

# Определение перечислений в предложениях на английском языке на основе их синтаксического анализа

Исполнитель: Клевцов В.А.

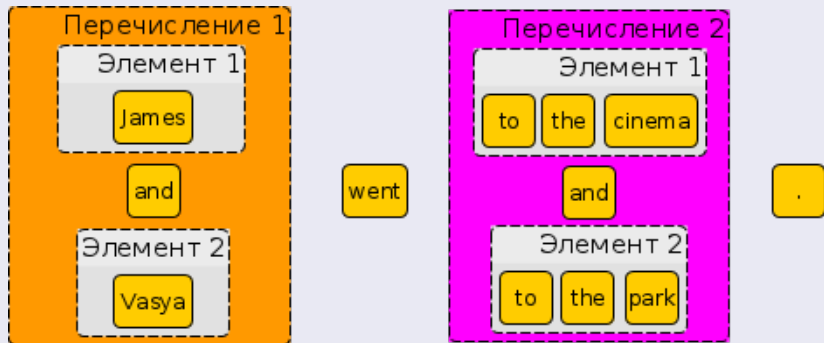
Волгоградский государственный технический университет

Научный руководитель: к.т.н. Сычев О.А.

## Вопрос

Составьте предложение из слов: «Vasya», «went», «park», «John», «and», «to», «the», «and», «cinema», «to», «the».

## Ответ



## Проблема

Трудоемкость создания вопросов, ответами на которые являются предложения с однородными членами и/или сложносочиненными предложениями.

## Подходы к решению

- 1) Отказ от использования вопросов с ответами содержащими однородные члены и/или сложносочиненные предложения.
- 2) Вручную вводить каждый вариант правильного ответа.
- 3) **Использовать автоматизированное определение перечислений**

# Постановка задачи

**Дано:**  $T = \langle A, S, E, t \rangle$

**T** - Эталонный ответ

**A** - Множество лексем эталонного ответа

**S** - Множество описаний лексем эталонного ответа

**E** - Множество описаний перечислений содержащихся в правильном ответе

**t** - Трудоемкость создания эталонного ответа на вопрос

$$t = t_A + t_S + t_E$$

**t<sub>A</sub>** - Трудоемкость ввода лексем эталонного ответа

**t<sub>S</sub>** - Трудоемкость ввода описаний лексем эталонного ответа

**t<sub>E</sub>** - Трудоемкость ввода описаний перечислений содержащихся в правильном ответе

**Цель:** минимизировать **t**.

**Решение:** Разработать функцию:  $F(A)$

**F** - Функция генерации множества описаний перечислений содержащихся в правильном ответе.

Новая трудоемкость будет определяться формулой:

$$t = t_A + t_S + t_{eE}, \forall A, t_E \geq t_{eE}$$

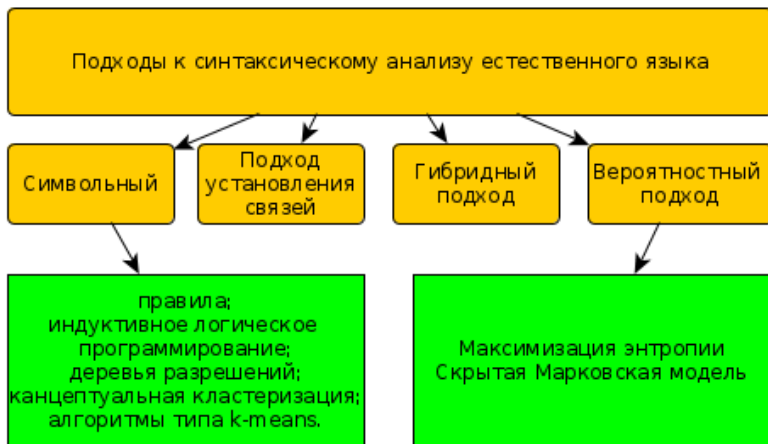
**t<sub>eE</sub>** - Трудоемкость редактирования автоматически определенных описаний перечислений содержащихся в правильном ответе



Целью работы является снижение трудоемкости создания вопросов типа Correct Writing для СУО Moodle за счет автоматического определения перечислений в английском языке на основе синтаксического анализа предложения.

