

---

ICANN72 | Неделя подготовки — последние сведения о правилах генерирования меток корневой зоны  
Четверг, 14 октября 2021 года, 11:00–12:00 по тихоокеанскому летнему времени (PDT)

ПИТИНАН КООАРМОРНПАТАНА (PITINAN KOOARMORNPATANA): Спасибо. Доброе утро, добрый день, добрый вечер всем! Приветствуем вас на заседании, посвященном информации о правилах генерирования меток корневой зоны, которое проводится в рамках недели подготовки к конференции ICANN72. Сегодня у нас такая повестка дня. Я просто кратко, на пять минут, расскажу о предыстории проекта правил генерирования меток корневой зоны, а затем мы перейдем к ключевым докладчикам нашего сегодняшнего заседания. У нас будет Матс Дуфберг (Mats Dufberg) от членов комиссии по выработке правил для латинского алфавита, Хиро Хотта (Hiro Hotta) из Японии, он председатель японской комиссии по выработке правил, а затем Мишель Суиньяр (Michel Suignard) из комиссии по выработке правил, который расскажет нам о следующей версии правил генерирования меток для корневой зоны.

Давайте я перейду к нашему первому пункту, это общие сведения о правилах генерирования меток корневой зоны. Итак, краткая предыстория этого проекта. Году примерно в 2010 или около того сообщество ICANN определило, что существует необходимость в вариантных доменах верхнего уровня, и исходя из этого сообщество организовало работу над комплексным отчетом о

---

**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**

---

проблемах. На то время задача заключалась в том, чтобы оценить возможные проблемы, которые могли бы возникнуть в случае реализации вариантных доменов в корневой зоне. То есть тогда это было на самом деле изучение проблемы, а не выработка решения.

В том отчете перечислялись шесть систем письма — арабская, китайская, кириллица, деванагари, а также греческий и латинский алфавит, которые вместе покрывали множество разных наборов символов, используемых в современных языках.

Затем этот отчет был подготовлен и опубликован в 2012 году, и в нем также говорилось о необходимости предусмотреть механизм, который позволял бы определять варианты написания для каждого набора символов. При этом такой механизм должен определяться из единого источника, потому что в конечном итоге все домены верхнего уровня попадают в корневую зону.

Сообщество определило, что в качестве такого механизма должны быть определены правила генерирования меток корневой зоны, а предложение [разработать] правила генерирования меток для корневой зоны было одобрено Правлением в 2013 году. За прошедшее с тех пор время сообщество сформировало комиссии по выработке правил и провело работу над правилами генерирования меток для соответствующих наборов символов.

Параллельно с этим корпорация ICANN также разработала рекомендации в отношении управления вариантными IDN-

---

доменами верхнего уровня. В 2019 году эти рекомендации были одобрены Правлением. Оно постановило, что GNSO и ccNSO должны использовать эти рекомендации для разработки политики и [неразборчиво] интеграции использования правил генерирования меток корневой зоны.

Затем, в 2020 году, Правление ICANN также приняло резолюцию, согласно которой GNSO и ccNSO должны учесть рекомендацию о техническом использовании правил генерирования меток корневой зоны, а в 2021 году была принята еще одна резолюция, согласно которой GNSO должна опубликовать отчет о последующих процедурах, применимых к новым gTLD, который должен также охватывать вопросы правил генерирования меток корневой зоны и следующий раунд ввода новых gTLD.

Итак, как работают правила генерирования меток? Правила генерирования меток, или LGR, представляют собой нечто вроде формулы, которая определяет, как именно должны формироваться метки. Поэтому они так и называются — правила генерирования меток. То есть когда мы интегрируем эти правила в корневую зону, они поддерживают машинное считывание, то есть специальный инструмент их читает, а затем мы можем автоматически определить, допустима ли та или иная метка или нет, и если она допустима, то можно также определить варианты ее записи, которые могут выделяться или организовываться поблочно.

---

Этот механизм, процедура использования правил генерирования меток, работает в два этапа. Первый этап — это работа комиссии по выработке правил, которая разрабатывает решение для того или иного набора символов, а затем, когда комиссии по выработке правил подготовят окончательные варианты решений для каждого набора символов, они передаются в комиссии по интеграции правил. Комиссия по интеграции правил рассматривает предложение комиссии по выработке правил, после чего эти две комиссии могут проводить множество раундов консультаций. И после того, как такое решение будет успешно рассмотрено и одобрено, оно будет интегрировано в корневую зону и его можно будет использовать для проверки допустимости меток доменов верхнего уровня.

Вот так в общих чертах выглядит работа комиссий по выработке правил. То есть для этих 28 наборов символов, которые у нас сейчас на экране, у нас есть 17 комиссий по выработке правил, которые на данный момент охватывают 26 наборов символов. Как вы можете видеть, эта работа была начата в 2014 году и продолжается уже много лет. На данный момент многие из комиссий по выработке правил уже закончили свою работу, сейчас еще продолжается процедура общественного обсуждения для японского и латинского набора символов. Мьянма сейчас еще только заканчивает работу над окончательной версией своего предложения, которое будет вынесено на общественное обсуждение.

---

Здесь показана временная шкала работы над правилами генерирования меток корневой зоны. Вот начиная с объявления о формировании комиссий в 2013 году, затем с 2014 года комиссии начинают активно работать. Со временем эта работа для того или иного набора символов заканчивается, правила окончательно формулируются и постепенно интегрируются в корневую зону.

