



Московский государственный  
технический университет  
имени Н.Э. Баумана



Кафедра ИУ5  
«Системы обработки информации  
и управления»

# **Модуль формирования концептуальных карт на основе анализа текста на русском языке**

**Студент группы РТ5-81Б  
Руководитель ВКРБ**

**Евсеев Георгий Андреевич  
Гапанюк Юрий Евгеньевич**

# Актуальность

В настоящее время наблюдается рост использования концептуальных карт. Самое частое применение – сбор и представление уже полученных знаний в форме, удобной для восприятия. Более того, концептуальные карты известны как эффективный инструмент для организации и навигации по большим объемам данных.

## **Проблемы при построении концептуальной карты:**

- Сложно составить карту с нуля по выбранной области
- Каркас схемы, созданный экспертом может облегчить задачу, но зачастую сложно найти такого человека.

# Цели и задачи

**В ходе выполнения квалификационной работы были определены следующие цели:**

- Исследование предметной области формирования концептуальных карт;
- Исследование области обработки естественного языка;
- Изучение, а так же проведение сравнительного анализа имеющихся, на данный момент, фреймворков и библиотек;
- Выбор модели машинного обучения, улучшающую предсказания;
- Программная реализация модуля;
- Тестирование модуля;

# Концептуальная карта

$$CM = \{C, R, T\}$$

$$C = \{c_1, c_2, \dots, c_n\}$$

$$R = \{r_1, r_2, \dots, r_m\}; r_i \in R = \{c_p, c_q, l_j\}; 1 \leq p \leq n; 1 \leq q \leq n; 1 \leq j \leq m;$$

