

# Глобальный анализ приема адресов электронной почты сайтами в 2019 году

Группа управления по универсальному принятию (UASG),  
Ассоциация компаний-разработчиков ПО Бразилии  
(Associação Brasileira das Empresas de Software, ABES),  
и Governance Primer

9 августа 2019



## СОДЕРЖАНИЕ

<b>Введение</b>	<b>3</b>
<b>Основные положения</b>	<b>4</b>
<b>Исследовательская группа</b>	<b>6</b>
<b>Методология</b>	<b>7</b>
Порядок работы	7
Схема метаданных	10
<b>Результаты</b>	<b>13</b>
<b>Анализ и корреляция</b>	<b>15</b>
Случаи неподтверждения выборки	15
Непринятие всех адресов электронной почты	15
Тестирование региональной корреляции	16
<b>HTML5: препятствие на пути к универсальному принятию</b>	<b>17</b>
<b>Можно ли автоматизировать эти отчеты?</b>	<b>19</b>
<b>Заключение</b>	<b>20</b>



## Введение

Цель обеспечения универсального принятия (UA) -- возможность использования всех доменных имен и адресов электронной почты всеми интернет-ориентированными приложениями, устройствами и системами. Это касается как новых доменов общего пользования верхнего уровня (gTLD), так и тех, в которых используется нелатинский алфавит. Хотя можно предположить, что эти домены функционируют аналогично историческим gTLD, это не так, и проблемы с совместимостью все еще более широко распространены, чем следовало бы.?

Наша цель заключается в том, чтобы этот адрес электронной почты:  
**测试1@server.technology**

и этот адрес:  
**دون@رسيل.السعودية**

имели такие же показатели принятия, как и этот:  
**user@test.org**

Настоящее исследование было заказано Группой управления по универсальному принятию (UASG) в продолжение аналогичного исследования 2017 года<sup>1</sup> в рамках общей инициативы по углублению понимания сообществом препятствий и основных проблем, связанных с повсеместной совместимостью существующих доменных имен.

Задача исследования – проанализировать показатели поддержки универсального принятия (UA-совместимости) в 1000 ведущих сайтов мира (по данным Alexa<sup>2</sup>) путем тестирования различных подходов и методологий в поле «электронная почта» в веб-формах и их проверки на практике. Эти формы электронной почты хорошо подходят для тестирования различных аспектов универсального принятия, поскольку дают возможность проверить разные критические точки, в том числе стандарты и реализацию HTML, а также использование JavaScript и других веб-ориентированных языков.

Очевидно, что многие разработчики не принимают во внимание новые сценарии использования. В итоге люди, желающие использовать инновационные адреса или писать на родном языке, вынуждены заводить запасной адрес электронной почты на случай, если обработка их приоритетного адреса окажется невозможной. Необходимо понять, где существуют эти проблемы с конечной целью их устранить.

В остальной части доклада приведена информация о методологии исследования и кратко описаны результаты, проблемы и возможные решения.

---

<sup>1</sup> <https://uasg.tech/wp-content/uploads/2017/09/UASG-Report-UASG017.pdf>

<sup>2</sup> <https://www.alexa.com/topsites>



## Основные положения

После того, как были отфильтрованы фиктивные домены, к тестированию оказались пригодными 527 из 1000 ведущих сайтов в списке Alexa. Группа продолжила отбор сайтов по списку до позиции 1922, чтобы иметь тысячу пригодных к тестированию сайтов. После этого была проведена оценка в ручном режиме для проверки принятия следующих адресов форматов возрастающей сложности:

ascii@ascii.newshort test@test.exp	ascii@ascii.newlong test@test.example	ascii@idn.ascii test@普遍接受-测试.org
unicode@ascii.ascii 测试1@test.org	unicode@idn.idn 测试5@普遍接受-测试.世界	arabic.arabic@arabic (RTL) دون@رسيل.السعودية

Затем результаты были помещены в базу данных MongoDB для удобства в виде метаданных по схеме, созданной при разработке методологии, а результаты были представлены в докладе, составленном на основе этой работы. С учетом различий между показателями принятия в 2019 и 2017 годах, мы наблюдаем обнадеживающую динамику роста этих показателей в отношении одной из частей миссии UA:

