
ICANN70 | Неделя подготовки — Свежая информация о программе создания IDN-доменов
11 марта 2021 года, 16:30–17:30 по EST

ПИТИНАН КООАРМОРНПАТАНА: Спасибо. Хорошо. Спасибо и добро пожаловать на заседание в рамках Недели подготовки к ICANN70, на котором будет представлена свежая информация о программе создания IDN-доменов. Хочу также напомнить: если вы не выступаете, выключите, пожалуйста, свой микрофон. Спасибо. Хорошо. Меня зовут Питинан Кооарморнпатана (Pitinan Koormornpatana), я буду вести сегодняшнее заседание.

Оно будет состоять из трех основных частей. Сначала о программе создания IDN-доменов расскажу я, а во второй части будут представлены отчеты от сообществ. Итак, на этом заседании присутствуют представители четырех сообществ.

Панагиотис представляет комиссию по выработке правил для греческого алфавита. Ким Кён Сок (Kim Kyongsok) представляет комиссию по выработке правил для корейского алфавита. Билл Джурис (Bill Jouris) представляет комиссию по выработке правил для латинского алфавита. И Инь Мэй У (Yin May Oo) представляет комиссию по выработке правил для мьянманского алфавита. Они расскажут о работе комиссий по выработке правил в рамках проекта LGR для корневой зоны. И в конце, надеюсь, у нас останется 10–15 минут, чтобы ответить на вопросы. Позвольте мне начать.

Итак, что касается цели программы создания IDN-доменов. Напомню, что цель заключается в том, чтобы обеспечить возможность внедрения доменных имен на местных языках и с использованием

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

местных алфавитов для глобальных сообществ. Разумеется, при этом необходимо соблюдать требования безопасности и стабильности. Для достижения этой цели у нас есть несколько проектов, для доменов как верхнего, так и второго уровня. Для доменов верхнего уровня есть правила генерирования меток корневой зоны, то есть проекты создания LGR для корневой зоны, в рамках которых сообщества совместными усилиями определяют, каким образом следует использовать конкретный алфавит для корневой зоны.

Проекты второй группы после определения правил использования — это проекты управления. У нас есть проект реализации вариантных строк TLD. И третья часть — это ускоренная процедура рассмотрения заявок на регистрацию IDN-доменов ccTLD. Затем, если перейти ко второму уровню, у нас есть проекты реализации руководящих принципов для IDN-доменов и справочные LGR, а также инструменты для проверки LGR или таблиц IDN. По каждому из этих пунктов я коротко расскажу предысторию для тех, кто не знаком с проектами, а затем дам свежую информацию.

Итак, по первому пункту, правила генерирования меток корневой зоны, или LGR для корневой зоны. Это процедура, которая была разработана сообществами для получения возможности выработать правила надлежащего использования каждого из алфавитов в корневой зоне. Для этого существует установленная процедурами структура: двухступенчатые комиссии. Первая ступень — это комиссия по выработке правил, формируемая пользователями алфавита. В состав каждой комиссии по выработке правил входят различные представители модели с участием многих заинтересованных сторон, среди которых много специалистов по техническим вопросам и

лингвистов. Кроме того, во многих комиссиях есть представители ccTLD, а также регистратур и регистраторов, работающих с этим алфавитом.

Таким образом, задача комиссий по выработке правил заключается в том, чтобы определить надлежащий метод использования алфавита с целью генерирования метки для корневой зоны, для чего необходимо установить три вещи. Во-первых, какие кодовые точки использовать? Во-вторых, каковы варианты, если таковые имеются? Об этом мы поговорим чуть позже. А также, какие необходимы правила?

После того, как комиссия по выработке правил завершит свою работу, результаты публикуются для общественного обсуждения. А затем, с учетом комментариев, поступивших во время общественного обсуждения, они передаются комиссиям по интеграции правил. При этом комиссии по интеграции правил обеспечивают надлежащую интеграцию всех наборов символов.

Кроме того, комиссии по выработке правил и комиссии по интеграции правил могут взаимодействовать друг с другом, если возникает необходимость дополнительного обсуждения. После завершения интеграции эти правила начинают действовать в корневой зоне. Итак, LGR для корневой зоны интегрируются поэтапно. Сейчас мы занимаемся четвертой версией LGR для корневой зоны, которая охватывает 18 наборов символов.

