

Краткое описание функциональных характеристик **ПО СКСА**

Аннотация

В данном документе приводятся общие сведения о программе Сервис каталогизации семантических активов, функциональные характеристики и назначение программы, описание функциональной структуры, необходимые программные и технические средства для функционирования программы, вызов и загрузка программы.

Термины и определения

1. Базовая модель данных – от англ, core model – информационная модель, отражающая понятия, применимые ко всем предметным областям.

2. Глоссарий – словарь узкоспециализированных терминов в какой-либо отрасли знаний с толкованием, иногда переводом на другой язык, комментариями и примерами. Глоссарий состоит из терминологических статей, моделируется как СА с учётом связей и отношений между терминами.

3. Домен (предметная область) – сфера знаний, влияния или деятельности, специализированная дисциплина, область знаний.

4. Карточка СА/нСД – это метаданные конкретного семантического актива (СА) или набора связанных данных (нСД) в каталоге. Карточка содержит следующие сведения (1) идентификационные метаданные (наименование, версия, описание, язык и др.), (2) классификационные признаки СА/нСД, (3) ссылки на элементы справочников объектных свойств; (4) связанные Wiki статьи; (5) информация о распространении.

5. Линейный справочник – справочник, элементы которого организованы в виде плоского списка без иерархической структуры, где каждый элемент имеет одинаковый уровень детализации и не содержит родительских/дочерних связей с другими элементами справочника.

6. Метаданные – данные, которые определяют и описывают другие данные. Семантически богатые метаданные формируются для описания показателей и наборов СД с использованием ИМ и СА.

7. Набор связанных данных (набор СД, нСД) – логически завершённая совокупность данных, опубликованная в формате связанных данных.

8. Набор связанных открытых данных (набор СОД) – набор СД, опубликованный и доступный в публичном каталоге в соответствии с открытой лицензией.

9. Общая модель данных – от англ, common generic model – информационная модель, которая отражает понятия, общие для конкретной предметной области, но не зависит от конкретной технологии или реализации. Общая модель данных также известна как концептуальная модель данных или модель данных предметной области (домена), представляет собой абстрактное высокоуровневое представление структуры данных и взаимосвязей в пределах определённого домена. Один из видов СА.

10. Онтология – концептуальная схема предметной области, содержащая совокупность терминов, выражений отношения и связанных с ними определений на естественном языке

вместе с одной или несколькими формальными теориями, предназначенными для отражения заданных интерпретаций этих определений. Один из видов СА.

11. Открытая лицензия – особый вид лицензионного договора, заключаемый в упрощённом порядке путём присоединения. Она позволяет автору или правообладателю сделать произведение доступным для неограниченного круга лиц, разрешая его использование, изменение или распространение на заранее определённых условиях.

12. Персистентность URI – неизменность и устойчивость ссылок на ресурсы, необходимая для обеспечения многократного (повторного) использования тех ресурсов, которые они идентифицируют, за счёт применения соответствующих рекомендаций и правил.

13. Портлет – модульный компонент веб-приложения, встраиваемый в страницу портала Liferay. Каждый портлет отвечает за собственный фрагмент интерфейса и логики (например, список карточек, настройка представлений), настраивается отдельно и может переиспользоваться на разных страницах.

14. Правила Semantic Web – набор основных правил представления данных в Интернет в виде, позволяющем адекватно интерпретировать их в информационных системах с учётом смыслового значения.

15. Связанные данные – данные, разработанные в соответствии с правилами Semantic Web. В 2017 г. Консорциум W3C рекомендует связанные данные в качестве наиболее эффективного способа открытия данных в Интернете.

16. Связанные открытые данные (Linked Open Data, LOD, СОД) – связанные данные, которые публикуются с использованием открытой лицензии, что способствует их повторному бесплатному использованию.

17. Семантические активы/модели (СА) – подготовленные для многократного использования описания данных: (1) метаданные, например, XML- и RDF- схемы; (2) базовые и общие модели данных, (3) онтологии, (4) тезаурусы, (5) справочные данные, например, списки кодов, таксономии, словари, глоссарии. СА каталогизируются и публикуются как открытые стандарты данных.

18. Стандарт данных – соглашение экспертного сообщества о представлении данных.

19. Токен доступа – параметр безопасности в виде уникальной последовательности символов, применяемый для идентификации субъекта и управления его доступом к функциям и ресурсам системы.

