

Сон Хо
Отделение паразитологии, Институт медицинского образования, Медицинский колледж Университета Халлим, Чунчон, Корея
shuh@hallym.ac.kr

Тхэ Джин Цой
Корейский национальный исследовательский фонд, Тэджон, Корея

Со Хён Ким
Корейский национальный исследовательский фонд, Тэджон, Корея

ПРИМЕНЕНИЕ XML-РАЗМЕТКИ для НАУЧНЫХ СТАТЕЙ, ОПУБЛИКОВАННЫХ В ЖУРНАЛАХ на КОРЕЙСКОМ ЯЗЫКЕ

Республика Корея занимает четвертое место по числу включенных в базу PubMed Central (PMC) журналов. По состоянию на сентябрь 2013 года в PMC включены 75 корейских журналов. С 2013 года различные агентства, финансирующие исследования, результаты которых должны быть опубликованы в корейской научной периодике, начали создавать полнотекстовые базы данных открытого доступа по медицине, естественным, социальным и гуманитарным наукам. В этих базах используется набор разметки журнальных статей (JournalArticleTagSuite (JATS) 1.0), так что статьи на корейском языке легко могут обрабатываться как полнотекстовые на языке расширяемой разметки (XML). Редакторы и издатели должны выпускать полнотекстовые XML-файлы на основе JATS 1.0.

Таким образом, статья освещает вопросы текущего применения JATS 1.0 в корейских научных журналах как на английском, так и на корейском языках, с упором на современные технологии, программы обучения и политику корейского правительства относительно общедоступного полнотекстового XML.

Опыт Кореи может стать моделью для создания в других странах полнотекстовых баз данных открытого доступа, основанных на JATS 1.0, на национальном языке. Подчеркивается пригодность JATS для научных журнальных публикаций не только во всех областях, но и на всех языках.

ВВЕДЕНИЕ

База PubMed Central (PMC) является основной возможностью для корейских медицинских журналов получить мировое распространение, так как это наиболее доступный путь к индексации в PubMed, важной в медико-биологических исследованиях базы данных. Другой способ – включение в базу Med Line, которая, в отличие от PubMed, принимает журналы на разных языках, однако за последние 10 лет в нее были включены очень немногие корейские медицинские журналы. Если корейские журналы отвергаются базой Med Line, единственной возможностью достичь международной аудитории для них остается изменение языка журнала на английский и использование PMC.

Включение англоязычных медицинских журналов в PMC не представляет трудностей, если качество расширяемого языка разметки XML соответствует критериям этой базы.

Рисунок отражает количество корейских журналов, включенных в PMC за период с 2008 по сентябрь 2013 г.

Повышенная посредством PubMed и PMC видимость дала многим медицинским журналам из Кореи возможность увеличить международную аудиторию, повысить количество и качество материалов и импакт-фактор. Этот выход за пределы государства включает, во-первых, индексацию в таких международных ссылочных базах, как Web of Science или Scopus; во-вторых, увеличившееся поступление материалов из-за пределов Кореи; и в-третьих, ежегодное улучшение журнальных показателей.

Когда в 2006 году в Корее был проведен первый тренинг по созданию XML для PMC, его организаторы не могли представить себе, что медицинские журналы действительно смогут достичь этих показателей, но во многих из них всего 7 лет спустя развитие движется полным ходом.

Корейская ассоциация редакторов медицинских журналов (The Korean Association of Medical Journal Editors, KAMJE) играет важную роль в производстве XML-файлов для PMC. Она создала в 2007 году полнотекстовую базу данных на основе PMCXML под названием Korea Med Synapse (http://synapse.koreamed.org/). Научные редакторы журналов в других областях стали обращать внимание на деятельность KAMJE. Деятельность KAMJE сподвигла ученых в Корее организовать Корейский совет научных редакторов (The Korean Council of Science Editors, KCSE), который был основан в сентябре 2011 года. Основная работа KCSE включает подготовку и обучение научных редакторов работе с текстом, публикационной этике и изданию вестников и журналов на международном уровне. KCSE также обучает редакторов и сотрудников полиграфических компаний адаптации информационных технологий к изданию журналов.

