

**Технология сохранения опыта эксплуатации ядерных установок с
использованием онтологического и датацентрического подходов в качестве нового
направления управления знаниями -
“Инженерия ядерных знаний”**

A Semantic-Based Approach
for Preserving Operational Experience of Nuclear Installations

А.Н. Косилов*, В.М. Куприянов*, И.А. Куприянова**, Максимов Н.В.*

*) НИЯУ МИФИ, Национальный центр Международной ядерной информационной системы
ИНИС МАГАТЭ.

***) НПО «ТАЙФУН»

Реферат

В работе представлено современное состояние одного из направлений деятельности для обеспечения сохранения и передачи знаний, возникающих в процессе эксплуатации объектов ядерной энергетики. Обсуждены проблемы, возникающие при разработке методов фиксации опыта эксплуатации ядерных установок для обеспечения безопасности процедур реализации их жизненного цикла. Предлагается подход, позволяющий использование простейших средств онтологического описания для целей идентификации сущностей фактологических описаний документов, возникающих в процессе реализации долговременных эксплуатационных процессов. Специальное внимание уделено особенностям использования национальных (русскоязычных) документов совместно с англоязычными ресурсами МАГАТЭ в сфере управления ядерными знаниями.

Обосновываются рекомендации необходимости развития специализированных средств и подготовки соответствующих специалистов для описания документов по ядерно-технологической тематике и создания базовых лингвистических ресурсов (терминологических словариков) для отражения формализованных связей между текстами документов и ядерно-технологическими знаниями в обеспечение информационной поддержки жизненного цикла отечественных проектов АЭС. Обосновывается новая компетенция в управлении ядерными знаниями - «Инженерия ядерных знаний» - «ИЯЗ» (Nuclear Knowledge Engineering) - NKE).

Цель работы

Проблемы управления знаниями интенсивно обсуждаются в ядерном сообществе, в частности, в МАГАТЭ с 2004 года [1]. Хотя основы этой дисциплины были созданы в 90-е годы прошлого века. Принято считать, существует три вида знаний. Явные - информация или знания, которые представлены на материальном носителе. Неявные: информация или знания, которые не содержащиеся в материальной форме, но могут быть сделаны явными. Скрытые - информация или знание, которые чрезвычайно трудно изложить в материальной форме [2]. Автор утверждает, что важно не сводить количество типов до двух: скрытое-неявное, потому что при этом теряется очень важная сфера деятельности – создание процедур для описания и идентификации знаний в предметной области. в частности, фиксация опыта эксплуатации [3]. Настоящий доклад посвящен именно технологическому аспекту фиксации сохранения знаний.

Особенности механизма идентификации каждой единицы хранения в системе сохранения знаний, являются базовыми принципами системы сохранения знаний. Обычно идентификация осуществляется путем определения атрибутов единицы хранения и фиксации их в качестве дополнительных имен (координат) в системе хранения. Основными координатами для знаний являются имена разделов тематических рубрик, которые и используются для индексирования записей.

Тематическая рубрикация – это процедура связывания каждого элемента сохраняемого объекта с одним из predetermined классов. Классы определяются единством некоторых понятий (в случае тематической рубрикации) или через совпадение структурных атрибутов (например, физические типы хранимых в системе объектов). Например, для предметной области “информационная поддержка жизненного цикла АЭС” возможна рубрикация понятий по режимам работы - “нормальный режим”, “нештатном режиме” или “режим перегрузки” и так далее. Второй тип рубрикации может определять вид описываемого документа, то есть, - “техническое задание”, “руководство по эксплуатации”, “информационное письмо”, и так далее.

Необходимо отметить, что базовая концепция управления явными знаниями связана с именно с процедурами тематической рубрикации (классификации), поскольку структурные аспекты могут быть определены с помощью более простых средств, например, по типу имени файла в компьютере, используемом для хранения знаний.

Как правило, в тематических рубриках недостаточно строго формализованы атрибуты объектов, необходимые для уникальной идентификации нового элемента, несмотря на существование predetermined рубрик в рубрикаторе. На практике специалист-индексировщик осуществляет рубрикации “вручную” для каждого вновь описываемого элемента хранения. Однако, в настоящее время возникла острая необходимость в создании информационной системы для вновь строящихся АЭС, поскольку требуется описать для сохранения огромное количество (более 100000 объектов) различных документов в информационном хранилище [4]. Это только одна из многих причин совершенствования методов автоматизации документа рубрикация для поддержки жизненного цикла АЭС.

