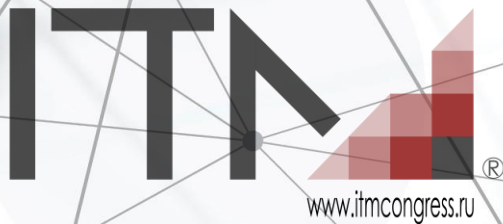


ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В МЕДИЦИНЕ



ПОДХОД К ПОСТРОЕНИЮ БАЗЫ ЗНАНИЙ НА ОСНОВЕ СЕМАНТИЧЕСКОГО АНАЛИЗА МЕДИЦИНСКОЙ ИНФОРМАЦИИ

Королева Ю.И., Раузина С.Е., Николаиди Е.Н.
РНИМУ им.Н.И.Пирогова Минздрава России

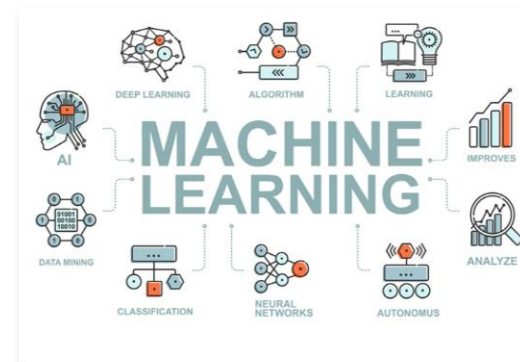


ИНСТИТУТ ЦИФРОВОЙ ТРАНСФОРМАЦИИ МЕДИЦИНЫ - СТРАТЕГИЧЕСКИЙ ПРОЕКТ 2030



«Создание научно-образовательной экосистемы для обеспечения здравоохранения принципиально новыми инструментами поддержки деятельности медицинских работников и исследователей, созданными и развиваемыми на основе инновационных цифровых технологий, искусственного интеллекта и Big Data и подготовки высококвалифицированных кадров с высокой степенью исследовательской и цифровой готовности»

Цель работы: разработка базы знаний на основе семантического анализа медицинской информации



ИСТОЧНИКИ МЕДИЦИНСКОЙ ИНФОРМАЦИИ

Метатезаурус Unified Medical Language System (UMLS) – 4.55 млн объектов, 90 млн связей

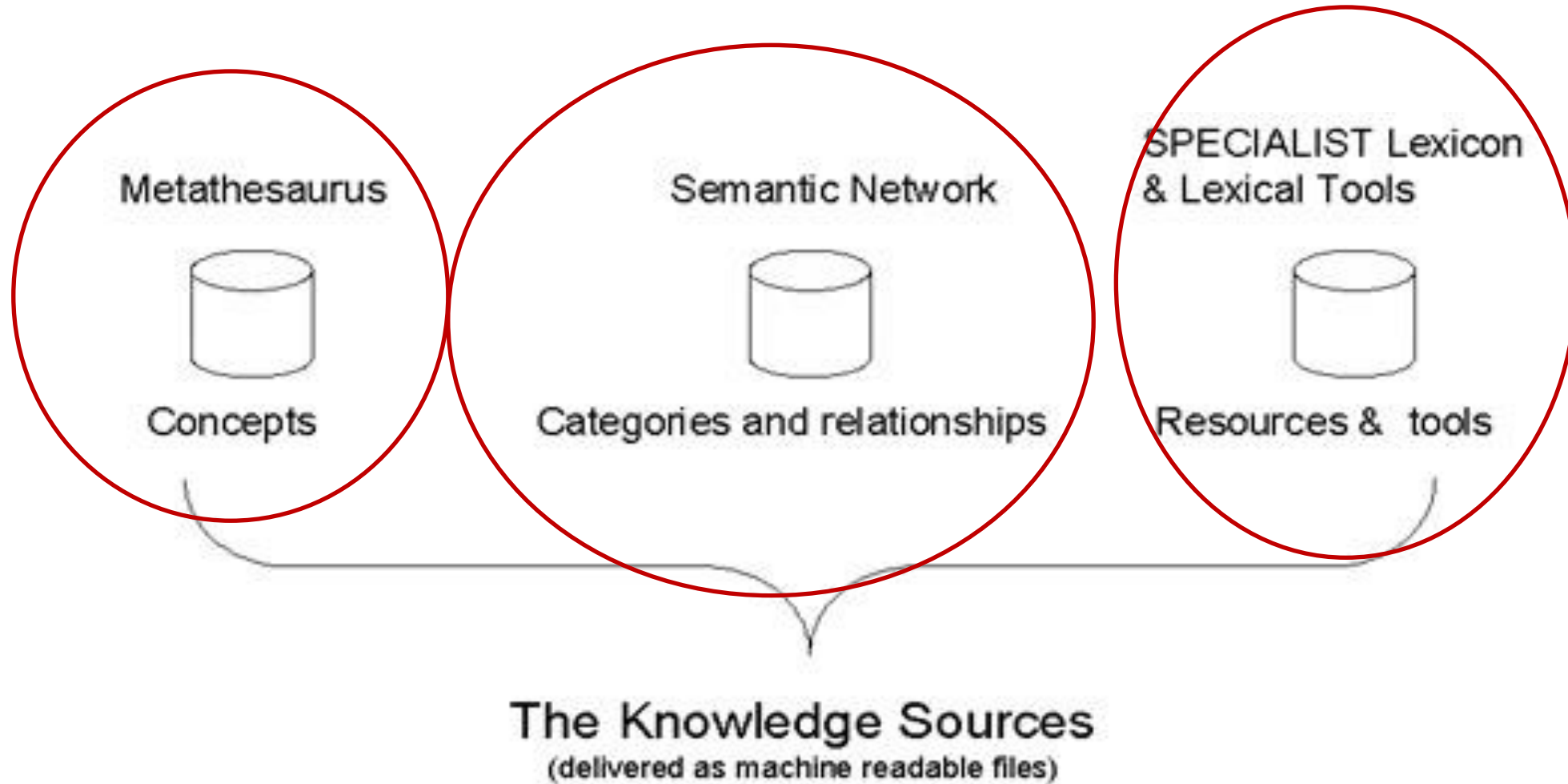
SemMed (база знаний Pubmed) – 420 тыс. объектов, 111 млн связей

