
ICANN73 | Неделя подготовки — Свежая информация о программе создания IDN-доменов
22 февраля 2022 года, 9:00 – 10:00 по AST

ПИТИНАН КООАРМОРНПАТАНА: Спасибо. Добро пожаловать на первое заседание недели подготовки к ICANN73. Доброе утро, добрый день, добрый вечер! Я Питинан Кооарморнпатана (Pitinan Kooarmornpatana) из группы, которая занимается программой IDN-доменов и универсального принятия.

Позвольте мне поприветствовать вас на этом заседании, где будет представлен отчет по программе IDN-доменов. Это будет 60-минутное заседание, которое охватывает следующие темы. Во-первых, мы познакомимся со свежей информацией об инструкциях по развертыванию интернационализированных доменных имен, а затем перейдем к теме дополнительных алфавитов для поддержки IDN-доменов, а также узнаем новости о состоянии LGR для корневой зоны. Поэтому я кратко предоставлю справочную информацию. И расскажу о результатах работы комиссий по генерации правил на бирманском, а также здесь присутствует член комиссии по интеграции правил, Марк Бланше (Marc Blanchet), который сообщит новости о LGR в корневой зоне версии 5.

На этом позвольте мне перейти к первому пункту. Инструкции по развертыванию интернационализированных доменных имен. Для

Примечание: Примечание. Следующий документ представляет собой расшифровку аудиофайла в текстовом виде. Хотя расшифровка максимально точная, иногда она может быть неполной или неточной в связи с плохой слышимостью некоторых отрывков и грамматическими исправлениями. Она публикуется как вспомогательный материал к исходному аудиофайлу, но ее не следует рассматривать как аутентичную запись

справки — это политика и практика регистрации IDN-доменов второго уровня, и цель — минимизировать риск для безопасности, а также путаницу для потребителей.

Что касается gTLD, это зафиксировано в договорах, а также упомянуто в спецификации 6 RA, раздел 1.4: оператор регистратуры должен соблюдать инструкции ICANN по развертыванию интернационализированных доменных имен. Кроме того, в Дополнительной статье 3 RAA упоминается, что регистраторы тоже должны соблюдать инструкции для IDN-доменов. Также ожидается, что эти инструкции будут соблюдать регистратуры национальных доменов, как указано в ускоренной процедуре рассмотрения заявок на регистрацию национальных IDN-доменов верхнего уровня.

Поэтому, в то время как эти инструкции сейчас обновляются, предыдущая версия инструкций была версией 3 от 2011 года, поэтому был проведен дополнительный анализ, собраны данные и опыт работы с IDN-доменами. Например, терминология IDN-доменов, представление таблицы, которая стала пригодной для машинного считывания в формате LGR в RFC 7940.

Максимальный стартовый набор или MSR, а также мы изучили пользовательский интерфейс силами SSAC. Кроме того, многие предложения проекта LGR для корневой зоны разработаны сообществами о том, как использовать тот или иной алфавит.

То есть у сообществ есть рабочая группа для обновления инструкций до версии 4. В версии 4 они охватывают в общей сложности семь тем и 19 инструкций. Таким образом, есть группа, которая занимается вопросами, относящимися к переходу, к теме формата таблиц IDN, к теме согласованности таблиц IDN и практики, меткам вариантов, а также сходству и возможности смешения меток, наряду с публикацией политики и правил, и, наконец, терминологии.

Эти инструкции составлены рабочей группой, а затем прошли период общественного обсуждения и были доработаны и опубликованы в мае 2018 года.

Тогда, в 2019 году, GNSO обратилась к корпорации с просьбой предоставить дополнительное время для изучения инструкций, прежде чем ввести их в действие. Этот вопрос был рассмотрен в Правлении с учетом указанного запроса. А затем, в мае прошлого 2021 года, GNSO инициировала ускоренный процесс формирования политики в области IDN-доменов для решения дополнительных вопросов, относящихся к IDN-доменам, не охваченным в отчете по SubPro. Это также включает в себя управление вариантами и обновление принципов для IDN-доменов в будущем.

В августе GNSO также попросила Правление отложить принятие инструкций до тех завершения этого EPDP из-за частичного совпадения. Затем Правление в октябре ответило, предложив

GNSO проанализировать и определить перекрывающиеся инструкции.

Таким образом, в январе GNSO ответила, поддержав предложение Правления, и определила следующие перекрывающиеся инструкции, поэтому существует пять дополнительных инструкций.

В настоящее время Правление рассматривает первый вопрос: как поступить в дальнейшем. Таков к данному моменту статус инструкций по развертыванию. Позвольте мне сделать паузу и посмотреть, есть ли какие-либо отзывы или комментарии, или вопросы, прежде чем мы двинемся дальше.

