

ФГАОУ ВО «Российский национальный исследовательский медицинский университет имени Н. И. Пирогова»
Минздрава России

Институт цифровой трансформации медицины

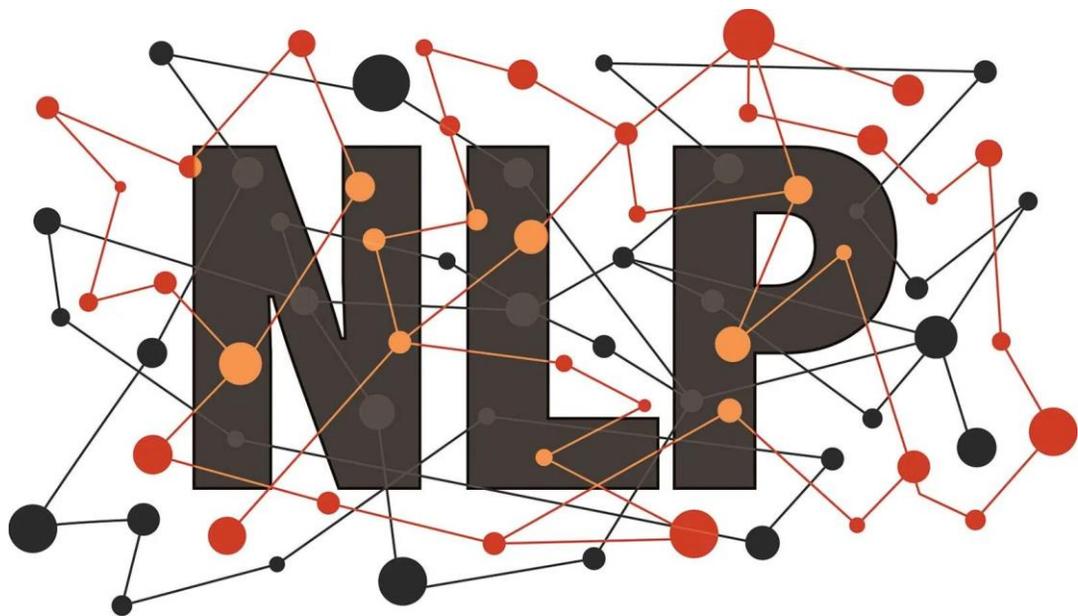
Перспективы развития и использования Унифицированной национальной медицинской номенклатуры

Астанин Павел Андреевич, аналитик лаборатории семантического анализа медицинской информации института цифровой трансформации медицины (ИЦТМ), ассистент кафедры медицинской кибернетики и информатики имени С. А. Гаспаряна (+79831580814; med_cyber@mail.ru)

ТЕРМИНОЛОГИЯ

ОСНОВА ДЛЯ ПОСТРОЕНИЯ НАШИХ СЕРВИСОВ

УНИФИЦИРОВАННАЯ НАЦИОНАЛЬНАЯ МЕДИЦИНСКАЯ НОМЕНКЛАТУРА



Цель

Обеспечение семантической интероперабельности медицинских информационных систем с применением единой терминологической платформы для разработки баз знаний, информационно-поисковых систем, систем поддержки принятия решений в медицине и здравоохранении.

UNIFIED MEDICAL LANGUAGE SYSTEM (UMLS)

Основой Унифицированной национальной медицинской номенклатуры (далее – Номенклатура) стала Единая система медицинского языка (Unified Medical Language System), разрабатываемая с 1986 г. в Национальной медицинской библиотеке США и объединяющая:

- ✓ **Metathesaurus (метатезаурус)** – 102 международных справочника терминов со связями между ними;



