ЧЕШСКАЯ РЕАЛИЗАЦИЯ URN: NBN СИСТЕМЫ





Зденек Васек,
руководитель отдела стандартов сохранности цифровой информации Национальной библиотеки Чешской Республики

Чешский URN: NBN идентификатор системы впервые был использован в большом проекте по оцифровке «Строительство национального проекта цифровой библиотеки» ЕС ВГД (первая часть была начата в 2012 – 2014 гг.)

Постоянные идентификаторы использовались в течение длительного времени в библиотечном сообществе. Наиболее известными из них являются ISBN и ISSN. Роль постоянных идентификаторов расширилась в цифровой вселенной. Идентификаторы цифровых документов предоставляют пользователям большое количество функций, которых никогда не было в традиционных идентификаторах. Появилась новая ключевая функция для идентификаторов путем посредничества и постоянного доступа к интернет-ресурсу. Главный недостаток использования новых адресов URL находится в его нестабильности.

С середины 1990-х годов, были разработаны несколько систем идентификаторов, чья техническая инфраструктура не строится на системе DNS (например, Handle System или DOI System). Инфраструктура этих систем строится на двух основных компонентах: реестр и распознаватель. Реестр представляет собой хранилище записей, назначенных для постоянных идентификаторов и связанных с адресами URL. Распознаватель обеспечивает постоянные идентификаторы разрешения, особенно перенаправление фактических адресов URL, где выявленные документы имеются в наличии.

Национальная библиотека Чешской Республики (Národní knihovna České Republiky) решила использовать URN: NBN систему. Автором этой системы является Юха Хакала (Национальная библиотека Финляндии) и система описывается как RFC 1388 Стандарт (Хакала, 2001). Стандарт описывает несколько правил, чтобы правильно использовать URN: NBNS (прежде всего модель синтаксиса). Национальные библиотеки уполномочены осуществлять URN: NBN систему только в своих странах.

При этом должна быть четко определена модель идентификации, в которой ресурсам присваивают идентификатор и какой уровень детализации можно будет представить.

Обзор

Чешская реализация системы идентификатора с помощью URN: NBNS системы называется ČIDLO (Чески systém pro IDentifikaci LOkalizaci dokumentů digitálního kulturního dědictví; Чешская система идентификации и локализации для документов цифрового культурного наследия).

Идентификатор системы ČIDLO был разработан как часть исследовательского проекта Национальной библиотеки Чешской Республики, основанной Министерством культуры. Результатом проекта является концептуализация системы идентификаторов национальной URN: NBN политики и программное решение технической инфраструктуры системы идентификаторов.

Чешский URN: NBN идентификатор системы впервые был использован в большом проекте по оцифровке «Строительство национального проекта цифровой библиотеки» ЕС ВГД (первая часть была начата в 2012 – 2014). Проект также включал создание системы LTP, которая использует идентификаторы. Цель проекта состояла, прежде всего в том, что оцифровка печатных книг и периодических изданий должна быть сохранена в системе LTP. Основная идея проекта заключается в обеспечении доступности и сохранности оцифрованных документов чешского культурного наследия.

Наша система работает с идентификатором с осени 2012. С тех пор она была расширена, для обеспечения больших функциональных возможностей с административной точки зрения.

Участники

Регистрирующим органом по присвоению

URN: NBNS в Чешской Республике является Национальная библиотека Чешской Республики. Национальная библиотека Чешской Республики предоставляет техническую и административную поддержку системы идентификатора для обеспечения её текущей надежности. Национальная библиотека также опубликовала подробную национальную политику URN: NBNS-системы, которая была утверждена в качестве национальной сертифицированной методологии Министерства культуры Чешской Республики в июне 2015 года.

Только государственные учреждения, а не отдельные лица могут быть регистратором системы идентификатора. URN: NBNS могут быть использованы только для идентификации документов, принадлежащих данному учреждению (то есть учреждения, обеспечивающие агрегацию контента просто не имеют права назначать URN: NBNS).

Участвующие учреждения могут участвовать в двух разных ролях – регистратора и архиватора. Регистратором должна быть либо библиотека, либо учреждение, которое производит цифровые документы и которые считаются культурным наследием (то есть книги, газеты, журналы, карты и подобного типа документы). В настоящее время мы одобрили около 100 регистраторов, в основном, это библиотеки.

Объект идентификации

Как уже упоминалось, документы, пригодные для идентификации должны соответствовать определенным критериям. Эти критерии включают в себя правила: документы должны быть дискретной единицей интеллектуального содержания, иметь значение национального наследия. Это ограничение связано с тем, что постоянные идентификаторы должны предоставлять пользователям долгосрочную выгоду и не должны использоваться для эфемерных работ.

Другие типы документов (диссертации и электронные публикации) запланированы на ближайшее время. Это потребует развития технической инфраструктуры.

Стандарт метаданных PREMIS в «НДК стандарт» – это национальный стандарт для метаданных и форматов данных, используемых при оцифровке в Национальной библиотеке Чешской республики и многих других чешских библиотек. NDK Стандарт включает в себя профиль метаданных и JPEG 2000 профиль для указания метаданных и изображения, используемые в проектах по оцифровке в библиотеках Чешской Республики. NDK Стандарт также используется в качестве стандарта для проектов по оцифровке, проводимых чешскими библиотеками, основанных Чешской национальной программой, которая является основополагающей для библиотек (VISK).