Тестовый пример	2017	2019
ascii@ascii.newshort	91%	97%
ascii@ascii.newlong	78%	84%
ascii@idn.ascii	45%	50%
unicode@ascii.ascii	14%	13%
unicode@idn.idn	8%	8%
arabic.arabic@arabic (RTL)	8%	7%

Некоторые расхождения можно объяснить применением разных наборов данных при проведении этих тестов, так как в них использовались списки ведущих сайтов с том виде, в котором они существовали на этапе сбора данных для соответствующего теста. Это значит, что показатели принятия некоторых тестовых примеров не понизились, а, скорее, не изменились. При этом несомненно улучшились показатели принятия новых доменов и интернационализированных доменных имен (IDN) на втором уровне, что представляет большой интерес.

Наиболее значимой проблемой, обнаруженной нами, был элемент `<input type="email">`, представляющий собой ключевое поле в HTML5 для обработки подобных данных, на которое опирается значительное количество сайтов. Факт: это поле не соответствует принципам универсального принятия. Мы считаем, что главным приоритетом соответствующих заинтересованных сторон должна стать работа над улучшением этого стандарта. Это поможет повысить показатели поддержки принятия адресов в целом, в особенности, если готовые к универсальному принятию сайты



смогут вместо этого использовать элемент `<input type="eaiemail">`, сигнализирующий о способности принимать подобные адреса.



## Исследовательская группа

**[Куратор] Пауло Миллиет Роке (Paulo Milliet Roque)** — ветеран технологической отрасли с опытом работы в сфере международной торговли и ведения переговоров с более 100 компаний в ряде стран. Соучредитель и вице-президент ABES, Ассоциации компаний-разработчиков ПО Бразилии. Ассоциация ABES была основана в 1986 году, является наиболее репрезентативной организацией сектора и объединяет около двух тысяч предприятий и компаний.

**[Координатор] Марк У. Датисгелд (Mark W. Datysgeld)** — посол UASG, имеет степени бакалавра и магистра в области международных отношений, специализируется на вопросах управления интернетом и влиянии информационных технологий на выработку государственных политик и политик частных компаний. Работая от имени бренда Governance Primer, он предоставляет консалтинговые услуги предприятиям и частным лицам по вопросам их участия в международных организациях и мероприятиях, связанных со сферой информационных технологий.

**[Ведущий тестер] Савио Винисиус де Мораис (Sávio Vinícius de Moraes)** — участник и посол программы NextGen ICANN, занимается исследованиями в области безопасности со специализацией на ИВ в среде SOHO (малый/домашний бизнес) с целью минимизации последствий массивных атак для ИВ-устройств. Активно участвует и обладает опытом работы в сфере разработки веб-приложений и системного администрирования.

**[Тестер] Эдсон Селио Феррейра Араухо (Edson Celio Ferreira Araujo)** — студент, занимается изучением компьютерной техники и разработкой систем в компании Grendene S/A. В свободное время принимает участие в проектах с открытым исходным кодом и работает в области управления интернетом. Выпускник программы Youth@IGF.

**[Тестер] Йонас Мендес Фиорини (Jonas Mendes Fiorini)** — молодой специалист в области информационных технологий, в настоящее время занимается изучением компьютерной техники в Федеральном университете Эспириту-Санту (UFES). Участвует в проектах, посвященных доступности цифровых технологий и организации сетей, с 2012 года — энтузиаст бесплатного ПО. Фиорини — выпускник программы Youth@IGF.

**Нам оказывают ценную поддержку:** Дон Холландер (Don Hollander), Аджай Дата (Ajay Data), Сармад Хуссаин (Sarmad Hussain), Дженнифер Чанг (Jennifer Chung), Нивалдо Клето (Nivaldo Cleto), Родриго де ла Парра (Rodrigo de la Parra), Даниэл Финк (Daniel Fink), Седа Акбулут (Seda Akbulut), Ванда Скартезини (Vanda Scartezini), Рубенс Кул (Rubens Kuhl), NIC.br.



