

ELAN — Электронная Антология физики частиц

Бунаков В.Э., Ежела В.В., Киселев В.В.,
Кожина И.В., Лазин Ю.А., Лесенко С.А.,
Луговский В.С., Луговский С.Б., Сухих Н.И.,
Ткаченко Н.П., Тюрин Н.Е., Филимонов Б.Б.
ГНЦ Институт физики высоких энергий,
г.Протвино Московской обл., Россия
E-mail: bunakov@mx.ihep.su, ezhela@mx.ihep.su

1 Введение

Становление современной научной картины мира неразрывно связано с историей физики элементарных частиц и теоретической физики. Эти отрасли знания не только составляют существенную часть научного багажа человечества, но оказали и продолжают оказывать значительное влияние на способы добывания новых научных знаний и на положение науки в обществе.

Современные технологии породили новые приемы научной коммуникации и влияют на политику крупных издателей. Электронизация публикаций в физике частиц продвинулась очень далеко, для примера укажем на Лос-Аламосские архивы препринтов ([3]) и их связь с ведущими реферируемыми журналами, которые издает Американское физическое общество ([4]).

Мы в ИФВЭ на протяжении ряда лет занимаемся технологиями баз данных и научных публикаций на Web. В качестве успешных проектов такого рода можно назвать Particle Physics Data System ([5]), на основе которой регулярно издается известный справочник [2], и электронный архив публикаций ИФВЭ ([6]). Эти работы частично поддерживаются нашим работодателем ([11]) и Российским фондом фундаментальных исследований ([12]). Основываясь на накопленном опыте, мы предлагаем проект создания электронной антологии важнейших работ по физике частиц, начиная с открытий рентгеновских лучей и электрона в конце прошлого века и вплоть до настоящего времени.

2 Принципы отбора работ. Первые результаты.

Отбор работ для подобной антологии может основываться только на экспертных оценках специалистов в данной предметной области: в нашем случае — на оценках физиков. Мы имеем успешный опыт совместной работы специалистов из разных стран, который привел к изданию Хронологии физики частиц — аннотированной библиографии классических работ, начиная с открытий Рентгена и Перрена в 1895 году. Хронология существует в виде книги [1] и в виде базы данных [7], доступной в режи-

ме тестирования. Хронология включает в себя около 600 расширенных библиографических описаний экспериментальных работ и теоретических работ, получивших экспериментальное подтверждение.

В настоящее время мы работаем над созданием архива полных текстов, с тем чтобы связать его с базой данных Хронологии. Для отработки технологий мы взяли несколько классических работ Бора, Резерфорда, Эйнштейна и Мандельштама, которые уже оцифрованы и размещены на Web. Мы предполагаем поддерживать двуязычный (русско-английский) интерфейс ко всем публикуемым работам, независимо от языка первоисточника, а также воспроизводить параллельные тексты самих первоисточников на русском и английском языках.

Отправной точкой для знакомства с проектом Антологии может служить посвященная ему Web-страница [8].

3 Технологии. Информационная структура проекта.

Текст первоисточника (это публикации в журналах или сборниках трудов) переводится в машинную форму с использованием программ оптического распознавания символов. Полученный ASCII-текст пропускается через программы проверки орфографии. Мы сохраняем также "сырые" сканированные версии оригинальных работ.

Многие тексты доступны и на английском, и на русском языках. Недостающие придется переводить. Наш проект открыт для сотрудничества с командами, желающими поддерживать параллельные иноязычные коллекции текстов, входящих в Антологию (в дополнение к русским и английским текстам).

На всех этапах необходимо вычитывать текст и вносить поправки. Необходимо не только техническое редактирование (сверять сверстанный текст с первоисточником), но и научное редактирование. Практика показывает, что некоторые классические работы содержат опечатки, в том числе содержательные — в формулах и т.д. Вероятно, большинство таких ошибок исправлялось и соответствующие публикации появлялись позже в научной периодике, но для работ 50- и 80-летней давности найти все эти исправления крайне затруднительно. Поэтому необходимы квалифицированные редакторы-физики, проверяющие содержательную сторону публикаций.