Pubmed – 20.5 млн статей медицинской тематики, 1 млрд связей

**Клинические рекомендации Минздрава РФ – 15 шт.
(заболевания органов дыхания)**

МЕТОДОЛОГИЯ

1. Анализ метатезауруса UMLS, выделение необходимых группы концептов и типов связей.
2. Формализация клинических рекомендаций раздела «Диагностика» неопухолевых заболеваний дыхательной системы.
3. Построение графов для каждой нозологической формы, отражающие связи заболевания с его признаками.
4. Обработка текстов нейронными сетями для извлечения объектов, связей и фактов из текстов Pubmed.
5. Обработка графов нейронными сетями для поиска объектов по «похожести» их окружения в семантическом графе.
6. Построение базы медицинских знаний UMLS+
7. Разработка и тестирование с помощью аналитического приложения алгоритма поиска релевантного окружения, в основе которого лежат методы интеллектуального анализа семантических сетей.



ХАРАКТЕРИСТИКИ UMLS v2022AA

Семантические группы: 15

- Activities & Behaviors – Деятельность и поведение
- Anatomy – Анатомия
- Chemicals & Drugs – вещества и лекарства

Семантические подгруппы: 127

Справочники: 222

Понятия: 4 553 796

Количество имен понятий: 16 857 345

tui	str_tui	cui	au	str	ts	stt	ispref	tty	sup...	lat	transle...	sab
T184	Sign or Symptom	C0010200	A30843576	Кашель	S	PF	Y	PT	N	RUS	F	MDRRUS
			854578	Кашель						RUS	F	MTD
			A30785816	Кашель	S	PF	N	LLT	N	RUS	F	MDRRUS
			A13514453	KASHEL'	S	PF	Y	SY	N	RUS	F	MSHRUS
			A8026561	КАШЕЛЬ	P	PF	Y	MH	N	RUS	F	MSHRUS
			854580	Покашливание						RUS	F	MTD
			A30732493	Покашливание	S	PF	Y	LLT	N	RUS	F	MDRRUS
			A20108967	Cough	P	VO	N	AB	Y	ENG	T	ICD10CM
			A0398304	COUGHING	P	VC	N	GT	N	ENG	T	CST
			A16981420	Cough	P	VO	N	AB	Y	ENG	T	ICD9CM
			A2880240	Cough	P	VO	N	PT	N	ENG	T	SNOMEDCT_US
			A0403132	Coughing	P	PF	N	SY	N	ENG	T	RCD
			A23015591	Observation of cough	S	PF	N	OAS	O	ENG	T	SNOMEDCT_US
			A31610513	Coughing	P	PF	N	PT	N	ENG	T	NCI_ACC-AHA
			A7873388	Coughing	P	PF	N	PT	N	ENG	T	SNOMEDCT_US
			A0398295	COUGH	P	VO	Y	FI	N	ENG	T	DXP
			A18619878	cough symptoms	S	VO	Y	SY	N	ENG	T	CHV
			A7876575	Coughing (observable entity)	S	PF	Y	FN	N	ENG	T	SNOMEDCT_US
			A32611470	Complaining of cough	S	PF	Y	OAS	O	ENG	T	SNOMEDCT_US
			A25692687	Cough	P	VO	N	LLT	N	ENG	T	MDR
			A32399985	Cough	P	VO	N	PT	N	ENG	T	NCI_FDA
			A13361485	Coughing - function	S	PF	N	IS	O	ENG	T	SNOMEDCT_US
			A18601373	cough symptom	S	VO	Y	PT	N	ENG	T	CHV

