



РНИМУ
имени Н. И. Пирогова



СЕМАНТИЧЕСКИЙ АНАЛИЗ НЕСТРУКТУРИРОВАННОГО ТЕКСТА – ФОРСАЙТ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В МЕДИЦИНЕ

Зарубина Т.В.
директор Института цифровой
трансформации медицины

ИЦТМ – ЧАСТЬ СТРАТЕГИЧЕСКОГО ПРОЕКТА «ПРИОРИТЕТ-2030»



Кафедра медицинской кибернетики и информатики всегда занималась разработкой СППКР

Год назад в рамках «ПРИОРИТЕТ-2030» создан ИЦТМ

Несколько направлений исследований, среди которых семантический анализ неструктурированной медицинской информации

Почему?

- Потребность в анализе неструктурированного текста во всех клинических профилях
- Разработка СЭМДов полноценно всех проблем не решит

Что дает структурирование и формализация медицинской информации?

- уменьшению врачебных ошибок - управление на основе данных
- внедрение в рабочие процессы врача - средств интеллектуального сопровождения при работе с медицинской и биологической информацией - систем поддержки принятия решений
- создание инструментов для решения задач классификации при создании наборов данных для обучения интеллектуальных систем на основе «больших данных»

ИСТОЧНИКИ МЕДИЦИНСКОЙ ИНФОРМАЦИИ

Основой анализа неструктурированной информации является база медицинских знаний (БМЗ), хранящая семантические закономерности, полученные из текста (концепты и связи). Наиболее эффективный способ представления БМЗ – графовая модель.

Метатезаурусы и справочники:

- Метатезаурус Unified Medical Language System (UMLS) – 4.55 млн объектов, 90 млн связей (SNOMED CT, LOINC, RxNORM);
- Справочники ФРНСИ

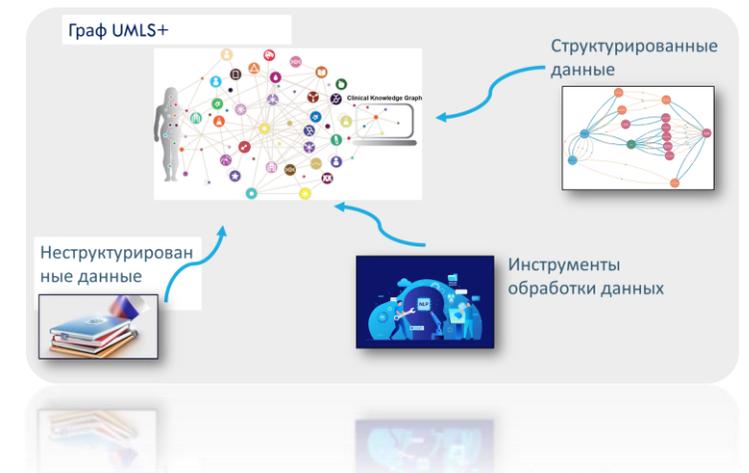
Медицинские публикации и нормативные документы:

- статьи, учебники,
- Клинические рекомендации (КР), порядки оказания МП, протоколы ведения пациентов

Документы ЭМК

МЕТОДОЛОГИЯ

1. Анализ метатезауруса (UMLS): выделение необходимых групп концептов, типов связей, входящих справочников, аналогов представления понятий, переводов на русский язык.
2. Применение математических алгоритмов, позволяющих выделять сущности или паттерны внутри графовых моделей (поиск релевантного окружения концептов в семантической сети UMLS).
3. Обработка текстов (статьи, реальные ЭМД) алгоритмами машинного обучения (в первую очередь, нейронными сетями) с целью получения недостающих связей, сущностей.
4. Построение адаптированной русскоязычной версии БМЗ, куда будут включены результаты обработки текстов, а в качестве мапированной терминологии будут использованы словари ФРНСИ.
5. Построение «образа» каждой нозологической формы (в виде графовой модели), отражающей связи заболевания с его признаками (симптомами, дифференциальными диагнозами, диагностическими методами, лечебными мероприятиями).
6. Реализация алгоритмов принятия решений с привлечением экспертов для построения СППКР.



Отечественные симптомчекеры: анализ по 5 нозологиям (заболевания органов дыхания) на основе симптомов, указанных в КР

Критерии сравнительной оценки:

- Соответствие введенного симптома имеющемуся в симптомчекере
- Возможность использования при вводе свободного текста
- Количество распознанных и нераспознанных симптомов
- Наличие искомого диагноза в топ-5 выдаваемых диагнозов
- Расчёт вероятности или ранжирование диагнозов в результатах
- Порядковый номер искомого диагноза в перечне итоговых диагнозов
- Помощь/ подсказки при выборе симптомов
- Обоснование диагноза (указанием значимых признаков)
- Наличие справочной информации о диагнозе
- Наличие рекомендуемых методов диагностики
- Удобство работы
- Дополнительные функции (например, отправить врачу)