Вот этапы, на которых находится каждая комиссия по выработке правил. У тех, которые уже интегрированы, сверху будет вот такая темная стрелка. Как видите, многие комиссии по выработке правил для других алфавитов очень скоро завершат работу. Конечно, есть

два алфавита, для которых еще предстоит сформировать комиссии. Но в целом этот проект уже достаточно близок к финальной стадии. Сегодня мы заслушаем отчеты комиссий по выработке правил.

Переходим к следующему проекту, TLD с вариантными IDN-доменами. «Вариант» — это то, что может быть воспринято пользователями алфавита одинаково. Однако определение одинаковости в разных сообществах различается, то есть одно и то же может иметь два разных значения. Так что, во-первых, это может затрагивать вопросы безопасности. Во-вторых, удобство пользования. Позвольте привести несколько примеров.

На этом слайде вы видите набранное шрифтом синего цвета слово «erіc» и еще одно точно-такое же слово. Человеческим глазом это воспринимается одинаково. Но за всем этим, в этой четырехзначной кодовой точке есть то, что распознают компьютеры. По сути, для компьютера эти две строки представляют два разных имени. И это может создать проблемы безопасности для пользователей, когда те захотят попасть в какое-то место, но окажутся на другом сайте. К примеру, это может создать проблемы, связанные с фишингом.

Поэтому соответствующее сообщество (в данном случае это комиссия по выработке правил для латиницы и комиссия по выработке правил для кириллицы) должно собраться и понять, что символ E здесь и похожий на него символ E здесь, являются вариантами, чтобы предотвратить проблемы с безопасностью для пользователей.

Второй аспект вариантов — удобство пользования. Например, вот китайские иероглифы. Вторые немного отличаются от первых,

потому что в китайском языке есть две версии — традиционная и упрощенная. Таким образом, представители китайского сообщества поймут, что эти два слова могут означать одно и то же, но в других странах это совсем другой иероглиф.

Поэтому, если вы владелец товарного знака, то хотели бы иметь возможность использовать эти два имени одновременно, чтобы клиент мог попасть на ваш сайт. В данном случае китайское сообщество должно собраться и определить, что это два варианта, и они необходимы для удобства пользования.

Вот похожий случай для арабского языка. Как видите, последние символы немного отличаются, но арабские пользователи во всей стране поймут, что это одно и то же. Это работа, которую должна проделать комиссия по выработке правил.

Итак, для каких алфавитов требуется работа над вариантами кодовых точек? В основном, как видите, вариантов нет только в четырех. В большинстве алфавитов есть проблемы с вариантами, которые необходимо обдумать. Так что это не простой вопрос. И работа комиссий по выработке правил очень важна.

Далее, как только у нас будет определение того, что такое варианты, встает следующий вопрос: как ими управлять. Для управления вариантами корпорация ICANN разработала рекомендацию, работа над которой завершилась в начале 2019 года и впоследствии, в марте 2019 года, она была принята Правлением. Правление просит ccNSO и GNSO учитывать эту рекомендацию при разработке соответствующей политики.

По состоянию на сегодняшний день и ccNSO, и GNSO имеют в своем составе рабочую группу по IDN-доменам. У ccNSO есть PDP по выбору и отмене выбора строк IDN-доменов ccTLD, в рамках которого также учитывается эта рекомендация в отношении вариантов.

У GNSO есть проект итогового отчета по SubPro, в котором также учтены некоторые из рекомендаций. Кроме того, в настоящее время начинается EPDP в области IDN-доменов, который находится на стадии разработки устава группы и также учитывает эту рекомендацию.

Что касается следующих проектов для доменов верхнего уровня, ICANN также занимается реализацией ускоренной процедуры рассмотрения заявок на регистрацию IDN-доменов ccTLD в части оценки строк. На сегодняшний день мы успешно оценили 62 строки из 43 стран и территорий, при этом у некоторых стран может быть несколько строк, исходя количества государственных языков.

Далее, что касается второго уровня, Руководство по внедрению IDN-доменов — это руководящие принципы, которые разрабатываем мы... разрабатывает сообщество, и цель которых заключается в уменьшении риска киберсквотинга и введения потребителей в заблуждение. Руководящие принципы применяются к регистратуре, предлагающей IDN-домены второго уровня. Эти руководящие принципы время от времени пересматриваются, и последней версией стала четвертая. Она была разработана и опубликована в 2018 году.

