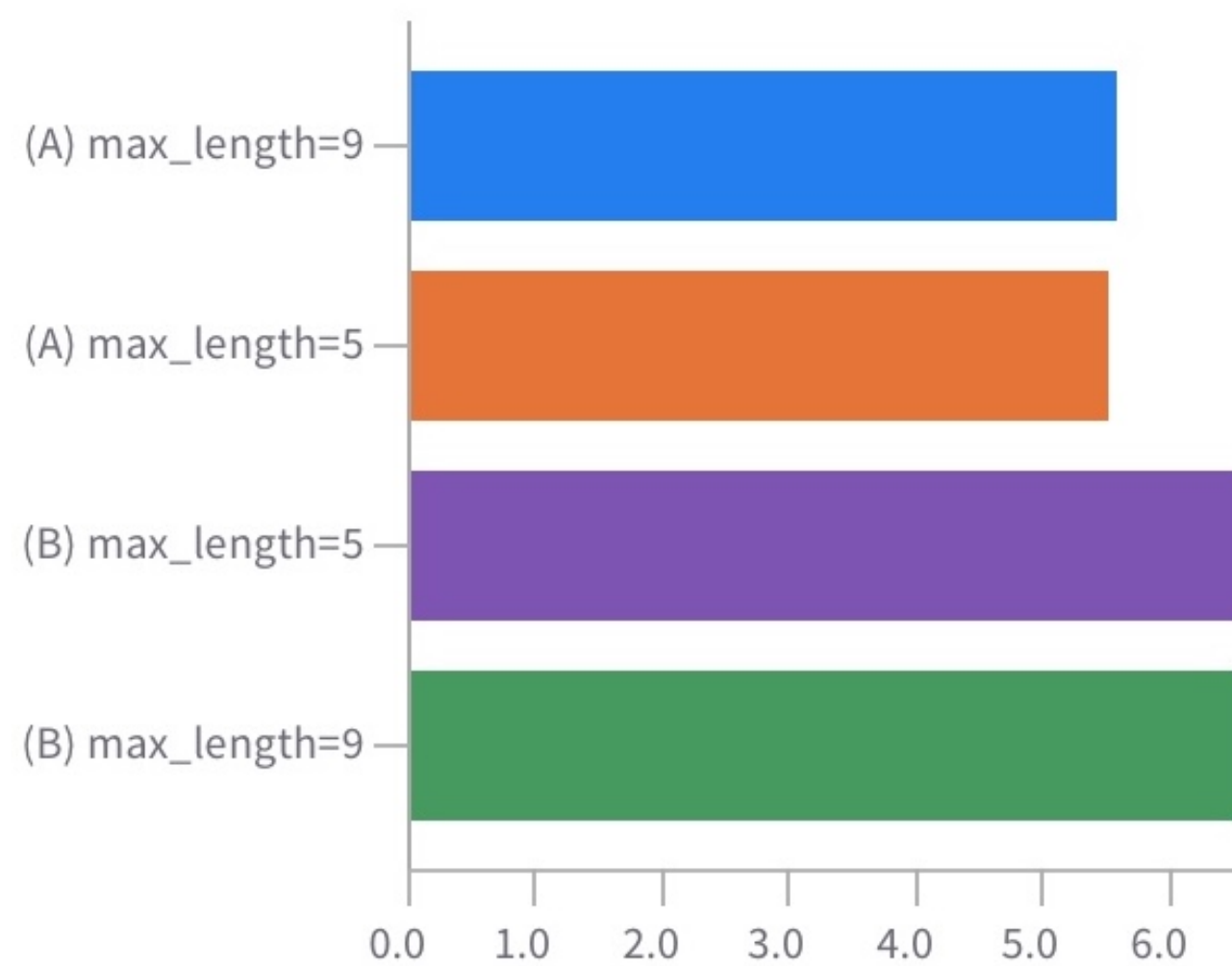
from_token → to_token:

Свойство → ТекущиеДанные

Top-5:

1. Код
2. Банк
3. Наименование
4. **Свойство**
5. Основной

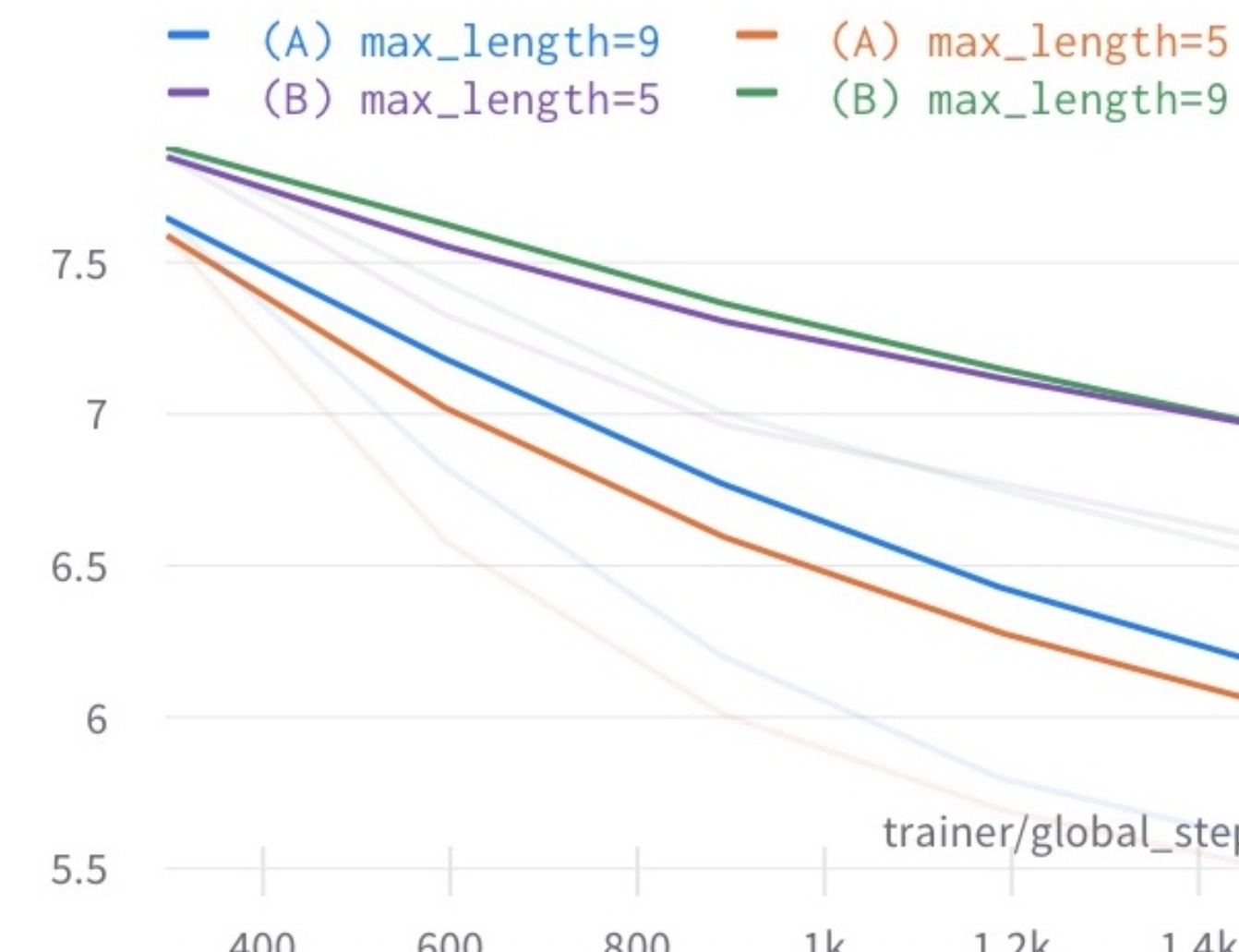
test/loss



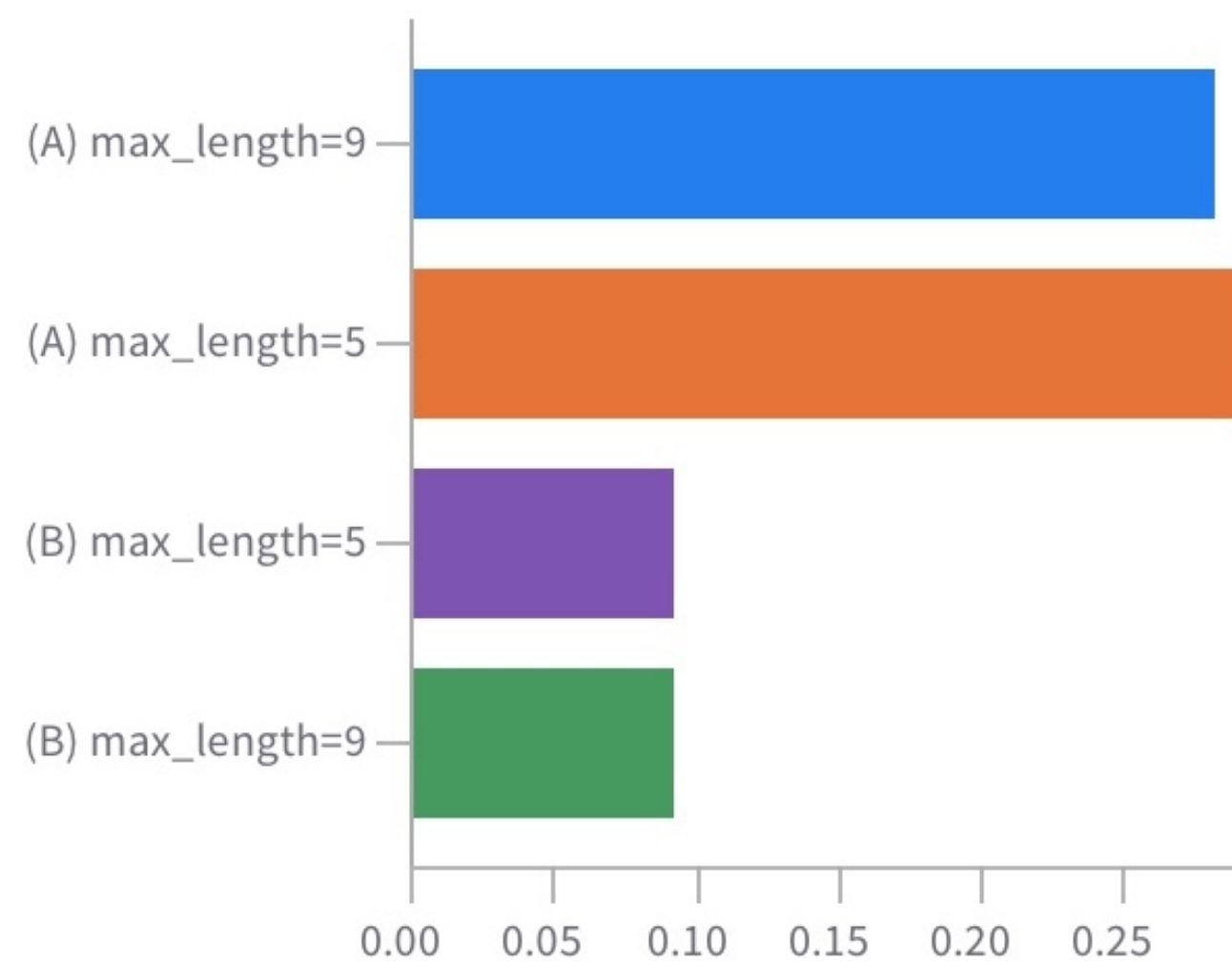
