

ЭКСПЕРТНАЯ ИНФОРМАЦИОННАЯ СИСТЕМА ОЦЕНКИ РЕЗУЛЬТАТИВНОСТИ НАУЧНО-ТЕХНИЧЕСКОГО ЗАДЕЛА В ОБЛАСТИ АВИАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

Общие положения

На современном этапе развития научно-технологического потенциала авиационной отрасли особую актуальность приобретает необходимость принятия управленческих решений, направленных на повышение эффективности научно-исследовательских и опытно-конструкторских работ в области разработки, изготовления и эксплуатации авиационной техники. Такие решения должны базироваться на основе использования объективных аналитических данных о тенденциях развития мировых и отечественных научно-технологических разработок. Эти данные могут быть получены путем содержательного комплексного анализа мирового потока научно-технической информации и анализа цифрового контента отрасли.

Источниками такой информации могут быть зарубежные и отечественные отраслевые и ведомственные порталы, сайты научных учреждений, редакций журналов, научных конференций и сообществ, электронные средства массовой информации, а также множество других источников научно-технической информации, размещенных в сети Интернет. Источником цифрового контента отрасли может быть научно-технологическая, проектная и нормативно-техническая документация. В процессе автоматического содержательного анализа этих разноязычных гетерогенных распределенных научно-технических источников информации могут быть выявлены перспективные зарубежные опытно-конструкторские разработки в отрасли, получены новые знания и определены приоритетные научные направления деятельности научно-исследовательских коллективов.

В высокотехнологичных отраслях особую актуальность приобретают задачи автоматической обработки, анализа и управления цифровым контентом при преобразовании больших объемов научно-технологической, проектной и нормативно-технической документации из бумажной формы

представления в их цифровое представление. Такие задачи также невозможно решить без использования современных семантических инструментов автоматической обработки и формализации цифрового отраслевого контента, основанных на технологиях ИИ и «больших данных».

Отсутствие на отечественном ИТ-рынке апробированных технологий и сервисов высокоскоростной автоматической обработки, формализации и анализа неструктурированной текстовой информации является существенным сдерживающим фактором для реализации этих задач. При этом необходимость создания промышленных технологий формализации смыслового содержания разнородных разноязычных научно-технологических текстов является чрезвычайно актуальной. Эти технологии должны базироваться на современных представлениях о смысловой структуре разноязычных текстов и на новых высокопроизводительных методах и средствах анализа неструктурированной текстовой информации.

В течение нескольких последних лет во ФГБУ «НИЦ «Институт имени Н.Е. Жуковского» ведутся работы по созданию отраслевой авиационной многофункциональной экспертной информационной системы, ориентированной на решение этих задач. Научные подходы, положенные в основу этой системы, являлись логическим развитием исследований и технологий, направленных на создание нового класса информационно-аналитических систем, основанных на современных высокоскоростных методах обработки текстовой информации, мультиагентном семантическом поиске в разноязычных гетерогенных источниках информации, содержательном анализе научно-технологического цифрового контента и извлечения новых знаний из этого контента. Эти решения соответствуют требованиям Программы фундаментальных научных исследований РАН на 2013-2020 гг. (Информатика и информационные технологии. Когнитивные системы и технологии, нейроинформатика и биоинформатика, системный анализ, искусственный интеллект, системы распознавания образов, принятие решений при многих критериях), а также находятся в сфере приоритетов

цифровых технологий, заложенных в Программе развития цифровой экономики России до 2035 г.

Цели и задачи создания ЭИС-авиа

Целью научных решений, положенных в основу ЭИС-авиа, является разработка моделей, методов, алгоритмов и технологий создания интеграционной автоматизированной информационной системы анализа и управления знаниями, основной задачей которой является реализация единого цифрового информационного пространства отрасли и обеспечение целостной, достоверной информацией и знаниями всех участников жизненного цикла объектов и процессов в сфере управления разработкой и эксплуатацией сложных инженерных объектов отрасли.

Архитектура ЭИС, показанная на рис. 1, представляет собой архитектуру распределённой информационной системы и включает следующие основные компоненты:

- центральное программное ядро системы;
- аналитическое ядро;
- база знаний;
- база данных;
- библиотека НТД и РНТД;
- подсистема безопасности (ПИБ);
- подсистема интерфейса пользователя.