САРМАД ХУСЕЙН:

У нас нет никаких вопросов. Спасибо, Питинан.

ПИТИНАН КООАРМОРНПАТАНА:

Спасибо. Мы также выделили некоторое время для сеанса вопросов и ответов в конце, так что мы сможем вернуться к этому. Хорошо, давайте перейдем к следующей теме — это дополнительные алфавиты для поддержки IDN-доменов.

Для справки, миссия ICANN охватывает работу по созданию доменных имен, которые могут быть использованы на родных языках. Итак, по сути — это IDN-домены. И стандарт для

интернационализованного доменного имени или IDNA на основе Unicode дает возможность использовать IDN.

Благодаря стандарту IDNA 2008. Он является [неразборчиво] механизмом для расчета кодовых точек, но также предлагает добавить дополнительные ограничения в зависимости от анализа. Однако там в заключение не проводится рассмотрение на основе алфавитов.

ICANN вместе с сообществами, использующими один и тот же набор символов, занимается разработкой для них безопасного и стабильного решения, и это основано на процедуре определения LGR для корневой зоны, которая до сих пор находится на первом раунде работы: мы определяем объем работ для 28 алфавитов — это является максимальным стартовым набором. Итак, они перечислены здесь от арабского, армянского, бенгальского и так далее к тайскому и тибетскому. Таким образом, это 28 алфавитов, идентифицированных для корневой зоны, и эти же 28 алфавитов используются для разработки эталонного LGR для второго уровня, а также для gTLD.

Но на самом деле, в самом Unicode, на данный момент в версии 14, есть 159 алфавитов. Он также включает в себя некоторые старые алфавиты, а также находящиеся в современном использовании, и в Unicode также определено, что не все эти алфавиты, возможно, подходят для идентификаторов, а доменное имя также является одним из идентификаторов.

То есть надо ответить на вопрос: какие среди этих 159 алфавитов должна поддерживать корпорация ICANN? Это будет рассмотрено на верхнем уровне, а также на втором уровне.

Итак, рассмотрим подробнее Unicode. Для этих 159 алфавитов в стандарте Unicode есть Приложение 31, где алфавиты разделены на три категории. Существуют алфавиты исключенные, для ограниченного использования и рекомендуемые.

Исключенных алфавитов 94. По сути, алфавиты, которые не являются общеупотребительными, могут быть историческими или используемыми в очень ограниченном или небольшом сообществе. И кроме того, у некоторых алфавитов есть нерешенные архитектурные проблемы. Например, проблемы визуализации и так далее. Примеры перечислены здесь: брахми, коптский, египетские иероглифы. Это исключенные алфавиты.

Другая категория — верхняя должна быть алфавитами с ограниченным использованием... это не правильно. Таким образом, для ограниченного использования, это современный алфавит, но он используется в ограниченном сообществе. Здесь, чтобы избежать проблем безопасности, в некоторой реализации он может быть запрещен, но все-таки может применяться. Примеры: балийский, слоговое письмо канадских аборигенов, чероки и так далее.

Последняя категория — рекомендуемые алфавиты. Их 29, а также некоторые общие наследуемые, как цифры и дефис. У такого

алфавита широко распространенное использование, он современный и используется повседневно, есть крупные сообщества, использующие его. Примеры: арабский, армянский и так далее, например, набор алфавитов, над которым мы уже работаем в проекте LGR для корневой зоны.

Итак, чтобы ответить на поставленный ранее вопрос, ICANN привлекла эксперта для проверки этих алфавитов и обобщила результаты в отчете. Таким образом, мы также можем привести ссылку на отчет из этой презентации, но в резюме данного отчета говорится, что рекомендуемый алфавит подходит для корневой зоны, а также на втором уровне. Если это алфавит с ограниченным использованием, то он может не подходить для корневой зоны, но его можно использовать на втором уровне. Однако решение следует принимать в каждом конкретном случае. И исключенные алфавиты не подходят ни для корневой зоны, ни на втором уровне.

Двигаясь дальше, оценка... оценка алфавита. Алфавит для ограниченного использования следует рассматривать в каждом конкретном случае из-за того, что такой алфавит все же может создавать некоторые проблемы для безопасности и стабильности. То есть сообщества, использующие один и тот же набор символов, действительно должны помочь выработать безопасное решение. Но из-за его ограниченного использования иногда у нас может отсутствовать возможность найти сообщество, чтобы проконсультироваться с ним для разработки правил генерирования меток. Здесь мы определили, что не располагаем

достаточной информацией для выработки эталонных данных. Поэтому без достаточной информации мы не сможем разработать эталон и не сможем эффективно поддерживать этот алфавит.

