

**Рекомендации по присвоению имен
интернационализованных адресов электронной
почты**

Оглавление

Введение	3
Глоссарий	3
Поддерживаемые системы письма	4
Длина имени почтового ящика в поле ввода	4
Сочетание разных систем письма	5
Предотвращение использования недопустимых последовательностей символов и последовательностей символов, отображаемых нестабильно	6
Предложения относительно псевдонимов и отображаемых имен	7
Символы и знаки	8
Нормализация символов Unicode	8
Вариантность и эквивалентность	9

Введение

Настоящие рекомендации предназначены для администраторов электронной почты, системных администраторов и ИТ-менеджеров. Их задача — собрать в одном месте и перечислить основные предложения относительно тех аспектов, которые необходимо учитывать при создании интернационализированных адресов электронной почты, а также используемых для них псевдонимов и отображаемых имен.

Настоящие рекомендации разработаны на основе адаптированного перевода на русский язык документа UASG028 “Considerations for Naming Internationalized Email Mailboxes”¹, подготовленного Группой управления по универсальному принятию (Universal Acceptance Steering Group).

Глоссарий

Имя почтового ящика: часть адреса электронной почты перед символом @ или 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 почтовой учетной записи», «имя пользователя почты», «локальная часть адреса 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. В адресе электронной почты часть адреса после символа @.

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

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

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 интернационализованного 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: система кодировки, включающая в себя ограниченный набор символов, состоящий из букв латинского алфавита без диакритических знаков, цифр и нескольких символов пунктуации.

Unicode: стандарт кодирования символов, охватывающий большинство существующих в мире систем письма и языков, имеющих письменность.

Кодовая точка: значение кодовой позиции в системе кодировки (ASCII, Unicode и т.д.).

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.

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

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

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

¹ <https://uasg.tech/wp-content/uploads/documents/UASG028-en-digital.pdf>

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, латинская система письма или латиница — это набор символов от A до Z. Расширенная латиница содержит 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. Кириллица – набор символов от А до Я, используемых в русском языке, а также дополнительные символы, использующиеся в украинском, белорусском, сербском языках, в различных языках малых народов северной части России и т.д.

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 остальными рекомендациями, 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

Учитывайте длину имени почтового ящика в символах. Определитесь какая минимальная и максимальная длина будет разрешена.

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. Если это так, то скорее всего 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.

Стоит учитывать, что в формате UTF8 один символ может быть представлен несколькими байтами, что приведет к фактическому увеличению длины имени почтового ящика. К примеру, в формате UTF-8 имя почтового ящика, состоящее из символов ASCII, использует такое же число байт, что и число символов в нем, кириллические символы кодируются двумя байтами, а символы других систем письма могут достигать и четырех байт.

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.

Сочетание разных систем письма создает проблемы с безопасностью и приводит к введению пользователей в заблуждение. К примеру, на следующем изображении показаны несколько похожих символов, имеющих различное значение кода в кодировке Unicode, которые могут вводить пользователей в заблуждение.

a a a a a a a a a a a
b Ъ Ь Ъ b b b b b b
c c c c c c c c c c c
d d d d d d d d d d d
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.

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 безопасности рекомендуется не допускать в именах почтовых ящиков смешения разных языков, использующих одну и ту же систему письма, 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 рекомендуется не допускать без необходимости.

К примеру, имя почтового ящика на русском языке стоит ограничить только кириллическими символами, а в имени почтового ящика на японском можно разрешить использовать кандзи, хирагана, катакана и латинские символы.

Пример 1: Адрес электронной почты на русском языке: *имя@домен.дву*

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 в пределах вашей политики построения адресов электронной почты возможные угрозы введения пользователей в заблуждение, ошибок или вредоносных действий в результате сочетания разных систем письма в пределах имени почтового ящика или в имени ящика и в 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 допустимые 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.

Рассмотрим для примера последовательности, записанные тайскими символами: ๓ и ๔ . Первая строка, $\text{๓ U+0E14(๓) + U+0E35 (๓)} + \text{U+0E35 (๓)}$ недопустима, потому что выглядит так же, как $\text{๓ U+0E14(๓) + U+0E35 (๓)}$. Сообщество пользователей тайского алфавита определило правило, согласно которому символ U+0E35 не может следовать после символа U+0E35.

Набор таких правил для определенной системы письма, используемый в системе DNS, называются Правилами генерирования меток² (Label Generation Rules, LGR). Возможно, будет уместно применить эти правила к почтовым адресам в вашей системе.

Однако Правила генерирования меток могут накладывать на вашу систему больше ограничений, чем это необходимо. К примеру, «Правила генерирования меток корневой зоны³» запрещают использовать цифры, однако нет ничего плохого в том, чтобы использовать цифры в именах почтовых ящиков. Для некоторых систем письма Правила генерирования меток еще не разработаны, поэтому опереться на них не получится.