LOINC
from Regenstrief



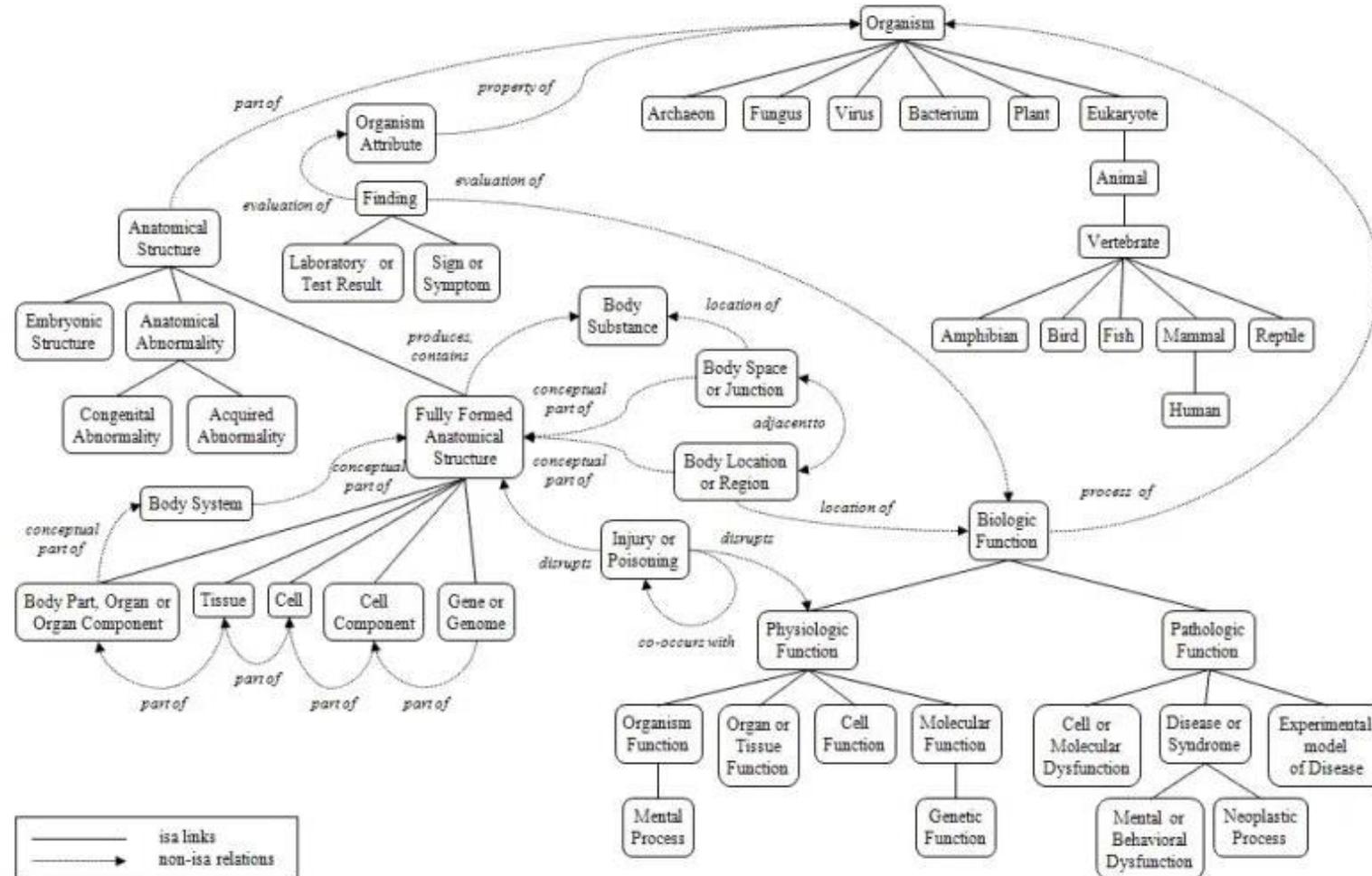
MedDRA

- ✓ **Semantic Network (семантическая сеть)** – семантические группы терминов, типы и атрибуты связей;
- ✓ **SPECIALIST Lexicon and Lexical Tools** (инструменты лексического анализа) – словари для обработки естественного английского языка.



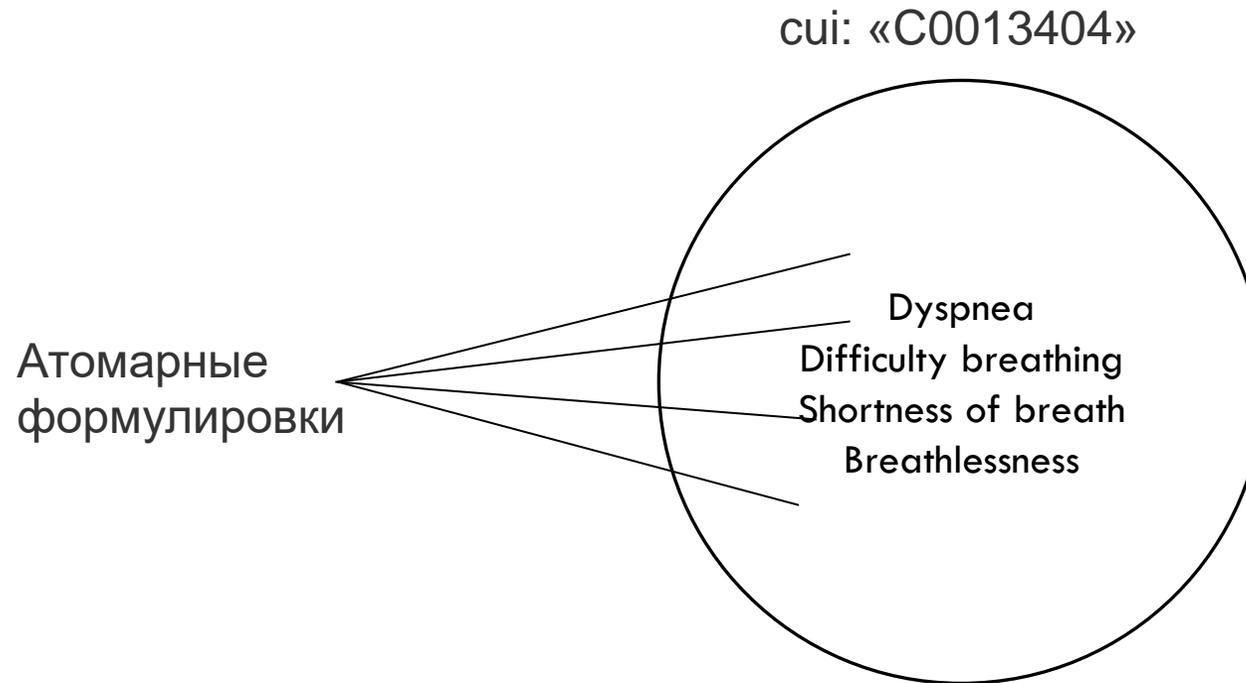
U.S. National Library
of Medicine

ГРУППЫ ТЕРМИНОВ UMLS



КОНЦЕПТЫ UMLS

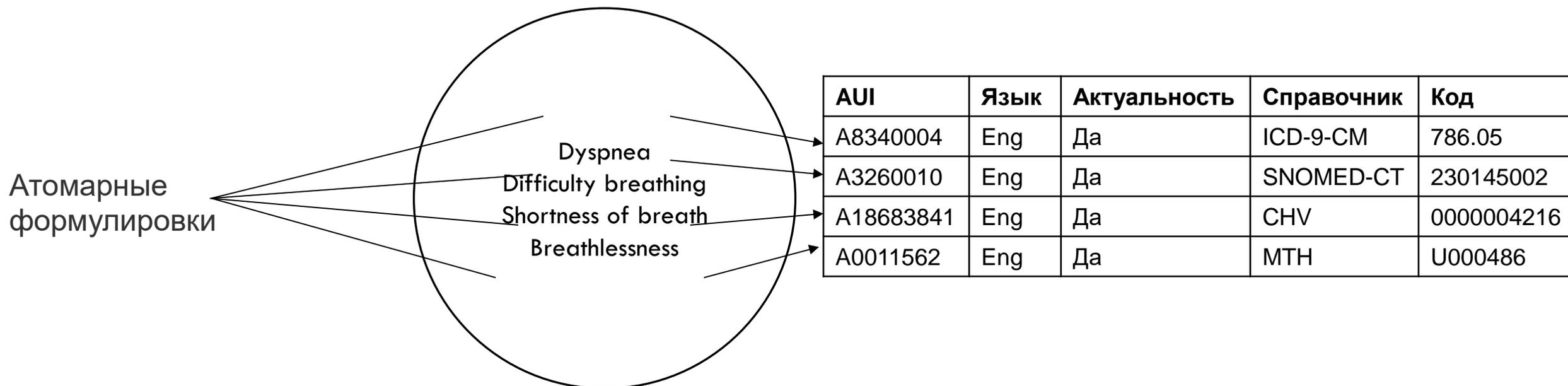
Концепт – это уникальная сущность, описываемая одним или несколькими вариантами произношения. Имеет идентификатор (concept unique identifier – cui), состоящий из буквы С и 7 цифр, и принадлежность к тематической группе (type unique identifier – tui), кодируемой конструкцией из буквы Т и 3 цифр.



tui: «Т184» – Sign or Symptom (Признак или симптом)

ФОРМУЛИРОВКИ UMLS

Элемент метатезауруса, имеющая уникальный идентификатор (atom unique identifier – aui), состоящий из буквы А и 6–8 цифр. Атомарная формулировка всегда привязана к одному и только одному концепту UMLS.