национальный исследовательский фонд Кореи финансирует основанную на XML полнотекстовую базу данных открытого доступа по общественным и гуманитарным наукам

Помимо естественных наук, Национальный исследовательский фонд Кореи финансирует основанную на XML полнотекстовую базу данных открытого доступа по общественным и гуманитарным наукам. Другие новые базы данных включают Korea PMC, создаваемую в этом году по образцу Canada PMC и Europe PMC Национальной медицинской библиотекой Кореи и агентством Корейского центра по контролю и профилактике заболеваний. Корейская федерация научных и технических обществ (The Korean Federation of Science and Technology Societies, KOFST) также тестирует полнотекстовую, основанную на XML базу данных открытого доступа.

Остальная часть этой статьи более подробно освещает не только набор меток журнальных статей (JATS) полнотекстовых, основанных на XML баз данныхв Корее, но и технологию, программы обучения и политику корейского правительства относительно полнотекстовых баз данных открытого доступа на основе XML.

KOREA PMC

Создание Korea PMC впервые было предложено Корейской ассоциацией редакторов медицинских журналов. В марте 2007 года KAMJE обсудила эту идею с Национальным центром по биологической информации США (The United States National Center for Biological Information, NCBI), который в конечном итоге стал подрядчиком договора на участие в Архиве NIH Pub Med Central (NIHPortfolio). Однако это соглашение было расторгнуто в июле 2007 года, когда NCBI изменила политику относительно клонирования PMC другими странами. Эти клоны, известные как международные PMC (PMCi),

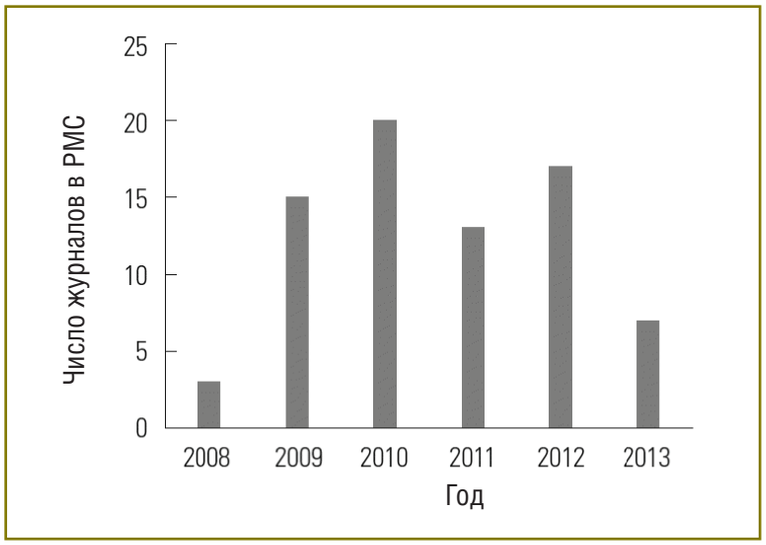


Рис. Число журналов из Кореи, принимаемых ежегодно к индексации в Pub Med Central (PMC)

должны были управляться следующим образом: во-первых, международная PMC должна поддерживаться правительствами каждой страны-участницы; во-вторых, должна соблюдаться политика открытого доступа к медико-биологическим журналам, финансируемым правительством. Так как KAMJE не поддерживается правительством и корейское правительство в то время не имело политики общественного доступа, приступить к реализации проекта оказалось невозможно. Между тем в 2011 году правительство Кореи приняло решение о создании новой организации – Национальной библиотеки медицины Кореи, подчиненной Корейскому центру по контролю и профилактике заболеваний. Строительство здания началось в 2012 году и будет завершено в 2014 году. Среди новых баз данных под управлением Национальной библиотеки медицины Кореи в стадии строительства находится также Korea PMC. Благодаря технической поддержке PMC Америки это уже устойчивая система, находящаяся в стадии тестирования. Korea PMC будет индексировать статьи, написанные на языках с нелатинскими алфавитами, например на корейском. Этот проект полностью поддерживается правительством Кореи. Korea PMC также будет принята американским Национальным центром по биологической информации в качестве третьей международной PMC, так как в 2012 году министр здравоохранения и социального обеспечения объявил политику открытого доступа для публикации исследований, финансируемых министерством. Korea PMC может появиться в начале 2014 года.