Позднее в 2019 году GNSO обратилась к Правлению с просьбой предоставить дополнительное время для изучения руководящих принципов, прежде чем вводить их в действие. В настоящее

время GNSO организует оперативную процедуру проверки этих руководящих принципов.

Далее, что касается проектов создания базовых LGR для второго уровня, это является продолжением на втором уровне. Таким образом, если регистратура хочет предложить метки IDN-доменов второго уровня, ей нужно отправить таблицы IDN в ICANN, а затем мы рассмотрим их на предмет проблем, связанных с безопасностью и стабильностью, прежде чем они попадут в репозиторий IANA. Это напоминает... это тоже правила генерирования меток, но для второго уровня.

То есть мы разрабатываем справочную таблицу IDN, чтобы регистратура могла использовать ее как основу. Они могут использовать ее напрямую, если захотят, или взять за основу при принятии решения по таблице IDN.

На данный момент мы опубликовали 42 справочные таблицы IDN. Многие из них основаны на наборах символов, а значит, могут охватывать гораздо больше языков, использующих определенные наборы символов, а также есть таблицы для отдельных языков. В настоящее время еще четыре таблицы находятся на этапе общественного обсуждения. Собственно, сегодня оно завершается. Это арабский алфавит, алфавит иврита, сингальский алфавит и язык иврит. Так что если кто-то из вас интересуется этими языками и хочет взглянуть, пожалуйста, загляните туда и оставьте свой комментарий.

Кроме того, что касается последнего проекта, это идет вместе со справочными LGR. Так что, если регистратура захочет использовать таблицу IDN, она не обязательно должна полностью совпадать со справочной. Могут быть отклонения, но необходимо обеспечить, чтобы

отклонения не вызывали проблем с точки зрения безопасности. Поэтому мы предоставляем инструмент, который призван в этом помочь. Функциональность инструмента заключается в том, что регистратура или обычный пользователь может загрузить таблицы IDN, а затем выбрать справочные LGR для сопоставления, и инструмент будет сравнивать и выявлять места, где следует провести проверку на наличие проблем с безопасностью. Мы надеемся, что это поможет сократить время проверки и повысить прозрачность и единообразие процесса. Инструмент будет доступен в апреле.

Итак, и это очень краткая информация о программах создания IDN-доменов. Давайте посмотрим, появились ли вопросы.

САРМАД ХУСЕЙН (SARMAD HUSSAIN): Здравствуйте! Сейчас в чате нет комментариев и вопросов. Спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Ладно. Спасибо. Тогда позвольте перейти к следующей части. Я хотела бы попросить комиссии по выработке правил представить свои отчеты. Итак, представитель первой комиссии, которая занимается греческим языком, вам слово. И, пожалуйста, говорите медленно и близко к микрофону. Спасибо.

ПАНАГИОТИС ПАПАСПИЛИОПУЛОС: Спасибо, Питинан. Рад видеть вас снова, здравствуйте, Сармад, здравствуйте, дорогие коллеги, друзья и переводчики. Меня зовут Панагиотис Папаспилиопулос (Panagiotis Papaspiliopoulos).

Я председатель комиссии по выработке правил для греческого языка. Это группа, которая занимается греческим алфавитом. Он используется в греческом языке, на котором говорят в Греции и на Кипре.

Наша комиссия была сформирована почти пять лет назад. В самом начале нам предстояло многое урегулировать. Следующий слайд, пожалуйста. Это ключевые моменты, которые я сегодня затрону. Следующий слайд, пожалуйста. Здесь вы видите информацию о греческом алфавите и греческом языке, мы используем четвертую версию MSR. Как я уже сказал, греческий алфавит используется для греческого языка, который является официальным языком в Греции и на Кипре. Следующий слайд, пожалуйста.

Поэтому в комиссию входят представители обеих стран. Я — председатель, г-н Сегредакис — заместитель председателя. Он представляет регистратуру. Члены комиссии представляют различные заинтересованные стороны. Среди нас есть представители государственных органов, регулирующих органов, органов стандартизации, специалисты в области лингвистики из обеих стран. Мы также оформили это официально, потому что работаем под эгидой компетентного Министерства цифрового управления Греции. Следующий слайд, пожалуйста.