УНИФИЦИРОВАННАЯ НАЦИОНАЛЬНАЯ МЕДИЦИНСКАЯ НОМЕНКЛАТУРА

Всего формулировок:

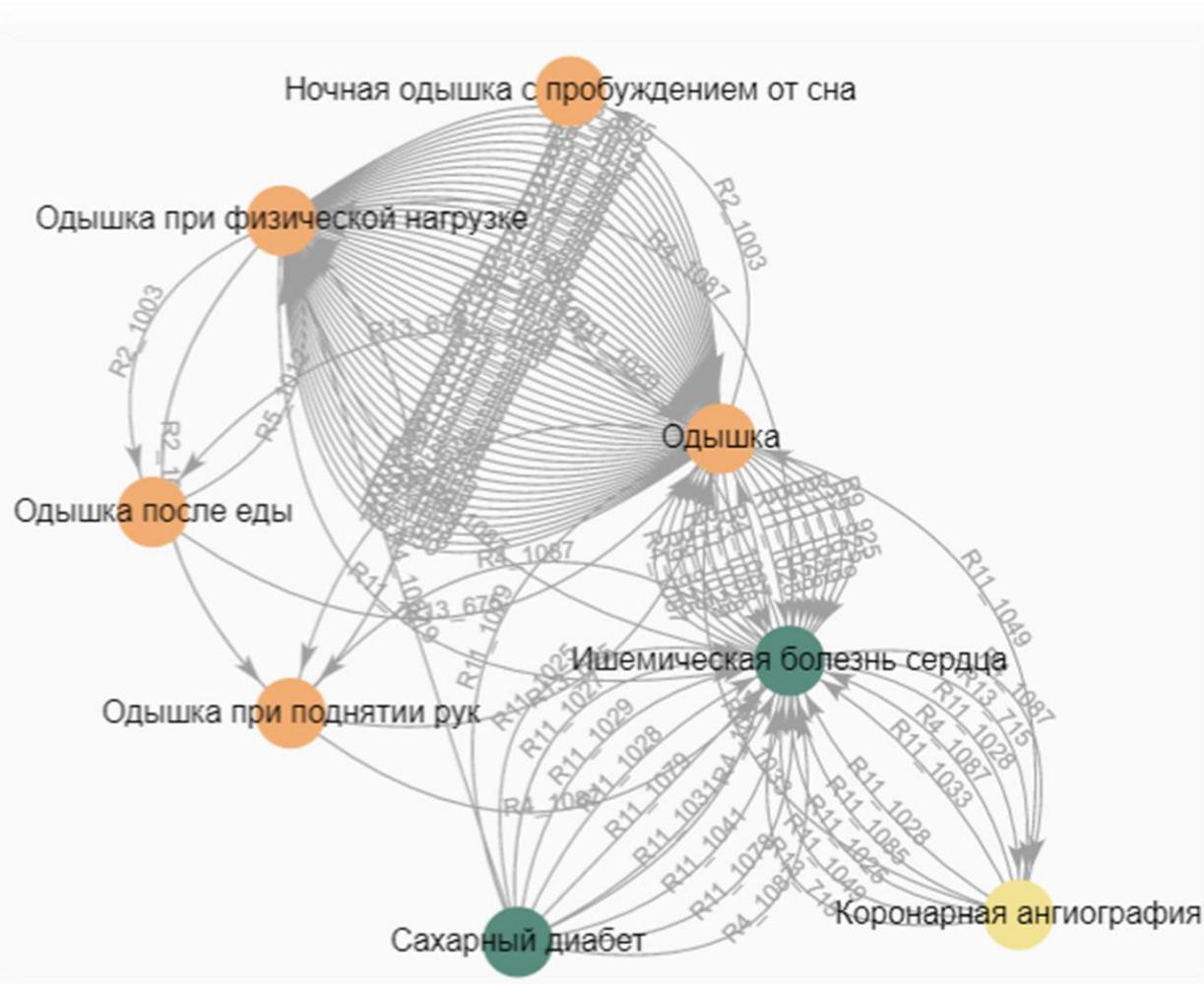
12,854,972

из них клинически значимых:

~1,5 млн

Всего концептов:

4,806,858



ТЕРМИНОЛОГИЧЕСКОЕ ПОКРЫТИЕ

По данным Newbury A., от 71% до 84% терминов из текстов научных статей находится в версии UMLS от 2023 г.;