В НАСТОЯЩЕЕ ВРЕМЯ NBN ИДЕНТИФИКАТОРЫ ИСПОЛЬЗУЮТСЯ ДЛЯ ОЦИФРОВАННЫХ МАТЕРИАЛОВ:



Мы создаем два экземпляра: отсканированный документ в ходе оцифровки, т.е. архивную копию предназначенную для архивации в цифровом хранилище и пользовательскую копию, которая должна быть доступна для конечных пользователей в системе цифровой библиотеки. URN: NBN присваивает идентификатор как архивным, так и копиям пользователя. Наша модель идентификации сосредоточена на определении содержания информации (понятие, введенное в ISO 14721), а не на другое представление данных одного и того же содержания.

Модель Синтаксис

Элементарные правила NBN синтаксиса определяются стандартом RFC 3188.

Есть две модели для URN: NBN идентификатора, описанные в стандарте.

Наша реализация стандарта RFC 3188 ис-

Стандарт метаданных PREMIS в «НДК стандарт» – это национальный стандарт для метаданных и форматов данных, используемых при оцифровке Национальной библиотекой Чешской республики и многих других чешских библиотек



пользует модель синтаксиса суб-пространства имен, которое можно сформулировать следующим образом:

Урна: NBN: CZ: < код регистратора > - <назначение NBN строки>

Код регистратора, как правило, «чешский код библиотеки» (используется в библиотечном сообществе, чтобы определить индивидуальную библиотеку), которая состоит из 6 буквенно-цифровых строк и NBN строка состоит из 6 алфавитно-цифровых символов. Например, «abg001» идентифицирует Городскую библиотеку Праги: NBN: CZ: abg001-000001. Другие учреждения имеют специальную серию суб-имен.

Регистраторы должны использовать URN: NBNS со строчными буквами (для различия между нулем и верхней буквы «O»).

URN: NBN политика в ЧР

Каждая библиотека имеет право стать регистратором, утверждение носит формальный характер. Запросы других учреждений оцениваются индивидуально регистрирующим органом. Критерии для утверждения небиблиотечных учреждений включают в себя тип документов, их культурную ценность и registrar's способность следовать NBN политике.

Каждая библиотека имеет право стать регистратором, утверждение носит формальный характер. Запросы других учреждений оцениваются индивидуально регистрирующим органом. Критерии для утверждения небиблиотечных учреждений включают в себя тип документов, их культурную ценность и registrar's способность следовать NBN политике

После утверждения регистратор получает код суб-пространства имен и учетной записи пользователя к системе идентификатора. URN: NBN генерируется системой идентификатора по требованию регистратора (через RESTful API).

Идентификатор должен храниться в метаданных длительный срок и записываться на любой последующей копии документа. Она также должна отображаться на цифровом сайте библиотеки, представляя данный документ, с тем чтобы дать возможность пользователям использовать идентификатор (для цитирования или связывания). Требуется, по крайней мере, одна копия документа доступной в Интернете для пользователей и зарегистрированной в OAI-PMH хранилище для того, чтобы сделать URN: NBN разрешимой. Также необходимо, чтобы по крайней мере одна копия документа находилась в архиве в долгосрочной системе хранения. Большинство регистраторов будет размещать свои данные в системе LTP НДК.

РЕГИСТРАТОР ОБЯЗАН:



Система идентификаторов собирает фактические адреса URL и сохраняет их в регистре идентификатора. Только URN: NBNS, будучи сгенерированные системой идентификаторов являются действительными.

Техническая инфраструктура

Техническая инфраструктура основана на системе программного обеспечения под названием CZIDLO. Главным разработчиком был Мартин Řehánek (в настоящее время внештатный программист). Чешский идентификатор системы состоит из следующих компонентов: распознаватель, регистр, генератор, бэк-офис и комбайн. Резольвер доступен на www.resolver.nkp.cz.

Есть несколько способов решения в нашей системе. Стандарт один для перенаправления на фактический адрес URL, где идентифицированный документ доступен. Решение также может отображать метаданные об идентифицированном документе или самих идентификаторов. Эти метаданные хранятся в системе идентификаторов. В случае отсутствия документа, пользователи могут просматривать записи в урне: NBN регистр. В этом случае они могут доказать, что документ, цитируемый ими, существовал в какой-то момент прошлого (таким образом, распознаватель помогает обеспечить достоверность цитат).

URN: NBN идентификатор не зависит от системы DNS домена. Это была основная идея лежащая в основе URN: NBN стандарта развития. Это делает URN: NBNS быть действительно постоянными идентификаторами. Тем не менее, есть также возможность использовать URN: NBN в URL, которую можно найти во многих национальных реализациях. Это делает решения более удобными для пользователей. Следует отметить, что такой URL может быть постоянным, пока URL-адрес распознавателя поддерживается. В нашем случае, URL может быть смоделирован следующим образом:

<URL-адрес распознаватель> <a URN: NBN>

Например https://resolver.nkp.cz/urn:nbn:cz:nk-000dju перенаправляет к URL-адрес системы презентации Национальной библиотеки.