- Procedures – процедуры

The screenshot displays the DataMonitor interface for the 'База медицинских знаний (UMLS)'. The main window is divided into several panes:

- Left Panel:** A navigation tree showing various categories like 'Activities & Behaviors', 'Anatomy', 'Chemicals & Drugs', etc., with sub-panels for 'tui1', 'sty1', and 'sty1_rus'.
- Top Panel:** Search and filter options for 'Group1', 'cui1', 'cui1_str', 'rel1', 'rel1', 'con...', 'Group2', 'cui2', and 'cui2_str'.
- Main Table:** A large table with columns for 'rel1', 'rel1', 'con...', 'Group2', 'cui2', and 'cui2_str'. It lists various medical terms and their relationships, such as 'abnormal_cell_affected_by_chemical', 'abnormality_associated_with_allergy', etc.
- Bottom Panel:** A summary table with columns 'sab1', 'code1', and 'cui1_str', listing ICD codes and their corresponding descriptions.

№	Название КР	ID	год
1.	Аллергический ринит	КР261	2020
2.	Острый синусит	КР313	2021
3.	Острые респираторные вирусные инфекции (ОРВИ) у взрослых	КР724	2021
4.	Хронический тонзиллит	КР683	2021
5.	Острый тонзиллит и фарингит (Острый тонзиллофарингит)	КР306	2021
6.	Острый обструктивный ларингит и эпиглоттит	КР352	2021
7.	Паратонзиллярный абсцесс	КР664	2021
8.	Грипп у взрослых	Проект	2021
9.	Бронхиальная астма	КР359	2021
10.	Внебольничная пневмония у взрослых	КР654	2021
11.	Хроническая обструктивная болезнь легких	КР603	2021
12.	Хронический бронхит	КР655	2021
13.	Эмфизема легких	КР656	2021
14.	Идиопатический легочный фиброз	КР677	2021
15.	Новая коронавирусная инфекция (COVID-19)	Врем.МР	2022

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.



КЛИНИЧЕСКИЕ РЕКОМЕНДАЦИИ

а
н
а
л
и
з

Концепт
нозологической
формы

Связь 1

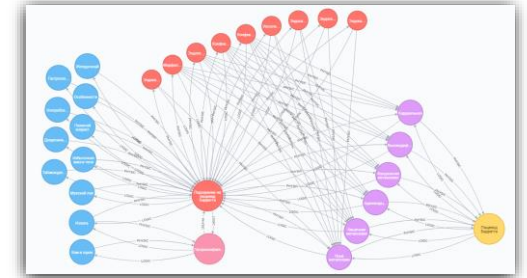
Связь 2

Связь 3

Концепт -
признак 1

Концепт -
признак 2

Концепт -
признак 3



ГРАФЫ ДЛЯ КАЖДОЙ НОЗОЛОГИИ

База медицинских знаний (UMLS)

tui1 = "T047" И cui1 = "C0008677"

Disorders

Group1

cui1	cui1_str
C0008677	БРОНХИТ ХРОНИЧЕСКИЙ

rel1	rela1	con...
AQ	associated_morphology_of	25
CHD	associated_with	75
PAR	classified_as	125
RB	classifies	50
RN	clinical_course_of	25
RO	clinically_associated_with	100
RQ	clinically_similar	25
SIB	co-occurs_with	50
SY	finding_site_of	25