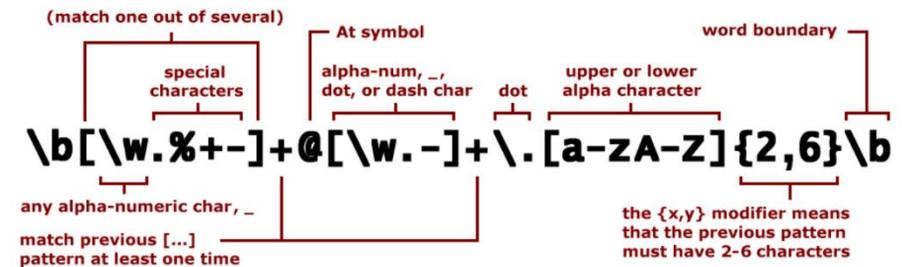
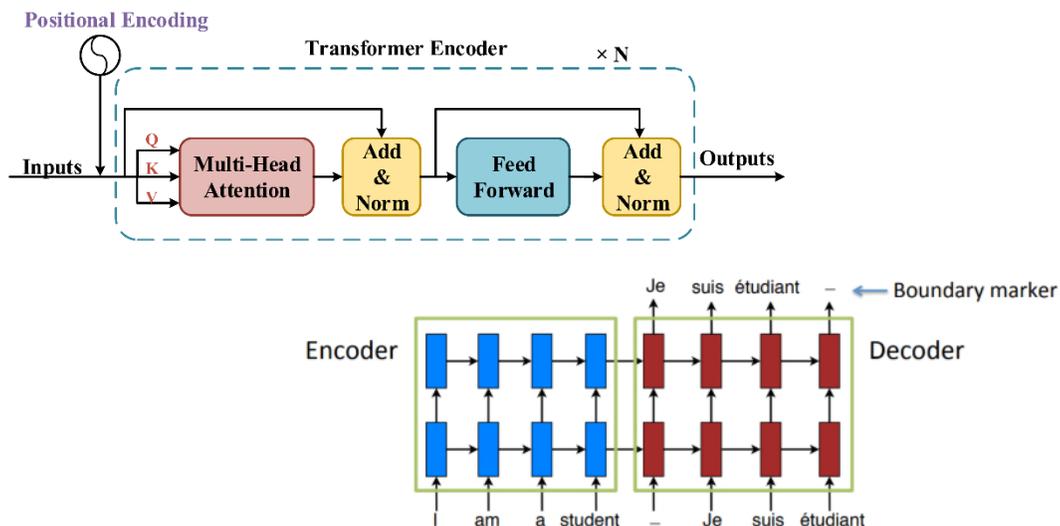
Около 80% терминов, не найденные в текстах, являются аббревиатурами или синонимами уже существующих терминов;

№	Клиническая область	Доля терминов (%), из них:			
		С полным совпадением	С частичным совпадением	Представленных в UMLS (степень покрытия)	Отсутствующих в метатезаурусе
1	Онкология	55	23	78	22
2	Неврология	49	28	77	23
3	Гастроэнтерология	49	32	81	19
4	Иммунология	43	30	73	27
5	Ортопедия	47	27	74	26
6	Нефрология	46	29	75	25
7	Ревматология	48	23	71	29
8	Травматология	47	37	84	16

АДАПТАЦИЯ UMLS НА РУССКИЙ ЯЗЫК

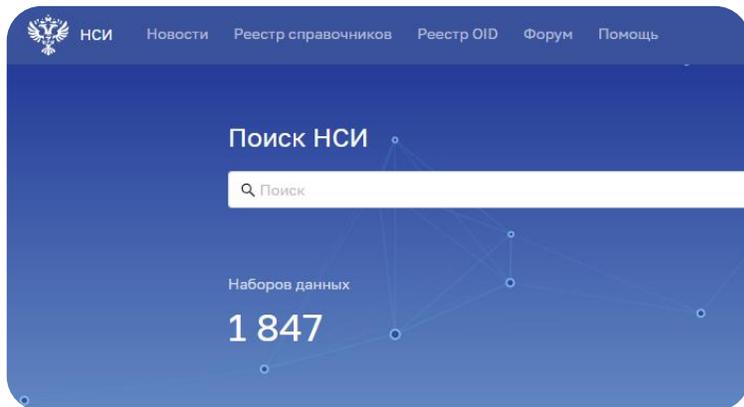
Исходно в UMLS было представлено 3% русскоязычных терминов (304 тыс.). Адаптация оставшихся терминов UMLS на русский язык велась по ряду параллельных направлений:

1) Гибридный перевод с использованием нейросети GNMT (энкодер Transformer + декодер RNN) и правил предобработки текста. Полученный технический перевод статистически значительно превосходит экспертный (4.685 ± 0.823 против 4.260 ± 1.105 ; $p < 0.001$);



АДАПТАЦИЯ UMLS НА РУССКИЙ ЯЗЫК

- 2) Экспертные уточнение и обогащение клинически значимых концептов (проработано около 200 тыс. из тематических групп «Симптомы», «Клинические находки», «Диагностические методы»);
- 3) Сопоставление со справочниками Федерального реестра нормативно-справочной информации Минздрава России. В настоящее время сопоставлено 178 тыс. (~39% от 456 тыс.) формулировок из 259 справочников (медианная доля сопоставления – 44%) с качеством, равным 87%.



C0009378	[T060 — diar — Diagnostic Procedure (Диагностические процедуры)]	D00441074	Мероприятия ПМО, диспансеризации и УД	29		Колоноскопия	Автоматическое сопоставление по справочников англоязычной формулировке (Алгоритмы автоматического анализа текста)
C0009378	[T060 — diar — Diagnostic Procedure (Диагностические процедуры)]	D00417042	Медицинские обследования для МСЭ	720		Колоноскопия	Автоматическое сопоставление по справочников англоязычной формулировке (Алгоритмы автоматического анализа текста)
C0009378	[T060 — diar — Diagnostic Procedure (Диагностические процедуры)]	A3124789	SNOMEDCT_US — Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT), United States Edition — Американская версия систематизированного свода клинических терминов	73761001	Endoscopic examination of colon	Эндоскопическое исследование толстой кишки	Гибридный перевод: Google Neural Machine Translation + правила ИЦТМ (Внутренние алгоритмы автоматического перевода)

АДАПТАЦИЯ UMLS НА РУССКИЙ ЯЗЫК

4) Создание собственных справочников на основе анализа научной литературы и данных реальной клинической практики, а также предоставленные экспертными сообществами:

- Справочник медицинских аббревиатур (16 тыс.);
- Справочник радиологических терминов (2,5 тыс.).