Helzy	https://helzy.ru/
Symptomchecker	https://symptomchecker.ru/
Medaboutme	https://medaboutme.ru/symptomes/
Health.mail	https://health.mail.ru/symptoms/
Diagnos.ru	https://www.diagnos.ru/
Docmate	https://docmate.site/
Symptom-checker.drclinic	http://symptom-checker.drclinics.ru/
Online-diagnos	https://online-diagnos.ru/diagnostics

НАШ ОПЫТ

1. Проанализирован метатезаурус UMLS, изучена структура, адаптированы переводы в части терминов, имеющих отношение к заболеваниям ДС.
2. Разработаны модули аналитической системы: поиск релевантного окружения концептов в семантической сети UMLS; расчет углового расстояния между концептами с использованием контекстной близости сущностей (векторных эмбедингов); поиск кратчайшего пути в графе между заданным корневым и концевым концептом. Проанализированы закономерности (наиболее часто встречающиеся группы терминов, типы и подтипы связей) между всеми участвующими узлами полученных графовых цепочек.
3. Формализованы клинические рекомендации в области диагностики неопухолевых заболеваний дыхательной системы. Выделенные из КР симптомы и диагнозы мапированы с терминами UMLS с помощью аналитической системы.
4. Разработан прототип ИПС, использующий все созданные аналитические алгоритмы, для решения задачи интеллектуального сопровождения при работе с медицинской информацией.
5. В прототипе ИПС реализован алгоритм обработки неструктурированных медицинских текстов с целью возможности ввода информации о заболевании свободным текстом.



СЕМАНТИЧЕСКИЙ АНАЛИЗ КР



1.4 Особенности кодирования заболевания или состояния (группы заболеваний или состояний) по Международной статической классификации болезней и проблем, связанных со здоровьем

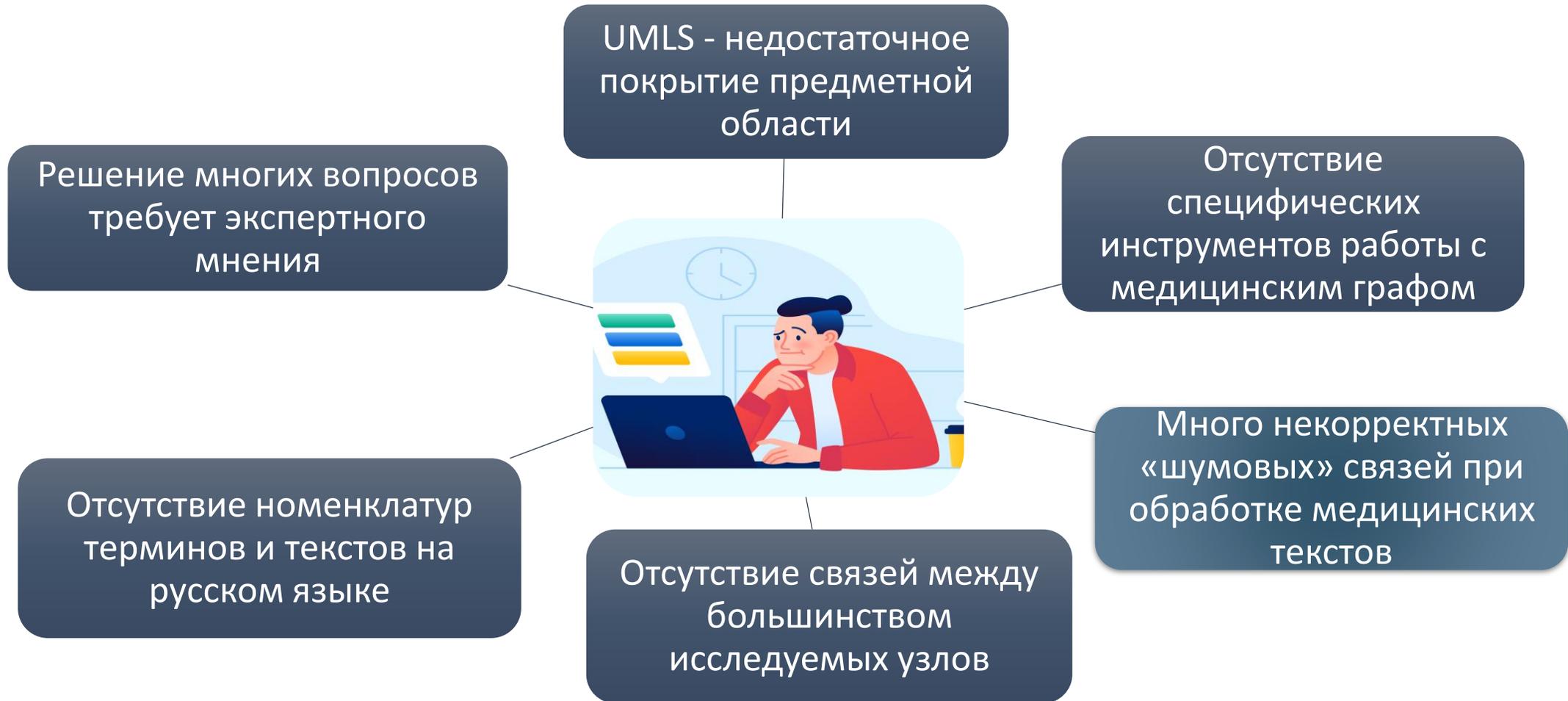
Перечень нозологических форм с указанием кодов по МКБ-10

2. Диагностика заболевания или состояния (группы заболеваний или состояний) медицинские показания и противопоказания к применению методов диагностики

Перечень признаков с атрибутами:

- Метод получения
- Тип (симптом, синдром, фактор риска, заболевание в анамнезе)
- Частота встречаемости
- Уровень убедительности и достоверности
- Условия появления и др.

