



Универсальное
принятие
Краткое руководство

Что означает «универсальное принятие»?



ПРИНЯТИЕ



ПРОВЕРКА



ХРАНЕНИЕ



ОБРАБОТКА



ОТОБРАЖЕНИЕ

Программное обеспечение и интернет-сервисы обеспечивают универсальное принятие, когда предлагают перечисленные выше возможности для всех доменных имен и адресов электронной почты.

Универсальное принятие (UA) — состояние, когда все интернет-ориентированные приложения, устройства и системы правильно и единообразно принимают, проверяют, хранят, обрабатывают и отображают все корректные доменные имена и адреса электронной почты.

Из-за быстрого изменения пространства доменных имен многие системы не распознают или не обрабатывают надлежащим образом новые доменные имена, главным образом потому, что эти имена могут быть длиннее трех символов или иметь отличный от ASCII формат. То же самое справедливо для содержащих эти новые расширения адресов электронной почты.

Созданная Интернет-корпорацией по присвоению имен и номеров (ICANN) Группа управления по универсальному принятию (UASG) — это возглавляемая сообществом общепромышленная инициатива по повышению информированности, а также выявлению и решению проблем, связанных с универсальным принятием доменных имен. Цель этих усилий в том, чтобы содействовать созданию единообразного интерфейса и положительного опыта взаимодействия с пользователями интернета во всем мире.

Дополнительные сведения о UASG и последних достижениях представлены здесь:

www.uasg.tech.

Обратите внимание на то, что принятие, проверка и обработка в настоящем документе рассматриваются как отдельные этапы. На практике эти функциональные возможности могут перекрываться.

ПРИНЯТИЕ



Принятие — это процесс получения адреса электронной почты или доменного имени в виде строки символов из пользовательского интерфейса, файла или API (интерфейса программирования приложений), используемого программным приложением или интернет-сервисом.

Рекомендации UASG

- Любой пользовательский интерфейс, требующий от пользователя ввести доменное имя или адрес электронной почты, обязательно должен поддерживать Unicode и строки длиной до 256 символов.
- Пользователям следует разрешать, но не требовать, вводить текст в ASCII-совместимой кодировке (также называемой Punycode) вместо эквивалента Unicode. Однако для отображения по умолчанию следует использовать Unicode, а текст в кодировке Punycode показывать пользователю только тогда, когда это приносит выгоду.

ПРОВЕРКА



Процесс, используемый для проверки синтаксической правильности полученного или отправленного адреса электронной почты или доменного имени. Многие программисты обучены выполнять проверку, следуя набору эвристических алгоритмов, которые требуют, чтобы в домене верхнего уровня было «правильное» количество букв или чтобы эти буквы входили в множество символов ASCII. Эти эвристические алгоритмы теперь неприменимы вследствие внедрения доменных имен, имеющих длину более трех символов и содержащих символы Unicode (отличные от ASCII).

Рекомендации UASG

- Проверка должна выполняться только в тех случаях, когда она необходима для функционирования приложения или сервиса. Это самый простой способ обеспечить принятие системой всех корректных доменных имен.
- Если проверка необходима, уделите внимание следующему:
 - ▶ Проверьте соответствие той части доменного имени, которая представляет TLD, официальной таблице:
 - <http://www.internic.net/domain/root.zone>
 - <http://www.dns.icann.org/services/authoritative-dns/index.html>
 - <http://data.iana.org/TLD/tlds-alpha-by-domain.txt>
 - ▶ См. также SAC070: <https://tinyurl.com/sac070>.
 - ▶ Отправьте в систему доменных имен (DNS) запрос для проверки доменного имени.
 - ▶ Для предотвращения опечаток требуйте повторного ввода адреса электронной почты.
 - ▶ Проверяйте символы в метках только в том объеме, который позволяет установить, что U-метка не содержит запрещенных элементов кода «DISALLOWED» или таких, которые не определены в этой версии Unicode. Посетите веб-страницу <https://tools.ietf.org/html/rfc5892>.
 - ▶ Ограничьте проверку меток небольшим количеством правил для метки целиком, сформулированных в документах «Запрос комментариев» (RFC). Посетите веб-страницу <https://tools.ietf.org/html/rfc5894>.
 - ▶ Если похожая на доменное имя строка содержит идеографический символ точки «.», перед проверкой его следует преобразовать в «.».