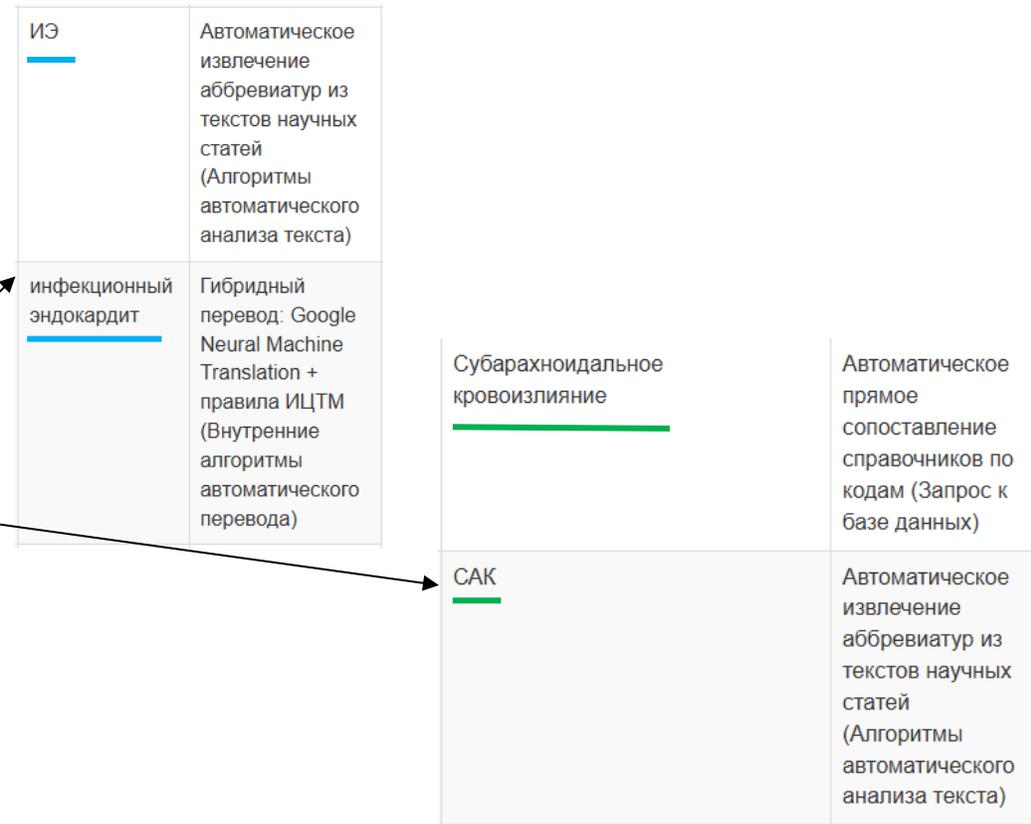
АННОТАЦИЯ:

Обоснование. Ишемический инсульт и церебральные микрокровоизлияния (ЦМК) - самые частые проявления неврологических осложнений инфекционного эндокардита (ИЭ), при этом их сочетание мало изучено.

Цель. Оценить паттерн сочетания кортикально-субкортикального инфаркта с долевым ЦМК или субарахноидальным кровоизлиянием (САК) у пациентов с «левосторонним» ИЭ.

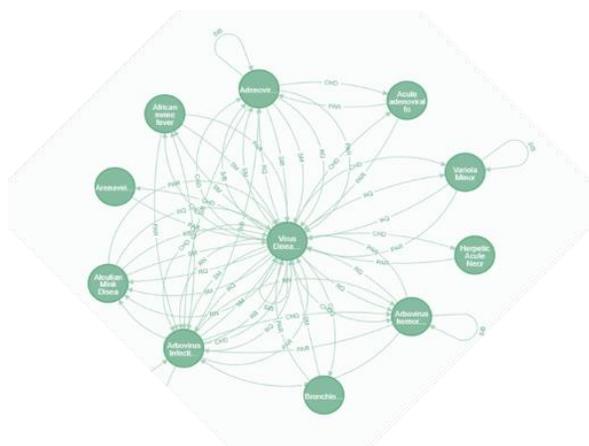
Материалы и методы. Проведено ретроспективное исследование по типу «случай-контроль», включающее пациентов с ИЭ, прошедших кардиохирургическое лечение в ФГБУ ФЦССХ. Критерии включения: возраст пациентов ≥ 18 лет; достоверный или вероятный (критерии Duke) ИЭ левых отделов сердца (аортальный и/или митральный клапаны); доступность для анализа данных предоперационной магнитно-резонансной томографии головного мозга. В контрольную группу включены пациенты без ИЭ, имеющие нелакунарный (предположительно эмболический) ИИ...

▼ Показать полностью

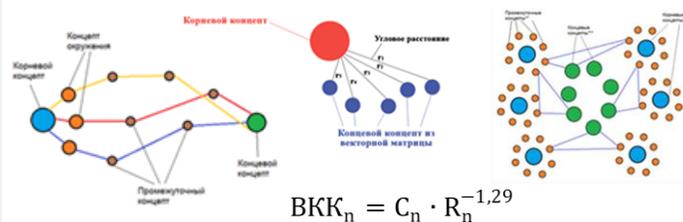


ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ НА ОСНОВЕ УНИФИЦИРОВАННОЙ НОМЕНКЛАТУРЫ

Адаптированные на русский язык термины + связи, исходно представленные в UMLS



Метрики и модели, построенные на основе концептов и связей



- ✓ Связность концептов в графе
- ✓ Векторные эмбединги
- ✓ Модели клинической значимости, связности, специфичности терминов

Связи из научной литературы (27 млн аннотаций PubMed) и реальных данных – Dataset MIMIC-IV (330 тыс.)



Лаборатория вычислительной физиологии Массачусетского технологического института

СЕРВИСЫ

НА ОСНОВЕ УНИФИЦИРОВАННОЙ НОМЕНКЛАТУРЫ И ДРУГИХ РЕСУРСОВ

ИЗВЛЕЧЕНИЕ ИМЕНОВАННЫХ СУЩНОСТЕЙ

В основе решения данной задачи лежит использование сервиса по Аннотированию медицинских текстов. Сервис является гибридным, использует нейронную сеть (BERT) для построения синтаксического графа неструктурированного текста с последующим поиском изоморфизмов по терминам Унифицированной номенклатуры. Валидационная точность инструмента – 93%.