train/loss



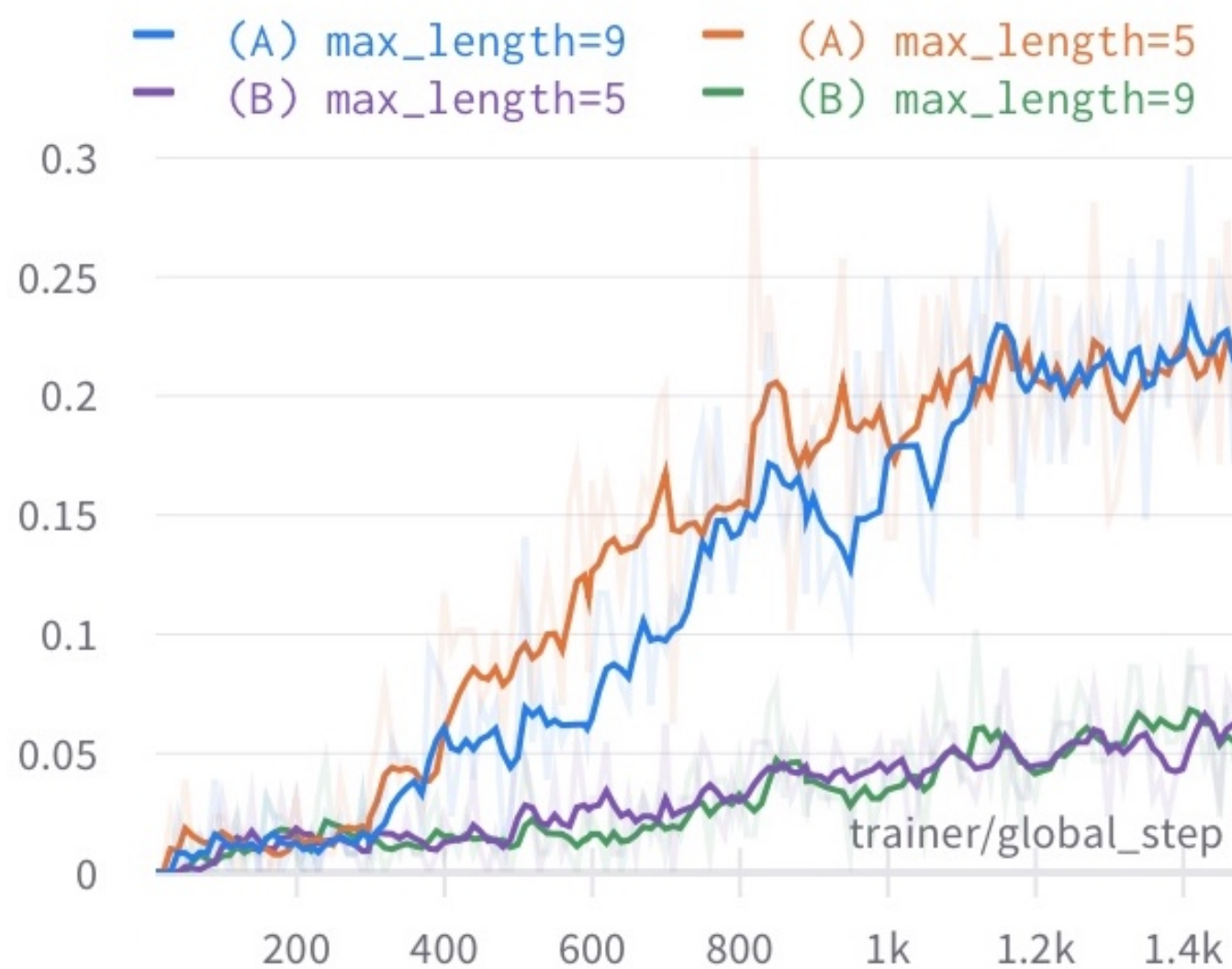
val/loss



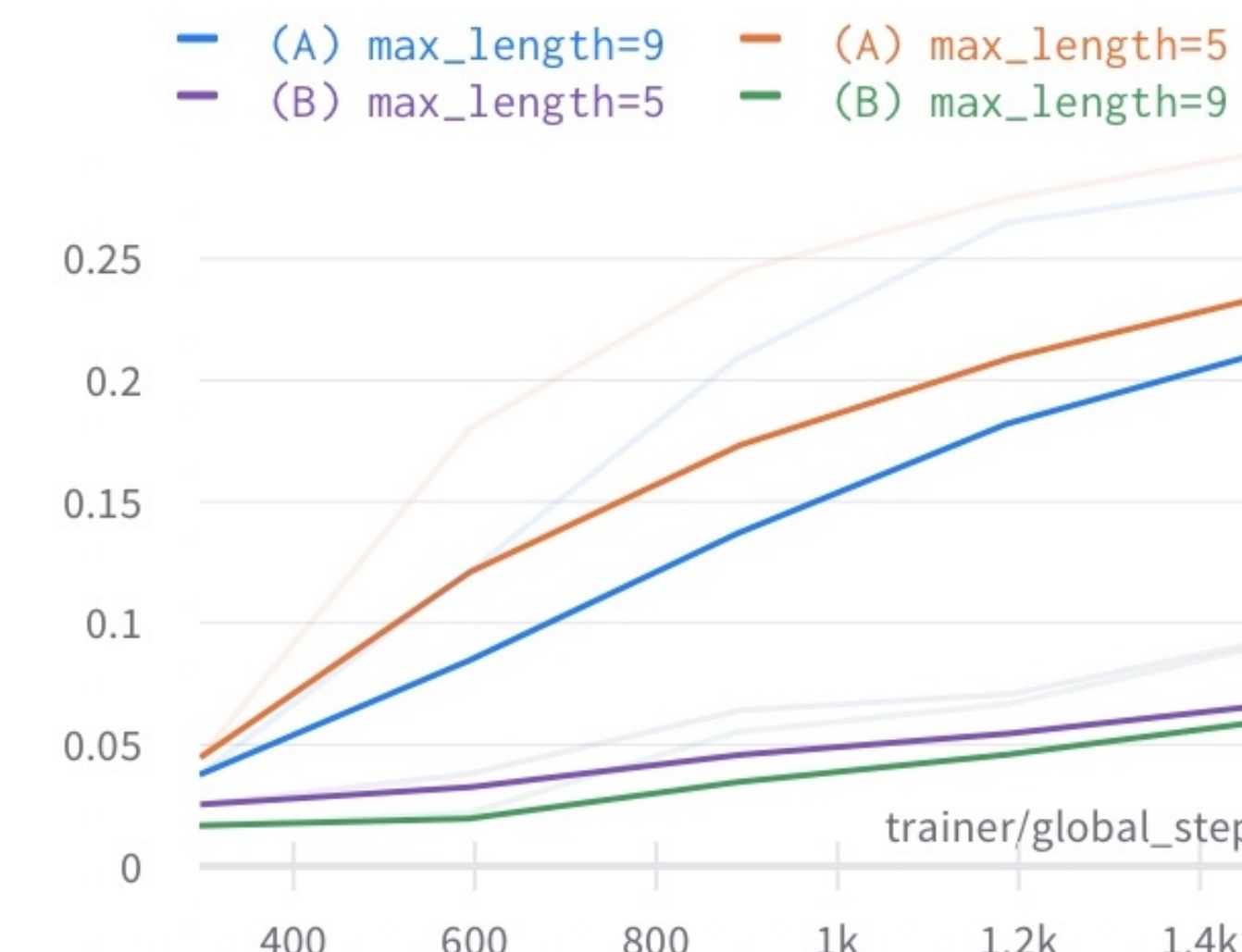
test/precision