## Методология

### Порядок работы

Наша методология основана на документе «UASG017. Анализ приема различных адресов электронной почты сайтами». Документ представляет собой первое широкомасштабное исследование, направленное на проверку соответствия принципам универсального принятия и проведенное в сентябре 2017 года. Настоящее исследование 2019 года проводится для дальнейшего изучения основных компонентов предыдущего с целью усовершенствования примененных методов.

Было принято важное методологическое решение не использовать пул сайтов исследования 2017 года. Хотя сравнение показателей соответствия сайтов принципам универсального принятия и имеет высокую ценность, использование того списка не было бы оценкой состояния сайтов, занимающих ведущие позиции в мире на текущий момент, а было бы совсем другим исследованием, не менее значимым.

Также были некоторые расхождения с точки зрения порядка составления списка, в том числе тот факт, что окончательный набор данных предыдущего исследования содержал приблизительно 750 единиц, а версия 2019 года – 1000. Было бы нереалистичным включать дополнительные записи в набор данных 2017 года задним числом, и наиболее целесообразным вариантом в этом случае было использовать новую базу данных. В будущем можно будет использовать любой из этих вариантов.

Первым шагом настоящего исследования было составить список URL-адресов 1000 сайтов, занимающих лидирующие позиции в мире в соответствии с конкурентным анализом Alexa, выполненным в начале 2019 года. Отобранные сайты должны были соответствовать следующим критериям: не быть вредоносным ПО и иметь на своих страницах поле ввода адреса электронной почты. Контент не принимался во внимание как фактор, то есть все сайты рассматривались как совершенно равные, вне зависимости от их темы.

Из первоначального списка из 1000 сайтов, занимающих ведущие позиции в списке 1000 топовых, только 527 соответствовали критериям исследования, вследствие чего группа продолжила перебирать список, чтобы включить в него дополнительных кандидатов и дошла 1922-го сайта, который был включен в базу данных последним. Было бы более точным назвать этот набор данных «1000 ведущих сайтов из списка Alexa, пригодных для тестирования».

Группа также проверяла национальную принадлежность сайтов при помощи актуальных на тот момент записей WHOIS, чтобы получить представление о географическом охвате исследования. Это помогло выявить тот факт, что большая часть сайтов из списка находятся в западном полушарии, в основном в США. В восточном полушарии наиболее представительными оказались Китай, Россия, Япония, Индия, Южная Корея, Тайвань и Гонконг.



Полная разбивка по кодам стран приведена ниже.

Обратите внимание, что всего в списке меньше 1000 единиц в связи с невозможностью определить происхождение небольшого числа сайтов.

AE	3	CN	47	HK	8	LV	1	AS	2	VC	1
AM	1	CR	1	ID	7	MA	2	SC	1	VE	1
AR	5	CY	8	IE	3	MU	1	SE	3	VG	4
AT	1	CZ	11	IL	6	MX	2	SG	2		
AU	13	DE	29	IM	1	NG	1	SI	1		
BA	1	DK	3	IN	16	NL	4	SK	1		
BD	1	DO	1	IR	7	NO	2	TH	2		
BE	1	EC	1	IT	14	PA	35	TN	1		
BG	1	EG	2	JP	24	PE	1	TO	3		
BR	13	ES	18	KE	1	PH	3	TR	9		
BS	10	FR	26	KR	11	PL	10	TW	12		
BY	1	GB	9	KZ	2	PT	4	UA	2		
CA	30	GI	1	LA	1	RO	2	UK	23		
CH	2	GR	3	LU	13	RU	33	US	381		

Затем был создан набор из шести почтовых ящиков для взаимодействия с полями форм отобранных сайтов по указанным URL-адресам, при чем уровень сложности каждого из них повышался по отношению к стандартным историческим доменам на основе ASCII. С начала и до завершения исследования использовались одни и те же адреса. Ниже приведен перечень почтовых ящиков:

Тестовый пример	Пример
ascii@ascii.newshort	test@test.exp
ascii@ascii.newlong	test@test.example
ascii@idn.ascii	test@普遍接受-测试.org
unicode@ascii.ascii	测试1@test.org
unicode@idn.idn	测试5@普遍接受-测试.世界
arabic.arabic@arabic (RTL)	دون@رسيل.السعودية

