

# Предложения по присвоению имен интернационализованны х адресов электронной ПОЧТЫ

2 ноября 2020 г.



## Содержание

<b>Введение</b>	<b>3</b>
<b>Основные положения</b>	<b>4</b>
<b>Вопросы политик, которые необходимо учитывать</b>	<b>5</b>
Поддерживаемые системы письма	5
Длина имени почтового ящика Строка	6
Сочетание разных систем письма	6
Предотвращение использования недопустимых строк и строк, отображаемых нестабильно	8
Соображения относительно систем письма справа налево	10
Соображения относительно псевдонимов и отображаемых имен	12
Символы и знаки	13
Нормализация символов Unicode	14
Соображения относительно того, что считается одинаковым	15
Прочие соображения	18
<b>Глоссарий</b>	<b>18</b>



## Введение

В настоящем документе, предназначенном для администраторов систем электронной почты, изложены предложения относительно порядка присвоения имен интернационализированных адресов электронной почты. Стандарты электронной почты [RFC6533](#), «[Интернационализация уведомлений о доставке и расположении сообщений](#)», и [RFC6430](#), «[Значение типа отчета об электронной почте: не спам](#)», в настоящее время позволяют использовать в именах почтовых ящиков (часть адреса электронной почты до символа @) и в доменных именах (часть после символа @) практически любой язык или систему письма; их запись более не ограничена символами A-Z латинского алфавита и арабскими цифрами 0-9.

Это усовершенствование важно для тех пользователей, которые предпочитают использовать языки, отличные от английского, а также для тех, с кем такие пользователи обмениваются сообщениями. Однако системным администраторам нужно предусмотреть для присвоения имен почтовых ящиков такие правила, которые позволили бы в полной мере использовать преимущества большего выбора допустимых символов, в то же время сведя к минимуму возможность ошибок, введения пользователей в заблуждение или возникновения возможностей для злоупотреблений.

Этот документ предназначен для администраторов систем электронной почты, которые хотят предоставить своим пользователям возможность использования интернационализированных почтовых ящиков и/или которым необходимо настраивать и управлять системами, поддерживающими интернационализированные адреса электронной почты. Он позволит администраторам задать правильные вопросы и сделать правильный выбор при введении своих политик.

Предполагается, что специалисты, читающие этот документ, знакомы с системами электронной почты и имеют опыт в том, что касается фазы 1 интернационализации адресов электронной почты (EAI) — возможности отправлять сообщения электронной почты на интернационализированные адреса и получать почту от таких адресов. В этом документе затрагиваются вопросы фазы 2 интернационализации адресов электронной почты (EAI), такие как присвоение имени и хостинг интернационализированных почтовых ящиков.

В разделе «Глоссарий» представлены определения специальных терминов и аббревиатур, используемых в этом документе.



## Основные положения

Этот документ предназначен для администраторов электронной почты, системных администраторов и ИТ-менеджеров. Его задача — собрать в одном месте и перечислить основные предложения относительно работы и политик, которые необходимо учитывать при создании интернационализованных ящиков адресов электронной почты, а также используемых для них псевдонимов и отображаемых имен.

Имена почтовых ящиков — часть адреса электронной почты перед символом @ — важный фактор того, насколько полезной, удобной для пользователя и безопасной будет та или иная система электронной почты, поэтому так важно, чтобы администраторы электронной почты определяли эффективные политики присвоения имен почтовых ящиков. Исторически адреса электронной почты записывались с использованием только букв латинского алфавита и цифр, однако эволюция технологий электронной почты привела к тому, что теперь как имена почтовых ящиков, так и доменные имена можно записывать почти на любом языке и с использованием почти любой системы письма. Из-за такой гибкости теперь сложнее определять правила. Этот документ призван помочь администратором электронной почты адаптировать свои политики в отношении присвоения имен почтовых ящиков таким образом, чтобы учитывать возможность записи адресов электронной почты с использованием символов, отличных от символов латинского алфавита и цифр. В конечном итоге именно стоящие перед системами электронной почты коммерческие задачи, наряду с языковыми потребностями пользователей и их корреспондентов, будут определять особенности политик и возможность использования того или иного алфавита (системы письма того или иного языка) для записи имен почтовых ящиков, а также допустимую длину таких имен.

Поскольку определенные варианты написания, сочетания символов, отдельные знаки или символы могут вводить пользователей в заблуждение, применяемые администраторами политики должны запрещать их использование (конкретные особенности зависят от языка, используемого для записи имени почтового ящика). Особого внимания заслуживают технические проблемы, связанные с тем, что некоторые имена могут выглядеть одинаково для пользователя, даже если в действительности они состоят из разных символов. С системами письма, в которых используется запись справа налево (Right-To-Left, RTL), связаны свои уникальные возможности существования имен, вводящих пользователей в заблуждение, так что если ваша система позволяет использование таких систем письма, необходимо учитывать ряд дополнительных аспектов политик.

Пользователи могут обмениваться электронной почтой с адресатами, испытывающими затруднения с чтением символов, которыми записан тот или иной адрес электронной почты. Чтобы уменьшить вероятность возникновения таких затруднений, рекомендуется использовать политики, которые позволяли бы пользователям выбирать имена для отображения (личные имена) и использовать псевдонимы (альтернативные адреса для того же пользователя).

Также в настоящем документе приводятся ссылки на ресурсы и материалы, содержащие более подробные сведения о некоторых из таких вопросов политик. Он также содержит глоссарий терминов, некоторые из которых могут оказаться новыми для администраторов, занимающихся реализацией поддержки различных языков и систем письма для записи имен почтовых ящиков.



