

О СОЗДАНИИ И УЧЕТЕ ЭЛЕКТРОННОЙ ДОКУМЕНТАЦИИ НАУЧНОГО ХАРАКТЕРА¹

Борисов А.В., Куриленко И.Е.

Московский Энергетический Институт (Технический Университет),
Россия

В настоящий момент во многих научных организациях применяются электронные системы хранения знаний и данных. В первую очередь они ориентированы на обеспечение сохранности и доступности электронных документов в рамках системы. Как правило, это электронные библиотеки или электронные каталоги. Разработаны и внедрены электронные системы государственной регистрации НИОКР, диссертаций, описаний алгоритмов и программ, предназначенные для учета результатов научной деятельности. Как правило, подобные системы представляют собой базы данных с возможностью поиска по ряду критериев. Системы такого класса не предназначены для работы с научным материалом, находящимся в стадии разработки или исследования, а ориентированы на сохранение подготовленных ранее документов. Кроме того, зачастую, возможности доступа к ним ограничены.

С другой стороны существует большое количество wiki-подобных систем, которые обеспечивают широкую доступность подготовленного материала и поддержку стадии разработки, но не обладают развитыми средствами экспорта материала в стандартную для отчета форму, механизмами упрощенного ввода и обновления информации, возможностью многокритериального поиска информации, ее оценки, а также не содержат развитых средств разграничения доступа. При работе с

¹ Работа выполнена при финансовой поддержке РФФИ

системами данного класса необходимо владеть специфическим для системы языком разметки, что часто удобно для IT специалиста, но может вызывать затруднения у исследователя, не являющегося профессионалом в этой области. Последнее особенно актуально при подготовке материала сложной структуры, содержащего таблицы, гистограммы, графики. Тем не менее, популярность систем данного класса в научном сообщества растет, т.к. они обеспечивают один из наиболее эффективных способов коммуникации между распределенными командами и предоставляют возможность публикации результатов исследований с целью вовлечения коллег в научную дискуссию или привлечения партнеров на начальных стадиях исследований.

В рамках данной статьи предлагается подход, позволяющий разработать систему работы с электронной документацией научного характера, сочетающую преимущества систем учета электронной документации и поддерживающую начальную стадию исследований, т.е. стадию разработки.

Коротко резюмируем ряд характеристик систем, которые представляются авторам существенными для дальнейшего рассмотрения.

Электронные базы и каталоги

- хранение документов в исходных форматах документа или в одном из общепринятых форматов (PDF, TeX и тд.);
- поддержка поиска по атрибутам;
- хранение финальной версии документа;
- ограниченный доступ к каталогу или библиотеке;
- высокая степень формализации;
- ориентация на хранение и учет электронных документов.

Wiki-подобные системы

- хранение полнотекстовой информации с поддержкой гипертекста;

- поддержка поиска по содержимому документа;
- хранение документов разной степени готовности с поддержкой версионности;
- поддержка этапа создания документа;
- низкая степень формализации;
- широкая доступность.

В силу специфики области применения электронные каталоги ориентированы на законченные работы, удовлетворяющие обязательных требований. Это не всегда удобно на стадии исследования, т.к. на этом этапе данные могут не иметь законченный вид. В тоже время часто полезно представить эти данные для обсуждения в научном сообществе.

С другой стороны, форма представления материала принятая в wiki-системах позволяет представить такой незаконченный материал. При формальном же оформлении и регистрации работ эти системы не предоставляют широкого набора функций, что приводит к необходимости выполнять работы по форматированию и экспорту в требуемый формат многократно. Последнее увеличивает объем рутинных работ, снижая привлекательность wiki-систем для работ, находящихся в высокой стадии готовности, что ведет к снижению интенсивности обновления и потери актуальности материала.

Важно также отразить ряд качеств рассматриваемых систем, которые не всегда в достаточной степени реализованы в существующих системах. Наличие этих качеств является, по мнению авторов, важным при практическом внедрении предлагаемого решения:

- представление и хранение знаний на базе выбранной формальной модели;
- упрощенное редактирование материала;
- упрощенный режим редактирования формул.

В данной работе предлагается применение систем накопления знаний (СНЗ), как основы для создания и модификации электронных материалов научного назначения в организациях, ведущих научную деятельность. СНЗ, построенная по описанному в [1] принципу, сочетает в себе преимущества wiki-подобных систем и возможности, необходимые для генерации отчета.

В рамках организаций-исследователей предлагается создать на базе СНЗ локальное хранилище знаний с внутренним модерированием. Новый материал попадает в такое хранилище на раннем этапе непосредственно при создании (см. Рис.1, Рис. 2.). В силу того, что система накопления знаний позволяет обеспечить доступ к материалам в локальной сети и через Internet эти материалы могут быть доступны научной общественности. После того как материал подготовлен и прошел стадию внутреннего модерирования, он может быть отправлен в центральное хранилище, являющейся частью информационной среды головной организации или ведомства. Данная стадия фактически соответствует регистрации отчета по НИР.