И это некоторые примеры риска, который может возникнуть из-за ограниченного использования алфавита. Итак, список алфавитов представлен здесь, все они с ограниченным использованием: чероки, кая, лису и так далее. Здесь приведены примеры меток. Некоторые из них могут выглядеть как латинские заглавные буквы, но на самом деле являются другими кодовыми точками. Поэтому, если вы видите здесь кодовые точки, они отличаются от латинской заглавной буквы А, и это может ввести пользователя в заблуждение или создать проблемы из-за фишинга.

Кроме того, некоторые из них выглядят как строчные латинские буквы, такие как CEO здесь и NEW и так далее. Но, по сути это только так выглядит, кодовые точки разные.

Другой пример, аналогичный, все еще из области визуального сходства. В этом алфавите с ограниченным использованием некоторые символы могут выглядеть как символы из рекомендуемого алфавита. Так, к примеру, новое письмо лы может выглядеть как бирманский или малаялам. Вы можете создать метку, используя только новое письмо лы, а другую метку на бирманском алфавите, и они будут выглядеть в значительной степени одинаково, но это разные кодовые точки. Кроме того, бенгальский, который является рекомендуемым алфавитом, уже

используемым в корневой зоне. При этом на силхетском нагари можно создать что-то похожее.

Другим примером риска является риск отображения. Из-за ограниченного характера использования у некоторых систем на самом деле может отсутствовать механизм надлежащего отображения или шрифты для управления этим алфавитом. Этот пример из алфавита неварского языка — одного из алфавитов с ограниченным использованием, и если мы скопируем этот символ из самой таблицы кода Unicode, 11420, и вставим его в программное обеспечение, то в этом случае на экране — это браузер Chrome на компьютере Mac — заметно, что не удается его верно отобразить, и здесь появляются вопросительные знаки. То есть применение алфавита с ограниченным использованием, у которого нет правильной поддержки отображения, также может создать проблему. И этот список — некоторые из алфавитов с ограниченным использованием, которые мы тестируем, и мы находим подобные проблемы.

Поэтому, возвращаясь к сделанному ранее выводу, очень общими резюмируя, что исключенный алфавит может оказаться непригодным как верхнего уровня, так и для второго уровня, алфавит с ограниченным использованием может подходить для второго уровня в зависимости от конкретного случая. Это отражено в отчете, который мы опубликовали для общественного обсуждения. Обсуждение было открыто 5 января, оно только что завершилось на прошлой неделе 15 февраля, а отчет будет

выпущен в начале марта. Так что сейчас мы принимаем вклад, полученный в результате этого общественного обсуждения, и учитываем его в рамках следующих шагов.

Позвольте мне на этом приостановиться. Есть вопросы или комментарии по этому поводу?

САРМАД ХУСЕЙН:

Есть еще несколько общих вопросов. Я зачитаю их вслух. Набил Ясин (Nabeel Yasin) спрашивает: «Считаете ли вы, что IDN-домены приведут к интернет-изоляции конечного пользователя и дефрагментации интернета?» Принс Эндрю Зута (Prince Andrew Zutah) задает уточняющий вопрос. «Как вы преодолеете языковой барьер при получении запроса на доменное имя IDN на другом языке?»

В чате уже есть ряд ответов. Также вопрос задает Раймонд Мамата (Raymond Mamattah). «Каковы основные отличия универсального принятия от IDN-доменов?» На эти вопросы уже отвечают в чате, но все равно, Питинан, если вы хотите кому-то ответить, вам слово. Спасибо.

ПИТИНАН КООАРМОРНПАТАНА:

Спасибо. Давайте я просто выберу один из них, и если что-то не охвачу, то мы сможем ответить во время сеанса ответа на вопросы. Разница между IDN-доменами и универсальным принятием в том, что IDN-домены скорее... мы готовим

технологии, поэтому раньше у нас, возможно, не было решения, разработанного сообществами о том, как правильно использовать конкретный алфавит. Прямо сейчас ICANN содействует работе сообщества, особым усилиям сообщества. Это решение теперь стало доступным.

Но затем следующий шаг, прямо сейчас, если этот алфавит используется, технически, может быть, программное обеспечение или другой поставщик услуг в интернет-экосистеме может не знать об этом и может считать... например, если у вас есть электронная почта в интернационализированных доменных именах, которые мы называем EAI, интернационализация адреса электронной почты. Так вот, если у вас есть EAI и вы используете свои электронные письма на некоторых сайтах, то когда вы захотите подписаться на что-то, сайт может отклонить запрос и сообщить, что это неправильный адрес электронной почты. Так что это следующий шаг помимо IDN-доменов, теперь, когда есть техническая готовность, как нам сделать его принятым всеми платформами, системами?