## Вопросы политик, которые необходимо учитывать

В этом документе рассматриваются вопросы, которые могут быть частью политики присвоения имен почтовых ящиков. В каждом разделе один из этих вопросов излагается следующим образом:

- Описание вопроса.
- Разъяснение предложений, касающихся аспектов систем электронной почты, основ работы с пользователями и их потребностей.
- Рекомендации в отношении политики (в некоторых случаях).
- Контрольный список действий, которые следует предпринять.

### Поддерживаемые системы письма

При настройке системы электронной почты администратор должен принять решение о том, какие алфавиты будут использоваться в этой системе по умолчанию.

Термин «алфавит» или «система письма» применяется здесь в лингвистическом смысле и означает набор правил и символов, которые используются для записи в одном или нескольких языках. К примеру, латинский алфавит — это набор символов от A до Z, а также ряд дополнительных символов, которые используются для записи текстов на английском, французском и множестве других языков в странах Европы, Северной и Южной Америки, а также Африки. Арабский алфавит используется для записи текстов на различных языках стран Ближнего Востока, Северной Африки и Южной Азии.

### Предложения:

- В большинстве систем для обеспечения обратной совместимости используются кодовые точки ASCII. Поддержка EAI позволяет администраторам использовать дополнительные системы письма.

### Контрольный список:

[ ] Узнайте, чего ожидают пользователи и целевая аудитория, чтобы понять, какие системы письма необходимы для записи обеих частей адреса электронной почты — имени почтового ящика и доменного имени.

[ ] Ознакомьтесь с этим документом, чтобы понять сложности, связанные с обеспечением поддержки дополнительных систем письма (например, вероятность неправильного прочтения, ошибок и возможностей для злоупотреблений), а также потенциальные последствия для ваших пользователей.

[ ] Исходя из ожиданий ваших пользователей, составьте список систем письма, которыми будет разрешена запись имен почтовых ящиков.



### Длина имени почтового ящика в поле ввода

Учтите длину имени почтового ящика в символах. Какая минимальная и максимальная длина будет разрешена?

#### Предложения:

- На поле ввода имени почтового ящика могут распространяться системные ограничения. Стандарты электронной почты допускают использования в имени почтового ящика до 64 байт, однако в некоторых системах длина этого поля ограничена меньшим количеством символов, чем предусмотрено стандартом.
- Короткие имена могут быть по-своему удобными для пользователей, потому что их проще запоминать и записывать. Длинные же имена могут быть удобными для пользователей тем, что они могут быть более конкретными и уникальными, то есть с меньшей степенью вероятности конфликтовать с именами других почтовых ящиков. Возможно, у вас уже есть политика, требующая, чтобы имена почтовых ящиков создавались на основе личных имен пользователей, например, из имени и фамилии пользователя, разделенных точкой.
- Возможно, вам имеет смысл оставить существующую политику, расширив в ней поддержку различных языков и наборов символов. Кроме того, подумайте о том, какие языки и наборы символов вашим пользователям будет проще понимать, запоминать и вводить в поле адреса электронной почты.
- При использовании для добавления языков и систем письма формата UTF8 один символ может быть представлен несколькими байтами, что приведет к увеличению длины имени ящика. К примеру, имя почтового ящика, состоящее из символов ASCII, для записи которых используется 20 байт, может увеличиться до 60 байт, если оно будет записано с использованием 20 букв другого алфавита.

**Рекомендации:** Отсутствуют.

#### Контрольный список:

Узнайте ограничения своей системы.

Убедитесь, что понимаете, чего ожидают ваши пользователи в том, что касается интернационализированных имен почтовых ящиков.

Рассмотрите возможность использовать для интернационализированных имен почтовых ящиков ту же или аналогичную политику длины, что и для традиционных имен почтовых ящиков.

### Сочетание разных систем письма

Сочетание разных систем письма — это использование для записи одного имени почтового ящика символов из разных алфавитов таким образом, каким они не используются в обычных условиях. В частности, некоторые символы одного алфавита могут быть похожими до степени смешения на символы другого алфавита, что открывает возможности для злоупотреблений или введения пользователей в заблуждение.



## Предложения:

- Сочетание разных систем письма создает проблемы с безопасностью и приводит к введению пользователей в заблуждение. К примеру, на следующем изображении показаны несколько разных, но похожих кодовых точек Unicode, которые могут вводить пользователей в заблуждение<sup>1</sup>.

а а а а а а а а а а а а  
b Ъ Ъ Ъ Ъ b b b b b b  
с с с с с с с с с с с с  
d d d d d d d d d d d d  
e e e e e e e e e e e e

