

AI-ready statistics: Опыт и перспективы применения больших языковых моделей в датахабе Статкомитета СНГ

Юрий Акаткин
Научный
руководитель

Елена Ясиновская
Главный
аналитик



SI ПЛАТФОРМА

© ООО «Электронное проектирование»
www.e-projecting.ru

СЕМАНТИЧЕСКАЯ ИНТЕЛЛЕКТУАЛЬНАЯ ПЛАТФОРМА

Уникальное сочетание современных технологий управления знаниями, семантических сетей и искусственного интеллекта для создания умных предприятий, повышения эффективности работы и эффективного внедрения инноваций

СОВМЕСТНАЯ РАБОТА ЭКСПЕРТОВ И ИТ- СПЕЦИАЛИСТОВ

Капитализация развития современного общества за счет синергии профессиональных сообществ по методикам, обменом знаниями, консультациями

[Подробнее](#)

СЕМАНТИЧЕСКИ- ОРИЕНТИРОВАННАЯ СИСТЕМА УПРАВЛЕНИЯ ЗНАНИЯМИ

Semantic KMS

Откройте новые возможности управления знаниями, сохраните инвестиции в условиях стремительной цифровой трансформации бизнеса. Semantic KMS обеспечивает систематизацию и точное понимание корпоративной информации. Семантические технологии, интуитивно понятные интерфейсы и инструменты анализа позволяют сохранять, извлекать и использовать знания с максимальной эффективностью, улучшая принятие решений и стимулируя инновации

[Подробнее](#)

КАТАЛОГ СЕМАНТИЧЕСКИХ АКТИВОВ

SA Catalog

Легко решайте сложные задачи управления моделями, описывающими данные. Принципы FAIR – (F) находимость, (A) доступность, (I) интероперабельность, (R) повторное использование – поддерживаются мощным инструментом для упорядочивания, поиска и распространения семантических активов, построенном на базе международных стандартов

[Подробнее](#)

ВИРТУАЛЬНЫЕ AI ЭКСПЕРТЫ

VIA experts

Используйте умных виртуальных экспертов в качестве надежных помощников, повышая эффективность и скорость работы сотрудников. Применение передовых алгоритмов ИИ, персонализированные рекомендации, консолидация экспертных знаний обеспечивают принятие обоснованных решений и реализацию сложных инновационных задач

[Подробнее](#)

СИСТЕМА УПРАВЛЕНИЯ ЗНАНИЯМИ И СОД В ДАТАХАБЕ СТАТКОМИТЕТА СНГ



ПРИМЕНЕНИЕ ИИ В СТАТИСТИКЕ

Full Conference programme published [...read more](#)



[Registration](#) [Programme](#) [Capacity building](#) [Exhibition](#) [Venue and location](#) [Contact](#)

Сентябрь 2025.
SDMX Global Conference 2025



Smarter Data
for Better
Insights



Innovation

SDMX empowering AI - AI empowering SDMX
Microdata and SDMX
SDMX applications outside the statistical world



Implementation

SDMX in action - practical use cases and success stories
Improving statistical business processes using SDMX



Инновации

SDMX усиливает ИИ —
ИИ усиливает SDMX

Искусственный интеллект в статистике сегодня — ключевое направление развития, открывающее новые возможности для анализа данных, повышения качества статистических (мета)данных, цифровой трансформации рабочих процессов



UNECE

modernstats
WORLD - WORK

Session 2: Innovation and AI – best practices in dissemination and communication

В эпоху стремительной цифровой трансформации как статистическим организациям коммуницировать более эффективно и содержательно? На этой сессии будет рассмотрено, как инновации и искусственный интеллект меняют стратегии распространения — от подготовки контента с помощью ИИ до улучшения пользовательского опыта, повышения доступности и оптимизации рабочих процессов. Мы обсудим, как лучше понимать и охватывать целевые аудитории, а также адаптироваться к меняющимся цифровым условиям.

Ноябрь 2025.

Expert Meeting on Dissemination and Communication of Statistics

РОЛЬ SDMX В СОСД

SDMX

Широко применяемый стандарт,
эффективное взаимодействие
между статистическими
организациями

Документация
Глоссарий
Код-листы

Информационная
модель

RDF DATA CUBE VOC

Семантическая
интероперабельность, гибкость,
кросс-доменное использование,
распространение в Semantic Web

RDF модели

СОСД

Богатые метаданные
Контекст предметной
области

↑
семантическая
СУЗ

Богатые семантика и контекст, среда
интерпретации, экспертное сообщество

БАЗИС ИНТЕРОПЕРАБЕЛЬНОСТИ

Открытая некоммерческая инициатива, направленная на преодоление технологических и организационных барьеров, препятствующих эффективному обмену и распространению связанных данных

Цель

Интеграция существующих стандартов обмена данными, классификаций и справочных систем в среду Semantic Web для достижения **устойчивой семантической интероперабельности** в широком спектре пользовательских сценариев

Платформа

Платформа Базиса Интероперабельности (IoBP)