$$c_q c_p \in C; p \neq q$$

$$T = \{t_1, t_2, \dots, t_s\}; t_{k-1} \leq t_k \leq t_{k+1}; k < s$$



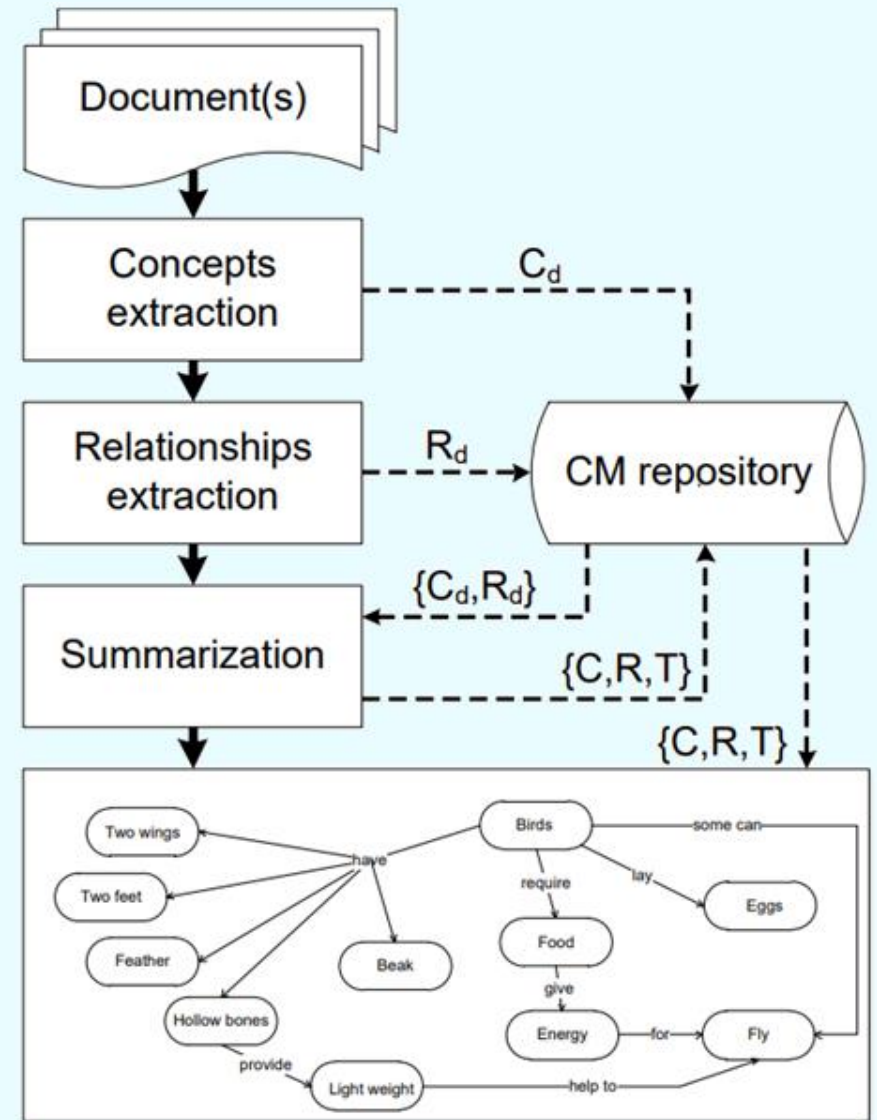
# CMM – Concept Map Mining

Concept map mining представляет собой процесс извлечения информации из одного или более документов для автоматического создания концептуальной карты. Созданная концептуальная карта является обобщенной сводкой исходного текста. (J. J. Villalon and R. A. Calvo, "Concept Map Mining: A Definition and a Framework for Its Evaluation," *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Sydney, NSW, Australia, 2008, pp. 357-360, doi: 10.1109/WIIAT.2008.387.)

$$D = \{C_d; R_d\}$$

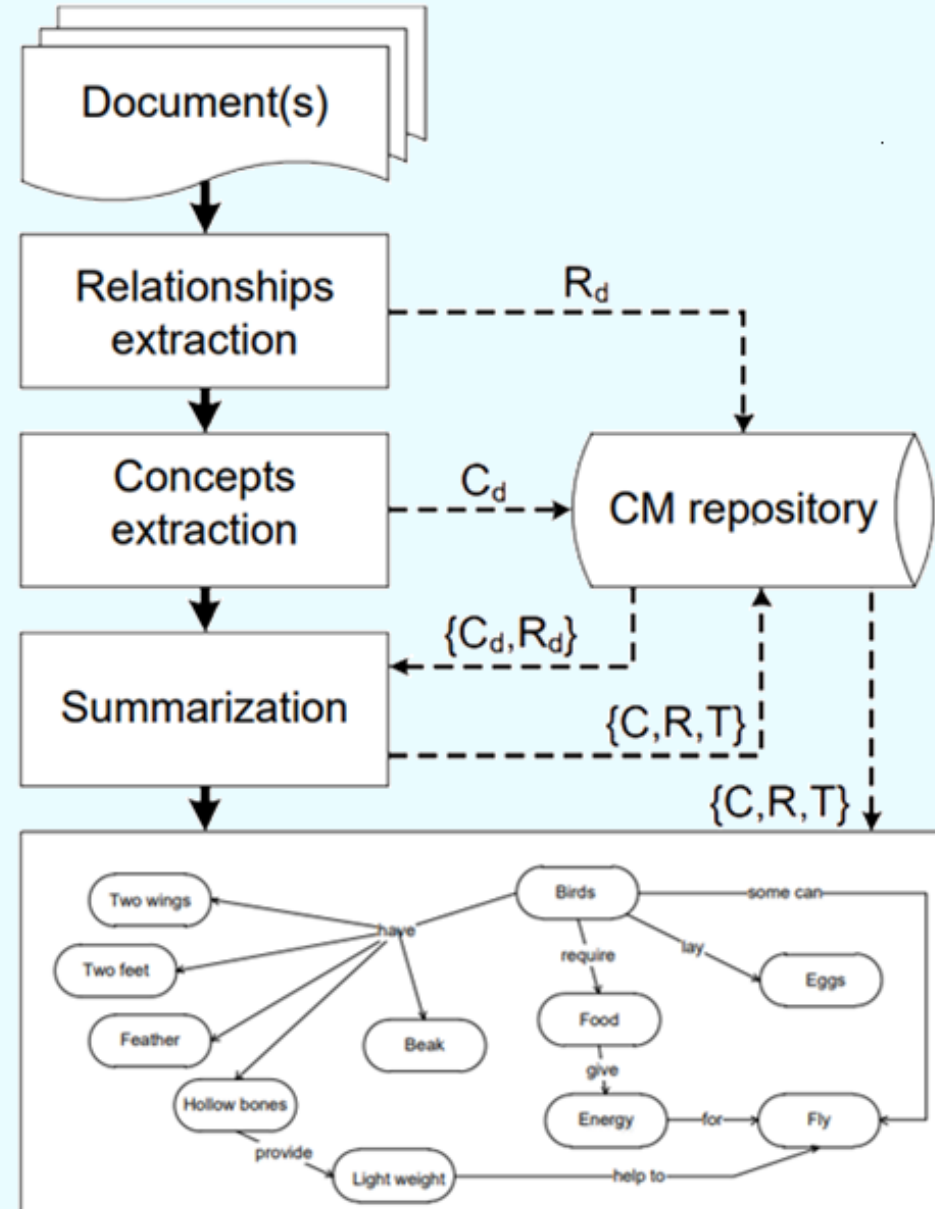
$$C_d = \{c_{d1}, c_{d2}, \dots, c_{dn}\}$$