- Те, с кем вы переписываетесь, будут вводить имя почтового ящика и доменное имя вместе как адрес электронной почты. Если доменное имя записано в той или иной системе письма, естественно, чтобы для имени почтового ящика использовалась та же система письма. Подумайте, можно ли вместе с системой письма, которая используется для записи доменного имени, использовать еще какой-то другой алфавит.
- Некоторые сочетания могут быть необходимы, в то время как другие могут создавать путаницу или представлять собой уязвимость с точки зрения вредоносных действий.
- В некоторых культурах для записи могут традиционно использоваться разные системы письма. К примеру, большинство текстов на японском языке представляют собой смесь из символов кандзи, хирагана, катакана и латинского алфавита. При этом арабские символы в японских текстах обычно не используются.
- В именах почтовых ящиков, записываемых латинскими символами, регистр символов обычно не различается. Символы верхнего и нижнего регистров считаются одинаковыми. Подумайте о том, есть ли еще какие-то символы, которые ваши пользователи рассчитывают видеть одинаковыми, для тех систем письма, которые вы используете. В целях ясности мы обычно не считаем сочетание верхнего и нижнего регистра смешением разных систем письма; подробнее об этом см. в разделе [Предложения относительно того, что считать одинаковым](#).
- В системе доменных имен (DNS) интернационализованные доменные имена (IDN) могут быть представлены двумя форматами: это U-метки и A-метки. Использование A-меток не рассматривается в настоящем документе как сочетание разных систем письма; подробнее об этом см. в разделе [Предложения относительно псевдонимов](#).

<sup>1</sup> Источник: <https://thenextweb.com/wp-content/blogs.dir/1/files/2014/08/confusables-3.png>



## Рекомендации:

- Для обеспечения максимального удобства использования рекомендуется для записи имени почтового ящика использовать ту же систему письма, что и для той части адреса электронной почты, которая обозначает доменное имя, если иное не требуется для целей вашей организации.
- Старайтесь не допускать в именах почтовых ящиков смешения разных языков, использующих один и тот же алфавит, или создания имен почтовых ящиков с визуально идентичными символами, использующими разные кодовые точки.
- В именах почтовых ящиков на том или ином языке следует разрешать только те символы, которые используются в сочетании друг с другом при записи обычных текстов на этом языке. Сочетание разных систем письма в пределах одного имени почтового ящика должно быть запрещено. К примеру, имя почтового ящика на арабском языке должно ограничиваться только арабскими символами, а в имени почтового ящика на японском можно разрешить использовать кандзи, хирагана, катакана и латинские символы.

Пример 1: Адрес электронной почты на арабском языке: `مستخدم@رسيل.السعودية`

Пример 2: Адрес электронной почты на японском языке с сочетанием разных систем письма: `誰か@日本語ドメイン名試験.jp`

## Контрольный список:

[ ] Рассмотрите возможности введения пользователей в заблуждение, ошибок или вредоносных действий в результате сочетания разных систем письма в пределах имени почтового ящика или в имени ящика и доменном имени.

[ ] Разрешите сочетать разные системы письма в тех случаях, когда это очевидно отвечает потребностям пользователей, обусловленным местными традициями и требованиями рынка. К примеру, пользователи коммерческой службы электронной почты могут быть заинтересованы в использовании имен с определенным сочетанием разных систем письма, а сотрудники организации могут не возражать против ограничения выбора имен почтовых ящиков и запрета смешения разных систем письма.

[ ] Если пользователям разрешается запрашивать имена своих почтовых ящиков, предложите им примеры.

## Предотвращение использования недопустимых последовательностей символов и последовательностей символов, отображаемых нестабильно

В некоторых системах письма символы могут сочетаться допустимым образом, обеспечивая эффективный обмен информацией, или недопустимым образом, вызывая путаницу и риски с точки зрения безопасности. Такие последовательности символов будет невозможно обрабатывать одинаково на разных платформах и в разных приложениях — они будут отображаться по-разному в разных системах или на разных экранах. Этот раздел посвящен различным способам предотвратить использование недопустимых последовательностей символов.

## Предложения:

- В системе DNS определены четкие правила в отношении того, какие символы можно использовать в тех или иных языках и системах письма, а какие нет.





В некоторых системах письма, в частности, в языках Юго-Восточной и Южной Азии, символы и обозначения могут сочетаться на разных уровнях.

- Рассмотрим для примера последовательности, записанные тайскими символами:  $\text{ด}$  и  $\text{ดี}$ . Первая строка,  $\text{ด} \text{U+0E14}(\text{ด}) + \text{U+0E35}(\text{ี}) + \text{U+0E35}(\text{ี})$  недопустима, потому что выглядит так же, как  $\text{ดี} \text{U+0E14}(\text{ด}) + \text{U+0E35}(\text{ี})$ . Сообщество пользователей тайского алфавита определило правило, согласно которому символ  $\text{U+0E35}$  не может следовать после символа  $\text{U+0E35}$ .
- Эти правила, используемые в системе DNS, называются Правилами генерирования меток (Label Generation Rules, LGR). Возможно, вы захотите применить эти правила к почтовым ящикам в вашей системе.
- Однако Правила генерирования меток могут накладывать на вашу систему больше ограничений, чем это необходимо. К примеру, Правила генерирования меток корневой зоны запрещают использовать цифры, однако нет ничего плохого в том, чтобы использовать цифры в имена почтовых ящиков. Для некоторых систем письма Правила генерирования меток еще не разработаны, поэтому опереться на них не получится.
- Вы можете использовать существующие [Справочные правила генерирования меток на втором уровне](#) (Second Level Reference LGRs) для проверки имен почтовых ящиков. Средство проверки на основе Правил генерирования меток доступно онлайн <https://lgrtool.icann.org/>. Исходный текст этого средства также доступен на платформе GitHub, откуда его можно загрузить для интеграции в вашу систему.