Group2

cui2	cui2_str
C2599447	История болезней:Нх:Рт:Член семьи:Имя:USS
C0006818	КАМПИЛОБАКТЕРИОЗЫ
C0010200	КАШЕЛЬ
C0010674	КИСТОЗНЫЙ ФИБРОЗ
C0241407	КЛЕЩЕЙ УКУСЫ
C0037278	КОЖНЫЕ БОЛЕЗНИ ИНФЕКЦИОННЫЕ
C0009186	КОКЦИДИОИДОМИКОЗ
C0014836	КОЛИ-ИНФЕКЦИИ
C0009663	КОНДИЛОМА ОСТРОКОНЕЧНАЯ

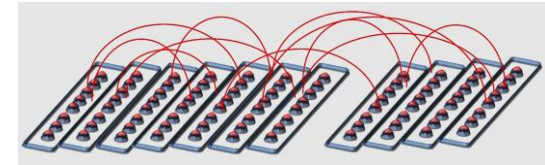
ВЫЯВЛЕННЫЕ ПРОБЛЕМЫ



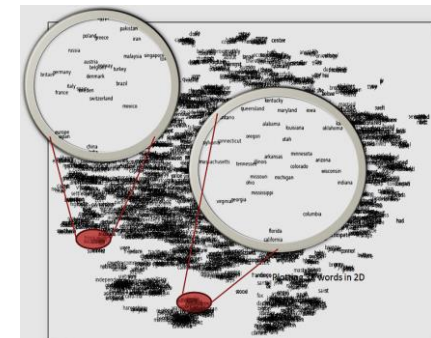
МЕТОДЫ ОБРАБОТКИ ДАННЫХ - НЕЙРОННЫЕ СЕТИ

- ✓ *Нейронные сети обработки текстов* – различные модификации сетей семейства Transformer. Обеспечивают:
 - извлечение объектов, связей и фактов из текстов;
 - категоризация объектов;
 - быстрый семантический поиск документов по смысловой схожести;
 - билингвальный перевод «на лету» текстов rus-en.
- ✓ *Нейронные сети обработки графов* – различные модификации сетей семейства TransE. Обеспечивают:
 - поиск объектов по «схожести» их окружения в семантическом графе;
 - кластеризацию (смысловую группировку) объектов в целях реализации аналитических сценариев.

Вектора термов с похожим контекстом становятся близки (похожи) по косинусной мере



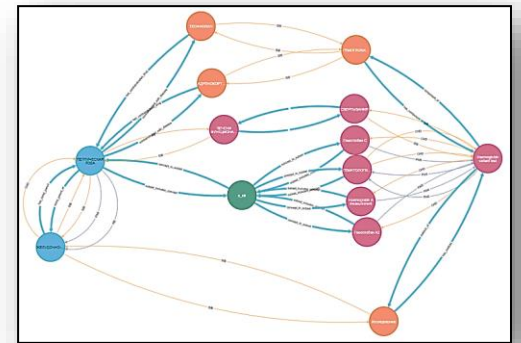
Похожие термы «стягиваются» друг к другу при кластеризации