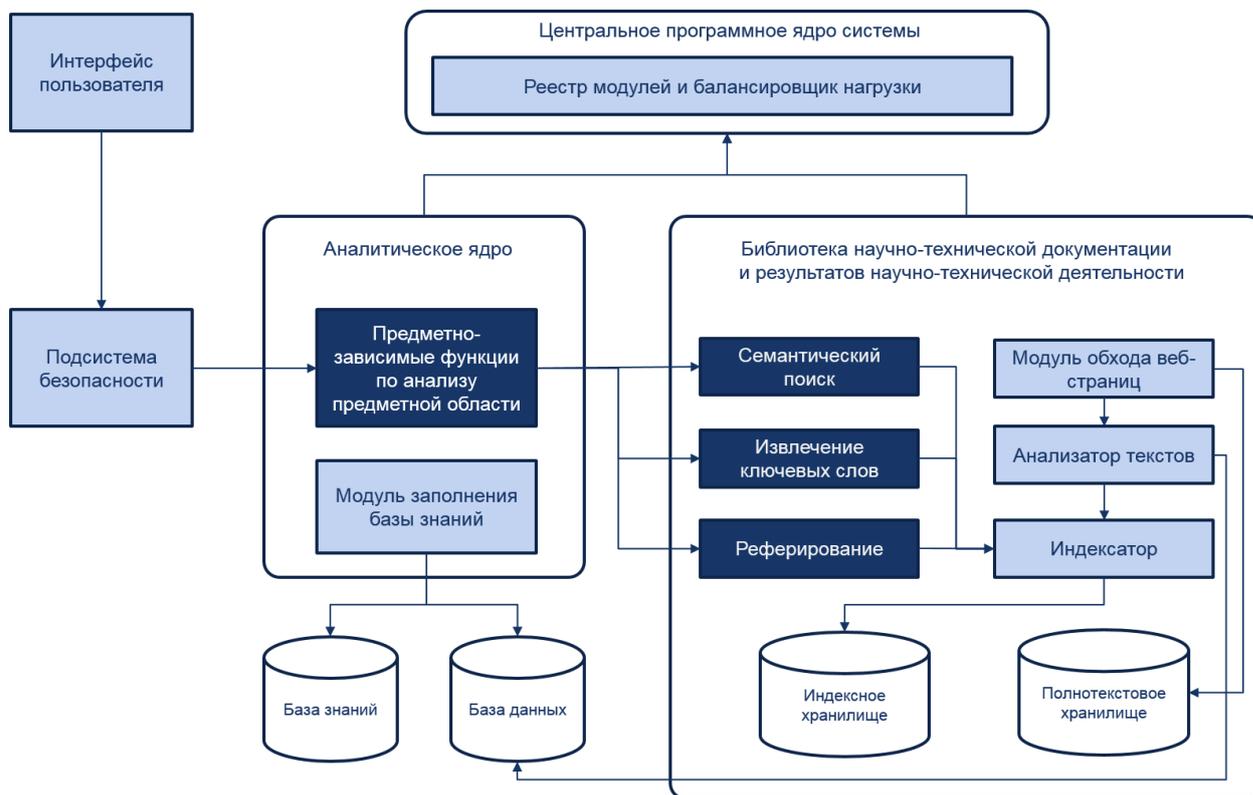


Рисунок 1-Архитектура ЭИС-авиа

Решаемые задачи:

- Разработка уточненных моделей, методов, алгоритмов и технологий автоматической обработки текстов, формализации их смыслового содержания и оперативного содержательного анализа разнородного разноязычного научно-технологического цифрового контента отрасли.

- Разработка унифицированного представления описаний результатов деятельности участников реализации процессов управления сложными инженерными объектами на разных стадиях их жизненного цикла.

- Обеспечение полноты и доступности конструкторской, проектной, эксплуатационной, научной и другой документации, относящейся к различным этапам жизненного цикла сложных инженерных объектов, в том числе за счет автоматизированного экспорта (или программной интеграции) данных из автоматизированных специализированных систем разного назначения, представляющих контент в различных форматах и с различными системами идентификации.

- Разработка уточненных интерфейсов сопряжения взаимосвязанных программно-информационных компонентов автоматического сбора, обработки, хранения, анализа и семантического поиска.