«Больной жалуется на хронический надрывистый кашель с гнойной мокротой, боли в грудной клетке, связанные с дыханием, которые в большинстве случаев сопровождаются лихорадкой, периодическое повышение АД. Отмечается снижение толерантности к нагрузке»

Главное меню Аналитические модули Тестирование информационно-поисковой системы Семантические анализаторы

Аннотирование медицинских текстов

Т184 - Признак или симптом
УНМН: T033 2246-14 - Хронический кашель
УНМН: T184 C0010201 - хронический кашель

Т033 - Клинические находки
УНМН: T033 C1271104 - артериальное давление

Больной жалуется на хронический надрывистый кашель с гнойной мокротой, боли в грудной клетке, связанные с дыханием, которые в большинстве случаев сопровождаются лихорадкой, периодическое повышение АД. Отмечается снижение толерантности к нагрузке

Примечание: вы можете оценить качество работы аннотации

Концепт (syn)	Атом (au)	Найденная в тексте формулировка	Идентификатор	Номер предложения в тексте	Правильность сопоставления фрагмента текста с термином	Правильность перевода языка	Текст
C1271104	A13562521	АД	УНМН: None 2141- УНМН: None 2142- УНМН: None 2855- УНМН: None 2956- УНМН: None 3367- УНМН: T031 C0038056 - мокрота УНМН: T033 C0159054 - Мокрота УНМН: T033 C0239134 - продуктивный кашель УНМН: T033 C0566528 - отхаркивается УНМН: T033 C0587141 - мокрота УНМН: T184 C0241235 - гнойная мокрота УНМН: T184 C3854658 - Гнойная мокрота	1			
C2598155	A16941898	Боли	УНМН: T029 C0817096 - Грудь УНМН: T029 C1527391 - передняя часть грудной клетки	1			
C0030193	A8016150	Боли	УНМН: T029 C0817096 - Грудь УНМН: T029 C1527391 - передняя часть грудной клетки	1			
C0030193	A9332925	Боли	УНМН: T029 C0817096 - Грудь УНМН: T029 C1527391 - передняя часть грудной клетки	1			
C0030193	A9333492	Боли	УНМН: T029 C0817096 - Грудь УНМН: T029 C1527391 - передняя часть грудной клетки	1			

Т031 - Продукты жизнедеятельности биологические ж
УНМН: None 2141-
УНМН: None 2142-
УНМН: None 2855-
УНМН: None 2956-
УНМН: None 3367-
УНМН: T031 C0038056 - мокрота
УНМН: T033 C0159054 - Мокрота
УНМН: T033 C0239134 - продуктивный кашель
УНМН: T033 C0566528 - отхаркивается
УНМН: T033 C0587141 - мокрота
УНМН: T184 C0241235 - гнойная мокрота
УНМН: T184 C3854658 - Гнойная мокрота

Т029 - Анатомо-топографические структуры
УНМН: T023 C1284162 - Грудная клетка
УНМН: T029 C0817096 - Грудь
УНМН: T029 C1527391 - передняя часть грудной клетки

Т169 - Функциональные свойства
УНМН: T033 C0424551 - сниженная толерантность к физической нагрузке
УНМН: T033 C0948372 - снижение толерантности к нагрузке
УНМН: T080 C1511741 - снижение
УНМН: T169 C1547177 - снижение

ПОДБОР ТЕРМИНОВ ПРИ ВВОДЕ ТЕКСТА

с/м Кернега

S0234178 (симптом Кернига)

симптом покалывания

**S0235634 (симптом
Пастернацкого)**

боль в голове

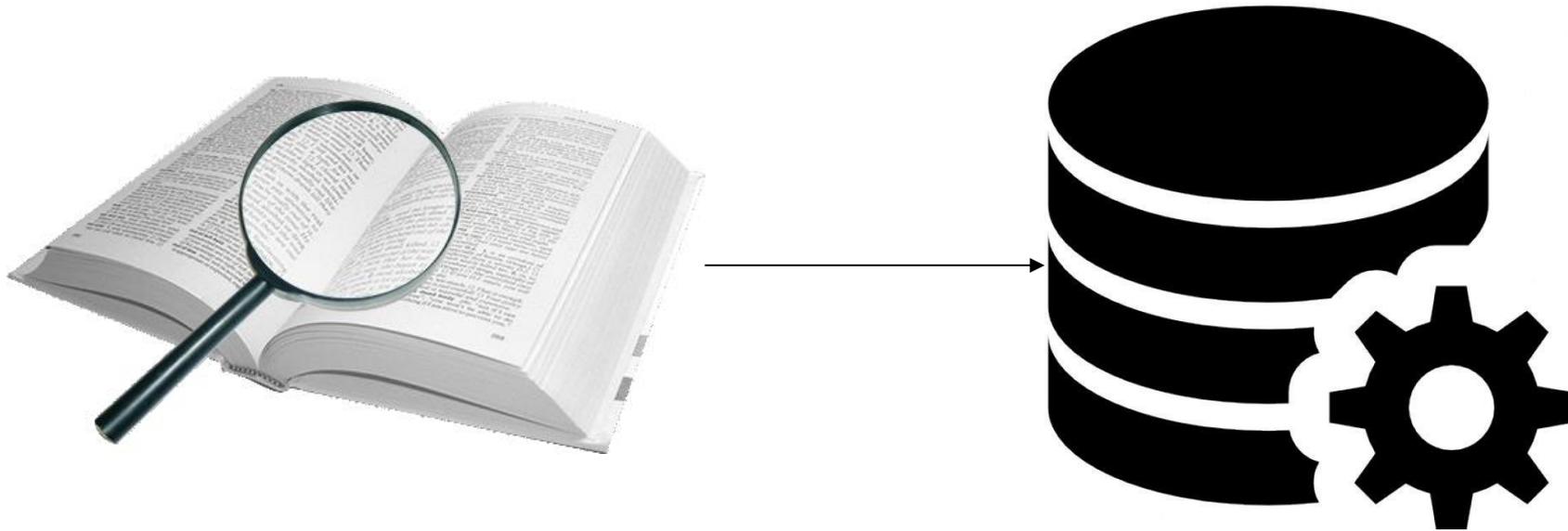
S0018681 (Головная боль)

ТОШНИТ

S0027497 (Тошнота)

ЕДИНЫЙ МЕДИЦИНСКИЙ ГЛОССАРИЙ

Гигантская справочная подсистема, представленная 299 тыс. дефиниций для 261 тыс. концептов. Дефиниции получены из международных справочников UMLS, текстов научных статей, а также с привлечением знаний врачей-экспертов. В настоящее время – один из крупнейших биомедицинских глоссариев в нашей стране.



СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ: СИМПТОМЧЕКЕРЫ

1) Система, осуществляющая постановку предварительного диагноза (по 199 социально-значимым нозологиям) с элементами информационно-справочной поддержки врачей первичного звена.