Граф UMLS+



Структурированные
данные



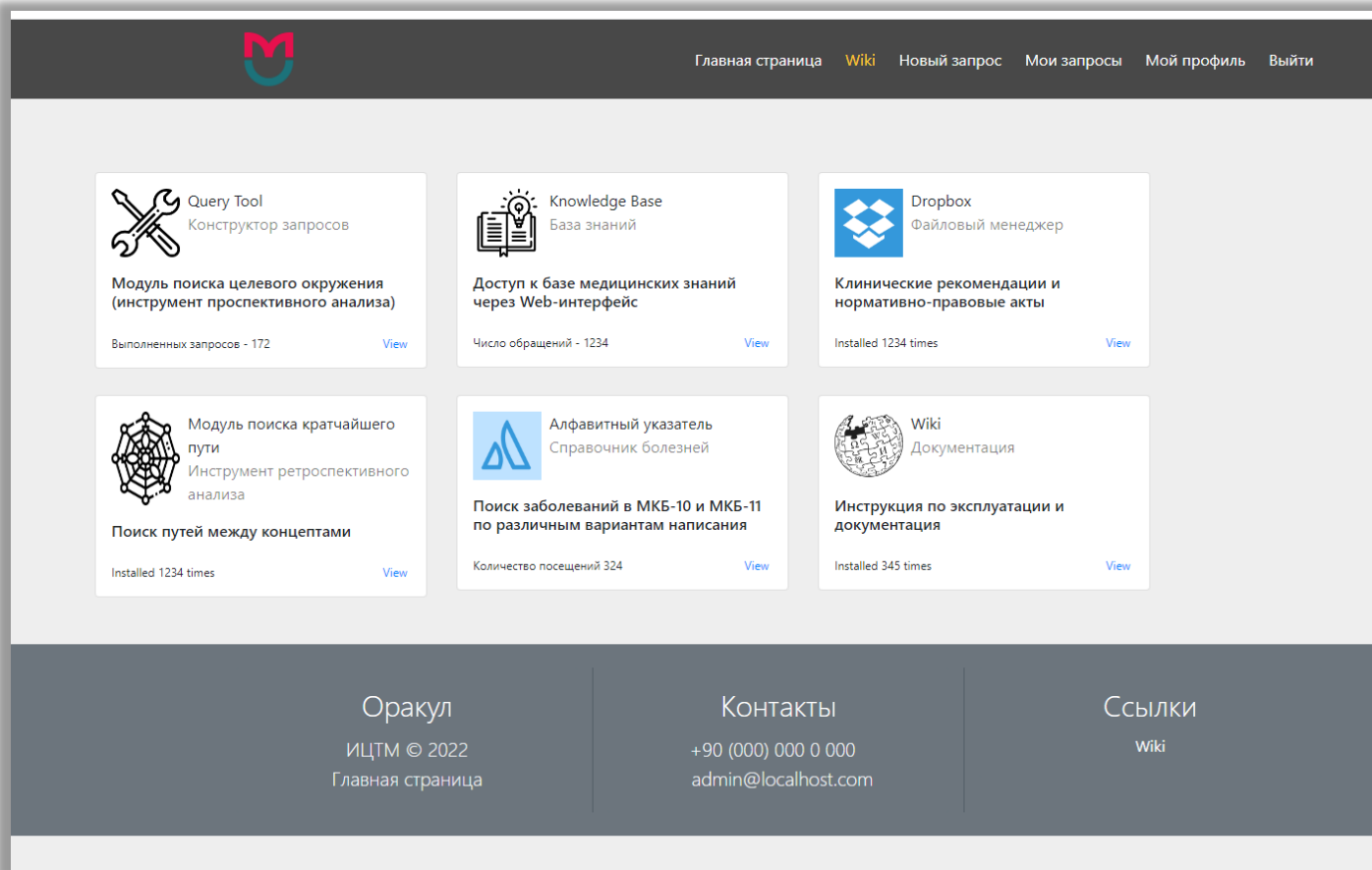
Неструктурирован
ные данные



Инструменты
обработки данных



АНАЛИТИЧЕСКАЯ СИСТЕМА

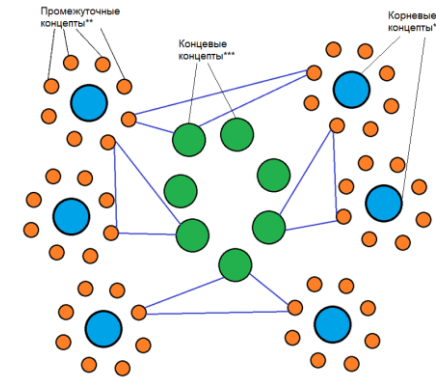


The dashboard features a navigation bar with links: Главная страница, Wiki, Новый запрос, Мои запросы, Мой профиль, and Выйти. The main content area contains six modules:

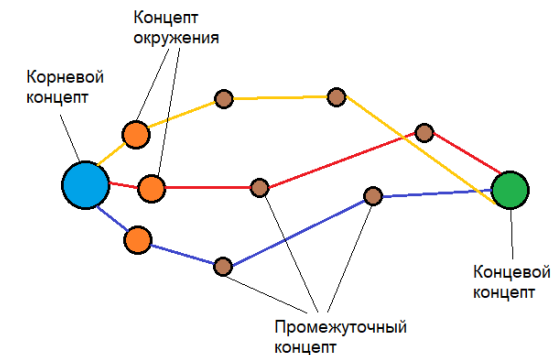
- Query Tool** (Конструктор запросов): Module for searching target environments (instrument for prospective analysis). 172 queries completed.
- Knowledge Base** (База знаний): Access to a medical knowledge base via a web interface. 1234 requests.
- Dropbox** (Файловый менеджер): Clinical recommendations and normative-legal acts. Installed 1234 times.
- Module for finding the shortest path** (Инструмент ретроспективного анализа): Search for paths between concepts. Installed 1234 times.
- Alphabetical index** (Справочник болезней): Search for diseases in МКБ-10 and МКБ-11 by different writing variants. 324 visits.
- Wiki** (Документация): Instructions for operation and documentation. Installed 345 times.