Сокращения

Сокращение	Описание
ИМ	Информационные материалы
нСД	Набор связанных данных
ПО	Программное обеспечение
СА	Семантический актив
СД	Связанные данные
СКСА	Сервис каталогизации семантических активов
СУБД	Система управления базами данных
СУЗиСМ	Сервис управления знаниями и семантического моделирования
DWH	Data Warehouse. Хранилище данных
ADMS	Asset Description Metadata Schema
DCAT	Data Catalog Vocabulary. Словарь каталога данных
HTTP API	Программный интерфейс для обмена данными между приложениями (или их компонентами), использующий протокол HTTP (или HTTPS) в качестве «транспорта»
JSON	JavaScript Object Notation. Универсальный текстовый формат для хранения и обмена данными
OLAP	On-line Analytical Processing – аналитическая обработка в реальном времени
OpenAPI	Открытый стандарт и спецификация для описания, создания и документирования веб-интерфейсов
RDF Data Cube Vocabulary	Словарь куба данных RDF
RDF	Resource Description Framework Модель для представления данных (метаданных) Разработанная Консорциумом W3C модель представления данных, и в особенности -метаданных. RDF представляет утверждения о ресурсах в виде, пригодном для машинной обработки
SKOS	Simple Knowledge Organization System. Система организации знаний. Общая модель данных для обмена и связи систем организации знаний через Интернет
SPARQL	Язык запросов к данным, представленным в соответствии с моделью RDF, а также протокол для передачи этих запросов и ответов на них
SDMX	Statistical Data and Metadata exchange. Международный стандарт обмена информацией, который разработан и ведётся в рамках международной инициативы по стандартизации и модернизации механизмов и процессов обмена статистическими данными и метаданными
XML	Extensible Markup Language (расширенный язык разметки) – самодокументируемый иерархический формат хранения и обмена данными и метаданными
URI	Uniform Resource Identifier. Унифицированный идентификатор ресурса Идентификатор ресурсов в соответствии со стандартом RFC3986, который определяет общий синтаксис URI и процесс разрешения ссылок URI, а также рекомендации и соображения безопасности для использования URI в Интернете

1. Общие сведения

1.1. Обозначение и наименование программы

Наименование программы: Сервис каталогизации семантических активов.

Обозначение программы: СКСА.

1.2. Назначение

Сервис каталогизации семантических активов (СА) обеспечивает сбор, публикацию и распространение подготовленных для многократного использования описаний данных для обеспечения семантической интероперабельности.

СКСА обеспечивает:

- формирование различных каталогов СА и управление правами доступа;
- настройка классификации СА и её применение для поиска и подбора СА;
- загрузка СА и его публикация в каталоге;
- создание/настройка/просмотр различных вариантов визуализации содержимого СА с учётом их специфики;
- автоматизация бизнес-процессов коллективной работы для решения задач экспертного сообщества;
- предоставление программных интерфейсов для взаимодействия с внешними сервисами.

1.3. Программное обеспечение

ПО СКСА разработан с использованием расширения европейского стандарта Asset Description Metadata Schema ADMS Workflow 1.0 на основе платформы Liferay 7.x.

ПО «Сервис каталогизации семантических активов» зарегистрировано Федеральным институтом промышленной собственности (ФИПС Роспатента), свидетельство о государственной регистрации программы для ЭВМ № 2022668514 от 07.10.2022г.

Сервис реализован по клиент-серверной архитектуре и может быть развернут как на одном сервере, так и в распределённой среде.

Для корректного функционирования программы СКСА на сервере и компьютерах клиентов должно быть установлено лицензионное программное обеспечение (ПО). Состав этого ПО и его установка описаны в документе «Руководство по установке программного обеспечения» СКСА.ЭП.РУ.01.

2. Функциональная структура

Функциональная структура включает следующие функциональные блоки:

- Блок основных функций в составе:
 - навигация в каталоге СА/нСД;
 - навигация в карточке СА/нСД;
 - управление файлами СА/нСД;
 - предоставление SPARQL-endpoint;
 - предоставление OpenAPI.

- Блок вспомогательных функций в составе:
 - ведение каталога;
 - настройка справочников объектных свойств;
 - настройка классификаторов СА/нСД;
 - ведение карточки СА/нСД;
 - настройка визуальных представлений;
 - печать и экспорт представлений в формат DOC и PDF.

- Блок сервисных функций в составе:
 - REST-сервис интеграции с внешними системами и инструментами разработки СА/нСД;
 - применение URL карточек СА/нСД и визуальных представлений;
 - разграничение доступа к функциям.