Рис. 1. Накопление знаний научного назначения

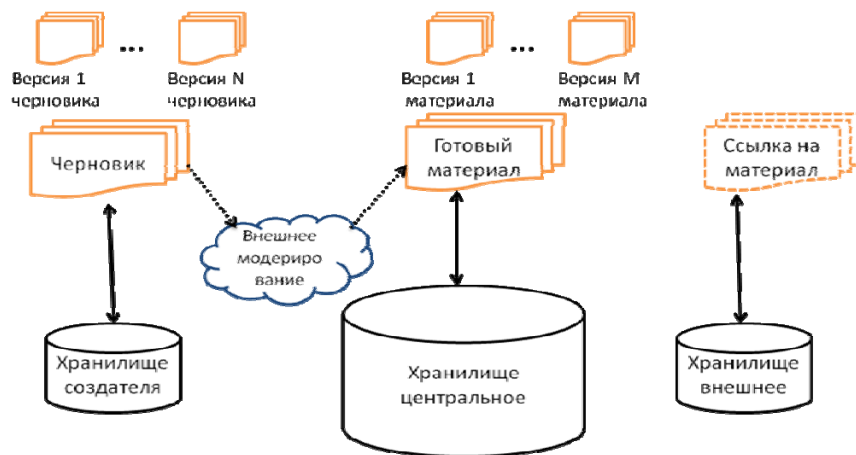


Рис. 2. Хранение и модификация научного материала

Предлагаемый подход помимо репликации данных предполагает изменение подхода к работе с научным материалом. Предлагается уже на начальных стадиях исследования делать новую информацию доступной в электронном виде как минимум в рамках организации-исследователя. Ключевым отличием подхода является ориентация на интегрируемость, доступность и целостность цикла ведения НИР на всех стадиях.

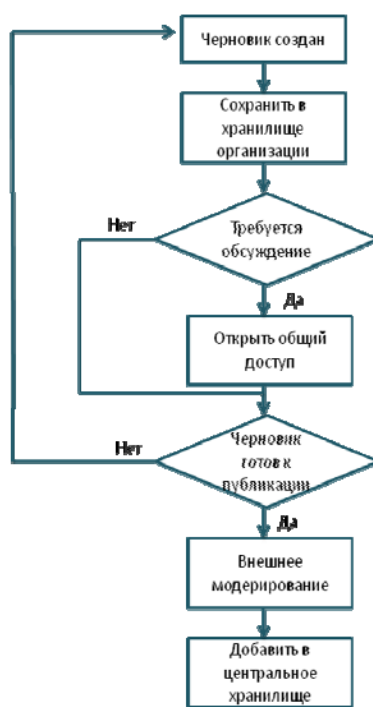


Рис. 3. Алгоритм внесения и модификации научного материала

На этапе, когда новые знания находятся в достаточной степени готовности для публикации в научной печати, они попадают в центральное хранилище. Именно в этот момент наряду с внешним модерированием возможно производить проверку на выполнение формальных требований и регистрацию новых созданных материалов в системе регистрации и учета. Центральное хранилище также представляет собой систему хранения знаний. Оно содержит только финальные и проверенные версии материала. При необходимости внешнего по отношению к организации-автору или центральному хранилищу размещения материала предлагается применение ссылок. Алгоритм внесения и модификации научного материала приведен на Рис. 3.

Так как исследования могут затрагивать разные области научной деятельности система должна обеспечивать релевантный поиск по материалам различных научных направлений и давать возможность простой работы, доступной неспециалистам в области информационных технологий. Первое может быть обеспечено за счет реализации на основе выбранной и обоснованной формальной модели представления знаний. Второе обеспечивается реализацией удобной среды редактирования.

Отсутствие этих качеств затруднит внедрение системы, в силу естественных сложностей при решении типовых для конечного пользователя задач. Именно в связи с этим выше в статье приведенные качества выделены как одни из самых важных.

Резюмируя вышесказанное, определим следующие основные принципы предлагаемого подхода:

1. доступность систем накопления знаний в сети Интернет;
2. возможность ограничения доступа к знаниям в случае необходимости или специальной направленности знаний;
3. внутреннее представление знаний в рамках выбранной формальной и обоснованной модели;

4. реализация удобного встроенного редактора, поддерживающего работу в различных научных областях;
5. хранение черновых версий материала в системе в удобном для создателя и данной предметной области формате;
6. хранение финальных версий материала в центральном хранилище с предварительным внешним модерированием;
7. наличие ссылок на хранилище организации-автора исследования при необходимости внешнего размещения материала для предотвращения дублирования материала.

Создание системы, интегрирующей в себе преимущества электронных каталогов и wiki-подобных систем, позволит сократить время подготовки электронной документации научного характера, обеспечив также доступ к научной информации на стадии разработки и исследования.

Предлагаемый механизм предполагает возможность двойного модерирования, что повышает уровень качества содержимого в сравнении с традиционными материалами, размещенными в открытых электронных энциклопедиях.

ЛИТЕРАТУРА

1. Борисов А.В., Куриленко И.Е. Информационная среда накопления и модификации знаний // Сб. док. IX международной конференции «Информатика: проблемы, методология, технологии» в 2 т. – Т.1 – 2009. – С. 430–433.