- Обеспечение стабильности во времени, полноты и точности поиска информации за счет формирования и динамической актуализации унифицированных лингвистических средств идентификации содержания документов.

При реализации разрабатываемой системы было осуществлено синтетическое соединение трех сложных программно-информационных компонентов:

- документальных информационно-поисковых систем и баз данных, обеспечивающих углубленный семантический поиск и анализ разнородной информации;

- лингвистического обеспечения, построенного на базе гибридных методов лингвистического, семантико-статистического и концептуального анализа текстов, обеспечивающих построение адекватных смысловому содержанию текстов документов и запросов их формализованных семантических образов, над которыми определены операции, в том числе корреспондирующие с процессами познания;

- человеко-машинных информационных интерфейсов, обеспечивающих персонализируемое представление познавательной траектории пользователя.

Это триединство рассматривается в контексте общего процесса синтеза, представления и поиска знаний, что предопределяет следующие задачи и направления работы:

- исследование процессов семантического поиска в области образования, производства и управления на основе системного подхода и когнитивных моделей этого процесса;

- разработка лингвистических декларативных средств и методов для обработки естественно-языковых документов и запросов, позволяющих

строить семантические поисковые образы, обеспечивающие возможность автоматического анализа и построения выводов на материале находимых по запросу документов (семантическое индексирование, картирование);

- разработка обобщенной объектной модели документа (как коммуникативной формы представления знания), обеспечивающей эффективное представление и обработку научного контента в процессах, относящихся к различным этапам жизненного цикла, а также формирование образа документа, идентифицирующего сам документ и его смысловое содержание на уровне базовых сущностей и отношений;

- разработка модели информационных ресурсов, обеспечивающей унификацию представления информации и метаинформации, а также построение сервисов доступа;

- разработка моделей «конструирования» вариантов представления предметной области (возможные «картины мира»), эволюция которых была бы предсказуемой;

- разработка моделей и механизмов поиска информации, основанных на динамической реформулировке запроса и адаптивной классификации текстов, реализующих управляемые процедуры формирования комбинаций документов, каждая из которых отражает возможную «картину мира» и имеет свое системообразующее основание;

- разработка системы идентификаторов для автоматического установления в текстах научно-технологической документации объектов жизненного цикла сложных инженерных систем и их параметрических характеристик методами семантико-синтаксического и концептуального анализа этих текстов;

- исследование и анализ контента отраслевых научно-технологических документов лингво-статистическими методами с целью выявления основных типов терминологических синтаксических конструкций и их лексических трансформаций при условии сохранения тождественности их смыслового содержания;

- разработка модели адаптивного визуального интерфейса, обеспечивающего сложные сценарии поискового взаимодействия, реализующие не только функцию «запрос/ответ», но также и аналитическую обработку, вплоть до интерактивных функций использования и актуализации «профилированного» лингвистического обеспечения, что по существу создает технологическую основу извлечения неявного знания.

- исследование и разработка методов отслеживания ключевой информации – по профилям мониторинга на основе контекстно-тематических многоязычных словарей.

- исследование и разработка методов обобщения многоязычных источников и выделения эксклюзивной информации.

- исследование и разработка методов анализа, сопоставления, систематизации, обобщения и выявления тенденций развития отечественных и зарубежных научных направлений и технологических разработок.

Практическими результатами явились следующие:

1. Базовая версия отраслевой промышленной интеграционной информационной системы мониторинга пространства Интернет и анализа цифрового информационного пространства отрасли (как основы цифровой платформы отрасли) для ситуационно-кризисных и аналитических центров высокотехнологических отраслей.

2. Базовый набор программно-информационных модулей интеграционной информационно-поисковой отраслевой системы и модулей их интеграции в основные технологические процессы отрасли.

3. Адаптированные под тематику научно-технологических и проектных документов отрасли декларативные средства ЭИС-авиа.

Технологические задачи, решаемые ЭИС-авиа

Потребителями предлагаемых технологических решений в отрасли может являться широкий спектр подразделений, в которых стоит задача сбора, обработки, анализа и управления научно-технологическим цифровым

контентом и выполняющих ряд частных научно-технологических задач, таких как:

- *Выявление и мониторинг центров компетенций авиационной науки.*

Информационное обеспечение процесса выявления и мониторинга центров компетенций авиационной науки.