САРМАД ХУСЕЙН:

Спасибо, Питинан. Я также поделился ссылкой на завтрашнее заседание, посвященное универсальному принятию. Есть вопрос от Денниса Тана (Dennis Tan). «Будут ли проблемы отображения немного связаны с универсальным принятием?»

ПИТИНАН КООАРМОРНПАТАНА: Да. Хороший вопрос, Деннис. Спасибо. Думаю, ответ на него нам нужно найти вместе. Да, по-моему, добиться принятия системой нетрадиционных доменных имен ASCII и электронных писем тоже является частью универсального принятия. Хорошо, тогда давайте двигаться дальше, а если в конце останется время, мы вернемся к этим вопросам.

Следующий пункт — свежая информация об LGR для корневой зоны. Я представлю краткий обзор проекта, а затем передам слово Инь Мэй У (Yin May Oo), сопредседателю комиссии по выработке правил для бирманского языка, чтобы представить отчет о работе над LGR в корневой зоне на бирманском, а затем мы передадим слово Марку, который расскажет о LGR для корневой зоны версии 5.

Итак, это просто обзор, краткая история проекта LGR для корневой зоны, все началось с требований сообщества о том, что есть необходимость в вариантах доменов верхнего уровня. В то время сообщество — это 2009–2010 годы — собралось вместе и работало над комплексным отчетом о проблемах, который определил, что необходимо сделать для получения возможности создавать варианты доменов верхнего уровня.

Требовалось выяснить две вещи. Во-первых, не было определения вариантов. По сути, это зависит от алфавита. Второй момент — отсутствие механизма управления вариантами.

Что касается первого, как определить варианты, сообщество также разработало процедуру LGR для определения правил генерирования меток корневой зоны. И эти правила, по сути, обеспечивают генерирование метки для корневой зоны, а внутри правил также вводится определение вариантов.

Эта процедура LGR была принята Правлением в 2013 году. С тех пор для каждого алфавита сообщество начало работать над своим собственным алфавитом и разрабатывать LGR для корневой зоны. Параллельно корпорация ICANN также приступила к разработке рекомендации о том, как управлять вариантами TLD.

Двигаясь вперед к 2019 году, Правление ICANN решило, что GNSO и ccNSO должны учитывать эту рекомендацию при управлении вариантами IDN-доменов, которые в этой рекомендации также интегрируют LGR для корневой зоны, а теперь принимает участие в разработке политики GNSO и ccNSO.

Кроме того, в 2020 году, при наличии различных предложений сообщества относительно LGR для корневой зоны для многих алфавитов Правление ICANN также решило, что GNSO и ccNSO должны принять к сведению рекомендацию о техническом использовании LGR для корневой зоны. В начале 2021 года GNSO также опубликовала отчет о последующем раунде создания новых gTLD, который также включает использование LGR для корневой зоны.

И это дает картину работы, проделанной сообществом. Для 28 алфавитов мы определили работу по созданию LGR в корневой зоне для этого раунда, мы закрываем 26 алфавитов, и вы можете видеть ряд работ, которые могут продлиться несколько лет, деятельность рабочей группы длится около пяти, уже шесть лет.

Но в начале этого года близится завершение деятельности всех активных комиссий по выработке правил или групп сообщества, которые мы называем комиссиями по выработке правил, и у нас еще остаются алфавит тана и тибетский, и мы надеемся, что сможем сформировать комиссию сообщества и тоже поработать над LGR для корневой зоны.

С этими словами я передаю микрофон Инь Мэй У, сопредседателю комиссии по выработке правил для бирманского, чтобы узнать новости об LGR в корневой зоне для бирманского. Вам слово.

ИНЬ МЭЙ У:

Приветствую всех. Это языки, которыми мы занимаемся на основе [неразборчиво], на них говорят от 550 млн до 500 000 носителей. Это список языков по количеству носителей в нашем сообществе, использующем один и тот же набор символов, бирманский алфавит.

Итак, бирманский алфавит использовался около 200 лет по всей стране, и мы заимствовали очень много словарных единиц [в сфере торговли и культуры], а также в том числе [неразборчиво] и

много вещей из разных языков, и это причудливая смесь на западном побережье Юго-Восточной Азии.

За прошедшие годы мы заимствовали множество [неразборчиво] от [неразборчиво] тибетского языка, который разработали, смешанные с [неразборчиво] носителями языка, тайского и многих других исторически существовавших языков, которые мы обнаружили.

Таковы истоки, и, как вы все знаете, в основе бирманского алфавита лежит система письменности абугида, которая распространена по всей Индии и Юго-Восточной Азии. Мы также приняли эту систему и заимствовали [неразборчиво] знаки и многое другое у разных этнических групп и культур. Таким образом, у нас в конечном итоге много согласных и диакритических знаков, больше одного. Используется до шести или семи диакритических знаков, чтобы сформировать один слог. Следующий слайд, пожалуйста.