Как я уже говорил, комиссия сформирована около четырех с половиной лет назад, и нам пришлось урегулировать множество моментов. Прежде всего, мы должны были убедиться, что греческий алфавит не используется ни в каких других языках. Мы провели исследование, связанное с языками, используемыми в сообществах,

которые раньше жили на территории Греции, но на самом деле это были [мертвые] языки, если можно так выразиться.

Существует еще один живой язык. Это [невнятно] язык. Это другой язык. Была попытка использовать буквы греческого алфавита для создания письменности, но после обсуждения с одним из исследователей этого языка, комиссия решила, что нет необходимости изменять уже полученные нами результаты или специально под него формулировать конкретные правила.

Проведя это исследование, мы начали анализировать, скажем так, наборы символов, связанные с греческим языком, такие как латиница, кириллица и армянский алфавит. Греческий язык — старый, древний язык, и греческая цивилизация оказала влияние на население всех стран, расположенных по соседству с Грецией. Поэтому нет ничего удивительного, что в латинице, кириллице и армянском алфавите оказались некоторые буквы, заимствованные из греческого алфавита, или буквы, очень похожие на греческие.

Так что нам пришлось проанализировать эти случаи, чтобы выявить все варианты пересечения и сходства алфавитов, или, если можно так выразиться, сходства в той степени, которая заставляет считать буквы вариантами, встречающимися в разных алфавитах. И, конечно же, в греческом языке мы используем знак ударения. Например, меня зовут Панагиотис. Я ставлю знак ударения над О. Есть случаи, когда ударение необходимо, чтобы различить значение слова, например, Афины — столица Греции, а Афина — древняя богиня мудрости, поэтому знак ударения на самом деле имеет значение. И поэтому мы пришли к

выводу, что есть варианты внутри набора символов. Скажем, основная форма буквы и форма со знаком ударения.

А еще у нас есть финальная сигма, то есть на самом деле у нас две формы сигмы, одна из которых может использоваться в середине метки, слова, или как финальная. После детального анализа в течение всех этих лет любых возможных ситуаций, связанных с латиницей, кириллицей и армянским алфавитом, мы предложили несколько версий. Что касается последней из них, мы провели обсуждения и при ценной поддержке со стороны Сармада и Питинан пришли к выводу, что, особенно для корневой зоны, можно задействовать механизм, ограничивающий присваиваемые варианты меток двумя, применяемыми для метки и базовой формы. Это означает, что в случае с моим именем, если бы это была метка, то это мог быть «Панагиотис» с ударением на О, но также «Панагиотис» без знака ударения.

Мы решили применить этот механизм, и у нас есть исключение для финальной сигмы. Так что максимальное количество присваиваемых вариантов исходной метки — четыре, если есть знак ударения и финальная сигма. Чтобы охватить случаи сходства и фактически поддержать любую будущую работу, в последней версии нашего отчета также есть приложение, указывающее на случаи подобного сходства, просто на будущее.

Итак, мы считаем, что добились большого прогресса. Это была нелегкая работа. Честно говоря, это можно было сделать немного быстрее, но мы считаем, что эффективно использовали время и выдали качественный результат и качественное предложение

комиссии по выработке правил, которое можно объединить с предложениями по латинице, кириллице и армянскому алфавиту.

Я вас долго не задержу. Это план. И он более или менее соответствует тому плану, который уже показала Питинан. Мы уверены, что добьемся всех поставленных целей в соответствии с графиком. Спасибо за то, что уделили время. Я готов ответить на любые ваши вопросы и замечания. Большое спасибо. Спасибо, Питинан.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Панагиотис. Итак, перейдем к следующей комиссии по выработке правил. Профессор Ким Кён Сок, вам слово. Спасибо.

КИМ КЁН СОК:

Спасибо. Меня зовут Ким Кён Сок, я представляю комиссию по выработке правил для корейского алфавита. В Корее сейчас раннее утро. Приятно с вами познакомиться. Спасибо. Я расскажу о корейских LGR. Следующий слайд, пожалуйста.