- *Информационная поддержка процессов жизненного цикла разрабатываемых проектов.* Информационное обеспечение процесса разработки сложного научно-технического изделия путем поиска и анализа документальной информации в локальных и внешних документальных ресурсах, а также организации явных и неявных знаний, относящихся к проекту.

- *Анализ и экспертиза содержания научно-технических документов (НТД).* Информационное обеспечение эксперта для автоматизированного анализа смыслового содержания НТД в процессе работы с конкурсной, контрактной, рабочей, проектно-технологической и отчетной документацией.

- *Работа с многоязычной нормативно-технической документацией и подготовка к сертификации.* Информационное обеспечение процесса подготовки документации и проведения международной сертификации сложных научно-технических разработок.

- *Поддержка процесса проектирования на уровне контроля требований.* Информационное обеспечение процесса разработки сложного научно-технического изделия по требованиям и проведения формальной инспекции требований.

- *Поддержка процессов продвижения, эксплуатации и сопровождения сложной научно-технической продукции.* Информационное обеспечение процесса продвижения сложной научно-технической продукции, подготовки эксплуатационной и маркетинговой документации, а также презентационных и справочных материалов для международных рынков.

В части информационной поддержки процессов жизненного цикла сложных инженерных объектов выполняется следующее:

- фактографический поиск;
- тематический поиск;
- проблемно-ориентированный поиск;
- поиск аналогов;
- поиск зависимостей (динамики);
- комплексный поиск (мультиобъектный, многоэтапный) для задач мониторинга проекта (достижение технических требований, согласованности (непротиворечивость) параметров, свойств); анализа заимствований; выявления оснований, ограничений; выявление потенциально возможных семантических связей.

В части анализа и экспертизы содержания научно-технических документов (НТД) проводится:

- оценка соответствия документации установленным критериям и требованиям (техническое задание, требования стандартов и т.д.);
- соотнесение предложений/результатов с современным уровнем науки и технологий и заданными стратегическими направлениями развития;
- поиск аналогичных и близких проектов (выявление заимствований, анализ новизны по отношению к предыдущим проектам и этапам, выявление дублирования работ в смежных проектах);
- создание информационных моделей упомянутых объектов по заданной совокупности документов.

В части работы с многоязычной нормативно-технической документацией и подготовка к сертификации осуществляется:

- обеспечение стандартизированного перевода терминологического аппарата предметной области;
- обеспечение возможности работы с иностранной нормативно-технической документацией на языке пользователя;

- соотнесение требований национальных стандартов и внутренних технических регламентов с требованиями международных стандартов;

- соотнесение положений проектной документации с требованиями стандартов;

- максимальное использование результатов сертификации отдельных элементов и предыдущих моделей, ранее прошедших международную сертификацию;

- подготовка выходной документации на иностранном языке (с учетом версионности разработки и замечаний участников процесса сертификации);

- обеспечение трассируемости документации в процессе валидации отдельных положений требований.

В части поддержки процесса проектирования на уровне контроля требований осуществляется:

- автоматизированное создание и ведение информационной модели жизненного цикла изделия в части терминологии отдельных компонентов на разных стадиях жизненного цикла;

- обеспечение сквозного контроля взаимного соответствия требований по свойствам и параметрам на всех этапах ЖЦ (от стандартов, технического задания, планов разработки, спецификаций требований до рабочей документации и отчетов о проведении формальной инспекции);

- контроль использования стандартизированной терминологии при подготовке документации;

- соотнесение смысловых фрагментов, относящихся к информационным описаниям объектов по всей совокупности документов проекта;

- разработка требований с использованием формально-логических шаблонов и автоматизированная замена и подстановка составных частей требований из сопутствующей документации;

- использование онтологического ресурса в процессе разработки и анализа требований;

- автоматизированное проведение верификации и формальных инспекций требований проектной документации по заданным опросным листам с использованием технологии семантического поиска.