Симптомчекер для врача первичного звена

ИПС-Тестер 1.

Укажите жалобы или симптомы

Сжимающие за грудиные боли и сердцебиение, проходящие после прекращения нагрузки. В анамнезе малоподвижный образ жизни. Вес избыточный. Ослабленные тоны сердца.

Поиск

Очистить

C0232290 давящая боль в груди ×
C0436824 сердечные тоны ослаблены ×
C0151826 боль за грудиной ×
C1532253 малоподвижный образ жизни ×
C0030252 Сердцебиение ×
C0497406 ИМТ от 25 до 30 ×

Возможно подходят:

C0235240 учащено бьющееся сердце
C0438716 сдавленность в груди
C0152174 боль, связанная с психологическими факторами
C0231807 функциональная одышка
C4724239 курение
C2051433 состояние пациента удовлетворительное
C0039231 Тахикардия неуточненная
C0235239 трепетание
C0476258 осознание сердцебиения
C0013404 Одышка
C0008031 Боль в груди неуточненная
C5225915 признаки и симптомы кровообращения
C0039070 Обморок [синкопе] и коллапс
C0015672 усталость
C0232286 Боль в области сердца

Подобрать диагноз

СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ: СИМПТОМЧЕКЕРЫ

Заболевания

Диагностические методы

EN

Сердечная недостаточность



Хроническая ишемическая болезнь сердца



Нарушение ритма и проводимости сердца



Хроническая обструктивная болезнь лёгких



Инфаркт миокарда



Показать всё ▾

Сердечная недостаточность

I00-I02, I05-I09, I34-I39, I42-I43, I50-I51, P29, Q20-Q28, R01, R06-R07, R18, R23, R34, R55

Сердечная недостаточность - это клинический синдром, вызванный нарушением структуры и/или функции сердца, приводящий к снижению сердечного выброса и/или повышению давления наполнения сердца в покое или при нагрузке.

[Узнать больше >>](#)

© ИЦТМ, 2024

Клинические
рекомендации

Методы диагностики

Научные публикации

Методы диагностики

- Рентгенография органов грудной клетки
- Компьютерная томография органов грудной клетки
- Пульсоксиметрия
- Общий анализ крови
- Определение СРБ
- Микроскопическое исследование мокроты
- Бактериологическое исследование мокроты

[Узнать больше >>](#)

© ИЦТМ, 2024

Клинические
рекомендации

Методы диагностики

Научные публикации

СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ: СИМПТОМЧЕКЕРЫ

2) Система поддержки принятия решений при боли в нижней части спины (узкоспециализированное решение с чёткими критериями применения).

Возраст:

Пол:
 Мужской Женский Не выбран

Оцените остроту боли в нижней части спины:
 Острая (развилась в течение нескольких минут/часов) Подострая (в течение нескольких дней/недель) Хроническая (в течение нескольких месяцев/лет) Не указано

Укажите длительность утренней скованности:
 Менее 30 минут 30 минут и более Не указано

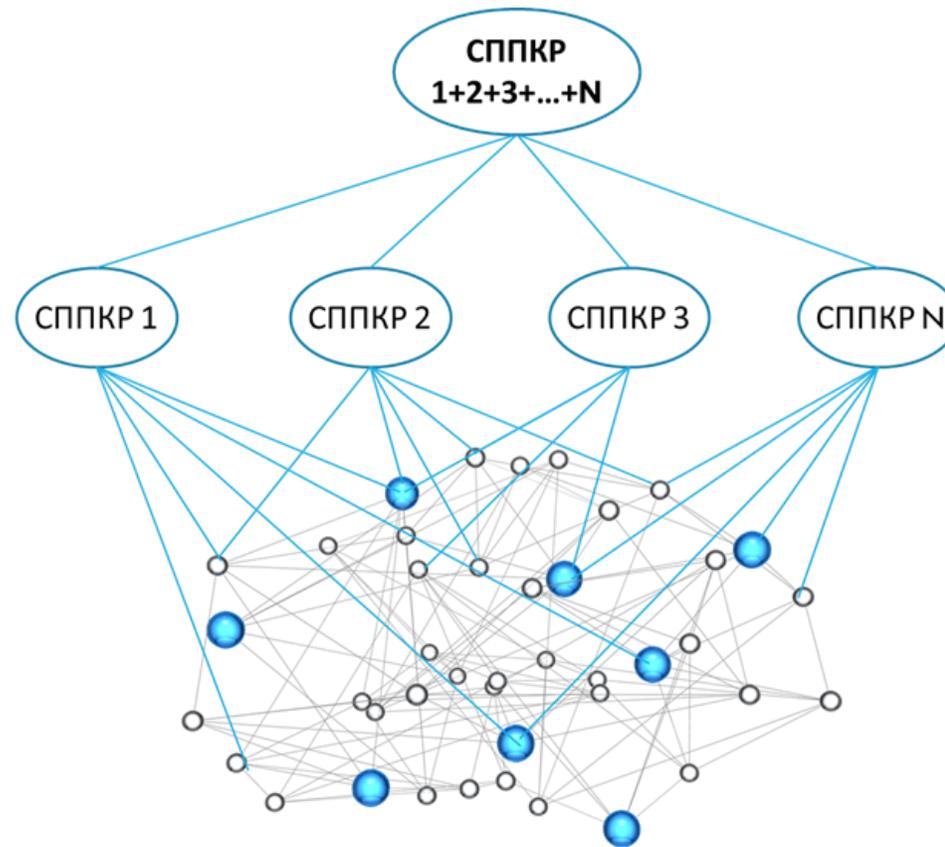
Характерно ли усиление боли в определённое время суток?
 Нет, боль не зависит от времени суток Да, боль усиливается к концу дня Да, боль усиливается к моменту пробуждения Не указано

Присутствует ли дистальная иррадиация боли (ниже колена)?
 Нет Да Не указано

Охарактеризуйте изменение интенсивности боли в покое и при физических нагрузках:
 Боль не зависит от физической нагрузки Боль ослабевает в покое, усиливается при физической нагрузке Боль ослабевает при физической нагрузке, усиливается в покое Не указано

Дополнительные жалобы, признаки и симптомы:

ОДНОПЛАТФОРМЕННЫЕ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ



ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА «КОНСУЛЬТАНТ ВРАЧА-РЕВМАТОЛОГА»

Подсистема поиска материалов по клинической специальности «Ревматология».