На данный момент у нас действуют правила генерирования меток корневой зоны версии 4, которых охватывают в общей сложности 18 наборов символов, а в 2022 году мы рассчитываем закончить интеграцию еще восьми наборов символов, таким образом, всего их будет 26. Вот такова ситуация у нас на сегодняшний день, сегодня с нами здесь члены комиссий по выработке правил для латинского и японского наборов символов, которые расскажут нам о своих решениях, а также пригласят всех желающих принять участие в общественном обсуждении.

Итак, я передам слово Матсу Дуфбергу из комиссии по выработке правил для латинского алфавита. Вам слово.

МАТС ДУФБЕРГ:

Большое спасибо. Следующий слайд, пожалуйста. Итак, это то, что относится к моей презентации, а я буду говорить о том предложении, которое мы выносим на общественное обсуждение. То есть это основные составляющие того, о чем я буду говорить. Следующий слайд, пожалуйста.

---

Итак, наше предложение для латинского алфавита вынесено на общественное обсуждение, комментарии будут приниматься до 23 ноября этого года. Здесь есть ссылка на это предложение, мы приглашаем всех желающих перейти по этой ссылке и ознакомиться с нашим предложением, а мы, конечно же, учтем все комментарии, большие и маленькие. Мы рады всем. Следующий слайд, пожалуйста.

В ходе этой презентации я пройду по нашему предложению, и мы надеемся, что после этого вам будет проще ознакомиться с ним самостоятельно и предложить нам свои комментарии. Итак, в фокусе нашего внимания основной документ и его взаимозависимые элементы, а файл XML правил генерирования меток — это нормативный документ, я не буду подробно о нем говорить.

Итак, я не буду обсуждать здесь краткие главы с общей информацией, потому что они не заслуживают обсуждения. В главе 2 этого предложения определяются ограничения набора символов, о котором идет речь в этом предложении. Первое ограничение — это то, что это предложение не может включать символы, которые не входят в т. н. максимальный стартовый набор. Максимальный стартовый набор, или MSR — это подмножество допустимых кодовых точек протокола IDNA, который, в свою очередь, является подмножеством символов стандарта Unicode. То есть это набор MSR, таким образом,

---

является подмножеством символов Unicode, и мы обязаны этого придерживаться в нашем предложении.

Этот максимальный стартовый набор был определен комиссией по интеграции правил. То есть в процессе работы мы предложили добавить в максимальный стартовый набор несколько дополнительных символов, и комиссия по интеграции правил добавила эти символы в новую редакцию MSR. Максимальный стартовый набор — это рабочее подмножество символов Unicode, что означает, что его можно как увеличить, так и, я думаю, уменьшить.

Мы в процессе работы предложили добавить несколько символов и они были добавлены. В главе 4 описывается процесс нашей работы. Мы должны для нашего предложения выяснить, какие символы следует использовать. Мы решили использовать т. н. шкалу EGIDS, это ступенчатая шкала межпоколенческой утраты языка Фишмана, она объясняется на следующем слайде. Это шкала от нуля, то есть это уровень международных языков, до четырех, то есть уровня образования. Мы еще добавили уровень 5, развивающийся, это еще по меньшей мере 1 миллион носителей. То есть это языки. Следующий слайд, пожалуйста.

Здесь мы видим шкалу EGIDS, это расширенная ступенчатая шкала межпоколенческой утраты языка Фишмана, которую мы использовали. По этой шкале оценивались почти все языки мира. Поэтому мы использовали ее для выбора языков, которые мы

---

рассматриваем. То есть это от нуля до четырех, а затем еще пятый уровень, который мы использовали для языков с количеством носителей от 1 миллиона человек. Есть еще другие уровни, но они сюда не включены.

В приложении 6 приводится полный список всех выбранных языков. Точного числа у меня здесь нет, но это довольно много языков. И для каждого из включенных языков у нас есть свое имя. Это 212 языков, которые мы включили в наш список. В этом списке есть название языка, его код ISO, а также уровень EGIDS для этого языка.

Для каждого из языков в этом списке мы определили все символы, которые используются для его записи, так что мы пришли к списку всех символов, которые используются в этих языках. То есть идея заключалась в том, чтобы сделать возможным создавать домены верхнего уровня, используя слова из этих 212 языков. Набор символов для каждого языка [в этом] отчете не описан, но в главе 9 приводятся ссылки на каждый из этих языков. Были также определены кандидаты, то есть потенциальные варианты символов. Подробнее об этом позже. Следующий слайд, пожалуйста.

То есть предложение по набору символов для латинского алфавита основывается, естественно, на кодовых точках Unicode, потому что таково ограничение. В простейшем случае кодовая точка — это просто символ, например, буква А. Кроме того,



---

кодовая точка может означать символ-модификатор. Тогда она используется для обозначения какого-то символа вместе с другой кодовой точкой. На этом слайде приведен пример буквы J со знаком тильды, которая записывается с использованием комбинации из двух кодовых точек.

Во многих случаях есть заранее составленные кодовые точки из базового символа и диакритического знака, например, буква А с диакритическим знаком акут, которая приводится на этом слайде. То есть это всего одна кодовая точка. Принципы, согласно которым тот или иной символ, используемый в языке, включается или не включается в предлагаемый набор, описаны во введении к главе 5. Следующий слайд, пожалуйста.

В предложении для латинского алфавита перечислены 218 символов. 197 символов записываются с использованием одной кодовой точки, то есть это простая запись. 21 символ составляются из последовательности из двух или более кодовых точек, то есть это может быть основной символ плюс один или два знака. То есть для каждого символа существует кодовая точка, из которой он состоит, она может быть одна или их может быть несколько.

Язык или языки, в которых этот символ используется для записи, это не обязательно все языки, но достаточно языков, чтобы показать, что этот символ используется в выбранных языках. Кроме того, в этом списке приводится также ссылка на алфавит

---

или алфавиты, которые используются в этих языках. Для символов от а до Z у нас не указаны подтверждающие языки, потому что они включены по умолчанию. Следующий слайд, пожалуйста.