В части поддержки процессов эксплуатации и сопровождения сложной научно-технической продукции проводится:

- подготовка эксплуатационной и прочей технической документации на основе рабочей документации;

- обеспечение возможности подготовки различных типов документов на сложные технические изделия, состоящие из множества варьируемых элементов;

- быстрая подготовка эксплуатационной документации на новые модели с измененными характеристиками;

- автоматизированное извлечение из текста документации свойств и характеристик изделий и автоматическая генерация справочных материалов по заданным шаблонам;

- автоматический поиск и извлечение сведений по изделиям и создание сравнительных таблиц характеристик изделий;

- автоматизированный перевод эксплуатационной документации и справочных материалов;

- обеспечение информационного взаимодействия по решению вышеуказанных задач с предприятиями-смежниками и производителями комплектующих.

Научная новизна предлагаемых методов, моделей, алгоритмов и технологий состоит в создании принципиально новых по своим технологическим возможностям компонентов отраслевой интеграционной информационной системы мониторинга пространства Интернет и анализа цифрового информационного пространства отрасли. Это симбиоз деятельностного и когнитивного подходов и новых высокоэффективных лингвистических, датацентрических методов комплексной автоматической

обработки текстовой информации, формализации ее смыслового содержания, содержательного анализа и семантического поиска информации.

В отличие от известных систем предлагаемый проект характеризуется тем, что:

- база знаний включает не только метаданные и полнотекстовые документы, но и соответствующее лингвистическое обеспечение, динамически актуализируемое в процессе жизненного цикла системы;

- имеет возможность навигации, быстрого поиска и отбора документов на основе таксономий знаний, а также вербальных выражений и интерактивной реформулировки запросов;

- обеспечивает глубокое семантическое индексирование, включая построение масштабируемого онтологического образа содержания документа (онтологическое представление формируется на основе автоматического анализа смыслового содержания, а не только для существенных понятий и связей);

- лингво-статистические методы построения и использования формализованного семантического образа комплексно учитывают закономерности представления смысла на естественных языках и особенности представления знаний в документальной форме (в дальнейшей перспективе - фрактальность знаковых описаний и квантовую природу коммуникаций) на основе семиотических систем;

- имеет средства, позволяющие анализировать базу знаний и отдельные выборки, включая выявление взаимосвязей в существующей системе знаний;

- включает методологию и технологию создания, наполнения и использования базы знаний, включая адаптацию лингвистического обеспечения к смежным задачам и отраслям знаний.

В целом особенностью предлагаемого проекта является системность и диверсификация при смысловом анализе разнородной разноязычной научно-технологической документации по всем стадиям жизненного цикла сложных

инженерных объектов, включая генерацию, применение, использование и
воспроизводство знаний.

Роль ЭИС-авиа в оценке перспективности направлений развития авиастроения и оценки НИОКР

В настоящее время ЭИС-авиа используется для информационного обеспечения поддержки принятия решений, включая поддержку деятельности экспертов и аналитиков, например, при экспертизе тематик или результатов НИОКР в рамках реализации мероприятий Государственной программы Российской Федерации «Развитие авиационной промышленности», в части:

- проведения оценки уровня готовности создаваемых технологий (technology readiness level) с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов;

- выполнения оценки уровня готовности производства (manufacturing readiness level) с использованием функции полнотекстового поиска ЭИС, построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов и логического вывода по базе знаний;

- выявления вариантов использования технологий с использованием методов извлечения информации из полных текстов научно-технических документов (в первую очередь патентов, как документов, содержащих описание вариантов применения технологий), построения ключевой лексики и резюме полнотекстовых научно-технических документов;

- проведения оценки соответствия направлениям исследований и тематикам ОКР, ведущимся за рубежом, с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов, в том числе базам ведомства по патентам и товарным знакам США (USPTO) и всемирной организации интеллектуальной собственности;

- осуществления оценки наличия задела, патентов и научных публикаций, тематически близких к теме НИОКР, у коллектива

исполнителей НИР или ОКР с использованием функции поиска похожих документов;

- определения соответствия тематики НИР или ОКР перспективным направлениям с использованием функции поиска тематически похожих документов;

- установления соответствия полученных результатов, представленных в отчете, декларируемой тематике НИОКР с использованием методов оценки тематического сходства полнотекстовых документов;

- выявления похожих или дублирующих НИОКР с использованием функции поиска нечетких дубликатов и текстовых заимствований в базах НИР и ОКР;

- поиска похожих научно-технических решений с использованием функции поиска тематически похожих документов по базам научно-технических решений;