#### Рекомендации:

- Старайтесь не допускать потенциальных проблем с безопасностью, которые возникают тогда, когда имена почтовых ящиков создаются на основе недопустимых и аналогичных им последовательностей символов.
- Если вы намереваетесь предлагать имена почтовых ящиков на одном языке, воспользуйтесь Правилами генерирования меток для этого языка. К примеру, если вы хотите предлагать почтовые ящики на немецком, воспользуйтесь Правилами генерирования меток для немецкого языка.
- Если вы намереваетесь предлагать почтовые ящики в написании алфавитом, который используется в нескольких разных языках, воспользуйтесь Правилами генерирования меток для этого алфавита. К примеру, если вы хотите предлагать почтовые ящики на немецком, французском и шведском языках, воспользуйтесь соответствующим подмножеством Правил генерирования меток для латинского алфавита.
- Если интересующая вас система письма не описана в [Справочных правилах генерирования меток на втором уровне](#), вы можете воспользоваться Правилами генерирования меток для корневой зоны или следовать общим принципам, описанным в стандарте [RFC6912](#). Правила использования цифр в именах, записываемых справа налево, см. в документе [RFC5893](#), «[Системы письма справа налево в интернационализованных доменных именах для приложений \(IDNA\)](#)».

#### Контрольный список:

[ ] Определите, отвечают ли требованиям к последовательности символов имен почтовых ящиков, которые вы хотите использовать, Справочные правила генерирования меток на втором уровне. Если отвечают, воспользуйтесь Справочными правилами генерирования меток. Если не отвечают в полной мере, доработайте их по необходимости.



[ ] Дайте своим пользователям возможность ознакомиться с Правилами генерирования меток — этот механизм может помочь вам вместе понять, какие символы и их сочетания будут считаться допустимыми.

[ ] Если это возможно, используйте инструмент проверки Правил генерирования меток для проверки последовательности символов имен почтовых ящиков вручную.

[ ] Подумайте, хотите ли вы использовать инструмент проверки Правил генерирования меток в готовом виде или же развернуть его у себя на внутреннем сервере для оптимизации рабочих процессов и защиты обрабатываемых им имен почтовых ящиков от передачи за пределы вашей сети.

### **Предложения относительно систем письма справа налево**

Среди распространенных в современном мире систем письма справа налево можно отметить арабский алфавит, а также иврит, фарси и урду. Исторически по мере распространения арабского письма первоначальный набор из 28 символов, которые использовались для записи текстов на арабском языке, дополнялся, чтобы добавить возможность записи звуков многих других языков, таких как пушту, фарси, урду и т. п. Алфавит, используемый в иврите, также применяется для записи других языков, таких как идиш.

В системах письма справа налево и сверху вниз писать начинают от правого края страницы, двигаясь по направлению к левому. В противоположность этому в системах письма слева направо начинают писать от левого края страницы, двигаясь вправо.

При этом необходимо учитывать, что в тексте, записанном справа налево, могут встречаться фрагменты, записываемые слева направо. Например, в тексте на арабском языке цифры и слова, записываемые латинскими буквами, пишутся слева направо. Поэтому такие системы письма иногда называют двунаправленными.

#### **Предложения:**

- Для систем письма справа налево характерны потенциальные возможности путаницы, которые отсутствуют в системах письма слева направо. Правила генерирования меток для таких систем письма отражают тщательную оценку таких возможностей. Отнеситесь серьезно к требованиям и запретам, предусмотренным Правилами генерирования меток, потому что за ними стоят веские, хотя и не всегда очевидные причины.
- При обмене электронной почтой между адресатами, записываемыми справа налево, и пользователями, которые не понимают соответствующую систему письма, гарантировать, что адрес будет понят, еще труднее. В таких случаях может быть особенно удобно использовать псевдонимы. Подробнее см. в разделе [Предложения относительно псевдонимов и отображаемых имен](#).
- В некоторых системах письма справа налево (например, в арабском языке и иврите) могут существовать «одинаковые» метки, отображаемые по-разному. Такие альтернативные варианты можно считать идентичными адресатами электронной почты, а не отдельными адресатами. Подробнее см. в разделе [Предложения относительно того, что считать одинаковым](#).
- Тексты, записываемые справа налево, часто бывают двунаправленными. К примеру, слова, записываемые латинскими символами (например, Cola) или арабскими цифрами (например, 123), будут писаться слева направо. Кроме того, некоторые символы пунктуации также влияют на направление



окружающего их текста. Это означает, что имена почтовых ящиков, содержащие элементы, записываемые слева направо, могут с большей вероятностью набираться с ошибками.