Итак, этот набор символов — это также одна из основных составляющих файла XML. Набор символов в главе 5 отсортирован по числовому значению кодовой точки, однако если вы захотите увидеть эти же символы, сгруппированные по начертанию, то есть, например, все символы А с различными диакритическими знаками, то такой список приведен в приложении С. То есть это тот же список, просто иначе отсортированный. Следующий слайд, пожалуйста.

В разделе 4 главы 5 мы перечисляем символы, которые не были включены. Это те символы, использование которых было подтверждено в по меньшей мере одном из выбранных языков, но которые нельзя включить в этот список, потому что они не входят в максимальный стартовый набор, или MSR. То есть MSR — это нечто вроде предварительного отбора, этот список определяет максимально возможный набор символов. Так что комиссия по выработке правил не может включать символы, которые не входят в максимальный стартовый набор.

Так что если вам попадетсa какая-то буква, которую, на ваш взгляд, стоило бы включить, но она не входит в набор символов MSR, то комиссия по выработке правил для латинского алфавита

---

ничем здесь помочь не сможет. Сначала нужно будет изменить максимальный стартовый набор. Следующий слайд, пожалуйста.

Итак, я говорил о наборе символов. Следующая часть посвящена наборам вариантов, которые составляют важную часть этого предложения. Этот вопрос раскрыт в главе 6. Итак, что такое вариант? Набор вариантов состоит из нескольких символов, которые по тем или иным причинам считаются одинаковыми и воспринимаются одинаково. Они могут совпадать по форме или использоваться как взаимозаменяемые в сообществе, использующем соответствующий набор символов, или в какой-то его части.

То есть если у вас есть набор вариантов, то вы не можете использовать эти символы независимо друг от друга. В большинстве случаев, если используется один из этих символов, то другие в этом же контексте блокируются. То есть для латинского алфавита большинство правил использования вариантов символов подразумевают блокирование остальных вариантов.

Наборы вариантов в пределах того или иного набора символов определяет сообщество пользователей, использующих данный набор символов, то есть в данном случае латинский алфавит, однако у нас есть также наборы вариантов, совпадающих между разными наборами символов, в основном это латинский и греческий алфавит, а также кириллица, а некоторые наборы

---

вариантов содержат совпадения как внутри одного, так и из нескольких наборов символов. Следующий слайд, пожалуйста.

Итак, в главе 6 и приложениях к ней от D1 до D9 описаны принципы определения наборов вариантов, а также анализ данных таких наборов вариантов и наборы-кандидаты. То есть вы можете изучить эту информацию, чтобы понять, почему мы пришли к тем выводам, к которым мы пришли в отношении этих вариантов символов.

Итак, два набора вариантов особенные. Они относятся к более старой версии IDNA от 2003 года, они также позволяют назначать другие варианты символов. И эти два набора относятся к... один из них — это символы эсцет (ß) и двойное s (ss), а другой — l с точкой и без точки. То есть эти два набора вариантов в этом предложении являются особенными. Следующий слайд, пожалуйста.

И наконец, в приложении E у нас приведен список визуально... наборов символов, визуально похожих до степени смешения. То есть это не варианты и мы не предлагаем их в качестве вариантов. Они отсутствуют в официальном файле XML правил генерирования меток и приводятся в приложении только для примера, который можно использовать при анализе визуальной схожести между разными наборами символов. Спасибо. Да, это процедура, сейчас мы вынесли это на общественное обсуждение, а итоговую версию мы надеемся представить в январе следующего года. Так что прошу вас, ознакомьтесь с этим документом и

---

подавайте свои комментарии к нашему предложению. Я не знаю, как у вас принято отвечать на вопросы.

**ПИТИНАН КООАРМОРНПАТАНА:** Есть вопросы в чате? Или же, если кто-то хочет высказаться, вы можете также поднять руку. Хорошо, я не вижу желающих, так что давайте сейчас продолжим, а если впоследствии у вас будут вопросы к Матсу или к группе по латинскому алфавиту, то мы сможем еще к этому вернуться. Следующим я хочу пригласить выступить г-на Хотту, который расскажет нам о правилах генерирования меток корневой зоны для символов японского языка. Хотта-сан, вам слово, прошу вас.

**ХИРО ХОТТА (HIRO HOTTA):** Спасибо, Питинан. Приветствую всех. У меня не очень устойчивое подключение. Прошу прощения, если связь прервется. Но, если что, я смогу продолжить через пару минут, я надеюсь. Хорошо Следующий слайд, пожалуйста.

Это члены комиссии по выработке правил для символов японского языка. Аналогично тому, как это бывает в других комиссиях по выработке правил, у нас есть члены с разными [взглядами]. Следующий слайд, пожалуйста.

Это общие сведения о символов японского языка и о самом языке. Уникальной, как мне кажется, особенностью японского языка является то, что в нем используются сразу три разных набора

---

символов, плюс еще символы ASCII и цифры, и эти три набора символов — это кандзи, хирагана и катакана. Эти символы могут сочетаться в любом порядке в пределах одного слова. Символы, к которым применяются эти правила генерирования меток, определены в японском промышленном стандарте JIS уровней 1 и 2, так что мы выбрали [их] в качестве набора символов для правил генерирования меток для японского языка.

Поскольку символов кандзи очень много, в наш набор входит около 6000 символов. Это, конечно же, максимальный стартовый набор, или MSR. Итак, у нас есть три набора символов. Один из них, кандзи, используется также в китайском и корейском языках. Это создает некоторые сложности при определении вариантов. Я объясню это позже.