The footer includes: Оракул (ИЦТМ © 2022, Главная страница), Контакты (+90 (000) 000 0 000, admin@localhost.com), and Ссылки (Wiki).

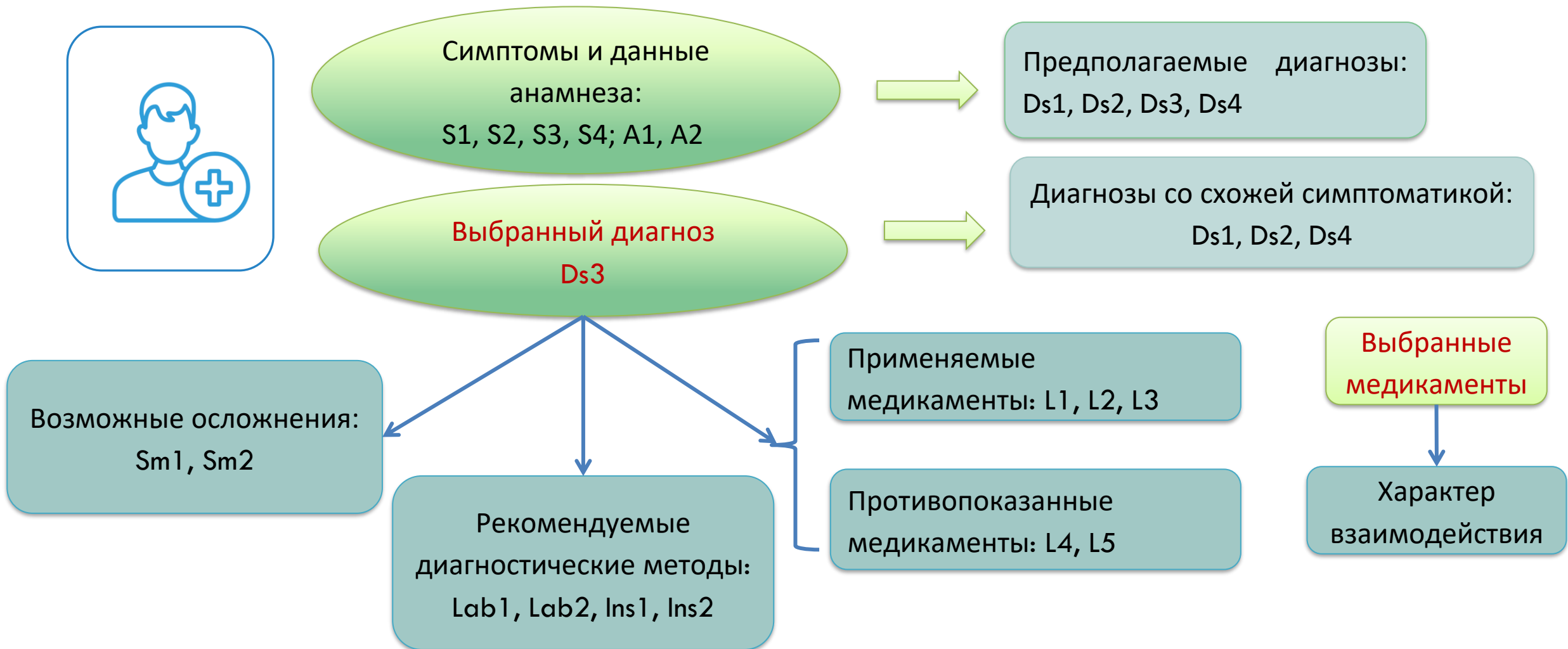
«Модуль поиска релевантного окружения концептов в семантической сети UMLS»



«Модуль поиска кратчайшего пути»



МОДЕЛЬ ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ



РЕЗУЛЬТАТЫ

Источник данных
на базе UMLS и
PubMed

Вычислительный
алгоритм и
методы NLP

Платформа для
информационно-
поисковой
медицинской
системы





БЛАГОДАРЮ ЗА ВНИМАНИЕ

koroleva_iui@rsmu.ru