- В арабском алфавите помимо цифр из набора символов ASCII используются также свои собственные цифры. Если в имени почтового ящика встречается цифра, подумайте о том, какого типа цифра должна использоваться и будет ли считаться идентичным имя, в котором будет использоваться цифра другого типа. Опасения, связанные с сочетанием разных систем письма, касаются имен, содержащих цифры обоих типов.
- В случае двунаправленного текста порядок символов может быть визуально неочевиден. Обычно текст вводится и сохраняется в том же порядке, в котором эти же слова произносятся в устной речи, однако в таком случае может не быть простого соответствия с положением записанного справа налево текста в отображаемом адресе. Это, как правило, не представляет проблемы для тех, кто знаком с данной системой письма и видит, когда направление текста меняется. Однако те, кто эту систему письма прочитать не могут, могут столкнуться с затруднениями, пытаясь определить порядок ввода символов, который соответствовал бы отображаемому адресу.
- На местах обычно используются клавиатуры, адаптированные для использования местных систем письма, поэтому с них может быть неудобно вводить некоторые из символов, использованных в том или ином адресе электронной почты. На некоторых клавиатурах могут отсутствовать клавиши для ввода точки (.) или символа электронной почты (@). Это затрудняет ввод адресов электронной почты. Учтите, какие клавиатуры будут использоваться для ввода названий ваших почтовых ящиков, и подумайте, стоит ли разрешать использование символов, которые может быть неудобно вводить.
- Во многих системах письма справа налево для обозначения гласных звуков и определенных особенностей произношения используются диакритические знаки. Их использование, по сути, необязательно. Если в имени почтового ящика будет разрешено использовать такие диакритические знаки, пользователи могут забывать их набирать. Проще разрешить имена без таких диакритических знаков.
- В арабском письме, где символы в каллиграфической записи обычно отображаются в слитном написании, есть специальные символы, позволяющие принудительно соединять символы или наоборот, запрещать такое соединение. Они называются «соединение нулевой ширины» ("zero width joiner") и «запрет соединения нулевой ширины» ("zero width non-joiner"). Они представляют собой относительно мало известный механизм, специфическую особенность кодирования текста на компьютерах, и многие пользователи могут не знать, как ими пользоваться. Имена почтовых ящиков, в которых используются символы соединения, пользователям может быть труднее набирать правильно.

#### Рекомендации:

- Использовать символы соединения, особенно в арабском письме, не рекомендуется. Например, используйте محمد вместо مُحَمَّد و شؤون вместо شؤُون .
- Имена почтовых ящиков, записываемые справа налево, не должны начинаться с цифры. Цифры в начале текста могут выглядеть неоднозначно с точки зрения правил направления письма.
- Старайтесь не допускать потенциальных проблем с безопасностью, связанных с использованием недопустимых и эквивалентных им последовательностей символов. Недопустимые и эквивалентные им последовательности символов особенно характерны для арабского письма. См. раздел [Предотвращение использования недопустимых последовательностей символов и](#)



последовательностей символов, отображаемых нестабильно выше и раздел Предложения относительно того, что считать одинаковым ниже.

- Если вы намереваетесь предлагать имена почтовых ящиков на одном языке, воспользуйтесь Правилами генерирования меток для этого языка. К примеру, если вы хотите предлагать почтовые ящики на арабском, воспользуйтесь Правилами генерирования меток для арабского языка. Подробнее см. В разделе Предложения относительно того, что считать одинаковым.
- Если вы намереваетесь предлагать почтовые ящики в написании алфавитом, который используется в нескольких разных языках, воспользуйтесь Правилами генерирования меток для этого алфавита. К примеру, если вы хотите предлагать почтовые ящики на арабском, пушту и урду, воспользуйтесь Правилами генерирования меток для арабского алфавита.
- Если интересующая вас система письма не описана в Справочных правилах генерирования меток на втором уровне, воспользуйтесь Правилами генерирования меток для корневой зоны или следуйте общим принципам, описанным в стандарте RFC6912.
- Правила использования цифр в именах почтовых ящиков, записываемых справа налево, см. в документе RFC5893, в котором изложены необходимые рекомендации по использованию систем письма справа налево в интернационализованных доменных именах для приложений (IDNA).

#### **Контрольный список:**

[ ] Старайтесь не допускать смешивания разных систем письма с письмом справа налево, поскольку это может вводить в заблуждение и создавать проблемы в области безопасности.

[ ] Используйте Справочные правила генерирования меток на втором уровне или Правила генерирования меток для корневой зоны.

[ ] Старайтесь не допускать использования двух почтовых ящиков с одинаковыми именами.

#### **Предложения относительно псевдонимов и отображаемых имен**

Некоторые системы электронной почты предоставляют пользователям удобные возможности заводить несколько разных имен почтовых ящиков электронной почты или несколько псевдонимов, перенаправляющих почту на один и тот же почтовый ящик. Такие псевдонимы и отображаемые имена служат нескольким разным целям и могут использоваться в разных сценариях, связанных с интернационализацией адресов электронной почты.

Примеры использования:

- Пользователи, нуждающиеся в адресах электронной почты, записываемых в разных системах письма, например, для двуязычных визитных карточек с использованием как латинского, так и местного алфавита.
- Обход проблем с несовместимыми системами при отправке и получении электронной почты.
- Возможность предоставить получателям более узнаваемые адреса электронной почты и псевдонимы, отвечающие данным в их адресных книгах.

#### **Предложения:**



- Оцените маршруты доставки электронной почты от ваших пользователей их адресатам и обратно. Помните о том, что почта, передаваемая по этим двум маршрутам, может обрабатываться разными системами, так что передача в одном направлении может работать без ошибок, а в другом — сталкиваться с проблемами.
- Чем более разнообразна база адресатов у ваших пользователей и чем больше разных маршрутов используется для доставки электронной почты из вашей системы и в нее, тем более вероятна возможность наткнуться на систему доставки, испытывающую проблемы с маршрутизацией почты, содержащей интернационализованные адреса электронной почты.
- Если разрешить использовать псевдонимы или отображаемые имена, использующие систему письма, отличную от той, в которой записано имя почтового ящика, это может привести к возникновению проблем в области безопасности; однако это может также быть удобно, когда интернационализованный адрес электронной почты используется в разных языковых сообществах.