Помимо кандзи у нас есть еще хирагана и катакана. Хирагана используется в основном для записи суффиксов, дополняющих полное значение слова, для наречий и союзов, а также для записи тех символов кандзи, которые могут вызывать затруднение, для упрощения их записи и чтения. К примеру, для детей дошкольного возраста и т. п.

Катакана используется в основном для другого рода слов. Катакана в основном используется для записи длинных слов, заимствованных из других языков, а также звукоподражательных слов.

---

В основном все японские символы считаются независимыми, однако в китайском и корейском языках — как я уже сказал, некоторые символы кандзи используются одновременно в китайском, корейском и японском языках — некоторые наборы символов кандзи считаются вариантами, если у нескольких символов совпадает название и прочтение.

И некоторые члены сообщества пользователей японского языка убеждены, что некоторые варианты символов китайского и корейского языков следует считать вариантами также и в японском языке. Следующий слайд, пожалуйста.

Это общие сведения о правилах генерирования меток корневой зоны для символов японского языка. Максимальный набор, как я уже сказал, включает 6000 символов, а варианты — в японском языке нет собственных вариантов, то есть тех, у которых было бы одинаковое название и прочтение. А по критерию визуальной схожести были определены 12 наборов вариантов, и это уникальный момент, потому что мы включили варианты кандзи, определенные в правилах генерирования меток для китайского и корейского языков. Думаю, несколько сотен таких вариантов мы позаимствовали из правил генерирования меток для китайского и корейского языков.

Это второй важный момент — не могут выделяться никакие варианты меток, кроме исходной метки. Это вопрос, в отношении

---

которого было труднее всего принять решение для японского языка, но в конечном итоге мы решили запретить варианты меток.

И еще правило WLE, это очень простое специфически японское правило. Метка не должна начинаться с маленьких знаков кана, а также знаков повтора или удлинения. Это правило одинаково для обычных слов японского языка. Следующий слайд, пожалуйста.

Особое внимание вопросу визуальной схожести. Итак, для японского языка мы определили несколько наборов визуально схожих символов. Первый — это, например, дефис и [знак удлинения тёон], это два символа, в списке Консорциума Unicode перечислены символы из разных наборов, схожие до степени смешения. На основе этого списка мы выбрали следующие десять пар символов в качестве кандидатов на визуально идентичные варианты. То есть это десять пар символов. В правом нижнем углу приведены слова, содержащие эти символы. Следующий слайд, пожалуйста.

И подтверждение визуальной идентичности этих 10 пар вариантов на основе списка Unicode. Мы провели собственное исследование, чтобы выяснить, идентичны ли они по зрительному восприятию. Мы провели тестирование с участием 40 человек. Из них 20 человек умели читать по-японски, и 20 не умели. Мы оценили каждую пару символов на предмет идентичности и в конечном итоге все пары получили оценку ниже 3,2. Это означает, что они



---

схожи до степени смешения. Поэтому мы решили, что их следует считать вариантами.

Следующее наше исследование было призвано выяснить, есть ли еще какие-то пары символов, схожих визуально, помимо десяти перечисленных выше пар. Мы обратились к около 176 разным респондентам, из которых ответили 73, и в результате ни одна из пар символов не ввела в заблуждение более чем 3% респондентов, поэтому мы решили считать вариантами, схожими до степени смешения, только десять пар из списка Unicode. Следующий слайд, пожалуйста.

Итак, это вторая особая тема. По сути, как я уже сказал, в метках на японском языке разрешается использовать любые сочетания символов. Эта особенность предопределяет большое количество вариантных строк. При этом необходимо учитывать, что определение многих вариантов было заимствовано из правил генерирования меток для китайского и корейского языков. К примеру, вот строка из четырех символов кандзи, которая читается как «кейо дайгаку», что значит «университет Кейо». У этой строки есть еще три варианта записи.

В действительности университет Кейо зарегистрировал и использует в качестве домена второго уровня все четыре варианта записи строки в зоне .jp, то есть это значит, что университет Кейо хочет использовать все эти варианты. Но если позволить университету Кейо использовать их все, это значит, что в

---

корневой зоне будут разрешены четыре домена TLD одновременно. Четыре — это не так много, но в случае с более длинными метками размер корневой зоны может стремительно вырасти до огромных размеров. Так что для предотвращения такого взрывного роста размера корневой зоны требуется сократить количество выделяемых вариантов метки. Следующий слайд, пожалуйста.

Так что комиссия по выработке правил для японского языка попыталась придумать различные способы сократить количество выделяемых меток, однако любой метод может сократить количество меток, менее пяти или около того. Поэтому мы решили разрешить только действительные метки, на которые подаются заявки, а все остальные варианты записи этих меток блокировать. То есть выделяться может только одна метка, та, на которую подается заявка, а все остальные ее варианты запрещаются, блокируются. Следующий слайд, пожалуйста.

В настоящий момент мы находимся на этапе общественного обсуждения, которое продлится до 16 ноября. Спасибо. У кого-то есть какие-то комментарии?

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Хотта-сан. Есть какие-то вопросы в чате?

---

**САРМАД ХУСЕЙН:** У нас есть вопрос, но он не касается непосредственно японского языка. Он более общий, поэтому мы его пока придержим, возможно, мы зададим этот вопрос комиссии по интеграции правил после того, как свою презентацию проведет Мишель.

**ПИТИНАН КООАРМОРНПАТАНА:** Хорошо Спасибо, Сармад. Спасибо, Хотта-сан. Мы переходим к комиссии по интеграции правил, Мишель Суиньяр проведет для нас презентацию, посвященную правилам генерирования меток версии 5. Мишель, вам слово.