Состав элементов и свойств каталога СКСА определён на основе словаря DCAT версии 3¹ (далее – DCAT3) и дополнен следующими классами и свойствами:

URI класса	Комментарий
http://www.w3.org/ns/adms#Asset	Экземпляры класса являются семантическими активами. Класс является потомком класса <code>dcat:Resource</code>
http://purl.semanticpro.org/ontologies/sksa/Wiki	Экземпляры класса являются описанием Wiki статьи
URI свойства	Комментарий
http://purl.org/vocab/vann/preferredNamespacePrefix	Предпочтительный префикс пространства имён, используемый при использовании элементов в документе XML
http://purl.org/vocab/vann/preferredNamespaceUri	Предпочтительный URI пространства имён при использовании элементов в документе XML
http://purl.semanticpro.org/ontologies/sksa/teamwork	Элемент коллективной работы
http://purl.semanticpro.org/ontologies/sksa/wiki	URL на Wiki статью
http://csi.semanticpro.org/adms/wf/maturityLevel	Уровень зрелости семантического актива
http://csi.semanticpro.org/adms-wf#blocked	Флаг блокировки изменения уровня зрелости
http://csi.semanticpro.org/2016/ns#disjunction	Имеет дизъюнкцию с
http://csi.semanticpro.org/2016/ns#domain	Домен
http://csi.semanticpro.org/2016/ns#equivalent	Имеет эквиваленты
http://csi.semanticpro.org/2016/ns#extend	Расширяет
http://csi.semanticpro.org/2016/ns#generalize	Обобщает
http://csi.semanticpro.org/2016/ns#import	Импортирует
http://csi.semanticpro.org/2016/ns#metadata	Метаданные
http://csi.semanticpro.org/2016/ns#specialize	Специализирует

¹ Data Catalog Vocabulary (DCAT) - Version 3. <https://www.w3.org/TR/vocab-dcat-3/>

3. Блок основных функций

3.1. Навигация в каталоге СА/нСД

Функция «Навигация в каталоге СА/нСД» предназначена для выбора карточки СА/нСД в каталоге и перехода к её отображению. Функция выполняется пользователем и позволяет ему просматривать, фильтровать и выбирать нужные карточки, используя различные критерии: наименование, первую букву наименования и классификационные признаки. Работа функции требует предварительной настройки каталога и классификаторов администратором. Фильтрация поддерживает языковой фильтр (ru, en) в соответствии с настройками веб-интерфейса. Операции выполняются через настроенные веб-интерфейсы СКСА.

Функция «Навигация в каталоге СА/нСД» реализуется следующими операциями:

- Просмотр информации о каталоге
- Просмотр списка карточек каталогизированных СА/нСД
- Фильтрация списка карточек СА/нСД
- Удаление условий отбора

3.2. Навигация в карточке СА/нСД

Функция «Навигация в карточке СА/нСД» предназначена для просмотра и анализа информации о СА и нСД. Функция выполняется пользователем и является удобным инструментом для ознакомления с информацией о семантических активах и наборах связанных данных, обеспечивая понимание их структуры и связей.

Функция «Навигация в карточке СА/нСД» реализуется следующими операциями:

- Просмотр карточки СА/нСД.
- Просмотр и проверка микроразметки.
- Просмотр списка доступных визуальных представлений.
- Просмотр визуального представления.

3.3. Управление файлами СА/нСД

Функция «Управление файлами СА/нСД» предназначена для распространения СА/нСД с использованием внутреннего хранилища Liferay и хранилища RDF-графов Virtuoso.

Функция выполняется администратором в веб-интерфейсах СКСА и позволяет загружать, удалять и просматривать файлы СА/нСД как в Virtuoso (когда требуется доступ к данным с использованием SPARQL-запросов), так и в Liferay (когда файл предназначен только для скачивания), используя стандартные инструменты этих систем.

Функция «Управление файлами СА/нСД» реализуется следующими операциями:

- Загрузка (импорт) файла СА/нСД в Virtuoso
- Удаление СА/нСД из Virtuoso
- Загрузка (импорт) файла СА/нСД в Liferay
- Скачивание файла из Liferay
- Удаление файла из Liferay

3.4. Предоставление SPARQL-endpoint

Функция «Предоставление SPARQL-endpoint» предназначена для обеспечения унифицированного доступа к именованным графам в хранилище RDF-графов СУБД Virtuoso посредством SPARQL-endpoint, реализованного на базе Virtuoso. Функция доступна пользователю, позволяет выполнять SPARQL-запросы как интерактивно через веб-интерфейс, так и информационным системам через HTTP API, соответствующий стандарту W3C SPARQL 1.1 Protocol. Это обеспечивает интеграцию с внешними информационными системами и получение данных в машинно-читаемых форматах.