Для одного слога это может быть одна согласная буква, которая может придвигаться к другой согласной вместе со средним элементом, когда мы можем добавить еще одну согласную, чтобы сместить ударение на другой звук, а затем добавить динамику тона, чтобы сделать его скрипучим или более высоким, например длинным звуком. Таким образом, существует много возможностей, и мы должны убедиться, что комбинации соответствуют каноническому порядку Unicode, чтобы не

сформировать каноническую визуализацию и убедиться, что метка надежна и безопасна. Эта метка [неразборчиво] человеческие возможности, одна согласная может быть изменена путем добавления одного [неразборчиво], а затем можно сделать гораздо больше. Следующий слайд, пожалуйста.

Как вы можете догадаться, в регионе используется более десяти алфавитов. Однако мы сократили их количество, чтобы сосредоточиться на шести языках, и исключили кодовые точки [неразборчиво], восклицательные знаки и устаревшие символы, так что для охвата шести языков в нашем наборе символов в конечном итоге есть 90+ кодовых точек. К сожалению, нам пришлось убрать ряд языков с меньшим количеством пользователей и [без ссылки на наличие.] Следующий слайд, пожалуйста.

После того, как мы проанализировали, какие кодовые точки позволяют создать метку, то изучили другие языки, чтобы отфильтровать возможные варианты и столкнулись с латинскими, греческими, кириллическими и армянскими буквами. Поскольку бирманский алфавит решено основывать на [неразборчиво] символах, это первое [неразборчиво], что касается этой буквы, мы обнаружили, что во многих языках по всему миру также используется эта полная окружность.

Итак, вот список, который мы получили в результате. Помимо полной окружности у нас есть окружности, открытые снизу, сверху

и сбоку. Среди них есть другая форма, которую мы обнаружили как вариант для других алфавитов. Следующий слайд, пожалуйста.

Таким образом, у нас есть варианты, основанные на символах полного круга, и поскольку эти варианты... кажется, что варианты ближе к западным языкам, и поскольку в нашем языке множество различных типов диакритических знаков, гласных, срединных элементов или [неразборчиво], мы должны маркировать, какая кодовая точка принадлежит к той или иной группе, и какие... в соответствии с группами они должны следовать в определенном порядке, чтобы формировать правильные [неразборчиво]. Таким образом, мы должны [неразборчиво] и составить правила, соответствующие поведению при формировании той или иной буквы. Следующий слайд, пожалуйста.

То есть мы должны составить начальные правила для каждого типа диакритических знаков. Какие диакритические знаки не должны следовать за другим типом диакритических знаков, чтобы не создавать путаницу или проблемы при отображении, чтобы мы могли устранить, а затем убедиться в отсутствии путаницы среди пользователей, поскольку некоторые пользователи... большинство пользователей, безусловно, знают, что написано, но могут быть люди, которые узнают это только через ограниченные ресурсы. В нашей стране в некоторых сельских районах могут быть люди, которые не знают полный набор символов языка, но все же знают общие термины, и они также могут захотеть использовать

интернет. Поэтому мы должны учитывать, чтобы наши метки не делали ничего, ухудшающего возможность их прочтения.

Вот примеры меток, и для каждого правила мы должны просмотреть весь список общих терминов и убедиться, что не формируется ничего опасного или неразличимого. Следующий слайд, пожалуйста.

Таким образом, начиная с 2018 года мы формировали наше GP, закончили составлять наше предложение и проводим общественное обсуждение, открытое в январе 2022 года, и мы все еще рассчитываем получить больше комментариев. Вопреки всем прогнозам и многим ситуациям, мы все еще надеемся, что сможем дотянуться до пользователей бирманского алфавита и людей, говорящих на бирманском языке, которые должны понять, что интернет не только для английского языка и символов ASCII, чтобы у них не было страха... если они боятся учить другие языки, то смогут использовать другие кодовые точки. Мы надеемся, что сумеем проинформировать сообщество о том, что интернет готов к снижению языкового барьера. Спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Инь Мэй. Есть вопросы или комментарии к Инь Мэю?

САРМАД ХУСЕЙН: Спасибо, Питинан. В чате есть вопрос, не имеющий прямого отношения к Инь Мэю, а относящийся к предыдущей части презентации. Вопрос задает Андрей. «Зачем адаптировать неиспользуемые исторические алфавиты? Если я правильно понял».

ПИТИНАН КООАРМОРНПАТАНА: Спасибо за вопрос, Андрей. Мы тоже задавали аналогичный вопрос во время общественного обсуждения. Спасибо за вопрос. Если у вас есть комментарий, будет хорошо, если вы им поделитесь. Не знаю, опубликован ли он уже, но не могли бы вы, Сармад, дать ссылку на общественное обсуждение. Это тоже было бы здорово. Спасибо. Итак, некоторые из подробностей обработки различных категорий алфавитов включены в отчет. Спасибо. Еще вопросы?