**МИШЕЛЬ СУИНЬЯР:** Хорошо Следующий слайд, пожалуйста. Итак, эта презентация будет посвящена четырем темам, которые вы можете видеть здесь. Первая из них — это интеграция, озабоченность тем, как мы провели интеграцию правил генерирования меток для китайского, японского и корейского языков, которые мы обрабатываем сегодня.

[неразборчиво] слайде, посвященном тому, как обрабатывать метки [неразборчиво], а также некоторые интересные моменты, касающиеся армянского, кириллического, греческого и латинского алфавитов, а затем о том, что охватывают правила генерирования меток корневой зоны версии 5. Итак, я подробнее остановлюсь на каждом из этих пунктов. Следующий слайд, пожалуйста.

---

То есть это, по сути, немного похоже на то, о чем только что говорил Хотта-сан. Одна из основных задач, стоящих перед нами сейчас, — это интеграция правил генерирования меток для китайского, японского и корейского языков. То есть работа над правилами для китайского, японского и корейского языков у нас уже завершена. У нас уже есть наборы правил для этих языков. На самом деле единственный язык, для которого еще продолжается общественное обсуждение правил генерирования меток, — это японский. Для остальных общественное обсуждение уже закончилось какое-то время тому назад. Это была очень сложная задача. Эти правила генерирования меток очень объемные, они содержат тысячи кодовых точек. Мы столкнулись с рядом довольно сложных проблем, касающихся вариантов... мы поговорим об этом сейчас или, возможно, позже, когда будут вопросы.

Так что, поскольку эти три набора правил у нас уже есть, мы работали над их интеграцией. Конечно же, [мы не приступали к] завершению этого процесса, но мы работали с разными комиссиями по выработке правил, чтобы понять, как это можно сделать.

Во-первых, интересно отметить, что в этих правилах генерирования меток варианты либо создаются сами по себе, либо наследуются через интеграцию. Такое наследование бывает двух видов. В одном случае у вас есть по меньшей мере две кодовые точки из результирующих правил генерирования меток, которые

---

определяются посредством интеграции из другого набора символов, что очень сильно затронуло правила для корейского и японского языков, потому что на самом деле больше всего вариантов было в китайских правилах генерирования меток. Это, как кому-то, возможно, известно, обусловлено существованием двух наборов символов китайского языка — традиционных и упрощенных. Многие символы [неразборчиво] могут быть представлены как символом в упрощенном начертании, так и символом в традиционном начертании. Там сложилось исторически.

Некоторые из этих упрощенных символов существуют также в японском или корейском языках, у них одинаковая форма, и множество вариантов также среди символов в традиционном начертании. Очень часто бывает так, что многие традиционные китайские символы рассматриваются как одно и то же, если использовать определение, которое уже использовалось.

То есть больше всего — намного больше — вариантов в китайских правилах генерирования меток, потому что во многих из этих наборов вариантов есть как минимум два аналогичных символа в правила генерирования меток для корейского и японского языков. Для них также необходимо было предусмотреть определение в этих правилах генерирования меток. На самом деле это означало, что нам пришлось, по сути, просить корейскую и японскую комиссии включить эти наборы вариантов, чтобы добавить как минимум две кодовые точки из существующих в

---

наборах вариантов в правилах генерирования меток для китайского языка.

То есть это была первая часть работы, когда нужно было, по сути, сделать так, чтобы все наборы вариантов... чтобы по меньшей мере два элемента, унаследованных из китайского набора вариантов, были также определены в правилах генерирования меток для японского и корейского языков. В этих двух языках они определяются как заблокированные, тогда как в китайских правилах генерирования меток допускается иногда их выделение. Это еще одна особенность этой системы.

В дополнение к этому правила генерирования меток для китайского языка на самом деле все же включают некоторые варианты, которые не были перечислены в списке в китайских правилах генерирования меток, которые были представлены изначально. Это объясняется в основном смешением в символах кана и хангыль, потому что в японских правилах генерирования меток, как вы уже видели, это приводило к возникновению в ряде случаев наборов вариантов между символами кандзи и кана. Разумеется, их необходимо было импортировать в китайский набор, потому что, конечно же, во многих случаях китайские символы включали также и эти, так сказать, эквиваленты кандзи. То же самое и в случае с корейским языком. Корейский немного отличается. В данном случае речь идет о символах хангыль и ханча, ханча — это, по сути, китайские символы, которые используются в корейском языке. Так что существуют

---

определенные совпадения между символами хангыль и китайскими символами, которые также нужно интегрировать в правила генерирования меток для китайского языка.

Для корейского и японского языков ситуация немного проще. Мы импортировали, как я уже сказал, всего одну или две кодовые точки, но в конечном итоге, конечно же, в корейские и японские правила необходимо включить наборы вариантов с одной из кодовых точек, связанных с другими китайскими символами, которые не входят в их собственные наборы. Так что получаются такие варианты из других наборов символов, как мы их называем, которые отображаются в правилах генерирования меток для корейского и японского языков.

В результате ожидается, что конфликт между метками будет определяться с использованием файла правил генерирования меток, полученного путем объединения. В правилах генерирования меток корневой зоны для каждого языка мы определяем т.н. элементарные правила генерирования меток, которые на самом деле описывают интеграцию правил генерирования меток, предложенных комиссиями по выработке правил, кроме того, у нас есть еще один файл, который называется объединенным, или общим файлом правил генерирования меток и представляет собой, по сути, описание того, каким образом определяется конфликт между меткой, на которую подается заявка, и уже существующими или делегированными метками. То есть нам нужен некий механизм, который позволял бы определять,

---

можно ли на самом деле создавать... доступна ли для делегирования метка, на которую подается заявка.