Функция «Предоставление SPARQL-endpoint» реализуется следующими операциями:

- Выполнение SPARQL-запросов.
- Использование шаблонов SPARQL-запросов.

3.5. Предоставление OpenAPI

Функция «Предоставление OpenAPI» предназначена для обеспечения унифицированного доступа к данным, хранящимся в RDF-графах СУБД Virtuoso (СКСА), через публичный REST API Semanticpro – HTTP-сервис только для чтения (read-only), который используется как через веб-интерфейс, так и внешними информационными системами.

Доступ осуществляется пользователем через интерактивный веб-интерфейс для просмотра и тестирования запросов или внешними информационными системами через программный интерфейс (REST API) для выполнения HTTP(S) запросы и получения данных в формате JSON.

Функция «Предоставление OpenAPI» реализуется следующими операциями:

- Использование OpenAPI Semanticpro.

4. Блок вспомогательных функций

4.1. Ведение каталога

Функция «Ведение каталога» предназначена для первичной настройки каталога СА/НСД, внесения изменений в информацию о каталоге, состав карточек каталога СА/НСД, а также для загрузки и выгрузки каталога в RDF-файл.

Функция выполняется администратором и реализуется следующими операциями:

- Загрузка типового шаблона каталога.
- Добавление и настройка портлетов СКСА с использованием интерфейса Liferay.
- Настройка информации о каталоге.
- Удаление каталога.
- Экспорт графа каталога.
- Импорт графа каталога.
- Создание карточки СА/НСД.
- Удаление карточки СА/НСД.

Настройка является обязательным этапом перед началом работы с каталогом.

4.2. Настройка справочников объектных свойств

Функция «Настройка справочников объектных свойств» предназначена для управления справочниками объектных свойств в каталоге СА/нСД. Управление осуществляется администратором через веб-интерфейсы СКСА, стандартные механизмы Liferay и WebProtege, а также вспомогательные функции СКСА.

Функция «Настройка справочников объектных свойств» реализуется следующими операциями:

- Создание и удаление элементов справочника.
- Настройка литеральных свойств элементов справочника.

4.3. Настройка классификаторов СА/нСД

Функция «Настройка классификаторов СА/нСД» предназначена для управления классификаторами СА/нСД в СКСА. Управление осуществляется администратором с использованием веб-интерфейсов СКСА, стандартных механизмов Liferay и WebProtege, а также вспомогательных функций СКСА.

Функция «Настройка классификаторов СА/нСД» реализуется следующими операциями:

- Создание и удаление классификатора
- Настройка литеральных свойств классификатора
- Связь классификатора с каталогом СА/нСД
- Создание и удаление элемента классификатора
- Настройка литеральных свойств элемента классификатора

4.4. Ведение карточки СА/нСД

Функция «Ведение карточки СА/нСД» предназначена для внесения изменений в карточки СА/нСД. Функция выполняется администратором с использованием настроенных веб-интерфейсов, а также стандартных механизмов и сервисов СКСА, Liferay и WebProtege.

Функция «Ведение карточки СА/нСД» реализуется следующими операциями:

- Настройка литеральных свойств карточки СА/нСД.
- Установка связей карточки СА/нСД.
- Создание и удаление карточки распространения СА/нСД.
- Создание и удаление связи с Wiki статьёй.

Перед началом работы с карточками СА/нСД требуется предварительная настройка каталога администратором, а также, при необходимости, настройка справочников и классификаторов с применением соответствующих функций СКСА.

4.5. Настройка визуальных представлений

Функция «Настройка визуальных представлений» предназначена для обеспечения визуализации СА/нСД с помощью визуальных представлений – способов показа данных (список, таблица, дерево, граф и т. п.). Функция выполняется администратором после создания карточки СА/нСД, в состав которой они входят.

Функция «Настройка визуальных представлений» реализуется следующими операциями:

- Создание визуального представления
- Редактирование настроек визуального представления
- Удаление визуального представления
- Копирование настроек визуального представления
- Копирование визуального представления

4.6. Печать и экспорт представлений в формат DOC и PDF

Функция «Печать и экспорт представлений в формат DOC и PDF» предназначена для печати HTML-страницы веб-интерфейса СКСА, а также экспортировать карточку или список карточек СА/НСД в файлы формата DOC или PDF.