> ***«уровень финансовой и материально-технической поддержки научных журналов в Корее и Японии уникален***

ЦЕНТРАЛЬНАЯ БАЗА ПО ОБЩЕСТВЕННЫМ И ГУМАНИТАРНЫМ НАУКАМ (SSHC)

Korea Citation Index (KCI) – это база данных, схожая с Web of Science или Scopus и включающая более чем 2000 корейских научных журналов, в том числе в области естественных наук, технологии, машиностроения, медицины, социальных и гуманитарных наук и искусств. Национальный исследовательский фонд Кореи также заказал создание центральной базы по общественным и гуманитарным наукам (SSHC) для корейских научных журналов. Для нее будут выбраны англоязычные журналы в области социальных и гуманитарных наук, и полнотекстовая база данных открытого доступа на основе JATS XML с числом статей около 100, весьма схожая с PMC, станет доступна в 2014 году. В мире не существует других международных полнотекстовых баз данных открытого доступа в области социальных и гуманитарных наук. SSHC будет первым шагом в этом направлении. Она будет имитировать PMC как наилучшую модель полнотекстовых, бесплатных или общедоступных архивов литературы; однако она будет принимать журналы на любом национальном или региональном языке мира, то есть на любом языке, кроме английского. Это предприятие стало возможным на основе декларации Министерства образования Кореи, в которой в декабре 2011 года была подчеркнута политика открытого доступа для финансовой поддержки научных публикаций. Центральная база по общественным и гуманитарным наукам может существенно увеличить видимость журналов из Кореи.

Из 34 стран – членов ОЭСР (Организация экономического сотрудничества и развития, OECD) Корея и Япония уникальны по уровню финансовой и материально-технической поддержки, оказываемой для публикации научных журналов. Корейские и японские власти в 2013 году выделили 8 млн и 5 млн долларов США соответственно. Корейская

федерация научных и технических обществ и Национальный исследовательский фонд Кореи, оба финансируемые правительством, желают конкурентоспособности для научных журналов целиком, а не только для отдельных публикаций, созданных гражданами Кореи. Кроме того, эти две организации желают повысить образованность общественности предоставлением бесплатного доступа к научному знанию. Именно по этой причине правительство Кореи поддерживает публикацию научных журналов и полнотекстовые базы данных.

SCIENCE CENTRAL

Science Central (http://e-sciencecentral.org/) представляет собой полнотекстовый архив научной литературы в открытом доступе Корейской федерации научных и технических обществ (КФНТО). Он находится в стадии тестирования, по состоянию на сентябрь 2013 года в него включены восемь журналов из Кореи и один из Хорватии. Из восьми журналов два издаются на корейском языке. Корейское правительство поддерживает научные, технологические и медицинские журналы через КФНТО и журналы по общественным и гуманитарным наукам и искусству – через Национальный исследовательский фонд Кореи. Хотя Федерация является объединением независимых научных обществ, работа обществ финансируется правительством Кореи. КФНТО поддерживает публикации журналов через ряд инициатив как для местных журналов, не проиндексированных в международных базах данных, так и международных журналов, как правило, англоязычных. Для ограниченного количества журналов доступна специализированная поддержка, включающая консультации или финансирование для создания импринта и инструкций для авторов; обучение по вопросам публикационной этики; развитие стиля и формата; дизайн обложки; верстку; разработку политики журналов; редактирование рукописей; корректуру англоязычных текстов; системы обработки рукописей; веб-сайт журнала; создание XML для JATS и Cross Ref; создание идентификатора ORCID, Cross Mark и Fund Ref; а также регулярное обучение редакторов или членов редакционной коллегии.