Итак, это план доклада. Дальше, пожалуйста. Корейские LGR охватывают набор символов корейского языка, то есть хангыль плюс ханча. Под корейским алфавитом обычно понимается хангыль. Однако в контексте корейских LGR набор символов корейского языка объединяет два элемента: хангыль и ханча.

У корейского языка долгая история, охватывающая более 2000 лет. Алфавит хангыль изобретен около 600 лет назад. Иероглифы ханча использовались до появления хангыля. Иероглифы ханча до сих пор используются в Республике Корея.

Корейский язык в основном используется в Республике Корея и Корейской Народно-Демократической Республике. На нем также говорят люди за пределами двух Корей. Следующий слайд, пожалуйста.

Это членский состав. Я это пропущу. А вот краткая информация о ходе работы. Версия 2.1 корейских LGR, у нас было около 11 000 слогов хангыля и ни одной группы вариантов среди них. А что касается ханча, то здесь имеется 4758 символов и 37 групп вариантов до расширения. После расширения у нас есть 283 группы вариантов. Среди слогов хангыля и символов ханча у нас семь групп вариантов. А среди иероглифов ханча после расширения у нас было 289 групп вариантов и показанный здесь набор из 4758 иероглифов ханча. По сути, это объединение двух наборов, один — KS X 1001, а другой — IICORE в ISO/IEC 10646. Следующий слайд, пожалуйста.

В 2018 году состоялось первое общественное обсуждение, и основной вопрос состоял в том, разрешить или не разрешать смешанные метки хангиль-ханча. При этом поступили как негативные, так и положительные комментарии. Следующий слайд, пожалуйста.

Итак, последний пункт. Мы обсудили, следует ли разрешить использование смешанных меток хангиль-ханча. С метками исключительно на хангыле и исключительно на ханча проблем не было. Так что комиссия по выработке правил для корейского алфавита (KGP) подтвердила, что достигнуто общее согласие разрешить использование меток исключительно на хангыле и

исключительно на ханча, но решила не разрешать использование смешанных меток хангиль-ханча. Следующий слайд, пожалуйста.

Затем мы сформировали четыре версии политики, 2.0, 2.1, 2.2 и 2.3. 2.3 — последняя, и в настоящее время мы ждем комментариев общественности. Следующий слайд, пожалуйста. Здесь указано, что ожидаются комментарии от комиссии по интеграции правил в отношении версии 2.3, но по сути комментарии уже получены и версия 2.3 K-LGR опубликована для второго раунда общественного обсуждения 2 марта. Он завершится в апреле. Следующий слайд, пожалуйста.

Это краткая история деятельности KGP. Она была организована в 2013 году, почти восемь лет назад, в 2015 году, мы опубликовали версию 0.1 K-LGR. В 2016 году корейское сообщество официально сформировало комиссию по выработке правил генерирования меток корневой зоны. В декабре 2017 года мы опубликовали версию 1.0, а в 2018 году, с января по март, проходил период общественного обсуждения версии 1.0 K-LGR, и мы провели порядка 39 заседаний KGP и несколько координационных заседаний CJK, то есть Кореи, Китая и Японии, на открытых конференциях ICANN. А также в Республике Корея, Китае и на Тайване. Следующий слайд, пожалуйста.

Это план, и как вы видите, в марте 2021 года у нас проходит второй раунд общественного обсуждения в отношении K-LGR 2.3, и если все пойдет хорошо, K-LGR будут интегрированы в очередную версию LGR для корневой зоны. Следующий слайд, пожалуйста. Спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Ким. Теперь я хотела бы попросить Билла Джуриса рассказать о работе комиссии по выработке правил для латинского алфавита.

БИЛЛ ДЖУРИС:

Доброе утро, добрый день, добрый вечер. Меня зовут Билл Джурис. Наш председатель, Мирьяна Тасич (Mirjana Tasic), борется с ковидом и пневмонией, поэтому сегодня с презентацией выступлю я. Следующий слайд, пожалуйста. Большинство алфавитов используется для одного языка. В лаосском языке используется лаосский алфавит. В грузинском — грузинский. Есть несколько случаев, когда родственные языки могут использовать один и тот же набор символов — например, арабский и фарси. И несколько крайних случаев, кириллица в Восточной Европе и [невнятно] алфавит в Южной Азии, где полдюжины языков используют один и тот же набор символов.