ХРАНЕНИЕ



Хранение относится к долгосрочному и/или кратковременному хранению доменных имен и адресов электронной почты. Независимо от периода существования данных, они должны (предпочтительно) храниться в форматах, определяемых стандартами RFC, или в других форматах, которые могут быть преобразованы в форматы RFC.

Рекомендации UASG

- Приложения и сервисы должны поддерживать применимые стандарты Unicode.
- Информацию следует хранить в UTF-8 (формат преобразования Unicode) по мере возможности. Кроме того, некоторым системам может потребоваться поддержка UTF-16, но в большинстве случаев формат UTF-8 предпочтительнее. Форматов UTF-7 и UTF-32 следует избегать.
- Примите во внимание все комплексные сценарии перед преобразованием A-меток в U-метки и наоборот во время хранения. Возможно, целесообразно хранить в файле или базе данных только U-метки, поскольку это упрощает поиск и сортировку. Однако преобразование может оказать влияние при взаимодействии с более старыми приложениями и сервисами, не поддерживающими Unicode. Рассмотрите возможность хранения в обоих форматах.
- Используйте четкие обозначения для удобного доступа к адресам электронной почты и доменным именам во время хранения. В тех случаях, когда адреса электронной почты и доменные имена хранятся в поле «author» (автор) документа или «contact info» (контактная информация) в файле журнала, исчезает понимание того, что это адрес.

ОБРАБОТКА



Обработка происходит всякий раз, когда адрес электронной почты или доменное имя используется приложением или сервисом для выполнения действия (например, поиска или сортировки списка) или преобразуется в альтернативный формат (например, при хранении ASCII как Unicode). Во время обработки также может осуществляться дополнительная проверка.

Доменные имена и адреса электронной почты можно обрабатывать неограниченным числом способов*, что увеличивает потребность в соглашениях, гарантирующих единообразие трактовки и классификации данных.

Рекомендации UASG

- Поскольку стандарт Unicode постоянно расширяется, следует проверить те элементы кода, которые не были определены в момент создания приложения или сервиса, и убедиться, что при взаимодействии с пользователем не будет «перебоев». Результатом отсутствия шрифтов в базовой операционной системе могут стать не воспроизводимые на экране символы (часто для их представления используется символ « »), однако такая ситуация не должна приводить к аварийному завершению работы.
- Используйте поддерживаемые API, совместимые с Unicode.
- Используйте для интернационализированных доменных имен (IDN) действующую редакцию документов протокола [<http://tools.ietf.org/html/rfc5891>] и таблиц [<http://tools.ietf.org/html/rfc5892>] интернационализированных доменных имен в приложениях (IDNA).
- По мере возможности выполняйте обработку в формате UTF-8.
- Убедитесь, что продукт или функция обрабатывает числа так, как это предусмотрено. Например, числовые символы ASCII и азиатские иероглифы, представляющие числа, должны обрабатываться как числа. [RFC5892, см. ссылку выше]
- Обновляйте приложения и серверы/сервисы одновременно. Если сервер использует Unicode, а клиентская часть нет, или наоборот, придется преобразовывать данные из одной кодовой страницы в другую при выполнении каждой операции передачи данных между сервером и клиентской частью.
- Выполняйте анализ кода, чтобы избежать атак переполнения буфера. При преобразовании символов текстовые строки могут существенно удлиняться или сокращаться.

* Примеры: идентификация людей в Новой Зеландии путем поиска в ccTLD .nz; идентификация фармацевтов путем поиска адресов электронной почты user@*.pharmacist.

ОТОБРАЖЕНИЕ



Отображение происходит всякий раз, когда адрес электронной почты или доменное имя обрабатывается в пользовательском интерфейсе. Отображение доменных имен и адресов электронной почты обычно не вызывает затруднений, когда используемые алфавиты поддерживаются базовой ОС, а строки хранятся в Unicode; однако в других случаях может возникнуть необходимость преобразований, зависящих от типа приложения.