Автоматизация тематической рубрикации документов эксплуатации АЭС позволит решить одну из наиболее актуальных задач информационной поддержки жизненного цикла АЭС – создание долгосрочной информационного хранилища для управления явными знаниями. Решении этой задачи является базовым условием для повышения безопасности эксплуатации АЭС в течение более чем 40 летнего горизонта срока эксплуатации.

Рубрикации и индексации

Реализация долговременного хранения знаний возможна только, когда хранимые объекты в минимальной степени зависят от «железа» компьютерной системы, ее операционной системы и программного обеспечения. Подход к реализации компьютерных систем с обеспечением такой независимости определяется как подход, ориентированные на данные (датацентрический). Реализация данного подхода возможна, если все необходимые сведения о свойствах хранимых

объектов сохраняются совместно с содержательными объектами в системе хранения [5]. Это возможно, если для каждой единицы хранения будет сделано дополнительное информационное описание - “знание о хранимых знаниях”, то есть “Мета-описание”. Мета-описание в среде хранилища обеспечивает связь между набором хранимых данных и содержанием фиксируемого знания. В большинстве случаев содержательная интерпретация мета-описания является основным способом описания хранимых знаний, необходимым для использования возможностей компьютера для решения задач, обычно решаемых интеллектом человека при тематическом анализе. Первым шагом для создания мета-описания является тематическая рубрикация.

Рубрикация может быть сделана с помощью компьютерной процедуры, если она может связать набор ключевых слов, извлекаемых из хранимого текста с соответствующей рубрикой рубрикатора. В настоящее время существует общепринятый во всем мире для ядерно-технической областей знаний тематический рубрикатор ИНИС [6] и тезаурус ИНИС как база дескрипторов (ключевых слов), связывающих тексты с рубриками. В России стандартом определен Государственный рубрикатор научно-технической информации ГРНТИ. Рубрикатор ИНИС определен как одноуровневая таблица, содержащая около 100 наименований тематических рубрик, которые охватывают все области применения ядерных знаний. В России определен Национальный стандарт для русскоязычных информационных систем Государственный рубрикатор научно-технической информации – GRNTI [7].

Однако, он содержит только состав рубрик и не имеет списка дескрипторов. Специалисты из Национального центра ИНИС Российской Федерации, вместе со специалистами Секретариата ИНИС провели работу по тестированию различных способов автоматизации рубрикации для процедуры тематической рубрикации, используя анализ результатов поиска ключевых слов в анализируемом тексте. Использовались типовые аннотации русскоязычных публикаций в качестве входной информации.

Процедура рубрикации

Специалисты Национального центра ИНИС РФ, вместе с персоналом кафедры английского языка НИЯУ МИФИ, начиная с 2012 года, пытались создать автоматизированную систему рубрикации технических текстов. Первый прототип был испытан в рамках проекта системы информационной поддержки для макета экспериментального жидкометаллического реактора на быстрых нейтронах. Реализовывалась структурная таксономия, используемая проектной организацией, которая была основана на рекомендациях, принятых по результатам разработки МАГАТЭ для создания системы описания знаний о быстрых реакторах. Рубрики верхнего уровня тематического рубрикатора для идентификации понятий были взяты из рубрикатора ИНИС. Скриншот фрагмента таксономии приведен на рис 1.