САРМАД ХУСЕЙН: У меня всего лишь еще пара вопросов, на которые тоже отвечено в чате. Мулуд Келиф (Mouloud Khelif) спрашивает: «Каковы ваши взаимоотношения с Группой управления по универсальному принятию?» Есть еще один вопрос: «Сколько делегированных IDN-доменов верхнего уровня в корневой зоне на данный момент?»

ПИТИНАН КООАРМОРНПАТАНА: Спасибо. Итак, Группа управления по универсальному принятию — это руководящая группа, возглавляемая

сообществом. Группа ICANN по программе универсального принятия IDN-доменов тесно сотрудничает с UASG. Таким образом, в то время как UASG [неразборчиво]... в UASG также есть несколько внутренних рабочих групп. Завтра, а также на основной неделе нами будут проведены заседания, посвященные именно универсальному принятию. Возможно, вам будет интересно тоже принять в них участие. Но, по сути, участники программы IDN-доменов и универсального принятия также работают с группой UASG, которую возглавляет сообщество.

Следующий вопрос — сколько IDN-доменов верхнего уровня... Около 150 — если считать IDN-домены ccTLD и gTLD. Я скоро найду точное количество. Я приведу ссылку на него. Хорошо, спасибо, Марк, их 154.

Хорошо, тогда давайте двигаться дальше. Я передам слово Марку Бланше (Marc Blanchet), члену комиссии по интеграции правил, чтобы он сообщил нам свежую информацию о следующей версии LGR в корневой зоне.

МАРК БЛАНШЕ:

Спасибо, Питинан. Итак, эта презентация о нашей работе в качестве комиссии по интеграции правил с момента проведения последней конференции ICANN. Наша цель — действительно сделать то, для чего создана комиссия, то есть выполнить интеграцию. Соответственно, мы берем LGR у комиссий по выработке правил и объединяем их, чтобы сформировать

корневую зону LGR, которая используется сообществом для TLD, корневых зон. Следующий слайд, пожалуйста.

В настоящее время мы работаем над пятой версией LGR в корневой зоне. Четвертая была опубликована осенью 2020 года. Правила генерирования меток версии 4 для корневой зоны содержали 18 алфавитов, таких как арабский, бангладешский, хань, деванагари, эфиопский, грузинский, гуджарати, гурмукхи, иврит, каннада, кхмерский, лаосский, малаялам, ория, сингальский, тамильский, телугу и тайский.

Мы намерены интегрировать шесть новых алфавитов: греческий, хирагана и катакана для японского, которые также содержат хань, но хань уже есть в LGR версии 4, корейский, который включает в себя собственный специфический алфавит, хангыль, а также какие-то символы хань, которые тоже находятся в LGR версии 4, латинский и мьянманский.

Интересно, что у нас было два алфавита, представленных нам много лет назад, армянский и кириллица, но мы решили отложить их. Это означает, что мы пока не интегрируем их в LGR для корневой зоны, потому что они тесно связаны с латынью и греческим. То есть мы ждем, чтобы получить все четыре алфавита вместе, финальные версии, чтобы мы могли интегрировать их, потому что они делятся вариантами, и мы должны, по процессу интеграции, фактически вводить различные перекрестные варианты алфавитов, один на другой. Поэтому нам нужно

подождать. Нам также пришлось ждать, поскольку мы хотели убедиться, что различные алфавиты согласованы друг с другом.

Соответственно, в LGR версии 5 теперь два набора связанных алфавитов, которые полностью интегрированы, CJK (китайский, японский и корейский) и латынь, греческий, кириллица и армянский. В некоторых отношениях LGR версии 5 будут самыми сложными за всю историю интеграции, и поэтому мы тратим много времени, чтобы проверить и убедиться в правильности интеграции. Следующий слайд, пожалуйста.

Эти LGR версии 5, над которыми мы работаем, будут содержать по существу 25 элементов LGR. Элемент LGR означает конкретный алфавит. У нас 25, потому что алфавитов 26, но в Японии для двух алфавитов один LGR, так что 26-1. И мы также выпускаем общий LGR, который механически генерируется элементами LGR.

Целевая дата публикации LGR в конечной зоне для общественного обсуждения — 24 марта, а это значит, что вскоре мы надеемся получить версию для мьянманского, только что представленную Инь, чтобы ее интегрировать.