$$R_d = \{r_{d1}, r_{d2}, \dots, r_{dn}\}$$



# Модифицированный метод СММ

Модификация заключается в перестановке этапов извлечения концептов с извлечением связи, в связи с чем сменятся и алгоритмы извлечения связей и извлечения концептов.



# Существующие решения для разработки модуля

spaCy



 Stanza

natasha



# Сравнительный анализ существующих решений

Критерии оценки и их коэффициент важности:

Легкость установки	0,156
Возможность токенизации	0,251
Легкость использования	0,174
Вид зависимости парсера	0,235
Возможность визуализации	0,195
Возможность переобучения моделей	0,275

Конечный результат сравнительного анализа:

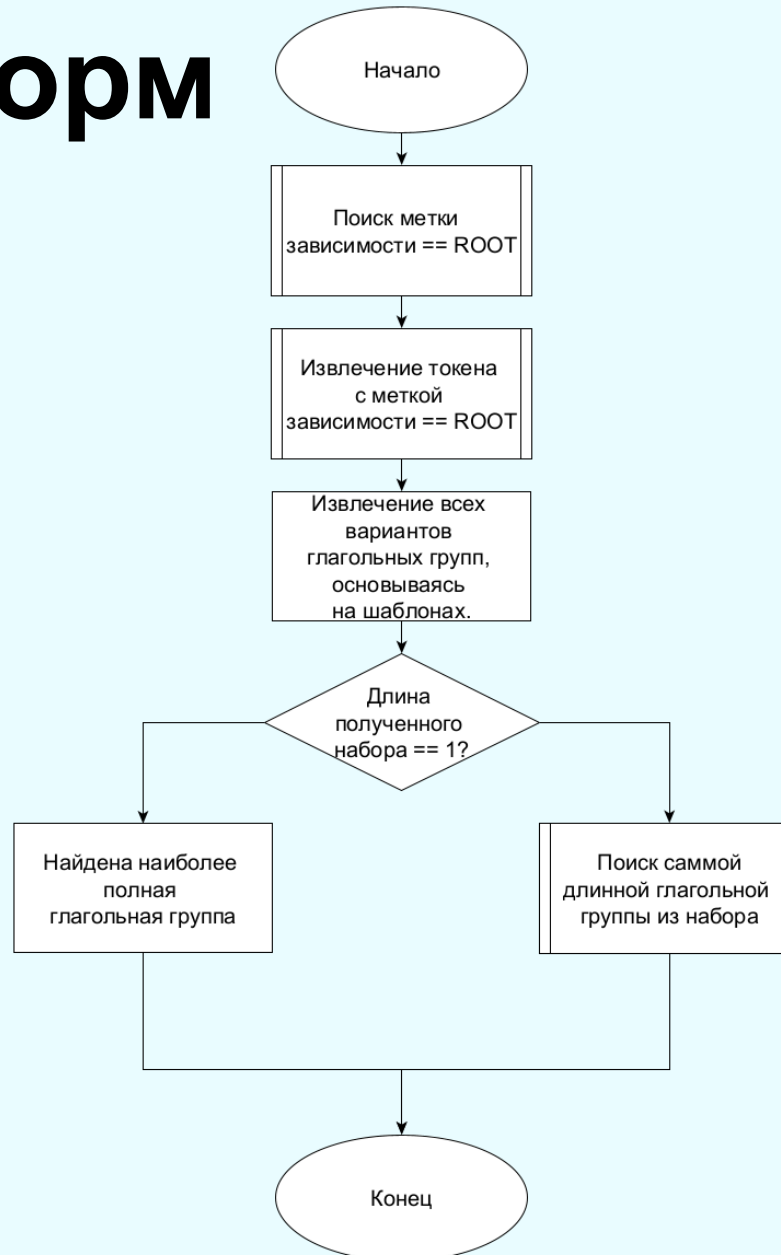
Spacy	Slovnet	Stanza	NLTK	StanfordNLP	DeepPavlov
1,290	0,695	0,888	0,645	0,721	1,290