Рекомендации UASG

- Отображайте все элементы кода Unicode, которые поддерживает базовая операционная система. Если приложение использует собственный набор шрифтов, необходимо обеспечить полную поддержку Unicode для коллекции шрифтов, доступных в операционной системе.
- При разработке приложения или сервиса, а также при управлении регистратурой принимайте во внимание поддерживаемые языки и обеспечивайте охват этих языков ОС и приложениями.
- Перед отображением выполняйте преобразование отличных от Unicode данных в формат Unicode. Например, конечный пользователь должен увидеть строку «everyone.みんな», а не строку «everyone.xn--q9jyb4c». (Это преобразование – пример обработки, обеспечивающей UA-готовность).
- По умолчанию отображайте Unicode. Используйте для пользовательского интерфейса текст в кодировке Punycode только тогда, когда это приносит выгоду. Для смягчения последствий дополните отображаемый Unicode текстом в кодировке Punycode при наведении курсора.
- Учтите, что адреса, в которых используется смесь алфавитов, будут встречаться все чаще. Некоторые символы Unicode могут выглядеть одинаковыми для человеческого глаза, но разными для компьютеров. Не следует считать строки, где используется смесь алфавитов, предназначенными для вредоносных целей, таких как фишинг, и если пользовательский интерфейс обращает внимание пользователя на такие строки, обеспечьте отсутствие при этом предвзятости по отношению к пользователям алфавитов, отличных от латинского. Дополнительные соображения безопасности Unicode изложены здесь: <http://unicode.org/reports/tr36/>.
- Обеспечьте при обработке Unicode совместимость с IDNA, поскольку это отвечает ожиданиям пользователей. Дополнительные сведения представлены здесь: <http://unicode.org/reports/tr46/>.
- Помните о наличии не назначенных и запрещенных символов. Дополнительные сведения представлены в RFC 5892: <https://tools.ietf.org/rfc/rfc5892.txt>.

Подготовьтесь к универсальному принятию

Проверка исходного кода и модульное тестирование

Процесс проверки исходного кода и подтверждения того, что используются только правильные методы программирования, программные библиотеки и интерфейсы (также называемые API). После этого администратор может проверить работоспособность приложения или сервиса, убедившись в наличии конкретных возможностей (принятие, проверка и т. д.), перечисленных выше. Этот способ обычно используется только разработчиками программного обеспечения и поставщиками интернет-сервисов.

В рамках своих усилий по повышению информированности, группа UASG устанавливает прямые контакты с разработчиками приложений и крупнейшими поставщиками интернет-сервисов, чтобы побудить их к анализу и тестированию исходного

кода на предмет универсального принятия, и распространяет перечень критериев, которые можно использовать для разработки сценариев тестирования.

Ручное тестирование

Требует выполнения многочисленных проверок новых и отличных от ASCII доменов, например, отправки адреса электронной почты при регистрации в интернет-сервисе и подтверждения его принятия. Поскольку существует бесчисленное множество интернет-сервисов с возможностью регистрации, а также потенциальных комбинаций новых адресов электронной почты, этот способ требует проверки приложений, сервисов, адресов электронной почты и/или доменных имен в различных сочетаниях, чтобы представить широкий спектр сценариев использования. Этот способ проверки доступен любому, но является наиболее трудоемким.

UASG также содействует внедрению данного способа, составляя список наиболее часто используемых веб-сайтов, приложений, адресов электронной почты и доменных имен, подходящих для тестирования.

Автоматизированное тестирование

Применение автоматизированных сценариев или директив для тестирования множества URL-адресов. Этот способ требует большего объема предварительной технической работы, но лучше поддается масштабированию при реализации крупных программ измерения и контроля. Практическим примером является недавнее исследование gTLD, выполненное Сетевым информационным центром стран Азии и Тихоокеанского региона (APNIC) по поручению ICANN: <https://tinyurl.com/new-gtld-ua>.

UASG изучает методы автоматизированного тестирования универсального принятия и проинформирует о полученных результатах.