Мы действительно [желаем] увидеть, как люди потратят свое время на рассмотрение LGR версии 5, потому что это, опять же, они самые сложные, а также, скорее всего, их нужно будет использовать для проведения новых раундов и других вещей. По существу, это завершит интеграцию для всех активных комиссий

по выработке правил LGR, которые мы получили. Следующий слайд, пожалуйста.

Что касается перекрестных вариантов алфавитов, играющих довольно важную роль в данном выпуске, они будут полностью отражены в общих LGR, которые, как я уже сказал, механически сгенерированы из элемента LGR, когда мы берем все перекрестные варианты между разными LGR и объединяем их вместе, применяя транзитивность и, соответственно, создавая все общие LGR. Общие LGR используются для проверки на наличие конфликтов между допустимыми метками.

А то, что показано ниже — на самом деле мы приняли решение с точки зрения обработки вариантов. Таким образом, если есть вариант, который определен только в элементе LGR, он будет [включен в список] и интегрирован. Для любого алфавита, соответственно, во [всем] наборе символов, варианты всегда будут перечислены в затрагиваемом и подходящем LGR. Варианты ASCII включаются в список всегда. Элемент LGR из латинского, греческого, кириллического и армянского наборов включается в список всех вариантов в наборе, но только латинские LGR будут перечислены в списке вариантов вне этого набора.

Для ранее отложенных алфавитов, кириллицы и армянского, добавлены варианты, налагаемые интеграцией, и несколько других элементов LGR, таких как малаялам, выбраны варианты, которые налагаются новыми алфавитами в LGR версии 5.

Обращаюсь с призывом к комиссиям по выработке правил, которые участвуют в этом. Например, кириллица, армянский, малаялам и другие.

Будет очень полезно, если вы подтвердите, что проделанная нами работа по интеграции, когда мы на самом деле добавляем новые варианты ваши LGR, не создает никаких ошибок для вашего элемента LGR или проблем с процессом интеграции. На самом деле это призыв ко всем GP, чтобы все сообщество перепроверила нашу работу, чтобы мы узнали, что все в порядке для следующих шагов относительно этих LGR для корневой зоны. Следующий слайд, пожалуйста.

Итак, это было представлено в прошлый раз, и я не буду слишком сильно углубляться, но хочу просто сказать, что обычно, когда вы хотите обработать или проверить метку, используя LGR для коневой зоны, сначала проверяется, что такая метка допустима для элемента LGR, который вы будете использовать, затем применяются общие LGR для выявления конфликтов, и если все в порядке, то элемент LGR алфавита используется повторно, чтобы вычислить доступный вариант, поэтому у вас будут различные метки, которые являются вариантами вашей первоначальной метки.

Вот так мы используем LGR для корневой зоны. Очевидно, что есть, например, инструмент LGR, созданный ICANN, который позволяет вам выполнить весь этот процесс. Следующий слайд, пожалуйста.

Следующие шаги комиссии по интеграции правил. Опять же, мы намерены опубликовать LGR для корневой зоны версии 5 к 24 марта, то есть примерно через месяц. Мы очень хотели бы, чтобы сообщество и особенно комиссии по выработке правил просмотрели и проверили нашу работу, и это очень полезный для нас комментарий, если вы не найдете проблем, но все-таки отправите сообщение на портал комментариев о том, что проверили свой алфавит, и все в порядке. Это очень ценно для нас. Поэтому, пожалуйста, найдите время, чтобы проверить, и дайте нам знать.

Итак, мы планируем завершить работу после общественного обсуждения ко 2-му кварталу 2022 года. Возможно, появится следующая версия LGR для корневой зоны, если комиссии по выработке правил для тибетского и тана сформируют и отправят нам свой LGR для окончательной интеграции. Однако, как мы поняли из презентации Питинана, эти комиссии еще не сформированы. Поэтому у нас нет даты выхода этой новой версии, которая, скорее всего, будет LGR 6. Мы не установили срок для этой работы. И, по-моему, это последний слайд, конец моей презентации. Спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Марк. Сармад, в чате еще есть вопросы?

САРМАД ХУСЕЙН: Да, у нас есть пара вопросов. Один задает Билл Джурис (Bill Jouris). Почему в списке нет перекрестных вариантов, найденных для кириллицы, греческого и армянского, даже если GP для латинского не считает их вариантами? Пожалуй, мы можем ответить на этот вопрос, а потом на следующий.

МАРК БЛАНШЕ: Хорошо, разработанная нами политика, по сути, отражена в общих LGR, все перекрестные варианты будут там. Однако, чтобы не получить все LGR с большим количеством вариантов, не связанных с самим алфавитом, мы фактически решили ограничить наложение некоторых элементов LGR с целью читаемости алфавитов. Но для получения полного списка перекрестных вариантов всегда будут использоваться общие LGR.