[Из этих трех наборов правил единственный, в котором такие метки выделяются, — это на самом деле только правила генерирования меток для китайского языка]. В них [есть] варианты. На самом деле варианты есть во всех правилах генерирования меток, просто в тех других они блокируются. То есть просто чтобы не было путаницы — нельзя сказать, что в правилах генерирования меток для японского языка нет вариантов, что они все блокируются. Мы все еще ожидаем завершения общественного обсуждения правил генерирования меток для японского языка, но мы не думаем, что в результате этого общественного обсуждения будут внесены какие-то значительные изменения. Однако время покажет. Следующий слайд, пожалуйста.

Итак, на этом слайде, или на этом изображении, также показана часть информации о правилах генерирования меток корневой зоны, так что на нем можно увидеть на самом деле больше деталей, но важно здесь то, что на самом деле при подаче заявки на делегирование метки в корневую зону есть два момента. Есть т. н. исходная метка и предлагаемые варианты метки. То есть это две разные вещи в этом процессе. Первое — это то, что мы обычно называем меткой, на которую подается заявка, то есть заявка касается именно ее.



---

Кандидат... мы, разумеется, используем элементарные правила генерирования меток, чтобы удостовериться, что метка, на которую подается заявка, допустима, то есть отвечает требованиям набора символов или правилам для этого набора, или каким-либо специальным требованиям, описанным в правилах генерирования меток, и если выясняется, что эта метка недопустима, то на этом все и заканчивается. Тогда мы генерируем вариант для индексирования, который представляет собой, по сути, такой механизм, в рамках которого создается некое уникальное представление этой метки, которое можно сравнивать с существующими метками, которые уже были делегированы или уже есть в корневой зоне. Если эта метка уникальна, то есть еще не используется, то мы ее принимаем. Если не уникальна, то отклоняем. Это значит, по сути, что уже делегирован другой вариант этой метки. То есть в таком случае мы ее отклоняем.

А вторая часть этого процесса выполняется, по сути, тогда же, в качестве возможного дополнительного этапа, то есть кандидат может предложить также варианты этой метки. Вы можете определить их, проведя собственное исследование, или же на самом деле сверившись с правилами генерирования меток, чтобы узнать, какие варианты могут быть у метки, на которую вы подаете заявку. Понятно, что прежде чем создавать эти варианты меток, нужно сначала убедиться, что эта метка допустима, то есть нужно еще раз свериться с правилами генерирования меток. Тогда нужно

---

еще раз убедиться, что ее можно выделять. Если ее нельзя выделять, тогда она блокируется и заявка также отклоняется, а если ее можно выделять, тогда она становится допустимым вариантом, который может быть делегирован в корневую зону.

Здесь есть тот важный момент, что выбирать, что будет основной меткой, а что вариантом, нужно осторожно, потому что сопоставление вариантов не симметрично. Для каждого элемента набора вариантов может быть разное сопоставление между основной меткой и ее вариантами. То есть иногда нужно тщательно выбрать такую исходную метку, чтобы для нее могли существовать также варианты, потому что в некоторые случаях, если в качестве исходной будет выбрана другая метка, то интересующий вас вариант, который был бы допустим в качестве исходной метки, может не быть допустимым для выделения в качестве варианта. То есть там не точная симметрия. Иногда нужно осторожно выбирать, что какую именно метку сделать исходной, чтобы можно было выделить и ее саму, и ее варианты. В таком случае сначала нужно свериться с определением в правилах генерирования меток и с сопоставлением, с типами сопоставления, которые используются для этих наборов вариантов. Следующий слайд, пожалуйста.

Итак, это, по сути... у нас есть четыре сценария, которые соотносятся между собой. Это армянский, кириллический, греческий и латинский наборы символов. Многие из них уже завершены. Кажется, правила генерирования меток для

---

латинского алфавита — единственные, по которым еще продолжается общественное обсуждение. Общественное обсуждение греческих правил закончилось недавно, но работа над правилами еще не завершена. Правила для кириллицы и армянских символов закончены уже давно, то есть на самом деле они, [так сказать, отложены].

Итак, краткое объяснение того, как это делается. Варианты внутри системы в полном объеме перечисляются в каждом из этих четырех наборов правил генерирования меток. Это, кстати, немного не так, как мы это делали для китайского, японского и корейского, но... То есть, по сути, варианты, которые соотносятся с взаимозависимыми элементами в каждом из этих четырех наборов символов, перечислены в правилах генерирования меток для каждого из них. То есть эти четыре набора правил генерирования меток можно рассматривать как некий согласованный набор... Если в одних правилах генерирования меток есть вариант, который соотносится с другим набором символов, то аналогичный вариант будет и в правилах генерирования меток для этого другого набора символов.

Мы сделали исключение для т. н. общих начертаний. То есть это такие символы, как О, то есть круг, или С. Они на самом деле есть и в других правилах генерирования меток, которые не имеют ничего общего с греческим [неразборчиво] в отличие от наборов символов для армянского, греческого и латинского языков, а также кириллицы. Они могут встречаться в системах письма,

---

которые используются в некоторых азиатских языках и которые уже включены в корневую зону, но там они никак не связаны с этими четырьмя алфавитами. То есть мы решили в качестве механизма интеграции включить их список только в правила генерирования меток для латинских символов, когда они будут интегрированы. Другие, конечно же, будут [неразборчиво] набора, но мы не хотели, чтобы они были в элементарных правилах генерирования меток, просто чтобы сократить уровень шумов, чтобы не создавать некоторые из этих наборов вариантов, потому что тогда все эти наборы правил были бы просто огромными, не неся никакой дополнительной полезной нагрузки.