Вы можете использовать существующие «Правила генерирования меток на втором уровне⁴» (Second Level Reference LGRs) для проверки имен почтовых ящиков. Средство проверки на основе Правил генерирования меток доступно онлайн <https://lgrtool.icann.org/>. Исходный текст

² <https://www.icann.org/resources/pages/lgr-proposals-2015-12-01-en>

³ <https://www.icann.org/resources/pages/root-zone-lgr-2015-06-21-en>

⁴ <https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en>

этого средства также доступен на платформе GitHub, откуда его можно загрузить для интеграции в вашу систему.

Если интересующая вас система письма не описана в «Правилах генерирования меток на втором уровне», вы можете воспользоваться «Правилами генерирования меток для корневой зоны» дополнив их при необходимости или следовать общим принципам, описанным в стандарте RFC6912.

Правила использования цифр в именах, записываемых справа налево, можно найти в стандарте 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, «Правила генерирования меток на втором уровне» или «Правилами генерирования меток для корневой зоны». Если отвечают, воспользуйтесь ими, 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.

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 псевдонимов и отображаемых имен

Некоторые системы электронной почты предоставляют пользователям удобные возможности заводить несколько разных адресов электронной почты или несколько псевдонимов, перенаправляющих почту на один и тот же почтовый ящик. Такие псевдонимы и отображаемые имена могут использоваться в разных сценариях, связанных с интернационализацией адресов электронной почты. Например, для двуязычных визитных карточек с использованием как латинского, так и местного алфавита. Использование псевдонимов, использующих систему письма, отличную от той, в которой записано имя почтового ящика, удобно, когда интернационализованный адрес электронной почты используется в разных языковых сообществах

Пример кириллического псевдонима для адреса электронной почты на английском:

имя@домен.дву [name@domain.tld](#)

Псевдоним Адрес электронной почты

Стоит рассмотреть возможность предоставления адресов электронной почты, записываемых 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 интернационализированным 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 т.д. 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.

Имеет смысл рассмотреть возможность применения принципа включения (см. Inclusion Principle дrafта ICANN “Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels”)⁵, который заключается в том, что каждый символ априорно исключен и должен быть рассмотрен, а затем включен в список разрешенных только при условии, что он необходим пользователям и его использование считается безопасным. Символы, которые вызывают проблемы безопасности в контексте конкретной реализации вашей системы, разрешать не следует. Также стоит установить правила, которые в одинаковой степени, понятно и единообразно регулировали бы имена почтовых ящиков, записанные символами латинского алфавита, и интернационализованные имена почтовых ящиков, и по возможности предусматривали бы одинаковые ограничения для них в плане разрешенных символов. Не следует, без весомой причины допускать использования математических символов, которые могут приводить к возникновению проблем в области безопасности, таких как # \$ % = “ * < , > +.

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 стран.

[] Следует изучить нужды своих пользователей. Если им необходимо использование каких-либо знаков и символов, то их можно добавить в набор разрешенных, при условии, что это не создает угрозы безопасности вашей системе и ее пользователям.

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. Формы нормализации с примерами можно найти на [сайте консорциума Unicode](#).

Для имен почтовых ящиков рекомендуется использовать форму нормализации, при которой все буквы по возможности объединяются в одну. Она называется композицией (Normalization Form C или NFC) и используется для доменных имен. В стандартном программном обеспечении нормализация Unicode обычно уже встроена в библиотеки. Администраторы

⁵ <https://www.icann.org/en/system/files/files/draft-lgr-procedure-20mar13-en.pdf>

электронной почты должны знать, какой тип нормализации используется в администрируемых ими системах, чтобы иметь возможность назначать почтовые ящики и обслуживать их.

Контрольный список

Узнайте, какой тип нормализации используется в вашей системе.

Убедитесь, что ваша программа электронной почты выполняет корректное сравнение имен независимо от формы нормализации.

Если возможно выбрать форму нормализации, рекомендуется выбрать NFC, если у вас нет особых причин использовать другие формы.

Вариантность и эквивалентность

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», 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.

Правила генерирования меток, включают в себя также идентификацию вариантных меток в разных системах письма. Ознакомьтесь с этими правилами, чтобы понять, использовать ли их в качестве основы для своего определения эквивалентных имен почтовых ящиков. Необходимо оценить 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. Однако их может быть настолько много, что ими будет трудно управлять. Поэтому следует подходить с осторожностью к выбору баланса между пользой и дополнительной нагрузкой.

При создании пользователю почтового ящика следует идентифицировать эквивалентные ему имена, чтобы отметить их как зарегистрированные и не выдавать их другим пользователям. Наиболее полезные эквиваленты имени почтового ящика можно сделать псевдонимами. Все варианты записи, которые возможно привести к одному и тому же варианту путем преобразования регистра, следует считать одним и тем же именем почтового ящика.

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 вашего продукта. В такой политике необходимо будет учитывать изложенные выше соображения относительно Правил генерирования меток, преобразования регистров, разделителей, цифр и символов.

Определите, соответствуют ли требованиям к определению эквивалентности имен почтовых ящиков, которые вы хотите использовать, Правилам генерирования меток. Если они не соответствуют, 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.