Накопив обширный опыт поддержки журналов, КФНТО поняла, что для повышения видимости научных журналов из Кореи необходима полнотекстовая платформа открытого доступа по типу PMC; поэтому было принято решение о создании Science Central – клона PMC для всех областей науки. Три существенных отличия от PMC: во-первых, охват всех областей науки, включая технологию, машиностроение, науки о жизни и многопрофильные дисциплины; во-вторых, включение журналов на всех языках мира, если аннотация, таблицы, иллюстрации и ссылки продублированы также на английском языке; в-третьих, принятие только журналов обществ или некоммерческих ведомственных журналов. Это новый эксперимент среди полнотекстовых баз данных открытого доступа. До сих пор неясно, сколько именно общедоступных, некоммерческих, ведомственных журналов или журналов обществ примут участие в Science Central. Поскольку стоимость производства JATS XML в наше время весьма умеренна, нагрузка на издателей будет разумной. Корейская федерация научных и технических обществ в состоянии полностью финансировать производство JATS XML для научных журналов из Кореи. Другая цель Science Central состоит в том, чтобы помочь местным или региональным языкам выжить в научной литературе.

Известно, что языком науки является английский; однако, если полнотекстовые версии на других языках будут представлены в формате JATS XML, читатели с другими родными языками смогут получить доступ к содержимому по мере развития технологий машинного перевода и, таким образом, местный язык как авторов, так и читателей сможет выжить в качестве инструмента распространения научной информации для человечества.

ОТОБРАЖЕНИЕ АЛФАВИТОВ КОРЕЙСКОГО И ДРУГИХ ЯЗЫКОВ ПОСРЕДСТВОМ JATS XML

Корейские авторы обычно пишут статьи на английском языке для иностранных читателей; однако при написании статей, рассчитанных на местных читателей, они пишут их на корейском языке и представляют в корейские журналы. В 2012 году JATS стал стандартом для XML научных журналов, аккредитованным Национальной организацией по стандартам информации (NISO). Основным преимуществом определений типа документа (DTD) в JATS является то, что они охватывают любой язык мира, который может быть закодирован в универсальном формате преобразования (UTF-8), стандарте кодирования ISO, позволяющем представлять каждый символ всех языков мира, так же как и другие специ-

Ключевые слова:
базы данных, национальный язык, наука и медицина, журнал открытого доступа, open access, XML

Keywords:
database, journal article tag suite, local language, science and medicine, open access, journal

альные символы. Архивирование информации на корейском языке благодаря JATS DTD больше не представляет серьезных технических или экономических проблем. Внедрять JATS в публикации на корейском языке очень просто, так как в JATS DTD существует система кодирования языка. Различные языки помимо корейского легко могут быть включены в нее путем добавления атрибута языка «xml: lang=Language". Например, нижеследующее описание, добавленное в XML статьи, может отображать корейские символы в этой статье.

```
<article
article-type="research-article"
dtd-version="1.0"
xml:lang="ko"
xmlns:mml="http://www.w3.org/1998/Math/MathML"
xmlns:xlink="http://www.w3.org/1999/xlink">
```

При замещении языков языковой код может быть изменен, например, xml:lang = "ja-Hani" (японский в написании кандзи [Hanzi, Ханджа, Хан]), "mn" (монгольский), "war" (Филиппины), "vi" (вьетнамский), "zh" (китайский), и "cz" (чешский). Код языка может быть вставлен во многих частях XML для JATS, чтобы отразить несколько языков в едином тексте статьи. DOI на корейском языке также без затруднений отображаются при преобразовании JATS XML в Cross Ref XML.

ПРОГРАММЫ ОБУЧЕНИЯ XML ДЛЯ JATS

Первый семинар по созданию XML для PMC был организован в августе 2006 года автором, Сон Хо (бывшим председателем Комитета по управлению информацией, KAMJE). После этого сотрудники KAMJE могли самостоятельно создавать XML для PMC. Поскольку XML для JATS имеет гораздо больше тегов, чем XML для PMC, необходимо обучение редакторов, сотрудников академических обществ, издательских и информационно-технологических компаний. Корейское общество научных редакторов провело семинар на тему «Применение информационных технологий в издании научных общедоступных публикаций» с 4 по 5 июля 2013 года в Сеуле. Участие в семинаре приняли почти 100 человек.