А еще есть латиница. 1700 или 1800 лет назад существовала империя и религиозные миссионеры, распространявшие алфавит в Западной и Центральной Европе, а затем 500 лет назад новые империи и еще большее количество миссионеров распространили алфавит еще шире. Таким образом, сегодня на шести континентах насчитывается более 400 живых языков, использующих латиницу.

При первом расширении к исходной 21 букве латиницы было добавлено еще несколько букв, но в основном новые фонемы образовывались путем использования диакритических знаков. А при втором расширении было добавлено еще больше диакритических

знаков, так что теперь в том или ином языке используется около 20 различных диакритических знаков. Следующий слайд, пожалуйста.

Эти семь членов комиссии принимали участие в проекте. С самого начала все мы были как минимум поверхностно знакомы с большинством европейских языков, где используется латинский алфавит, что позволило нам заняться всеми этими вопросами. Следующий слайд, пожалуйста.

Как видите, мы разработали ряд критериев для определения того, какие языки следует включить в созданный нами набор. Мы включили все языки, имеющие статус официальных на уровне государства, провинции, штата или внутреннего региона страны, плюс мы добавили языки, имеющие 5 баллов по шкале EGIDS, иными словами, не являющиеся официальными, а также те, на которых говорит более миллиона человек. И в итоге у нас получилось чуть больше 200 языков.

Мы занимались этим в течение первых двух лет проекта, но, наверное, стоит отметить, что только в сентябре прошлого года мы обнаружили еще пару языков, которые тоже соответствуют указанным критериям и которые мы изначально упустили из виду, так что мы не удивимся, если в ходе общественного обсуждения найдется еще парочка.

Хотя сейчас в наборе латинских символов больше 200 кодовых точек, в любом конкретном языке используется менее четверти из них. Часто гораздо меньше четверти, а это значит, что пользователи никогда не сталкивались со всем множеством возможных кодовых точек латинского алфавита.

Как уже отмечалось, у нас есть около 20 используемых диакритических знаков, но в любом конкретном языке используется лишь несколько из них, и, как правило, если пользователь знает полдюжины различных языков, то может быть знаком практически с половиной диакритических знаков.

Еще интереснее, что в латинице также широко используются различные шрифты, поэтому пользователи привыкли не обращать внимания на небольшие различия в форме букв. Это значит, что если кто-то ожидает увидеть конкретный диакритический знак или букву, а увидит нечто другое, не встречавшееся ему раньше, скорее всего, он предположит, что это просто другой шрифт, и не поймет, что на самом деле эта разница важна. Так что здесь много возможностей для путаницы.

Комиссия по выработке правил для латиницы использовала очень узкое определение варианта, исходя из предположения, что для новых TLD комиссия по анализу схожести строк выявит все противоречия. Несмотря на это, вариантов оказалось гораздо больше, чем в большинстве алфавитных наборов символов, и у нас есть еще большее количество того, что мы назвали «неоднозначностями» — это пары кодовых точек, которые некоторые члены комиссии (даже в некоторых случаях большинство членов комиссии) не сумели отличить друг от друга, даже напрямую сравнивая их. При этом такие пары не вполне соответствовали нашему очень строгому определению варианта, поэтому их просто включили в виде приложения.

Хочу отметить, что, как уже упоминала Питинан, недавно опубликованы руководящие принципы, касающиеся доменных имен второго уровня, где говорится о вариантах и о том, как с ними работать, но для доменных имен второго уровня ручная проверка, включенная в процесс создания TLD, не представляется возможной. В рамках проекта создания IDN-доменов, возможно, потребуется серьезно рассмотреть вопрос о том, следует ли считать выявленные нами неоднозначные символы вариантами для целей SLD. Как вариант, мы, конечно, можем просто решить, что неправильное использование DNS не особо нас волнует, и проигнорировать эту проблему. Следующий слайд, пожалуйста.

Мы проработали несколько версий нашего документа и каждый раз получали комментарии от комиссии по интеграции правил. Последний раз — в декабре прошлого года. Следующий слайд, пожалуйста. Сейчас мы почти закончили отвечать на комментарии комиссии по интеграции правил и собирать тестовые наборы данных. Мы надеемся, что эти материалы будут готовы для передачи комиссии по интеграции правил в конце текущего или в начале следующего месяца. Следующий слайд, пожалуйста.