Опять же, как я уже сказал, определение конфликта между метками подразумевает использование объединенного файла, т. н. общих правил генерирования меток. Допустимость проверяется всегда каждым элементарными правилами генерирования меток, это без изменений. Сложилась такая ситуация, что правила генерирования меток для двух наборов символов, армянского и кириллицы, были отложены. Потому что до тех пор, пока не будут готовы все четыре из них, мы не можем знать, как сложится ситуация, особенно если будут включены варианты в пределах одного набора символов, потому что в таком случае, по сути, один набор символов потребует соответствующего изменения других, когда речь будет идти о разных вариантах в том или ином наборе символов... это на самом деле довольно распространенная ситуация. У нас уже так бывало.

---

К примеру, правила генерирования меток для латинских символов потребовали включить в правила для набора символов кириллицы определенные варианты, которых там изначально не было, их не было в исходном предложении правил генерирования меток для кириллицы. Поэтому нам, по сути, пришлось добавить эти варианты в пределах одного набора символов в правила генерирования меток для кириллицы.

То есть это значит, что мы видим, что в отложенные правила генерирования меток пришлось внести некоторые изменения, но мы обсуждали это с комиссиями по выработке правил, по крайней мере с членами [официальных] комиссий по выработке правил, чтобы убедиться, что это допустимо. И эти добавленные варианты сопоставления будут частью общей интеграции, то есть это будет отличаться от того, что было в исходных правилах генерирования меток для армянского языка и кириллицы. Но мы согласовали это с членами комиссий по выработке правил, чтобы убедиться, что это допустимо. Конечно же, в отношении полностью интегрированного набора правил будет проводиться общественное обсуждение, для правил генерирования меток корневой зоны версии 5. Следующий слайд, пожалуйста.

То есть правила генерирования меток корневой зоны версии 5 будут содержать 18 наборов символов, которые уже есть в правилах генерирования меток корневой зоны версии 4. Мы добавили два новых набора символов для китайского, японского и корейского языков, потому что китайский у нас уже был, а теперь

---

мы добавили еще японский и корейский. И сейчас мы добавляем два новых алфавита — латинский и греческий. И, как я уже сказал, у нас есть еще два отложенных алфавита — кириллица и армянский. Остался еще последний, по которому работа еще продолжается, это бирманский язык. Мы надеемся, что Мьянма [неразборчиво] на общественное обсуждение в ближайшее время. Я не хочу... это скорее то, что должно быть сказано от имени ICANN, а не мной.

То есть планируется, что в правилах генерирования меток корневой зоны версии 5 будет 25 наборов символов. Два набора символов еще не готовы, это тана и тибетский. Я видел, что на каком-то из предыдущих слайдов речь шла о 28 наборах символов. Я не уверен. Я насчитал только 27, то есть один где-то потерялся. Или, может, 28 было на другом слайде. Не знаю, у меня получилось всего 27.

ПИТИНАН КООАРМОРНПАТАНА: Мишель. У нас есть два для японского языка, катакана и хирагана.

МИШЕЛЬ СУИНЬЯР: Вот почему. Хорошо Да, потому что мне стало интересно, как это у нас цифры не сошлись. Как бы то ни было, это завершение. У меня есть еще один слайд, но он просто для справки. У меня все. Думаю,

---

сейчас время перейти к вопросам, если они у нас есть, для всех нас.  
Спасибо.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Мишель. Сармад, вы хотите вернуться к тому вопросу, который у вас был, или на него уже отчасти ответили?

САРМАД ХУСЕЙН: Спасибо, Питинан. Я зачитаю этот вопрос. Мишель, возможно, отчасти на него уже ответил, но мы дадим ему возможность что-то добавить, если он захочет. Это вопрос от Донны Остин (Donna Austin). Был ли такой же вывод о недопущении вариантов — это в контексте правил генерирования меток для японского языка — сделан в отношении какого-то еще набора символов? Таков вопрос. Был ли такой же вывод о том, чтобы не разрешать варианты, сделан в отношении какого-то еще набора символов?

МИШЕЛЬ СУИНЬЯР: Я хочу еще раз сказать, что это не обязательно... варианты есть. Во многих других правилах генерирования меток варианты есть. Но большинство из них заблокированы. Возможность выделения вариантов — это сложный вопрос, потому что если разрешить выделять варианты, то это не так просто с точки зрения DNS. Люди больше... То есть мы в комиссии по интеграции стараемся сделать так, чтобы, если вы определяете какую-то схему выделяемых вариантов, чтобы вы были должны как-то свести их количество к

---

минимуму. Для китайского языка могут существовать, кажется, не более пяти вариантов, включая исходную метку. То есть если вы подаете заявку на какую-то метку, то вместе с ней у вас может быть всего четыре выделяемых варианта. Это максимум.

И для латинского алфавита, кажется, тоже в правилах генерирования меток есть какая-то защита, ограничивающая количество выделяемых вариантов. Для большинства из них можно относительно просто прийти к чему-то разумному. Японский для нас стал очень сложной задачей, потому что мы вместе с комиссией по выработке правил для японского языка очень много работали над тем, чтобы ограничить число выделяемых вариантов для метки, на которую подается заявка. Никакого удачного решения там нет. Мы и так смотрели, и этак, мы пытались что-то придумать.

У нас не получилось найти решение, которое бы... потому что в худшем случае может получиться взрывной рост объема, могут быть тысячи... если будет создана какая-то плохая метка... под плохой я имею в виду метку, которая будет создана специально для того, чтобы реализовать этот нехороший сценарий, когда может получиться так, что, даже если у вас есть инструменты, эти инструменты могут не... получится такое большое число, все настолько усложнится, что ваши инструменты будут работать не один день, а прийти к какому-то решению так и не получится.