> ***«местные журналы, издаваемые научными обществами, испытывают затруднения в привлечении качественных статей***

Три докладчика были связаны с университетами, двое – с информационно-технологическими компаниями, двое – с компаниями редактирования рукописей и один – с университетской библиотекой больницы. Темы заседаний и название каждого модуля были следующими:

- **Заседание 1: Журнальный XML**
  1. Journal Article Tag Set (JATS) 1.0: Новый стандарт ISO для журнального XML.
  2. Создание полнотекстового XML для журнальных статей на английском и корейском языках на базе JATS 1.0.
  3. Кодирование XML в JATS 1.0.
  4. Построение XSL и/или CSS для таблицы стилей XML-файлов на основе определений типа документов (DTD).
  5. Отображение в интернете HTML-файлов из документов JATS XML.
- **Заседание 2: Cross Ref XML**
  6. Основы Cross Ref XML.
  7. Кодирование в Cross Ref XML.
  8. Преобразование JATS 1.0 XML в Cross Ref XML.
  9. Добавление гиперссылок и индекса цитирования с помощью DOI.
  10. Применение ORCID и Fund Ref с помощью Cross Ref XML.
  11. Применение Cross Mark.
- **Заседание 3: Сетевые инструменты и инструменты просмотра**
  12. Использование сервисов социальных сетей и QR-кодов.
  13. Pub Reader: новый читательский инструмент, удобный для использования в интернете, на планшетах и смартфонах.
  14. Создание ориентированных на мобильные устройства сайтов журналов.
  15. Связанный экран (N-Screen) для доступа к научным журналам на многих устройствах из единого источника.

Целью семинара было повышение международной конкурентоспособности каждого журнала. Разумеется, важнейшим фактором для достижения этой цели является публикация статей высокого качества; однако большинство высококачественных статей в итоге попадают в журналы с высоким импакт-фактором. Вследствие этого местные журналы, издаваемые научными обществами, испытывают затруднения в привлечении качественных статей. Даже при не самом научном содержании видимость таких журналов может быть улучшена, если они максимизируют другие факторы, например, качество веб-сайта и доступность через международные общедоступные базы данных. В стремлении к повышению видимости основ-

ным фактором реализации является XML для JATS. Из XML для JATS документы могут быть легко преобразованы в другие форматы XML и отформатированы для просмотра. Лучший пример – Pub Reader.

Для построения полнотекстового, общедоступного веб-сайта журнала или базы данных редакторам очень важно понимание XML для JATS. Если журнал издается международной коммерческой компанией, редакторам нет необходимости изучать его. Однако большинство научных журналов в Корее публикуются обществами, и их редакторы должны быть в курсе новых информационных технологий, применимых к их журналам. Если редакторы понимают, как создать XML для JATS и как конвертировать его в другие XML или просмотрщики, они обретут уверенность, необходимую для применения этих технологий в своих журналах. Это реалистичное для Кореи предположение, так как редакторы журналов, как и корейцы в целом, обладают относительно продвинутым пониманием информационных технологий.

В продолжение учебных мероприятий стало очевидно, что корейские редакторы научных журналов могут без труда понять принципы и перейти к практике XML-кодирования и преобразования на своих собственных ноутбуках. Участники подтвердили, что XML для Cross Ref и цитирования с помощью DOI, ORCID, Cross Mark и Fund Ref XML легко создаются из XML для JATS. Они узнали о принципах работы QR-кодов и о преимуществах журнальных приложений. Pub Reader был популярной у редакторов темой, потому что это один из инструментов, используемых для связанного экрана (N-экрана). Связанный экран (N-экран) – это клиентоориентированное приложение, в котором пользователь является совладельцем нескольких приложений и имеет к ним доступ в любое время и в любом месте с помощью новейшей смарт-системы. Связи между такими устройствами, как смарт-телевизоры, смартфоны, планшеты, ноутбуки и аудиоустройства, легко реализуются с помощью различных интеллектуальных устройств и облачных служб. Когда редакторы поймут эти новые стандарты и методы XML, они смогут более эффективно управлять инженерами, которые будут внедрять эти службы.