Реестр является ядром базы данных системы идентификаторов. Он хранит все назначенные URN: NBNS, сопровождающие метаданные, фактические адреса URL, а также все предыдущие URL-адреса. Генератор напрямую подключен к реестру системы.

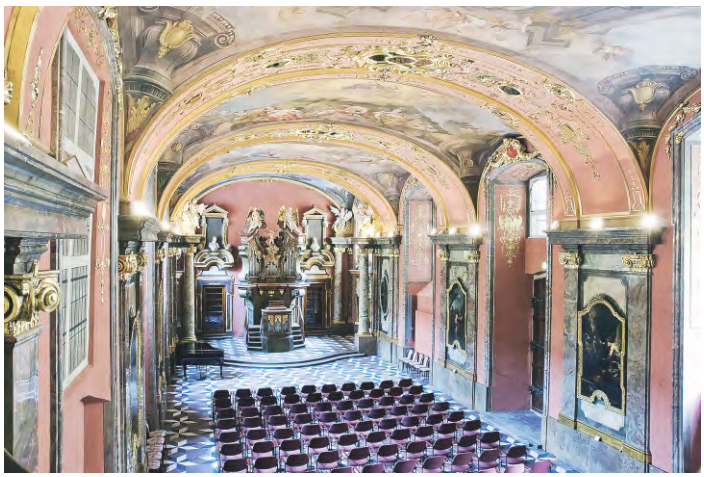
Каждая национальная реализация URN: NBN системы имеет различные технические решения и различные административные правила. Например, есть веб-сервис для генерации уникальных идентификаторов в Финляндии, которые в момент создания не связаны с реестром, используемого распознавателя. Реестр осознает назначенный URN: NBNS не скоро, чем после сбора их из хранилища регистратора. Многие страны (например, Германия или Нидерланды) решили доверить генерации NBN строк с их регистраторами.

Комбайн обеспечивает механизм для доставки всех текущих URL-адресов из registrar's системы представления в реестр. Комбайн периодически запрашивает OAI-PMH депозитарий системы представления. Бэк-офис является компонентом для администрирования системы идентификаторов и управления рабочих процессов (регистрации регистраторов, назначая регистраторов кодов, определение политики доступа и т.д.).

Программное обеспечение построено в виде набора веб-приложений Java для приложений контейнера (например, Apache, Tomcat). Поиск осуществляется (через JDBC, спящий режим) базу данных PostgreSQL. Frontend состоит из HTTP API RESTful (Джерси), OAI-PMH поставщика данных (Servlet) и веб-приложений AJAX. Внешние процессы обеспечиваются Кварц планировщиком заданий.

Заключение

Данная система была признана международным библиотечным сообществом, что обеспечивает постоянный доступ к документам цифрового наследия по своей сути и предполагает цифровую сохранность и постоянное выявление этих документов.



Постоянные идентификаторы предлагают возможность обеспечить постоянный доступ в случае, если они хорошо определяются и управляются с помощью системы идентификаторов. Национальный URN: NBN системы доказали свою надежность.

Национальные библиотеки кажутся надежными хранителями систем идентификаторов, поскольку они имеют долгосрочную традицию в области стандартизации. Чешская реализация URN в: NBN стандарта является довольно сложной. Она была предложена в качестве системы идентификации для оцифрованных материалов, чтобы покрыть большие масштабы оцифрованного контента собственного производства многих чешских библиотек. Система полезна не только для обеспечения постоянного доступа, но и для поддержки долгосрочного сохранения определенных документов. Реестр предоставляет пользователям метаданные, достаточные для установления достоверности цитат. Даже если документ больше не существует, пользователи могут найти его записи в реестре и использовать его в качестве доказательства своего прошлого существования.

Существует веб-сервис для генерации уникальных идентификаторов в Финляндии, которые в момент создания не связаны с реестром используемого распознавателя. Реестр осознает назначенный URN: NBNS не скоро, чем после сбора их из хранилища регистратора. Многие страны (например, Германия или Нидерланды) решили доверить генерации NBN строк с их регистраторами

Зденек Васек, Чехия Республикаси Миллий кутубхонасининг Рақамлаштириш бўйича ахборот стандартлари сақлаш бўлими раҳбари

Мақола муаллифи доимий идентификатордан (ноёб объектни фарқланувчанлиги) фойдаланиш кутубхона соҳасида кенг тарқалган тизим сифатида ISBN ва ISSN системаларида кенг жорий этилганлиги ҳақида маълумот берган. Яъни аввалда ҳужжатли кўринишга эга бўлган, рақамлашган идентификатор функциялари эндиликда Интернет ҳужжатлардан ҳам доимий равишда фойдаланиш имкониятига эга бўлди. Бу системадаги фойдаланишдан кўзланган асосий мақсад Чехия маданий меросидаги босма нашрларни рақамлаштирилган ҳолда сақланишини таъминлашдир. ■