САРМАД ХУСЕЙН: Спасибо, Марк. Есть еще один вопрос от Вадима Михайлова. «Я понимаю правильно, латынь в LGR версии 5 — это не только базовая латынь, но и расширенная? Если это так, то какие расширения, А, В, С из стандарта Unicode были включены?»

МАРК БЛАНШЕ: На самом деле этот вопрос скорее относится комиссии по выработке правил для латыни, и я думаю, что Билл здесь, поэтому он может ответить. В целом, каждая комиссия по выработке правил сама определяет все языки с некоторым уровнем

использования, применяя конкретный известный стандарт, который называется [неразборчивым], и определяет объем своей работы применительно ко всем языкам, которые будут поддерживаться этими LGR, алфавитными LGR и, соответственно, создает наборы кодовых точек, которые рассматриваются.

В то же время они также должны смотреть на MSR, максимальный стартовый набор. Обычно он более широкий, и из него исключены все исторические символы, которые больше не используются. Соответственно, это создает начальный набор кодовых точек, подлежащих рассмотрению. Так что в целом это все поддерживаемые латинские символы, я не знаю, около 200 языков, использующих латинский алфавит.

САРМАД ХУСЕЙН:

Спасибо. Еще один вопрос задает Сатиш Бабу (Satish Babu). «Версия 5 будет полностью обратно совместима с версией 4»?

МАРК БЛАНШЕ:

Да. Опять же, окончательная версия еще не подготовлена, но должна быть полностью совместимой, и в данный момент нет никаких оснований считать, что она не будет обратно совместимой. И если это когда-либо произойдет, люди напишут об этом во время общественного обсуждения.

САРМАД ХУСЕЙН:

Спасибо. Был еще один вопрос, на который также ответили в чате, но я его зачитаю. Вопрос задает Набиль Ясин (Nabeel Yasin). «Системы DNS, такие как BIND, способны транслировать домены IDN в IP-адреса?» Был комментарий в чате, где сказано, что в DNS-системах все метки находятся в формате A-Label, поэтому они, как я полагаю, в конечном итоге преобразуются в символы ASCII для внутренней обработки при разрешении имен. Марк, не хотите что-нибудь добавить. Спасибо.

МАРК БЛАНШЕ:

Что же. Этот вопрос больше относится к IDNA в целом. Есть две формы доменных имен и интернационализованных доменных имен. Одна — этот алфавит, реальное значение, которое увидит пользователь, и именно эта строка символов выводится пользователю, например, в браузере. Однако в системе DNS на самом деле кодируется другая строка символов. Система DNS все еще работает с ASCII. Мы разработали это, чтобы не менять все системы DNS в мире.

Соответственно, строка символов, написанных пользователем на своем алфавите, фактически переводится, как написал Михаил, в то, что мы называем A-Label при помощи алгоритма PunyCode. Он начинается с xn--, а затем идут остальные символы. Как правило, пользователь их не видит. Но именно это DNS использует для разрешения имен в IP-адреса, например.

САРМАД ХУСЕЙН: Спасибо. Вопрос задает Реймонд Маматтах (Raymond Mamattah). Выполнена ли какая-то работа над символами языка эве? Это один из основных языков в Западной Африке, особенно в Того и Гане.

МАРК БЛАНШЕ: Я об этом не имею никакого понятия. Я не уверен, знают ли об этом Питинан или Сармад, но я не могу ничего вспомнить. Хорошо, Михаил говорит, что это было рассмотрено на латыни. Хорошо. Спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо. Еще вопросы или комментарии? У нас осталась одна минута. Хорошо, чтобы просто подвести итоги, как краткое резюме, что касается обновления инструкций, мы еще рассматриваем следующие шаги.

Что касается общественного обсуждения поддержки дополнительных алфавитов в IDN-доменах, которое совсем недавно закрылось, если вы все еще хотите отправить комментарий, пожалуйста, не стесняйтесь сделать это, но в ближайшее время мы подготовим итоговую версию полученных комментариев. А также, что касается общественного обсуждения бирманского алфавита, оно открыто до 3 марта. Так что, хотя, кажется, это вас не очень сильно касается, для перекрестных вариантов также включена форма «O» и некоторые

распространенные формы. Если интересно, вы тоже можете взглянуть.

Наконец, общественное обсуждение LGR для корневой зоны версии 5 планируется начать к концу марта. Мы хотели бы получить комментарии, и было бы хорошо, если вам удастся это просмотреть. Думаю, у меня на сегодня все. Пожалуйста, не стесняйтесь обращаться к нам по адресу IDNprogram@icann.org с любыми вопросами и предложениями.

На этом я благодарю всех за участие в этом заседании. Надеюсь, что неделя для вас будет плодотворной, а конференция ICANN успешной. Спасибо.

[КОНЕЦ СТЕНОГРАММЫ]