---

То есть я еще раз скажу, что сложилась такая ситуация, что мы не могли позволить слишком много выделяемых вариантов меток. Именно поэтому мы вместе с комиссией по выработке правил для японского языка пришли к выводу что единственным решением должно стать блокирование всех вариантов. Именно так мы пришли к тому, что получилось. Но я повторюсь — они есть, просто заблокированы.

Так же, кстати, обстоят дела сейчас со многими правилами генерирования меток. Во многих правилах генерирования меток заблокированы варианты. Это не уникальная особенность правил генерирования меток для японского языка.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Мишель. Есть еще какие-то вопросы?

САРМАД ХУСЕЙН: Есть еще один вопрос от Роберта Нкамбве (Robert Nkambwe). «Является ли эта работа подмножеством работы над универсальным принятием, и наоборот?. Извините за этот вопрос, я еще только разбираюсь в новой для меня информации об ICANN». Итак, вопрос о том, являются ли правила генерирования меток корневой зоны подмножеством работы над универсальным принятием, или наоборот.

И в чате кто-то уже отвечал, но все равно, Питинан, возможно, вы хотите на это ответить.

---

ПИТИНАН КООАРМОРНПАТАНА:        Разумеется. Спасибо. Спасибо за этот вопрос, Роберт. Да, они на самом деле связаны. IDN-домены в определенном смысле представляют собой механизм, позволяющий создавать имена, записываемые символами, отличными от ASCII, или интернационализированные доменные имена, то есть универсальное принятие — это как раз о том, чтобы имя, записанное символами местного языка, не просто выделялось в техническом смысле, но чтобы такие IDN-домены также принимались системами, сайтами и приложениями в Интернете.

Еще один аспект универсального принятия касается не только IDN-доменов. Тут есть определенная взаимосвязь. Что касается универсального принятия, отчасти задача состоит в том, чтобы гарантировать правильную обработку интернационализированных доменных имен приложениями Интернета, но есть еще один [неразборчиво], это новые gTLD, новые длинные или новые короткие TLD. Возможно, вы видели домена вида [точка-что-то-новое], теперь помимо обычных доменов верхнего уровня есть такие же домены с более длинной частью после точки, такие как .photography, а также новые короткие домены верхнего уровня. такие как [.run] и т. п. То есть универсальное принятие касается как интернационализированных доменных имен, так и неинтернационализированных доменных имен. Так что это взаимосвязано довольно тесно в каком-то смысле. То есть

---

универсальное принятие и IDN-домены — это связанные понятия.  
Это ответ...

САРМАД ХУСЕЙН:

Спасибо. Да. У нас на самом деле есть еще один комментарий, однако я думаю, что было бы хорошо обсудить его с Мишелем, если это возможно. Это вопрос также от Роберта Нкамбве. Он спрашивает: «Я ожидал, что прежде чем интегрировать правила в корневую зону, будет проводиться доскональное тестирование. Я не вижу этого на графике процесса». Так что если Мишель... у нас осталось мало времени, но не могли бы вы поделиться с нами какой-то информацией о тестировании, каким образом оно проводится для предложений правил генерирования меток корневой зоны перед их интеграцией. Спасибо.

МИШЕЛЬ СУИНЬЯР:

Да, это то, что мы просим предоставить каждую комиссию по выработке правил, — набор каких-то меток для тестирования. Нам этот набор меток нужен на самом деле для применения этих правил генерирования меток к набору символов, в отношении которого предлагаются эти правила. Мы также проверяем, отвечают ли требованиям предложенных правил генерирования меток те домены верхнего уровня, которые уже есть в корневой зоне. Потому что если нет, то это тоже как бы не здорово. Поэтому мы на самом деле проверяем существующие метки, которые уже делегированы в корневую зону.

---

Мы также тестируем правила генерирования меток применительно к данным, по сути, существующим... мы прогоняем их по очень большим объемам данных для каждого набора символов. Их можно найти в Интернете. Затем мы проверяем на соответствие этим правилам генерирования меток большинство обычных часто используемых слов. И когда мы видим какое-то отклонение, мы спрашиваем у комиссии по выработке правил о его причинах. Иногда это бывает в тех случаях, когда какие-то изменения были внесены в сам набор символов [неразборчиво]. Иногда это объясняется какими-то историческими причинами. И каждый раз в таких случаях мы получали от комиссий по выработке правил ответы, которые нас удовлетворяли.

Так что да, мы действительно тестируем их с использованием разных механизмов. Обычно такое тестирование проводится перед интеграцией, потому что интеграция — процесс на самом деле по большей части автоматический, по сути, правила генерирования меток приводятся в тот вид, в котором они пригодны для интеграцию в корневую зону. Но это на самом деле не... техническая работа делается задолго до этого комиссиями по выработке правил, в ходе взаимодействия между комиссиями по интеграции и комиссиями по выработке правил.

ПИТИНАН КООАРМОРНПАТАНА: Спасибо, Мишель. Я думаю, что это все, о чем мы хотели поговорить. Мы уже превысили наше время, так что на этом я

---

закрою наше заседание. Спасибо всем за участие и обмен мнениями. Пожалуйста, ознакомьтесь с материалами двух общественных обсуждений, я надеюсь, вы сможете их как-то прокомментировать. На этом я объявляю наше заседание закрытым, хорошего вам дня и плодотворной работы в остальных заседаниях в рамках этой конференции ICANN. Надеюсь увидеться с вами на следующей неделе. До свидания.

**[КОНЕЦ СТЕНОГРАММЫ]**