Формулы и прочая диакритика переводятся в формат TeX, рисунки сохраняются и обрабатываются отдельно. TeX-файл через цепочку DVI — Postscript переводится в формат PDF (Portable Document Format). Конеч-

Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ,
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург

ным продуктом являются четыре файла: два в формате TeX (русский и английский) и два сгенерированных из них — в формате PDF. Для просмотра текстов на Web можно использовать соответственно TeX Explorer ([13]) и Acrobat Reader ([14]). Готовые файлы переносятся на сервер и связываются с базой данных Хронологии, для которой, в свою очередь, подготавливаются параллельные русское и английское расширенные библиографические описания. Расширенное библиографическое описание включает в себя имя автора, заглавие, ссылки на первоисточники публикации, аннотацию, приставительный список литературы, обозначения рассмотренных в работе ядерных реакций и элементарных частиц, дополнительные комментарии. Для лауреатов Нобелевской премии приводится фотопортрет и формула (за что присуждена премия).

Процесс обработки текстов для Антологии отражен на рисунке 1.

В обеих версиях, TeX и PDF, работают гиперссылки внутри текста отдельной статьи — на формулы, рисунки, библиографию. Мы надеемся реализовать гиперссылки между разными документами нашей коллекции, а также связать нашу коллекцию, через механизм цитирования, с внешними базами данных библиографии и другими электронными коллекциями. Наибольший интерес в этом отношении представляют база данных HEP-SPIRES ([9]), поддерживаемая библиотекой Стэнфордского линейного ускорителя, база данных Science Citation Index, поддерживаемая Институтом научной информации в Филадельфии ([10]), электронные версии журналов Американского физического общества ([4]) и архивы препринтов в Лос-Аламосе ([3]).

Информационная структура нашего проекта отражена на рисунке 2.

4 Заключение

Общедоступная электронная Антология физики частиц стала бы серьезным подспорьем для историков науки и системы высшего образования. Антология представляла бы собой своеобразный "зал славы" нескольких поколений физиков и имела бы общекультурное значение. Кроме того, поскольку современные научные исследования требуют значительных ресурсов, Антология решала бы важную социальную задачу, информируя общество о реальных достижениях и способствуя поддержанию статуса фундаментальной науки.

Побочными, но интересными результатами создания Антологии были бы отработка современных технологий на обширном массиве "сложных" публикаций с большим количеством специальных символов и графики, а также приобретение опыта управления подобными ретроспективными проектами.

В настоящее время мы заняты обкаткой технологий на массиве реальных первоисточников и поиском спонсоров. В качестве одного из вариантов мы рассматриваем возможность параллельного издания Антологии в Интернет и в виде печатного многотомника с приложениями на компакт-дисках.

Библиография

[1] V.V.Ezhela et al. *Particle physics: one hundred years of discoveries: an annotated chronological bibliography*. — AIP Press, 1996.

[2] V.V.Ezhela et al. *A Guide to Experimental Particle Physics Literature*. — Lawrence Berkeley National Laboratory. LBL-90 revised.

[3] Лос-Аламосские архивы электронных препринтов. [<http://xxx.lanl.gov>]

[4] Журналы Американского физического общества. [<http://publish.aps.org/>]

[5] Particle Physics Data System — индексированная библиография экспериментальных работ по физике частиц и числовые данные, извлеченные из публикаций. [<http://wwwppds.ihep.su:8001/ppds.html>]

[6] Архив публикаций ИФВЭ — пример коллекции научных публикаций на Web с двуязычным интерфейсом. [<http://dbserv.ihep.su/~pubs>]

[7] Электронная версия Хронологии физики частиц. [<http://dbserv.ihep.su/compas/index.html>]

[8] Домашняя страница проекта Антологии. [<http://dbserv.ihep.su/~elan>]

[9] База данных HEP-SPIRES. [<http://www-spires.slac.stanford.edu/find/hep>]

[10] Институт научной информации. [<http://www.isinet.com/>]

[11] Государственный научный центр Институт физики высоких энергий. [<http://www.ihep.su/>]

[12] Российский фонд фундаментальных исследований. [<http://www.rfbr.ru/>]

[13] TeX Explorer — программа (plug-in) для просмотра TeX-файлов в браузере Интернет. [<http://www.software.ibm.com/network/techexplorer/>]

[14] Acrobat Reader — приложение и plug-in для работы с PDF-файлами. [<http://www.adobe.com/prodindex/acrobat/readstep.html>]