Функция «Печать и экспорт представлений в формат DOC и PDF» выполняется пользователем и реализуется следующими операциями:

- Печать HTML-страницы СКСА
- Экспорт карточек СА/НСД

5. Блок сервисных функций

5.1. REST-сервис интеграции с внешними системами и инструментами разработки СА/НСД

Функция «REST-сервис интеграции с внешними системами и инструментами разработки СА/НСД» предназначен для обеспечения детального доступа администратора к элементам каталога через API, основанного на HTTP в формате JSON, а именно JSON Web Services Liferay

Функция «REST-сервис интеграции с внешними системами и инструментами разработки СА/НСД» реализуется следующими операциями:

- Получение данных
- Обновление данных
- Загрузка файла в Virtuoso
- Получение именованного графа

5.2. Применение URL карточек СА/НСД и визуальных представлений

Функция «Применение URL карточек СА/НСД и визуальных представлений» предназначена для предоставления пользователю URL карточек СА/НСД и визуальных представлений для встраивания во внешние системы.

Функция «Применение URL карточек СА/НСД и визуальных представлений» реализуется следующими операциями:

- Применение URL карточки СА/НСД
- Применение URL визуального представления
- Встраивание в HTML-страницы

5.3. Разграничение доступа к функциям

Функция «Разграничение доступа к функциям» предназначена для управления доступом к инструментам системы (Liferay, Virtuoso, WebProtege) на основе ролей, разрешений и членства в группах пользователей. Управление доступом к Liferay и WebProtege реализовано через Keycloak с поддержкой механизма единой аутентификации (Single Sign-On — SSO). Управление доступом к Virtuoso осуществляется на уровне пользователей и ролей СУБД.

Функция «Разграничение доступа к функциям» выполняется администратором и реализуется следующими операциями:

- Управление пользователями и разграничение прав доступа
- Применение правил разграничения прав доступа на основе ролей и разрешений Liferay
- Применение правил разграничения прав доступа на основе пользователей и ролей Virtuoso
- Применение правил разграничения прав доступа на основе пользователей и уровней доступа WebProtege

6. Используемые технические средства

Программа СКСА может работать на однопроцессорных и многопроцессорных компьютерах, построенных на платформе Intel x86 с установленным ПО Liferay Portal.

Для общесистемного ПО используются операционные системы на базе Linux, например, Ubuntu LTS 22.04, имеющую расширенный срок поддержки.

Для компоненты управления пользователями используется открытое ПО, поддерживающие протокол авторизации OAuth 2, например «Keycloak» – открытое ПО для идентификации и управления правами пользователей. компоненты хранения структурированных данных пользователей используется открытое ПО – СУБД «PostgreSQL».

Для компоненты хранения графовых данных используется открытое ПО – СУБД «Virtuoso» (Virtuoso Server Open Source Edition).

Для компоненты экспертного редактирования и сопровождения подготовленных для многократного использования описаний данных используется открытое ПО – WebProtege.

Для компоненты поддержки персистентности универсальных идентификаторов ресурсов используется открытое ПО веб-сервер Nginx.

СКСА функционирует по клиент-серверной архитектуре и требует выделенного сервера под управлением операционной системы семейства Linux. Минимальные требования к серверным техническим средствам приведены в Таблице 1.

Таблица 1. Требования к серверным техническим средствам

Параметр	Требование
Архитектура процессора	x86_64
Процессор	не менее 4 вычислительных ядер, 2.4 ГГц
Оперативная память	не менее 16 ГБ
Дисковая подсистема	SSD
Свободное дисковое пространство	не менее 100 ГБ
Сетевой интерфейс	1 Гбит/с
Операционная система	Ubuntu LTS 22.04

Загрузка (инсталляция) программы СКСА выполняется на сервере приложений в соответствии с Руководством по установке программного обеспечения (СКСА.ЭП.РУ.01).

Доступ пользователей к функциям СКСА осуществляется через веб-браузер без необходимости установки дополнительного специализированного программного обеспечения на рабочие станции пользователей.

Минимальные требования к клиентским техническим средствам приведены в Таблице 2.

Таблица 2. Требования к клиентским техническим средствам

Параметр	Требование
Процессор	не менее 2 вычислительных ядер
Оперативная память	не менее 4 ГБ
Разрешение экрана	не менее 1366×768
Веб-браузер	Google Chrome, Mozilla Firefox, Microsoft Edge (актуальные версии)
Сетевое подключение	доступ к серверу СКСА по HTTP/HTTPS

Для корректной работы пользовательского интерфейса СКСА в веб-браузере должна быть включена поддержка JavaScript и разрешено использование файлов cookie.