ВЫЯВЛЕННЫЕ ПРОБЛЕМЫ



ДАЛЬНЕЙШИЕ ШАГИ

Проектирование
ИПС

Подключение
дополнительных
источников
информации
(статьи, ЭМК)

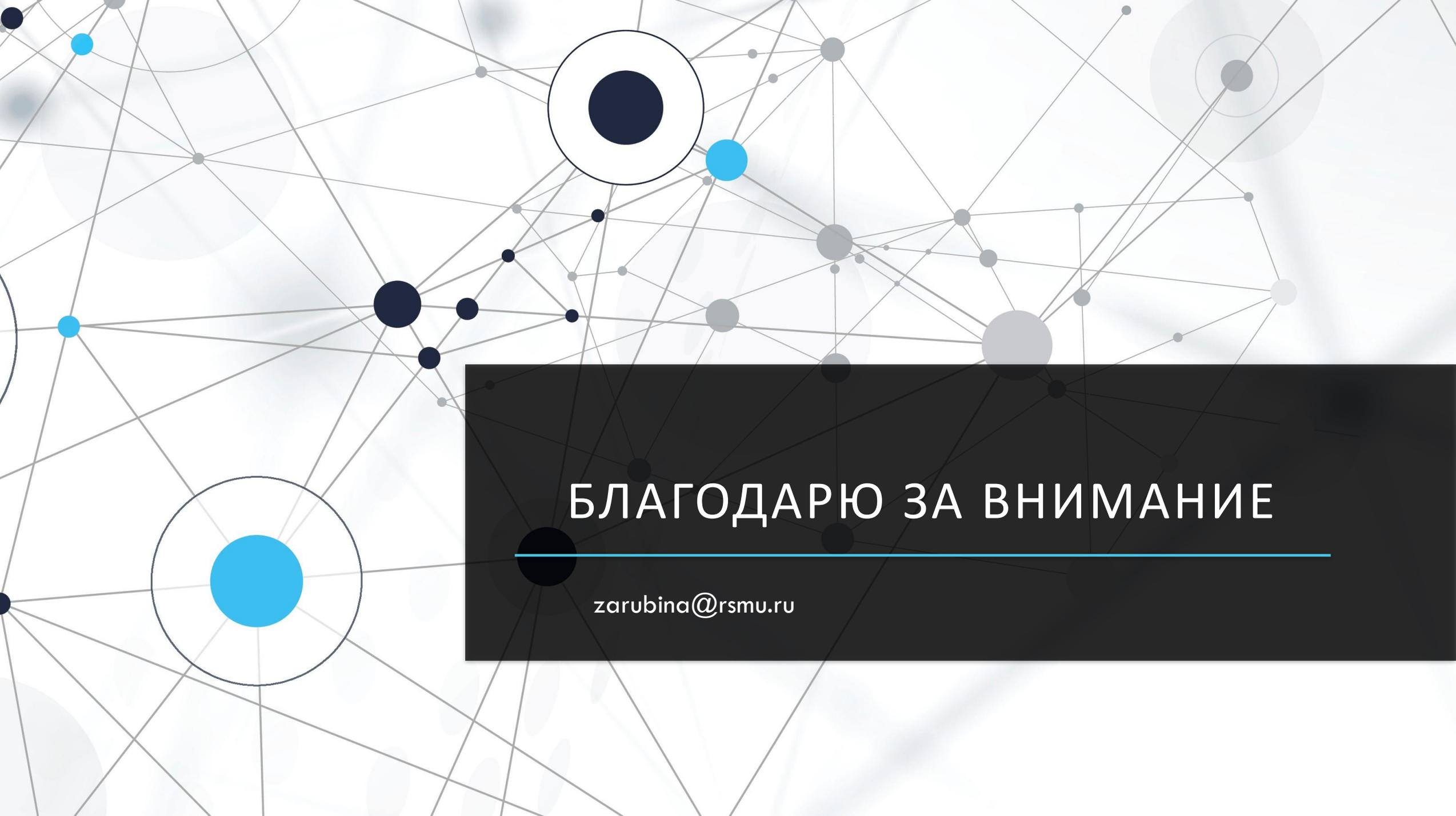
Создание БМЗ

Построение
«образов»
нозологий с
привлечением
экспертов

Разработка
СППКР

Промышленная
реализация (с
привлечением
индустриального
партнера)

Успешный семантический анализ неструктурированного текста может стать форсайтом искусственного интеллекта в медицине как систематический процесс построения видения будущего, нацеленного на повышение качества принимаемых решений



БЛАГОДАРЮ ЗА ВНИМАНИЕ

zarubina@rsmu.ru