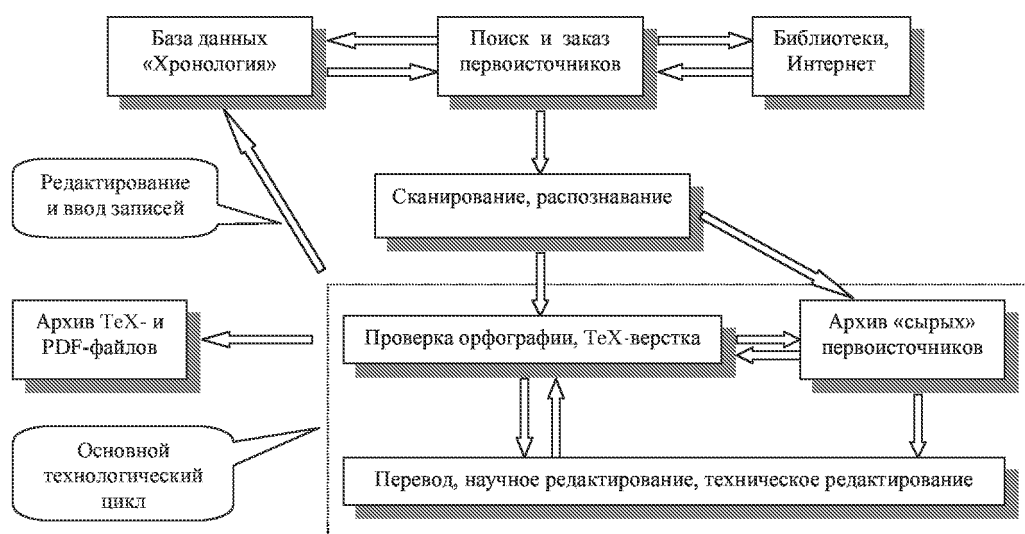


Рис. 1: Процесс обработки текстов

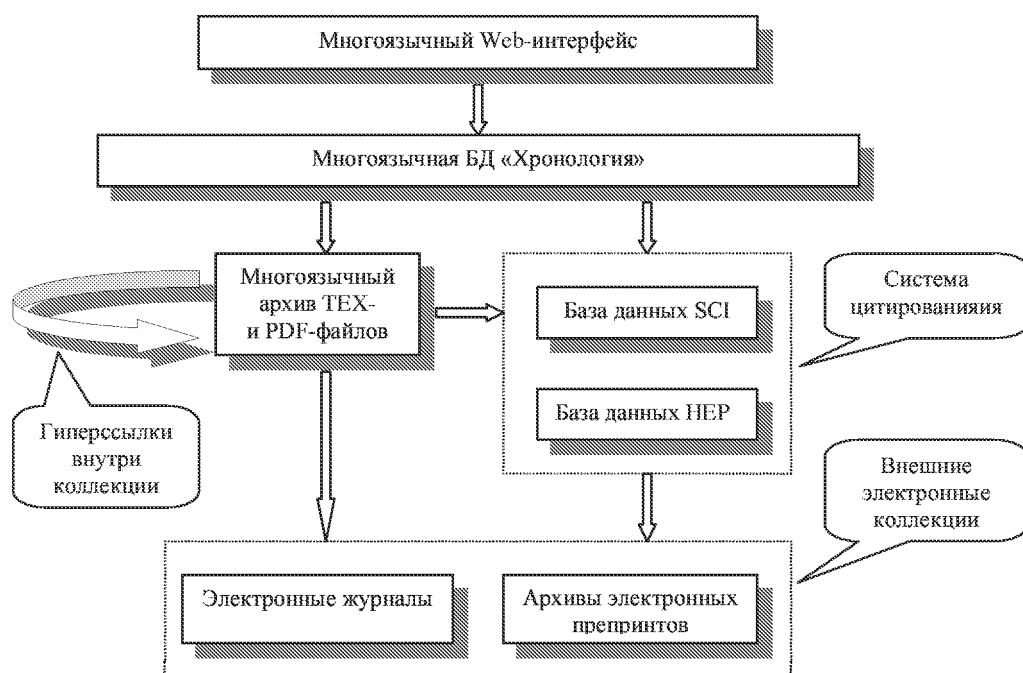


Рис. 2: Информационная структура проекта