То есть мы надеемся передать это комиссии по интеграции правил в апреле. Если предположить, что будет поднято не очень много существенных вопросов, то мы оптимистически надеемся в начале июня опубликовать отчет для общественного обсуждения. Если в процессе общественного обсуждения будет поднято мало существенных вопросов, то мы надеемся, что будем готовы опубликовать итоговый документ в сентябре. Но это, конечно, в лучшем случае. Я помню, как однажды комиссия по выработке

правил для корейского алфавита провела общественное обсуждение, а потом вернулась назад и занялась разработкой новой версии, прежде чем вновь вынести свои результаты на общественное обсуждение. И поэтому вполне возможно, что у нас будет так же. Думаю, все члены комиссии надеются и уповают на то, чтобы до такого дело не дошло, но это не исключено.

Если есть вопросы, пожалуйста, дайте мне знать. Спасибо.

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

САРМАД ХУСЕЙН:

Да. Спасибо, Питинан. Спасибо, Билл. В чате один вопрос. Вопрос задает Шивасубраманиан. Вопрос в том, по каким сценариям владелец домена подает заявку на доменное имя с пересечением алфавитов? Выбирает ли владелец домена четыре алфавита с армянской клавиатуры, покупает греческую клавиатуру, выбирает три алфавита с греческой и собирает доменное имя для регистрации? Пожалуйста, объясните, почему так много внимания уделяется дополнительным вариантам в наборах символов.

БИЛЛ ДЖУРИС:

Не знаю, как ответить на первую часть. Причина внимания к вариантам пересечения алфавитов, или вообще к вариантам, как объясняла ранее Питинан, заключается в том, что если пользователи не видят разницу между символами, используемыми в доменном имени, это вызывает проблемы с безопасностью, это вызывает

проблемы с удобством пользования, и именно поэтому мы уделяем этому так много внимания.

На ту часть вопроса, которая касается процедуры, я не знаю, как ответить. Возможно, Сармад поможет.

САРМАД ХУСЕЙН: Спасибо, Билл. Думаю, вы ответили на вопрос. Больше вопросов в чате я не вижу. Спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Сармад. Хорошо. Спасибо, Билл. Перейдем к следующей комиссии по выработке правил. Инь Мэй У, пожалуйста, вам слово. Пожалуйста, говорите чуть громче и медленнее. Спасибо.

ИНЬ МЭЙ У: Приветствую всех. Это Инь Мэй У из комиссии по выработке правил для мьянманского алфавита, я сопредседатель комиссии. Наш председатель в настоящее время находится в Мьянме, и во время блокировки интернета она не может присутствовать на заседании. Поэтому презентацию продолжу я. Следующий слайд, пожалуйста.

Это сегодняшняя повестка дня. Следующий слайд, пожалуйста. Итак, в комиссии по выработке правил для мьянманского алфавита мы занимаемся мьянманским набором символов. Мьянманский набор символов используется во многих языках, в основном в языках, на которых говорят пользователи IDN-доменов в Мьянме, но в мире есть и много других людей, использующих эти символы. При этом большинство использует бирманский язык, так как он является

государственным языком в течение многих лет, и у нас есть шанский язык, араканский, каренские языки сго, мон и пао. Так что этим набором символом пользуется большинство этнических групп, и у всех этих языков больше 5 баллов по шкале EGIDS. Следующий слайд, пожалуйста.

Как видите, один и тот же набор символов у нас используется во многих языках. Географически наша страна расположена в восточной части Бенгальского залива и вдоль реки, мы начали использовать эту письменность с [1811-1812 года] и с тех пор она претерпела развитие. Таким образом, наш язык, наша письменность, основаны на деванагари, и мы адаптировали иероглифы племен [хмонги], которые жили в районе Андаманского моря в 11 веке. Мы испытываем влияние как северных, так и южных племен.

Живя вдоль реки Иравади, мы на протяжении многих лет развивали культуру и язык. И около 600 лет назад мы усовершенствовали использование языка. Сначала мы просто записывали [первый и второй язык], а затем приспособились к использованию большего количества диакритических знаков в [невнятно] эпоху, примерно в [1718] году. После этого мы приноровились к разговорному бирманскому.

