

Дипломная работа

Механизмы обеспечения связности в системах динамической генерации текстов

Автор:

Лукьяненко Анастасия Юрьевна

Научный руководитель:

к.ф.-м.н. Бордаченкова Елена Анатольевна

Общение человека с компьютером на естественном языке



Задача распознавания текстов на естественном языке

- Морфологический анализ
- Синтаксический анализ
- Семантический анализ

Задача синтеза текстов на естественном языке

- Статическая генерация текстов
- **Динамическая генерация текстов**

✓ Связность – важный критерий качества синтезируемых текстов

Постановка задачи

1. Изучить лингвистические средства достижения связности в текстах, производимых человеком в условиях динамичной среды
2. Проанализировать существующие системы динамической генерации текстов
 - С точки зрения применяемых в них алгоритмов генерации
 - С точки зрения связности производимых ими текстов
3. Выбрать лингвистические механизмы обеспечения связности, применимые при динамической генерации, но пока не реализованные
4. Создать экспериментальное программного обеспечения для динамического синтеза текстов
 - Предметная область - Биатлон
 - Реализовать некоторые выбранные недостающие механизмы обеспечения связности

Связность текста с точки зрения ЛИНГВИСТИКИ

1. Смысловая связность

- Регулярное повторение ключевых слов

2. Коммуникативная связность

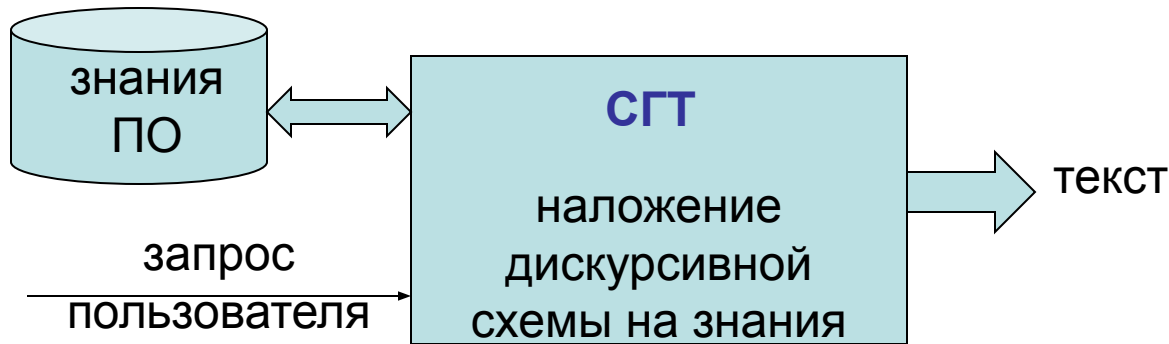
- Организация тема-рематических потоков

3. Структурная связность

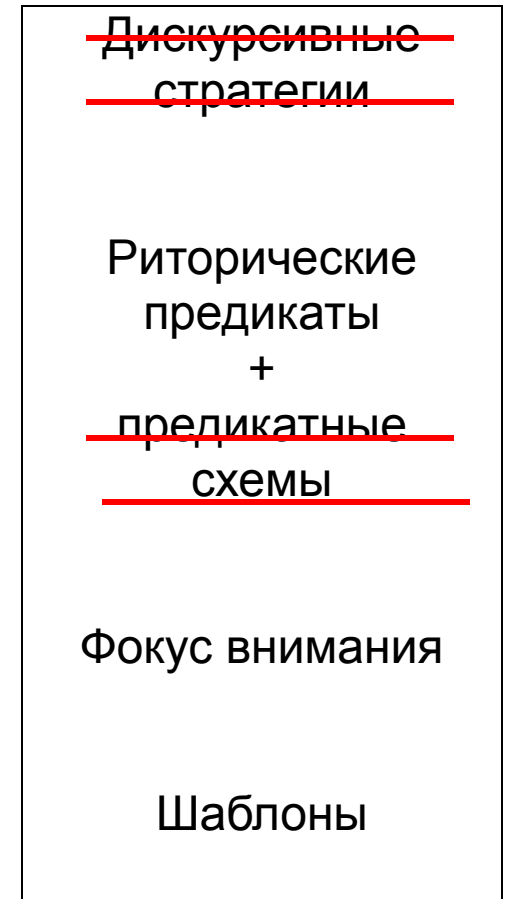
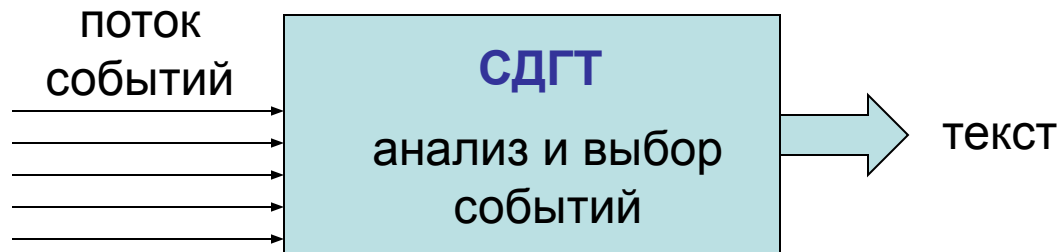
- Грамматические средства связи
- Композиционно-структурные средства связи
- Ритмикообразующие средства связи
- Стилистические средства связи