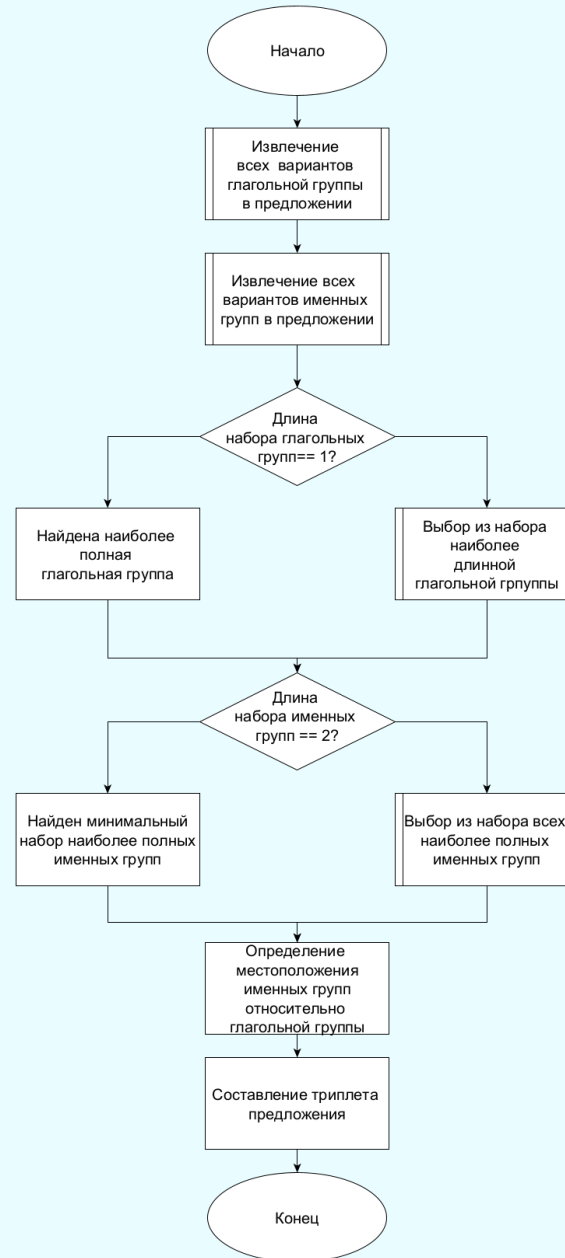
# Алгоритм работы модуля формирования концептуальных карт на основе анализа текста для русского языка