Таким образом, у нас много языков, но большинство символов, согласных букв и диакритических знаков в них общие. Итак, в общей сложности, есть 99 кодовых точек, которые мы использовали в MSR. Следующий слайд, пожалуйста.

Что касается природы мьянманского набора символов, в большинстве языков девангари одна согласная, и она может иметь

диакритический знак спереди, как [невнятно] слева прямо за центральной согласной. Но у нашего языка есть еще одна особенность: мы можем использовать другую согласную в качестве [усиления гласной], как [невнятно], так что мы можем усилить звук, а можем также добавить больше диакритических знаков, как [невнятно]. Таким образом, мы можем взять два или более [невнятно], чтобы произнести одно и то же, а поскольку у нас гораздо более сложные диакритические знаки, то должны позаботиться о правилах.

И кроме того, у нас есть [сцепление слогов]. Это означает, что одно слово может быть произнесено по-разному. Такое письмо существовало до [1180] года, но некоторые слова мы используем и сейчас, так что это нельзя игнорировать. Поэтому нам приходится рассматривать все письменности с точки зрения разных языков и адаптировать все в одном [невнятно]. Следующий слайд, пожалуйста.

Итак, мы проанализировали и рассмотрели многое для вариантов в наборе символов. У нас есть односимвольные комбинации, которые выглядят так же, как трехсимвольные, как вы видите в таблице. И мы проанализировали весь набор символов и все языки. Нам нужно рассмотреть, какие могут быть варианты в наборе символов.

Что касается вариантов, строка номер четыре, пять, шесть, семь, помимо того, что это омоглифы, они также могут быть омофонами и иметь одинаковое смысловое значение. Поэтому мы добавили в правила, что эти символы не должны присутствовать в одной метке, чтобы [невнятно] или анализ случаев пересечения алфавитов. Следующий слайд, пожалуйста.

У нас более трех языков, где есть один и тот же символ, поэтому с первой до шестой строки это [невнятно], которые мы проанализировали, и есть аналогичные символы в грузинском алфавите. Кроме того, у нас есть письменность ория, и есть диакритические знаки ория. Мы рассматриваем это, потому что большинство символов мьянманского алфавита имеют округлую форму, округлую каллиграфию, поэтому, проводя сравнение с другими наборами символов, мы просто перечисляем возможные комбинации, которые могут пересекаться с мьянманским алфавитом.

Итак, это у нас пересечение алфавитов, а это результат, полученный нами после проверки всех языков, которые могут иметь сходство [невнятно] с мьянманским, а остальные просто указаны как неоднозначности. Следующий слайд, пожалуйста.

Это просто более широкий взгляд на то, что мы называем неоднозначностями, поскольку местные жители могут [оценивать] эти символы одинаково, так как разница очень мала. Следующий слайд, пожалуйста.

Итак, это информация о том, как обстоят дела на данный момент. Наш комиссия по выработке правил сформирована в июне 2018 года, последний отчет до этого заседания был подготовлен в 2020 году, так что мы надеемся опубликовать правила для общественного обсуждения в середине 2021 года, например, к концу мая, и завершить создание LGR до конца 2021 года, будем надеяться, в сентябре. Большое спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Инь Мэй. У нас есть несколько минут до конца заседания, перейдем к вопросам и ответам. Есть вопросы или комментарии, Сармад?

САРМАД ХУСЕЙН: На этот раз вопросов в очереди нет.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо. Я видела в чате активное обсуждение, так что если кто-то хотел выступить [невнятно] возможно. Ладно. Не вижу, чтобы у нас были вопросы или комментарии для обсуждения на этом заседании, так что если у вас появятся какие-либо дополнительные вопросы или комментарии, не стесняйтесь обращаться к нам. Адрес электронной почты idnprogram@icann.org, мы ответим на ваш вопрос.

Итак, я полагаю, что выступили все. У вас есть заключительные замечания перед тем, как мы завершим эту телеконференцию? Хорошо. Поскольку все молчат, думаю, что телеконференцию можно закрыть. Спасибо за участие в заседании и за ваше внимание, а также за дискуссию. Теперь можно остановить запись. Всем хорошего дня и, пожалуйста, будьте осторожны. До свидания.

[КОНЕЦ СТЕНОГРАММЫ]