- проведения ранжирования центров компетенций авиационной науки согласно техническому заданию на НИР или ОКР;

- выполнения оценки динамики публикационной активности научного сообщества в рамках заданной тематики тематической области с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов, в том числе построения диаграммы распределения научных публикаций и авторефератов диссертаций по годам;

- осуществления оценки уровня освоенности направления патентообладателями с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов, в том числе построения диаграммы распределения патентов по правообладателям;

- проведения оценки актуальности результатов НИР или ОКР с использованием функций полнотекстового поиска ЭИС по коллекциям патентов и заявок на патенты с целью выявления соотношения патентных

заявок к действующим патентам, содержательно и тематически похожих на полученные результаты;

- выполнения оценки при развитии тематики, соотношения количества промышленных образцов и программ для ЭВМ и количества изобретений и полезных моделей с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов, в том числе построения диаграммы распределения патентов по их видам;

- выявления интереса к направлению со стороны держателей патентов в России и за рубежом с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов, в том числе построения диаграммы кумулятивной патентной массы по анализируемому направлению, при этом оценивается прирост кумулятивной патентной массы по анализируемому направлению развития авиастроения;

- выявления интереса к направлению со стороны СМИ с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых публикаций профильных СМИ, в том числе построения диаграммы распределения публикаций СМИ по годам;

- прогнозирования оценки ожидаемого уровня конкуренции в РФ со стороны зарубежных компаний с использованием функций полнотекстового поиска ЭИС и построения агрегированной статистики по коллекциям полнотекстовых научно-технических документов, в том числе построения диаграммы распределения патентов по правообладателям и диаграммы государственной принадлежности патентообладателей с вычислением доли зарубежных промышленных компаний среди патентообладателей;

- проведения оценки наукометрических показателей исполнителей коллектива с использованием функций полнотекстового поиска и данных наукометрических баз;

– выявления терминологии, используемой в научно-технических документах, для обозначения ключевых элементов создаваемых технологий с использованием методов извлечения информации из полных текстов, построения ключевой лексики научно-технических документов и их коллекций, построения резюме полнотекстовых документов;

– выявления технологий, созданных в рамках направления, с использованием методов извлечения информации из полных текстов патентов (как научно-технических документов, содержащих описание технологий и их применения) и логического вывода по базе знаний ЭИС;

– выявления вариантов использования технологий, созданных в рамках направления, в авиастроении с использованием методов извлечения информации из полных текстов патентов (как научно-технических документов, содержащих описание технологий и их применения), построения ключевой лексики патентов и их коллекций, построения резюме.

В зависимости от конкретных задач, связанных с оценкой перспективности направлений развития авиастроения или оценкой НИР и ОКР, эксперты и аналитики могут применять различные подмножества перечисленных выше (или других) вариантов использования ЭИС.

В настоящее время во ФГБУ «НИЦ «Институт имени Н.Е.Жуковского» также проведен комплекс исследований, направленных на решение частных задач сбора, обработки, поиска и анализа цифрового контента отрасли, а также создан научно-технологический потенциал для решения этих задач в промышленных масштабах. Предлагаемая ЭИС-авиа позволит обеспечить создание единого цифрового информационного пространства отрасли путем реализации унифицированных технологий автоматической обработки и формализации смыслового содержания нормативно-справочной, научно-технологической и проектной документации. Эти технологии также могут обеспечить сопряжение различных сторонних информационных систем на уровне обмена

документов, обработанных по единым стандартам семантического представления документов, полученных из различных видов информационных источников.

Основным конкурентным преимуществом предлагаемой концепции является то обстоятельство, что разработка этой системы будет базироваться на современных теоретических представлениях о смысловой структуре разноязычной текстовой информации, обеспечивающей реализацию перспективных технологий автоматической обработки нормативно-справочной, научно-технологической и проектной документации, семантического поиска по отраслевому массиву этих документов, содержательный анализ и механизмы извлечения знаний из цифрового контента отрасли.

В рамках мероприятия получено свидетельство о государственной регистрации программы для ЭВМ № 2021612534 «Экспертная информационная система – (ЭИС-АВИА)». Дата государственной регистрации в Реестре программ для ЭВМ 18 февраля 2021г.