Ниже приведен пример арабского отображаемого имени для адреса электронной почты на английском:

"مشرف@اختيار-سجل.مصر" <[adminmail@datamail.asia](mailto:adminmail@datamail.asia)>  
Отображаемое имя      Адрес электронной почты

#### Рекомендации:

- Рассмотрите возможность предоставлять адреса электронной почты, записываемые символами ASCII, в качестве псевдонимов для интернационализованных адресов электронной почты ваших пользователей. Такой псевдоним должен доставлять сообщения по тому же адресу, что и интернационализованный адрес. У пользователя должна быть возможность выбирать, отправлять ли сообщение с интернационализованным адресом или же с адреса-псевдонима в формате ASCII.

#### Контрольный список:

[ ] Выясните, поддерживает ли ваша система псевдонимы.

[ ] Если ваша система поддерживает псевдонимы, добавьте в пользовательский интерфейс возможность создавать псевдонимы в процессе выбора имени почтового ящика. Пользователи будут уведомляться о том, что им следует завести псевдоним в формате ASCII, при этом они должны иметь возможность создавать такие псевдонимы и привязывать их к своим интернационализованным именам почтовых ящиков.

[ ] В интерфейс пользователя можно также добавить возможность создавать дополнительные псевдонимы позже, если ваша система поддерживает такую функциональность.

[ ] Если вы присваиваете пользователям имена почтовых ящиков, а не позволяете им выбирать их самостоятельно, присваивайте псевдоним в формате ASCII тогда же, когда вы присваиваете им имя интернационализованного почтового ящика.

#### Символы и знаки



В именах почтовых ящиков в формате ASCII разрешается использовать некоторые символы и знаки, например, точку (.), подчеркивание (\_), дефис (-) и плюс (+). Принимая решение о выборе интернационализованного имени почтового ящика, администратор также должен учитывать то, какие знаки или символы Unicode следует разрешить к использованию в поддерживаемых системах письма.

#### **Предложения:**

- Следует рассмотреть возможность применения принципа включения (см. [процедуру правил генерирования меток](#)). Принцип включения заключается в том, чтобы начать с чистого листа, а затем включить в список разрешенных только те символы, которые необходимы пользователям. Символы, которые вызывают проблемы безопасности в контексте конкретной реализации вашей системы, разрешать не следует.
- Разрешайте использовать только те символы и знаки, которые могут читать и вводить ваши пользователи и их корреспонденты. Имя почтового ящика должно быть возможно ввести с помощью обычно используемых редакторов ввода (IME).
- Установите правила, которые в одинаковой степени, понятно и единообразно регулировали бы имена почтовых ящиков, записанные символами латинского алфавита, и интернационализованные имена почтовых ящиков, и предусматривали бы одинаковые ограничения для них.

#### **Рекомендации:**

- Старайтесь не допускать использования математических символов, которые могут приводить к возникновению проблем в области безопасности, таких как # \$ % = " \* < , > .
- Не следует разрешать использование символов эмодзи (отсылка к пояснению SSAC ICANN [SAC95](#)).
- Не следует разрешать сочетания символов, нарушающие принцип осуществимости ввода, например, диакритический знак арабского алфавита хамза (U+0654).

#### **Контрольный список:**

[ ] По умолчанию математические символы должны быть запрещены.

[ ] По умолчанию следует избегать использования знаков и символов, отсутствующих на клавиатурах/устройствах ввода ваших пользователей. Некоторые символы и знаки отсутствуют на устройствах ввода, предназначенных для определенных рынков.

[ ] Вы должны понимать своих пользователей и ожидания на местном рынке. Если для вашего рынка необходимо использование знаков и символов, в набор разрешенных знаков можно добавить такие популярные символы, как точка (.), подчеркивание (\_), дефис (-) и плюс (+).

[ ] При необходимости рассмотрите возможность использовать дополнительные символы и убедитесь, что это не приводит к проблемам в области безопасности для конечных пользователей, в т. ч. таким, как введение пользователей в заблуждение, использование символов из разных систем письма и прочие возможные проблемы.

## **Нормализация символов Unicode**



В стандарте Unicode определена процедура нормализации текста, которая заключается в замене эквивалентных последовательностей символов таким образом, что любые два эквивалентных текста могут быть сведены к одинаковой последовательности кодовых точек.

Unicode определяет нормализованные формы символов, которые могут кодироваться по-разному, поскольку существуют разные типы стандартизированных форм, например, форма нормализации D (NFD) и форма нормализации C (NFC).

Пример: Å может кодироваться как 212B или как 00C5. В обоих случаях такая запись должна обозначать один и тот же почтовый ящик, а не два разных ящика, и может быть нормализована в одинаковые формы NFD и NFC.

Примечание: в этих примерах используются символы латинского алфавита, однако аналогичные примеры существуют и для других систем письма.

Более подробные сведения о других типах нормализации с примерами см. на [сайте консорциума Unicode](#).

## Рекомендации

- Для имен почтовых ящиков должна использоваться форма NFC, потому что она же используется для доменных имен, а также потому, что строки NFC в большинстве случаев получаются короче, чем строки NFD.
- В стандартном программном обеспечении нормализация Unicode обычно уже встроена в библиотеки. Администраторы электронной почты должны знать, какой тип нормализации используется в администрируемых ими системах, чтобы иметь возможность назначать почтовые ящики и обслуживать их.
- Для доменных имен используется форма нормализации C (NFC). При создании системы убедитесь, что используемые библиотеки поддерживают форму NFC для обеспечения единообразия.

## Контрольный список

[ ] Узнайте, какой тип нормализации используется в вашей системе.

[ ] Убедитесь, что ваша программа электронной почты выполняет сравнение имен независимо от типа формы нормализации.