Статьи Сервисы поддержки принятия решений Калькуляторы Клинические рекомендации

лечение анк. спондилита

Искать ответ

1

Научная статья
<http://localhost:5000/articles/40>
**Symptomatic therapy for ankylosing spondylitis;
Симптоматическая терапия анкилозирующего спондилита;**
Представлена характеристика **анкилозирующего спондилита (АС)** – одного из основных воспалительных заболеваний позвоночника, при котором в патологический процесс вовлекаются не только крестцово-подвздошные, периферические и корневые (плечевые и тазобедренные) суставы, но и аорта, клапанный аппарат сердца и миокард, бронхолегочная система, почки, нервная система, орган зрения. Указано, что относительно высокая распространенность АС, развитие его в молодом возрасте, прогрессирующее снижение подвижности позвоночника и ранняя инвалидизация больных требуют проведения активной противовоспалительной терапии. Существенный прогресс в лечении АС связан с внедрением в клиническую практику генно-инженерных биологических препаратов. В то же время отмечена важная роль при АС симптоматической терапии (нестероидные противовоспалительные препараты – НПВП, **глюкокортикоиды – ГК**, простые анальгетики, миорелаксанты). Миорелаксанты активно воздействуют не только на основные проявления АС (боль и ригидность), но и на функциональное состояние опорно-двигательного аппарата и исход заболевания в целом. Снижая мышечный тонус и уменьшая выраженность ригидности, они способствуют нарастанию амплитуды движений в позвоночнике и периферических суставах. Среди препаратов этой группы в России наиболее широко используются миорелаксанты группы бензодиазепинов. В настоящее время в России наиболее широко используются миорелаксанты группы бензодиазепинов. В настоящее время в России наиболее широко используются миорелаксанты группы бензодиазепинов.

ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА «КОНСУЛЬТАНТ ВРАЧА-РЕВМАТОЛОГА»

Интерактивная работа с глоссарием через текст найденной научной статьи.

Искать термин

1

Концепт
<https://yandex.ru/>
С0017710
Глюкокортикоидный гормон; Препарат, содержащий глюкокортикоиды; Глюкокортикоидный препарат; Глюкокортикоиды; Глюкокортикоид; Терапевтический глюкокортикоид; ГК

Группа кортикостероидов, влияющих на углеводный обмен (глюконеогенез, отложение гликогена в печени, повышение уровня сахара в крови), ингибирующих секрецию адренкортикотропного гормона, обладающих выраженной противовоспалительной активностью. Они также играют роль в жировом и белковом обмене, поддержании артериального давления, изменении реакции соединительной ткани на травму, уменьшении количества циркулирующих лимфоцитов и функционировании центральной нервной системы.

группа кортикостероидов, влияющих на углеводный обмен, ингибирующих секрецию кортикотропина и обладающих выраженной противовоспалительной активностью; они также играют роль в жировом и белковом обмене, поддержании артериального давления, изменении реакции соединительной ткани на повреждение, уменьшении количества циркулирующих лимфоцитов и функционировании центральной нервной системы.

Кортикостероид, который поддерживает различные сердечно-сосудистые, метаболические, иммунологические и гомеостатические функции и специфически связывается с глюкокортикоидным рецептором (GR, NR3C1). Связывание глюкокортикоидов с ГР приводит к транспортировке активированного рецептор-глюкокортикоидного комплекса в ядро, после чего он связывается с глюкокортикоидными ответными элементами (GRE) и регулирует экспрессию ГР-зависимых генов.

Любое стероидоподобное соединение, способное существенно влиять на промежуточный метаболизм. Глюкокортикоиды также являются клинически полезными противовоспалительными средствами.

Синтетически полученные формы встречающихся в природе глюкокортикоидов.

Астанин Павел Андреевич

Активация Windows
Чтобы активировать Windows, перейдите в раздел "Параметры".

ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА «КОНСУЛЬТАНТ ВРАЧА-РЕВМАТОЛОГА»

Интерактивная работа с глоссарием через текст найденной научной статьи.

Глоссарий

Сервисы подключения, принятая оценка | Калькуляторы | Клинические рекомендации | Журналы | Статьи | Диссертации | Глоссарий

Искать термин

1

Концепт
<https://yandex.ru/>
C1334899
GRL; GCCR; Глюкокортикоидный рецептор; Подсемейство ядерных рецепторов 3, элемент 1, группа C; Ядерный рецептор подсемейства 3 (элемент 1) группы C; GR; NR3C1; GCR; GP
Ген, участвующий в активируемой лигандом регуляции транскрипции. Мутации в этом гене связаны с резистентностью к глюкокортикоидам.
Астанин Павел Андреевич

Активация Windows
Чтобы активировать Windows, перейдите в раздел "Параметры".

СТЕК

ТО, НА ЧЁМ РАЗРАБАТЫВАЮТСЯ НАШИ СЕРВИСЫ

КОМПОНЕНТЫ СИСТЕМЫ

Базы данных:

- ❑ Реляционная (PostgreSQL);
- ❑ Графовая (Neo4j);
- ❑ Векторная (Chroma);
- ❑ Документоориентированная (MongoDB).



Web-приложение (Flask – Python 3.9):

- ❖ 77 микросервисов с API для удалённой работы;
- ❖ Пользовательские решения для аналитика, врача-эксперта.



Элементы информационной безопасности:

- Хеширование данных и параметризация запросов;
- Изоляция пользователей, отслеживание и логирование транзакций.

Оркестраторы процессов:

- Apache Airflow



**СПАСИБО ЗА
ВНИМАНИЕ!**

ФГАОУ ВО «Российский национальный исследовательский медицинский университет имени Н. И. Пирогова»
Минздрава России

Институт цифровой трансформации медицины

Перспективы развития и использования Унифицированной национальной медицинской номенклатуры

Астанин Павел Андреевич, аналитик лаборатории семантического анализа медицинской информации
института цифровой трансформации медицины (ИЦТМ), ассистент кафедры медицинской кибернетики и
информатики имени С. А. Гаспаряна (+79831580814; med_cyber@mail.ru)