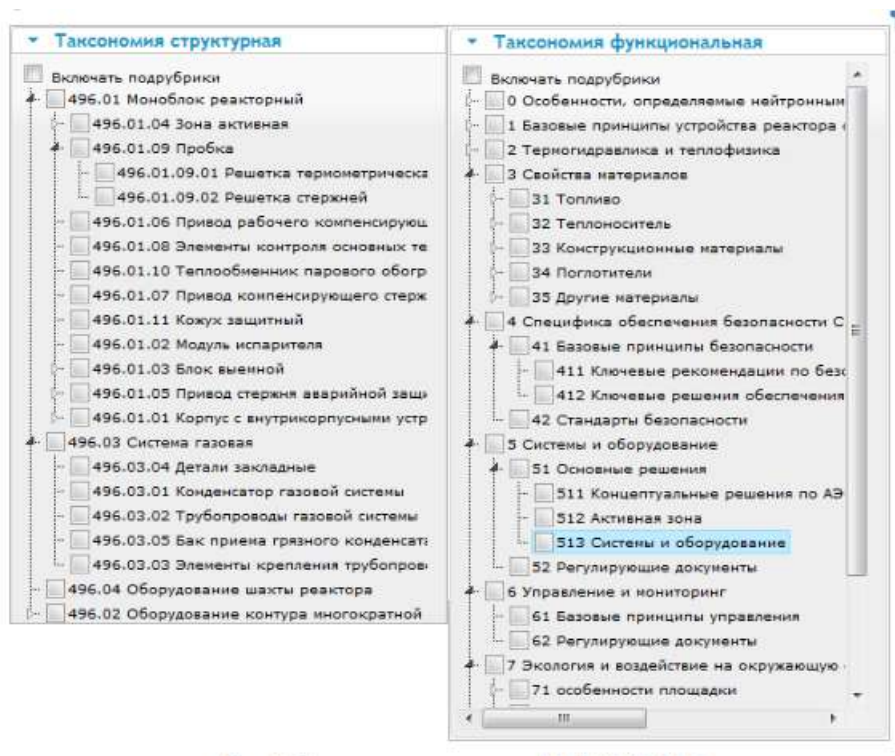


Рис. 1. Систематика физическая структура, и концепция таксономии для быстрых нейтронов установки.

На первом шаге создания процедуры, было составлено около 100 ключевых слов на русском языке для каждой рубрики реального рубрикатора системы, эти ключевые слова были отобраны из типичных текстов ядерно-энергетической сферы. Они были определены в качестве дескрипторов из рубрики. Затем, осуществлялась рубрификация с помощью специальной компьютерной программы сравнения последовательности символов каждого дескриптора с словами рубрицируемого текста. Количество совпадений между дескрипторами и ключевыми словами текста были использованы в качестве индикатора рубрики. Однако, этот простой алгоритм обеспечивал очень низкую стабильность для тематически сходных текстов, поэтому оно было отклонен.

На втором этапе было принято решение создать максимально простую онтологическую модель для создания более надежных связей между субъектами тематических рубрик и ключевых слов в исследуемом тексте. В ее основу была положена модель связей данных, обеспечиваемая ссылками от ключевых слов к рубрикам, сформированная по аналогии с моделью, описанной в рекомендации стандарта ISO 15926:2. За основу этой модели была взята совокупность определений иерархии тезауруса ИНИС [8]. Некоторые новые типы ссылок были включены в качестве дополнительных элементов.

Была реализована модель связей данных, которая обладает следующими свойствами.

1. Модель позволяет выделить связи как между физическими объектами, так и тематическими понятиями.

2. Модель поддерживает не только ссылки типа “сужающий термин” - “расширяющий термин”, но таких как “нечто есть часть чего-то”, “нечто содержит в себе что-то”, “нечто подобно чему-то” и некоторые другие, как для физических объектов и тематически понятий.

3. Ссылки позволяют определить для объекта рефлексивное отношение.

Затем, был определен набор уникальных атрибутов для каждой рубрики из рубрикатора. После

этого, опытные специалисты-индексировщики разместили связи типичных ключевых слов с атрибутами рубрик с помощью ссылок ссылкам модели данных. Используемая на сегодняшний день Вся модель имеет более 50 типов ссылок.

Схема алгоритма рубрикации показана на рис 2.

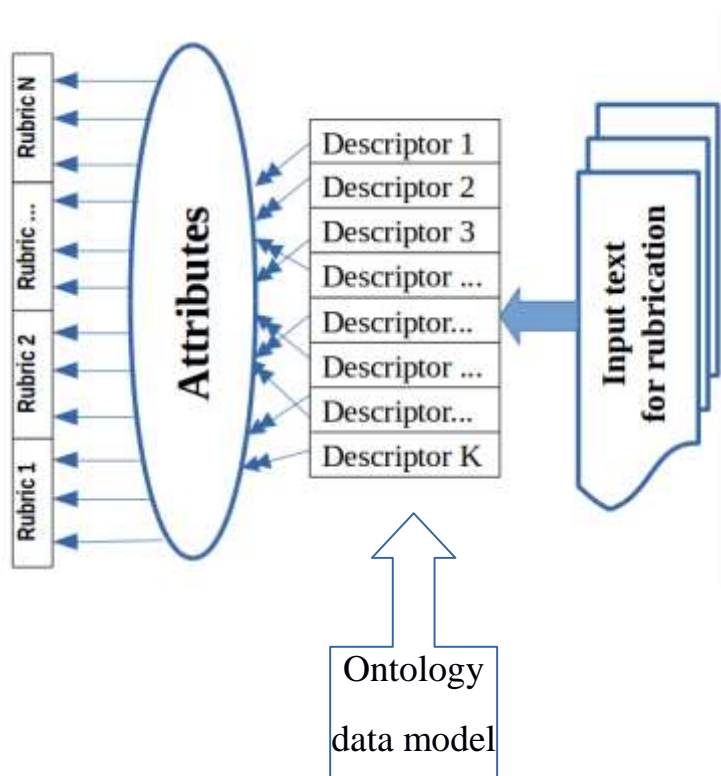


Рис. 2. Схема онтологической рубрикации

Результаты.