[ ] Если возможно выбрать тип нормализации, выберите NFC, если у вас нет особых причин использовать другие формы.

## Предложения относительно того, что считать одинаковым

Когда метки доменных имен считаются сообществом «одинаковыми» или «взаимозаменяемыми», они называются «вариантными» метками доменных имен. Когда последовательности символов в имени ящика электронной почты считаются сообществом «одинаковыми», в настоящем документе они называются «эквивалентными» именами почтовых ящиков.

Иногда эквивалентные имена полезны, их можно использовать в качестве псевдонимов для основного имени почтового ящика. В других случаях они могут вводить в заблуждение и вызывать проблемы в области безопасности, если они не определены или присвоены другому пользователю.



## Предложения:

- Некоторые службы электронной почты могут считать эквивалентными символы в верхнем и нижнем регистре в адресах электронной почты в формате ASCII.
- Сообщество DNS проделало большую работу, чтобы определить четкие Правила генерирования меток (LGR), включающие также идентификацию вариантных меток в разных системах письма. Ознакомьтесь с этими Правилами генерирования меток, чтобы понять, захотите ли вы использовать их в качестве основы для своего определения эквивалентных имен почтовых ящиков.

## Примеры:

1. Эквиваленты традиционного и упрощенного китайских алфавитов:

名称 — означает «имя», (U+540D U+79F0)

名稱 — означает «имя», (U+540D U+7A31)

名称 (минчэн, означает «имя») используется в КНР, а 名稱 (минчэн, означает «имя») на Тайване и в Гонконге, однако любой представитель китайского сообщества понимает, что это одно и то же.

2. Эквиваленты в арабском алфавите (слово, означающее «сеть» по-арабски):

شبكة (Unicode U+06C3 U+06A9 U+0628 U+0634)

شبكة (Unicode U+0629 U+06A9 U+0628 U+0634)

شبكة (Unicode U+0629 U+0643 U+0628 U+0634)

3. Символы полной и половинной ширины в японской и китайской системах письма. J I M и JIM, 1 2 3 и 123, アイオ и 7イ.

4. Эквиваленты в приведении к единому регистру:

John.Callen

john.callen

John.callen

john.Callen

5. Варианты имени и разделителя (также могут считаться эквивалентными):

john.callen

johncallen

- В контексте системы электронной почты вы можете рассмотреть целесообразность того, чтобы доставлять все сообщения, адресованные эквивалентным именам, в один и тот же ящик электронной почты. Это позволит избежать введения пользователей в заблуждение, использования символов из разных систем письма и прочих проблем в области безопасности.
- Дополнительные эквиваленты можно присваивать одному и тому же почтовому ящику в качестве псевдонимов, однако их может быть очень много (сотни или





даже тысячи), тогда ими будет трудно управлять. Поэтому следует подходить с осторожностью к выбору баланса между пользой и дополнительной нагрузкой.

- Помните, что варианты написания той части адреса электронной почты, которая обозначает доменное имя, могут иметь такой же эффект с точки зрения пользователей. Обсуждение доменных имен и вариантов их написания выходит за рамки настоящего документа.
- Вы должны знать о тех ошибках или недоразумениях, которые могут возникать в связи с вводом пользователями адресов электронной почты для вашей системы. В этом документе затрагиваются возможные ошибки, связанные с той частью адреса, которая обозначает почтовый ящик.
- Путаница может возникать также в отношении правильного написания или нормализации той части адреса, которая обозначает доменное имя. Вопрос вариантов доменных имен выходит за рамки настоящего документа.

### Рекомендации:

- После выдачи пользователю почтового ящика следует идентифицировать эквивалентные ему имена, чтобы отметить их как зарегистрированные и не выдавать их другим пользователям.
- Рассмотрите возможность обрабатывать наиболее полезные эквиваленты имени почтового ящика как его псевдонимы (см. раздел [Предложения относительно псевдонимов](#) настоящего документа) и доставлять почту, адресованную таким эквивалентным ящикам, в исходный почтовый ящик. При этом, однако, необходимо учитывать то, что управление таким ящиком может быть затруднительно, поэтому нужно осторожно и с учетом реальных потребностей проанализировать сравнительные за и против — как удобства такого подхода, так и создаваемую в таком случае нагрузку.
- Рассмотрите практические языковые правила для обработки эквивалентных имен почтовых ящиков. К примеру, для имен почтовых ящиков в формате ASCII практически удобно считать одинаковыми варианты записи в верхнем и нижнем регистре. Для имен почтовых ящиков, записываемых упрощенным письмом китайского языка, таких как 中国, что значит «Китай», может иметь смысл считать эквивалентным запись того же слова традиционным письмом — 中國.
- Все варианты записи, которые возможно привести к одному и тому же варианту путем преобразования регистра, следует считать одним и тем же именем почтового ящика.
- Когда в одном наборе символов (множестве кодовых точек, которые можно использовать для создания имен почтовых ящиков) используется сразу несколько вариантов цифр, для таких цифр нужно выполнять сопоставление, которое позволяло бы считать разные кодовые точки одинаковыми. К примеру, администратор почтового ящика, записываемого на языке хинди, может рассматривать строки abc123 и abc१२३ как эквивалентные варианты записи имени почтового ящика:

Имя почтового ящика: abc123

Эквивалентное имя почтового ящика: abc१२३

### Контрольный список:

[ ] Разработайте политику, которая позволяла бы определять «одинаковые» или эквивалентные имена почтовых ящиков и которую можно было бы применять к вашей



системе электронной почты с учетом используемой системы письма, ожиданий пользователей и технических возможностей вашей реализации. В такой политике необходимо будет учитывать изложенные выше соображения относительно Правил генерирования меток, преобразования регистров, разделителей, цифр и символов.