# Алгоритм работы этапа извлечения глагольных форм

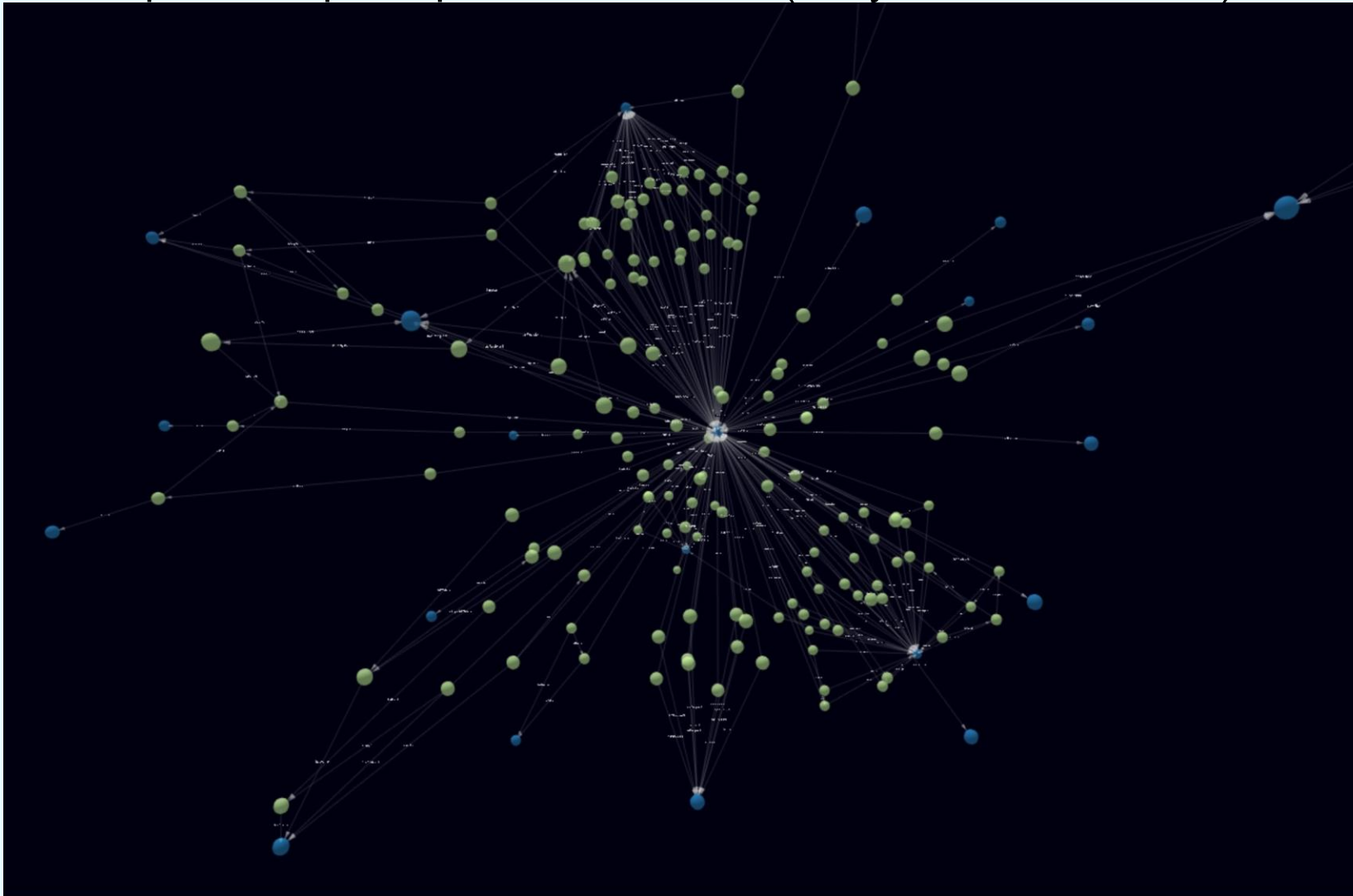


# Алгоритм работы этапа составления триплетов



# Предложение/замечание

В качестве предложения была высказана идея учета времени как элемента рассмотрения характеристики объекта (получение контекста).



# Оценка точности созданной концептуальной карты

Оценка будет производиться с использованием концептуальных карт экспертов, которые будут приняты как “Золотой стандарт”. (Hearst M.A. The debate on automated essay grading //IEEE Intelligent Systems and their Applications, vol. 15, no. 5, pp. 22-37, Sept.-Oct. 2000, doi: 10.1109/5254.889104.)

$$HMA - HIAA$$

где HMA – human-machine agreement, а HIAA – human inter-annotator agreement.

$$HIAA = k = \frac{(p_o - p_e)}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

где а  $p_o$  – относительная согласованность между экспертами, а  $p_e$  – Гипотетический вероятность шанса согласия между экспертами, если все эксперты сделали карту совершенно случайно.

# Выводы

**При разработке модуля формирования концептуальных карт на основе анализа текста для русского языка были достигнуты следующие результаты:**

- Исследована предметная область обработки естественного языка;
- Исследована предметная область формирования концептуальных карт;
- Был проведен сравнительный анализ имеющихся, на данный момент, фреймворков и библиотек;
- Выбрана модель машинного обучения, улучшающую предсказания;
- Разработан модуль формирования концептуальных карт на основе анализа текста для русского языка;
- Успешно проведены тесты разработанного модуля;