Генерация текстов

- Статическая генерация



- Динамическая генерация



Сравнение естественных и машинных текстов

Человек

1. Референция

- ✓ Непосредственный повтор
- ✓ Местоимённая замена
- ✓ Синонимическая замена
- ✓ Эллиптический повтор

2. Связность

- ✓ Анафора, близкие связи и дальние
- ✓ Катафора, близкие связи и дальние
- ✓ Соединительные слова
- ✓ Сложные предложения и обороты
- ✓ Указание временных параметров
- ✓ Разнообразие форм глаголов

3. Эмоциональность

- ✓ Повторы
- ✓ Восклицания и вопросы
- ✓ Обращения

СДГТ

1. Референция

- ✓ Непосредственный повтор
-
-
- **Эллиптический повтор**

3. Связность

- ✓ Анафора, близкие связи
- ✓ Катафора, близкие связи
- **Соединительные слова**
- ✓ Сложные предложения
- **Указание временных параметров**
- **Разнообразие форм глаголов**

4. Эмоциональность

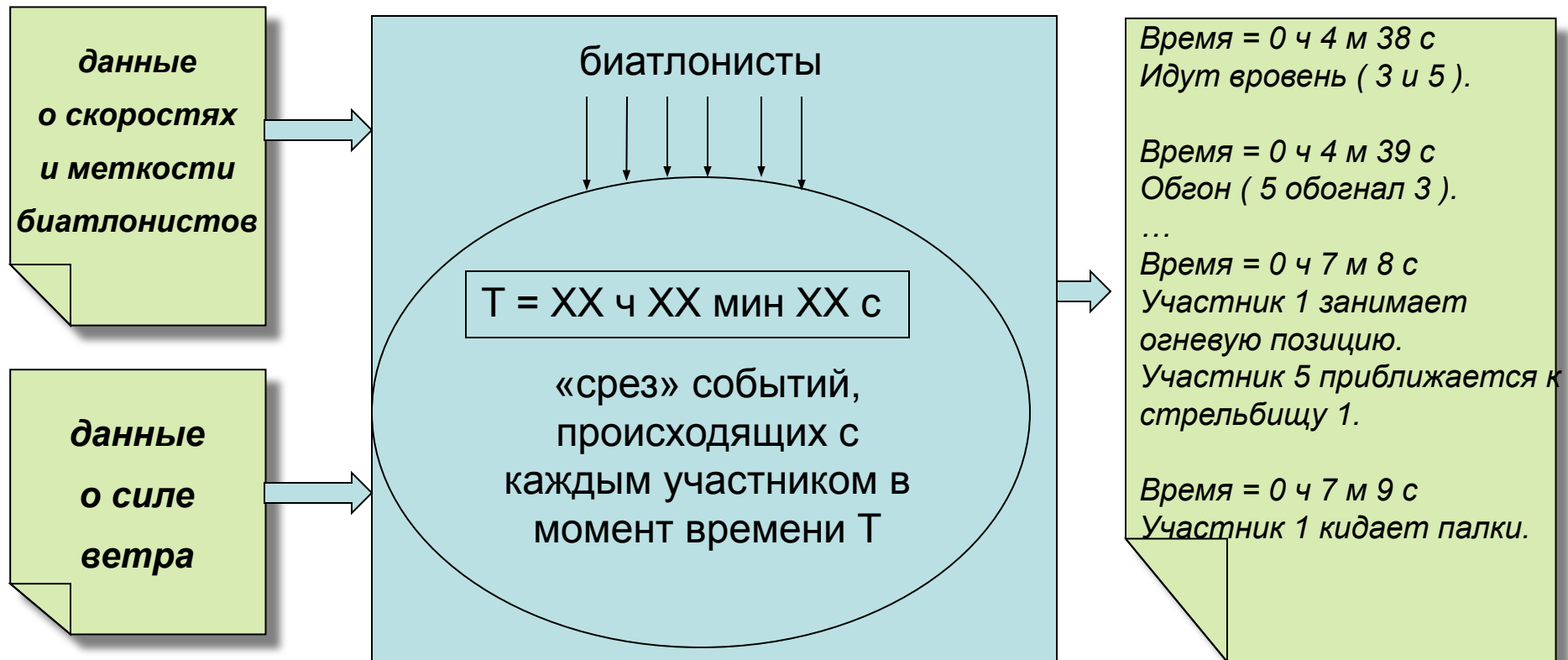
- **Повторы**
- ✓ Восклицания
- **Обращения**