Тестовые исследования были реализованы на совокупности русскоязычных эксплуатационных документов АЭС. Получены следующие результаты.

1. Тщательный выбор оптимальных атрибутов в качестве признаков для каждой рубрики является очень важным этапом создания системы рубрикации. Оптимизация можно быть обеспечена только высококвалифицированными профессиональными индексировщиками (например, из персонала Национального центра ИНИС для область знаний «эксплуатация АЭС»).
2. Онтологическая модель данных дает более стабильные и надежных результаты рубрикации текста, чем простой учет ключевых слов в рубрицируемом тексте. Особенно, важно разделить ссылочные связи отдельно на физические объекты и тематические понятия сущности. Это позволяет установить различие между описаниями “быть частью чего-л.” и “быть более узким понятием чего-либо”. Очевидно, это требует специальной дополнительной разметки свойств дескрипторов и атрибутов рубрик.
3. Очень важно оптимизировать рабочий список дескрипторов для каждой подобласти знаний. Тестирование показывает, что избыточное количество дескрипторов для какой-либо темы дает менее надежные результаты рубрикации, чем “оптимальный набор”. Для русского языка величина 50-70 дескрипторы для каждой рубрики оказалась приемлемой. Кроме того, дескрипторы должны быть “общеупотребительными терминами” обычного текста. Таким

образом, оптимальный список описателей (как атрибутов, так и дескрипторов) может быть выбран и назначен только “вручную” силами высококлассных специалистов – индексировщиков с большим опытом работы

4. Все отмеченное выше позволяет утверждать, что создание автоматизированной системы рубрикации и системы мета-описаний эксплуатационных текстов требуют специалистов, имеющих уникальные компетенции - “инженеров по ядерным знаниям”. Они должны знать:

- предметная область АЭС;
- базовые лингвистические понятия;
- принципы семантического анализа;
- структуру и содержание различных рубрикаторов (ИНИС, GRNTI – для русского языка);
- терминологию английского языка в ядерной области;
- основные понятия и принципы построения современных онтологических описаний;
- базовые принципы компьютерной информатики (в т.ч. проектирование баз данных).

Перечисленные компетенции и знания определяют облик нового специалиста «Инженер по ядерным знаниям – «ИЯЗ». Сохранение явных ядерных знаний для современных установок требует значительного количества таких специалистов. Подготовка таких специалистов – должна стать задачей профильных учебных организаций, в частности, НИЯУ МИФИ.

¹ IAEA-TECDOC-1399, The nuclear power industry’s ageing workforce: Transfer of knowledge to the next generation, 2004

² KOENIG M. E. D., Information services and downstream productivity, *Annual review of information science and technology*, 1990, vol. 25, pp. 55-86 [32 page(s) (article)] (11 p.)

³ I. Kitaev, V. Kupriyanov, Y. Yanev, *Preserving nuclear knowledge in CIS countries: its current status and future prospects* *Int. J. of Nuclear Knowledge Management* / 2009 Vol.3, No.4, pp.423 – 430
A. Kosilov, Y. Yanev, T. Mazour, Knowledge management for a new nuclear power infrastructure - *Int. J. of Nuclear Knowledge Management* 2009 - Vol. 3, No.4 pp. 431 - 440

⁴ Information and technological solutions applied for knowledge-management tasks / Golitsina, O.L., Kupriyanov, V. M., Maksimov, N.V. // *Scientific and Technical Information Processing*. – 2015. – 210. - p. 150-161

⁵ Standard ISO 15926-1-2008 (1-11) Industrial automation systems and integration -- Integration of life-cycle data for process plants including oil and gas production facilities

⁶ Subject Categories and Scope Descriptions, Rev. 1 (August 2010), IAEA-INIS/ETDE-02, 2010, IAEA, Vienna, Austria

⁷ GRNTI (ГРНТИ - ГОСУДАРСТВЕННЫЙ РУБРИКАТОР НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ 2007 — 2016), <http://www.grnti.ru>

⁸ NIS Thesaurus, IAEA-INIS-01 (2015/12); INIS Thesaurus Supplement, IAEA-INIS-01 (2015/12.Supp), IAEA, Vienna, Austria