- 1) исследовать подходы, методы и средства обработки естественных языков, выбрать программное средство для осуществления синтаксического анализа;
- 2) исследовать правила английского языка, выбрать структуры языка, изменение порядка записи которых, не изменяют семантику предложения;
- 3) разработать метод выделения перечислений в английском языке на основе синтаксического анализа;
- 4) реализовать и интегрировать метод в плагин Correct Writing, провести тестирование и эксперимент, оценить достижение цели.

# Обзор существующих методов анализа естественного языка

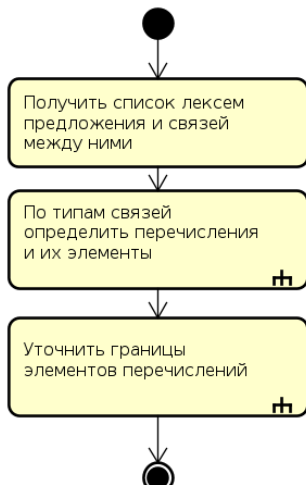




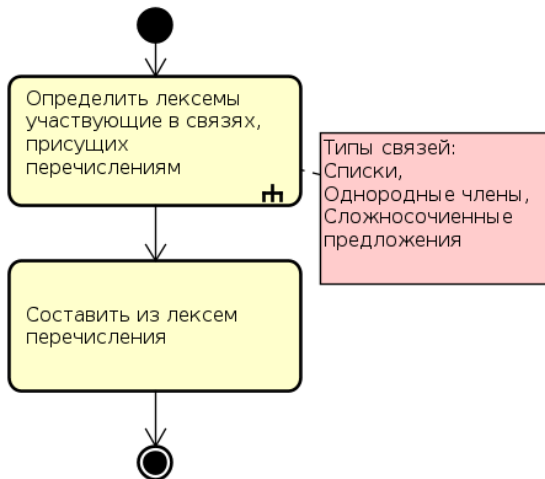
# Выбор средства синтаксического анализа

	Stanford Parser	BLLIP reranking parser	Berkeley Parser	Enju parser
тесты	+	+	+	-
значение лексем	+	+	+	+
лицензия	GPLv2	Apache v2	GPLv2	GPLv2
язык	Java	Python	Java	—
связь с Moodle	отдельный сервер	вызов на сервере	вызов на сервере	отдельный сервер
зависимости между лексемами	+	-	-	+
подход к синтаксическому анализу	гибридный.	вероятностный.	символьный. k-best алгоритм	установка связей. HPSG.

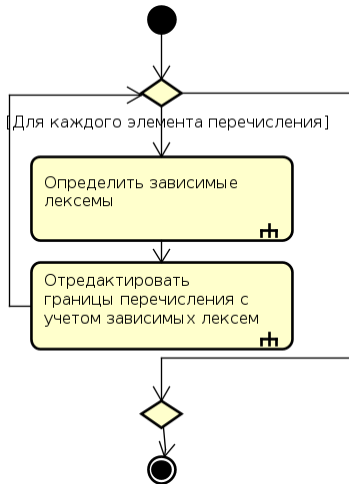
# Метод определения перечислений на основе синтаксического анализа



# Метод определения перечислений на основе синтаксического анализа. Определение перечислений по типам связей лексем предложений.



# Метод определения перечислений на основе синтаксического анализа. Уточнение границ элементов перечислений.



- 1) Написание 3 главы.
- 2) Статья по результатам глав 1 и 2.

- 1) Создано средство ввода тестов.
- 2) Начат ввод тестов.
- 3) Готовится статья по результатам 2-х семестров.
- 4) Написана 1 глава.

## Научная новизна работы

- 1) Использование метода определения перечислений на основе синтаксического анализа в новой области естественный языков.
- 2) Разработанное метод позволяет с высокой точностью определить перечисления в предложении на естественном языке.

## Практическая ценность работы

Реализованное программное средство позволяет проводить оценку ответа студента на естественном языке, с учетом содержания в ответе однородных членов и сложносочиненных предложений.

- 1) Готовится статья "Учет перечислений с произвольным порядком элементов при сравнении порядка лексем в тестовых вопросах с открытым ответом" в журнал "Вестник компьютерных и информационных технологий".
- 2) Определение перечислений в предложениях на английском языке на основе их синтаксического анализа, Клевцов В.А., Внутривузовская научная конференция ВолгГТУ, Волгоград, 2016 г.



Спасибо за внимание!

L<sup>A</sup>T<sub>E</sub>X