поддерживает семантическое выравнивание, обогащение и публикацию существующих стандартов с использованием системы управления знаниями, инструментов моделирования, контроля пространств имён и инфраструктуры устойчивых URI



INTEROPERABILITY BASIS PLATFORM

basis.semanticip.org

Log-in



SDMX

Agencies

Code Lists

Glossary

Documents

Library

English

For Developers

УЧАСТИЕ В ГЛОБАЛЬНОМ РАЗВИТИИ СЕМАНТИКИ SDMX

Разработка онтологии SDMX

- Активное участие в рабочей группе по разработке онтологии SDMX (Task Force SDMX Ontology – LOD)

Локализация и семантическое обогащение SDMX

- Совместная работа по локализации SDMX на сербский язык со статистической службой Республики Сербия

Выравнивание семантики SDMX

- Определение подходов совместной работы с MOT и техническим комитетом SDMX (SDMX Technical Working Group)

HLG-MOS. Проект AI-ready Dissemination

[Project Proposal for 2026 AI-Ready Dissemination.docx](#)

UNECE High-level Group for the
Modernisation of Official Statistics



Project Proposal on **AI-Ready Dissemination: Optimizing Statistical Products for Third-Party AI Consumption**

WP1. Обмен знаниями и кейсы

систематизация инициатив и обмен практиками применения ИИ в статистике

WP2. Качество данных

адаптация и развитие методов оценки и обеспечения качества статистических данных для ИИ

WP3. Технические подходы

разработка решений и стандартов для оптимального использования данных ИИ-системами

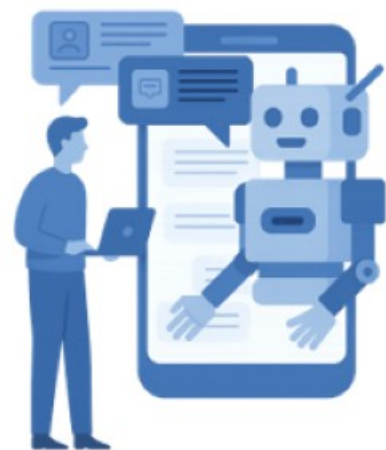
WP4. Взаимодействие с внешними участниками

сотрудничества с технологическими компаниями, госорганами и другими заинтересованными сторонами

СЕМАНТИЧЕСКАЯ СУЗ И ИИ: СИНЕРГЕТИЧЕСКИЙ ЭФФЕКТ

- **Формализованные знания предметной области**
- **Подготовленные и верифицированные экспертами статистические домены**
- **Семантически богатые метаданные**

семантическая
СУЗ



- **Надежные и заслуживающие доверия ответы LLM**
- **Точная интерпретация и корректное понимание пользователем**
- **Расширение и углубление знаний за счёт применения ИИ**

ОПЫТ ПРИМЕНЕНИЕ ИИ ДЛЯ РАЗРАБОТКИ ОКЭД-4

Первичный отбор LLM

- 6 моделей
- Корпус из 20 позиций NACE 2.1.
- Автоматизированная оценка на основе эталонного перевода

Выбор модели ИИ для перевода

- 3 модели
- Корпус из 100 позиций NACE 2.1.
- Слепая экспертная оценка (4 эксперта)

Перевод NACE 2.1

- Выбранная модель Gemma 3
- Улучшение результатов работы LLM
- Таблицы сопоставления (NACE 2.0, ОКЭД-3)

Выделение 5-го знака классификатора с применением ИИ

- Предложения ИИ по формированию 5-го знака
- Анализ и проверка экспертами
- Уточнение по результатам проверки

Подготовка проекта модельного классификатора экспертами

- Корректировка результатов перевода LLM (стиль, терминология, соответствие статистической практике)
- Анализ корректировок для улучшения результатов работы LLM

DONE



IN PROGRESS

СЦЕНАРИИ ПРИМЕНЕНИЯ ИИ В ДАТАХАБЕ СТАТКОМИТЕТА СНГ

ДЛЯ ЭКСПЕРТОВ СУЗ

- Подготовка проектов аннотаций для материалов СУЗ
- Анализ документов и выявление терминов-кандидатов
- Подготовка проектов определений и пояснений терминологических статей

ДЛЯ СТАТИСТИКОВ

- Подготовка справочных метаданных и методических разделов в соответствии с международными стандартами (Единая интегрированная структура метаданных, SIMS)



ПЕРСПЕКТИВЫ

Фокус на формировании экосистемы, в которой статистические знания, данные, метаданные и ИИ интегрированы с использованием семантических технологий

Статистические данные, ориентированные на применение ИИ

Обеспечение качества статистических данных уровня AI-ready за счёт развития СОСД и семантически богатых метаданных

Использование ИИ-ассистентов, основанных на формализованных знаниях предметной области

Внедрение ИИ в статистическую практику с использованием знаний, аккумулированных в СУЗ

Повышение качества аналитики и удобства работы пользователей за счёт семантики

Обеспечение точных и контекстно-зависимых ответов LLM, интерпретируемых для различных групп пользователей

**СПАСИБО ЗА
ВНИМАНИЕ!**

www.e-projecting.ru