ПОЛИТИКА ОТКРЫТОГО ДОСТУПА В КОРЕЕ

Korea PMC базируется не на требовании того, чтобы издаваемые в Корее журналы

> ***редакторы должны быть в курсе новых информационных технологий, применимых к их журналам***

были общедоступными. Скорее, результаты исследований, финансируемые правительством Кореи, должны публиковаться в изданиях открытого доступа. Центральная база по общественным и гуманитарным наукам все еще является предварительным проектом, поэтому журналы по общественным и гуманитарным наукам в Корее не обязаны предоставлять полный текст на основе XML для JATS в Национальный научный фонд Кореи.

В отношении Science Central Корейская федерация научных и технических обществ применяет более авторитарную политику в пользу финансирования журналов открытого доступа. Так как Корейская федерация научных и технических обществ предоставляет оборотные фонды для научных журналов, она может рассмотреть использование XML для JATS среди прочих критериев оценки журналов, конкурирующих за ограниченные средства. Соответственно ожидается, что большинство редакторов научных журналов сделает искреннюю попытку выполнить все критерии, рекомендованные KOFST, для получения финансирования. Это может быть первый в мире случай политики журналов открытого доступа, установленной правительством страны. Однако до укоренения этой политики размещение файлов XML для JATS в Science Central останется добровольным.

ЗАКЛЮЧЕНИЕ

В статье освещенотекущее состояние и будущее направление полнотекстовых баз данных на основе XML для JATS в Корее; а также технологии, учебные программы и политика корейского правительства относительно полнотекстовых баз данных открытого доступа на основе XML. На основании представленной информации читателю должно стать ясно, почему в Корее сейчас создаются подобные базы данных. Описанная здесь программа обучения редакторов может быть полезным образцом для других стран или других областей, желающих поощрить доступность локальных исследований. Полезно также будет продолжить наблюдение за работой трех новых баз данных журнальных статей в Корее – Korea PMC, SSHC и Science Central в будущем.

Конфликт интересов

Сообщений о возможных конфликтах интересов, относящихся к содержанию статьи, не поступало.

Благодарности

Работа выполнена при поддержке исследовательского гранта Национального исследовательского фонда Кореи (политические исследования-2013-003 – продвижение академической инфраструктуры) и Корейской федерации научных и технических обществ (KOFST-2013) правительства Республики Корея.

Источники

1. The Korean Association of Medical Journal Editors. Korea Med technical workshop, 2006 [Internet]; Seoul: The Korean Association of Medical Journal Editors; 2006 [cited 2013 Oct 25]. Available from: http://kamje.or.kr/workshop/2006/0811/1.pdf.

2. *Huh S.* Science Central: open access full-text archive of scientific journals based on Journal Article Tag Suite regardless of their languages. Biochem Med 2013;23:235–6.http://dx.doi.org/10.11613/BM.2013.029.

Sun Huh
Department of Parasitology and Institute of Medical Education, College of Medicine, Hallym University, Chuncheon, Korea

Tae Jin Choi
National Research Foundation of Korea, Daejeon, Korea

So-Hyeong Kim
National Research Foundation of Korea, Daejeon, Korea

Using Journal Article Tag Suite Extensible Markup Language for Scholarly Journal Articles Written in Korean

The Republic of Korea is the fourth ranking country for the number of PubMed Central (PMC) journals. As of September 2013, 75 journals from Korea are included in PMC. Starting in 2013, several research funding agencies for scholarly journal publications in Korea began to establish open access, full-text databases in the fields of medicine, science, and social sciences and humanities. In those databases, Journal Article Tag Suite (JATS) 1.0 is used so that articles written in the Korean language can be easily manipulated as full-text extensible markup language (XML). Editors or publishers must produce full-text XML files based on JATS 1.0. Thus this paper surveys the current state of the application of JATS 1.0 to Korean scholarly journals, including both those in English and those in Korean, focusing on the current technology, training programs, and the policy of the Korean government on open access, full-text XML. This experience in Korea can be a model for constructing mother-tongue, open access, full-text journal databases based on JATS 1.0 in other countries. The usefulness of JATS in scholarly journal publications not only of all fields, but also in all languages, is stressed.

Оригинал публикации:
http://www.escienceediting.org/journal/view.php?number=6