Экспериментальное программное обеспечение, синтезирующее текст

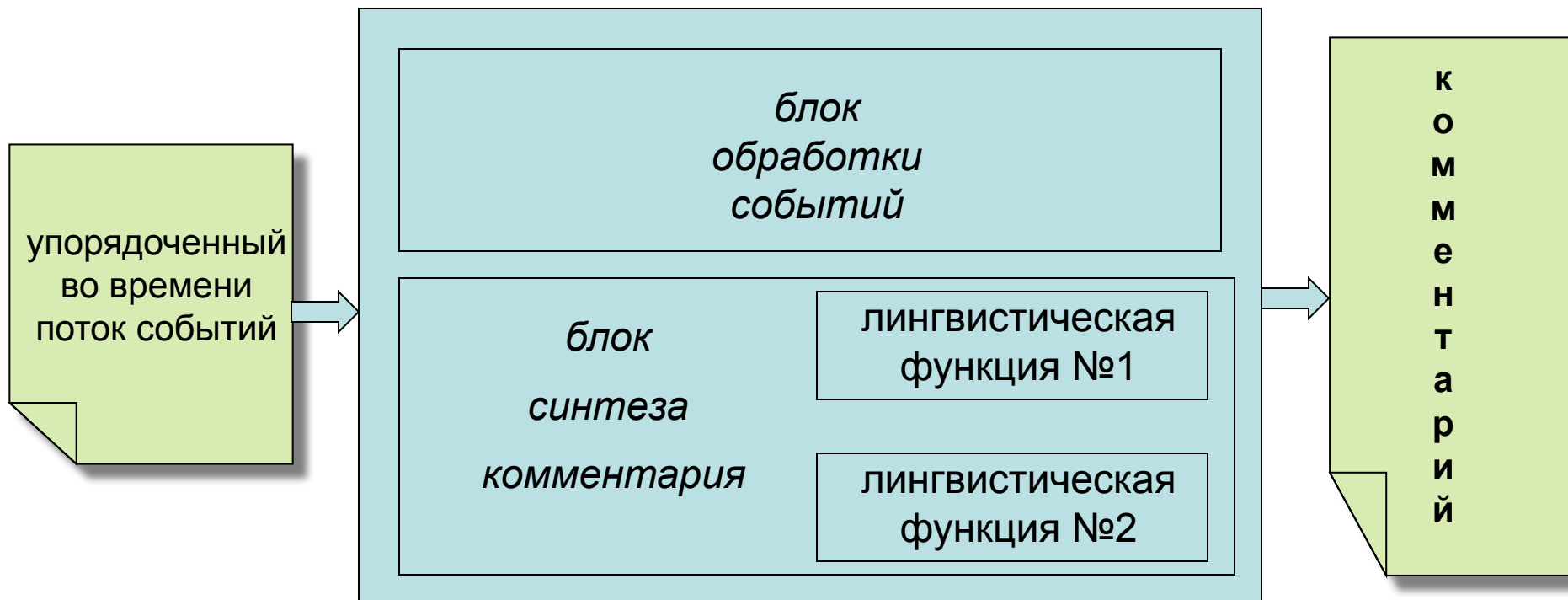


Генератор индивидуальной гонки по биатлону

✓ Анализ и классификация событий реальных соревнований по биатлону



Генератор динамического связного комментария



Реализация двух лингвистических механизмов

- ✓ Использование слов-связок в начале предложения
- ✓ Использование глаголов прошедшего времени

Пример работы автоматического комментатора

Вульф быстрым движением разбирается с предохранителем.

Важный момент, прицеливается перед выстрелом Вульф.

Стреляет.

Меткий выстрел!

Вульф прицеливается.

Выстреливает по мишени.

А теперь промахивается.

Прицеливается.

Бьёт по мишени Вульф.

Прямо в яблочко!

...

Вульф выстреливает по мишени.

Но на этот раз мимо!

...

Стреляет.

Да, снова выстрел не попал в цель.

Закончил стрельбу Вульф.

Одевает ружьё.

Надевает перчатки.

Хватает палки.

Вульф получает 3 штрафные минуты, стабильный результат, на прошлом рубеже был такой же.

Основные результаты

В рамках данной работы были получены следующие результаты:

1. Изучены тексты на естественном языке, человеческие и машинные, с точки зрения используемых в них механизмов достижения связности
2. Выявлены те лингвистические механизмы, реализация которых возможна в рамках систем динамической генерации текстов
3. Создано экспериментальное программное обеспечение, динамически генерирующее тексты, включающее реализацию следующих лингвистических механизмов
 - ✓ Использование соединительных слов в начале предложений
 - ✓ Использование разнообразных форм глаголов-сказуемых

Спасибо за внимание!