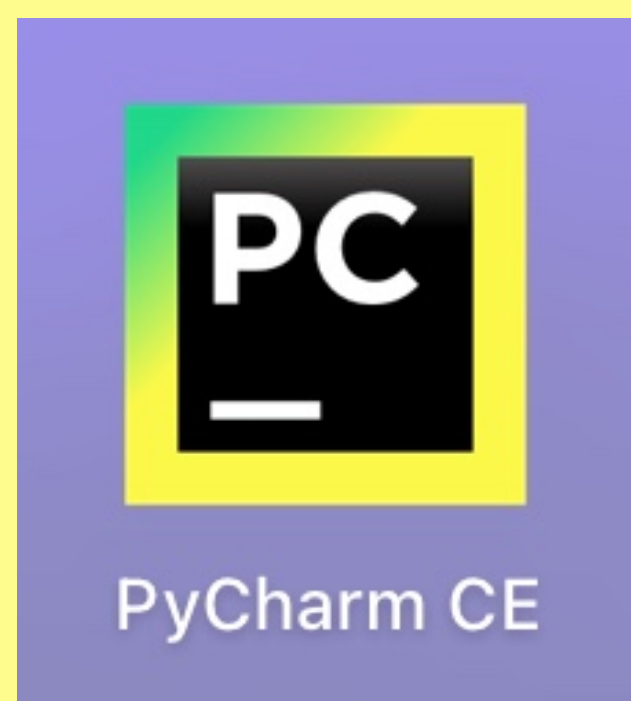
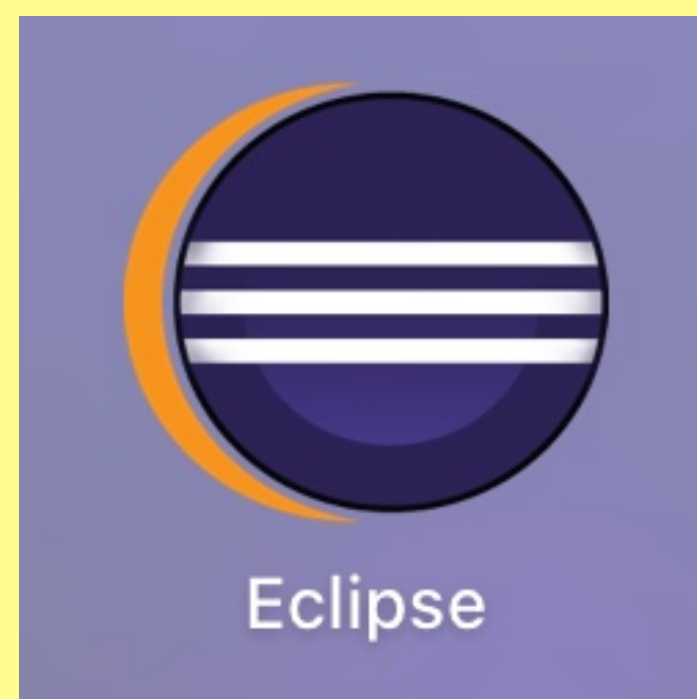
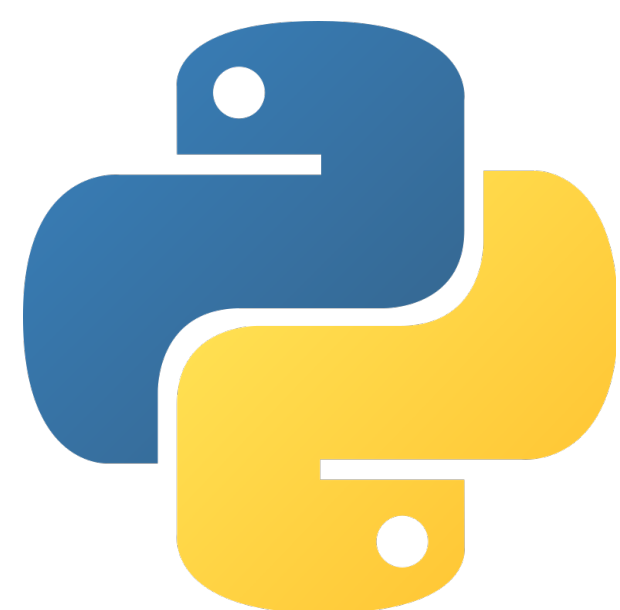
train/precision



val/precision



Использованные технологии



**Машинное обучение для формирования
контекстной подсказки (контент-ассист)
при разработке на платформе
1С:Предприятие**

Спасибо за внимание!

Студент 2 курса магистратуры, М05-912е гр.:
Каразеев А. А.

Научный руководитель:
Дзюба М. В.