[ ] Определите, отвечают ли требованиям к определению эквивалентности имен почтовых ящиков, которые вы хотите использовать, [Справочные правила генерирования меток на втором уровне](#). Если они не отвечают вашим требованиям, доработайте их по необходимости.

[ ] Дайте пользователям возможность ознакомиться с вашими правилами, чтобы они могли понимать, какие символы и сочетания будут считаться допустимыми и какие последовательности символов могут быть эквивалентны друг другу.

[ ] При использовании эквивалентных последовательностей символов подумайте, хотите ли вы использовать инструмент проверки Правил генерирования меток в готовом виде или же развернуть его у себя на внутреннем сервере для оптимизации рабочих процессов и защиты обрабатываемых им имен почтовых ящиков от передачи за пределы вашей сети.

### Прочие предложения

Та часть адреса электронной почты, которая обозначает интернационализированное доменное имя, может записываться символами Unicode или же соответствующей им строкой из символов ASCII, начинающейся с «xn--».

### Предложения:

- Вариант записи доменного имени в виде строки «xn--» — это низкоуровневая техническая особенность. В идеале пользователю вообще не нужно ее видеть. Однако в переходной период, когда не все системы электронной почты поддерживают интернационализированные адреса, запись «xn--» можно использовать как полезный прием, позволяющий обойти эту проблему.

### Рекомендации:

- Записывайте доменные имена в виде интернационализированных строк, а не строк ASCII. Старайтесь не отображать альтернативный вариант записи в виде строки «xn--».
- Некоторые клиенты электронной почты не могут автоматически сопоставлять формы имен ящиков электронной почты в виде U-меток и A-меток, так что убедитесь, что в вашей системе эти метки сопоставляются друг другу.

## Словарь

*Псевдоним:* адрес электронной почты, перенаправляющий сообщения на тот же ящик, что и другой адрес электронной почты. Обычно псевдоним — это адрес электронной почты в формате ASCII, представляющий собой альтернативный вариант записи интернационализированного адреса электронной почты на тот случай, если корреспондент не может отправить сообщение на интернационализированный адрес, потому что система доставки не поддерживает интернационализацию адресов электронной почты.



*ASCII*: ограниченный набор символов, состоящий из букв латинского алфавита без диакритических знаков, цифр и нескольких символов пунктуации. Термин ASCII — это используемая в компьютерной отрасли аббревиатура.

*Отображаемое имя*: имя, которое отображается рядом с адресом электронной почты.

*Доменное имя*: последовательность имен, которые используются для обозначения того или иного веб-сайта или сервера электронной почты, например, uasg.tech или icann.org.

*EAI*: интернационализация адреса электронной почты.

*Интернационализация адреса электронной почты*: практика записи адресов электронной почты с использованием всех возможных языков и систем письма, а не только символов ASCII или латинского алфавита.

*Эквивалент*: две или более кодовые точки с одинаковым значением или признанные местными властями одинаковыми с точки зрения языка и образующие т. н. «варианты символов». К примеру, такие варианты встречаются в упрощенном и традиционном китайском письме и потенциально могут вводить в заблуждение пользователей, использующих эти системы письма.

*IDN*: интернационализированное доменное имя.

*IME (редактор ввода)*: IME — это приложение, позволяющее использовать стандартную клавиатуру (например, со 101 клавишей с раскладкой США) для ввода символов и знаков, которые не представлены непосредственно на самой клавиатуре.

*Интернационализированное доменное имя*: доменное имя, записанное символами, отличными от символов ASCII.

*Правила генерирования меток*: правила использования символов в доменных именах.

*LGR*: правила генерирования меток.

*Имя почтового ящика*: часть адреса электронной почты, обозначающая почтовый ящик в пределах того или иного доменного имени. Может использоваться в ряду других взаимозаменяемых терминов с тем же значением, таких как «имя почтового ящика», «имя учетной записи», «имя пользователя», «часть UTF-8», «локальная часть».

*Punycode*: алгоритм, позволяющий закодировать строку Unicode с использованием ограниченного набора символов ASCII, используется для записи имен хостов Интернета.

*Фаза 1 EAI*: система электронной почты, позволяющая получать и отправлять сообщения электронной почты на интернационализированные адреса, но не поддерживающая создавать и размещать (хостинг) почтовых ящиков с интернационализированными именами.

*Фаза 2 EAI*: система электронной почты, позволяющая успешно создавать и размещать почтовые ящики с интернационализированными именами, а также



получать и отправлять сообщения электронной почты на интернационализованные адреса.

*Преобразование*: процесс, посредством которого компьютерная система превращает последовательность символов в их графическое изображение на экране.

*Система письма*: набор правил и символов, которые используются для записи в одном или нескольких языках.

*Script.txt*: файлы данных, которые используются в базе данных символов Unicode, для обозначения значений свойств всех кодовых точек алфавита (<http://www.unicode.org/Public/12.1.0/ucd/Scripts.txt>).

*Unicode*: стандарт кодирования символов, охватывающий большинство существующих в мире систем письма и языков, имеющих письменность. См. <http://unicode.org>.

*Вариантные имена*: имена, которые считаются сообществом «одинаковыми».