Исключение почтового ящика «ascii@ascii.idn» привело к возникновению значительного расхождения по сравнению с результатами теста 2017 года. Это изначально не входило в намерения группы, поскольку согласно соответствующим RFC этот почтовый ящик должен был поддерживаться, как один из вариантов использования. Однако после неудачных попыток зарегистрировать тестовый адрес электронной почты, используя этот формат в пяти различных регистратурах из-за того, что этот формат не поддерживается, мы совместно с руководством UASG пришли к выводу, что при проведении настоящего исследования этот ящик будет исключен в связи с тем, что он не принимается в достаточных масштабах.



Один элемент, унаследованный из предыдущего исследования, тестовый пример «arabic.arabic@arabic», который в случае записи с помощью надлежащей системы знаков, становится тестовым примером записи справа налево (RTL). Мы в некоторой степени исправили неправильное название, добавив RTL к метке, однако недостаточное тестирование записей на иврите, персидском языке, урду, синдхи, идише и других подобных языках помешало полностью перейти на это название. Для учета этих факторов необходимы дополнительные тесты.

Следующим шагом было практическое тестирование форм электронной почты путем ввода адресов тестовых примеров по одному и отправки форм сайтам на обработку. Некоторые формы были пропущены на этой стадии из-за необходимости проходить СМС-подтверждения или решать нерешаемые задачи CAPTCHA на клавиатуре западного стандарта. В этом втором случае, хотя их количество было недостаточно высоким, чтобы исказить результаты, есть вероятность того, что это могло до некоторой степени повлиять на результаты исследования.

При успешном прохождении процесса, сайту присуждались шесть оценок «принял» или «не принял» в соответствии со следующими критериями:

Оценка «принял», если:

- ✓ Отправка завершилась сообщением об успешном выполнении операции.
- ✓ Подача принята без сообщений об ошибках.
- ✓ Отображено сообщение об «уже зарегистрированном» адресе электронной почте<sup>3</sup>.

Пометка «не принял», если:

- ⊗ При вводе адреса электронной почты сайт выдал сообщение об ошибке.
- ⊗ Отправка завершилась сообщением об ошибке.
- ⊗ Был получен отказ в отправке.

В результате была получена следующая таблица:

Сайт	Адрес электронной почты 1	Адрес электронной почты 2	Адрес электронной почты 3	Адрес электронной почты 4	Адрес электронной почты 5	Адрес электронной почты 6
test.org	принял	принял	принял	не принял	не принял	не принял
ページ. 日本	принял	принял	принял	не принял	не принял	не принял

При тестировании первых 100 сайтов у группы уходило в среднем по 10 минут на выполнение каждого теста. По ходу тестирования среднее время проверки одного сайта снизилось приблизительно до пяти минут, и мы считаем, что можно предположить, что в будущем при повторном применении этой методологии этот параметр сохранится.

<sup>3</sup>Это происходило в случаях, когда у сайта была общая база данных с другим, ранее зарегистрированным сайтом.



После выполнения тестов код HTML каждой страницы сохранялся на местном компьютере, чтобы коды подтверждения из поля электронной почты можно было извлечь для анализа. На этот процесс каждый раз уходило разное количество времени в зависимости от типа решения и применяемых технологий. Если подтверждение выполнялось со стороны сервера или в HTML5, на определение конкретного варианта уходила одна минута или меньше. При этом при использовании других технологий, таких как JavaScript, на каждый сайт могло уходить до 15 минут, в частности, из-за того, что программный код зачастую минифицировался<sup>4</sup>.

Вместо того, чтобы просто поместить эти результаты в цифровую таблицу, группа решила организовать их в базу данных MongoDB, платформу, ориентированную на JSON, которая требует моделирования. Это решение оказалось адаптируемым, легким, и каждый тестер в любой момент имел доступ к наиболее актуальной версии тестов. Данные было легко визуализировать, и с ними могли работать даже менее опытные пользователи. В целом, мы рассматривали этот вариант как более совершенный по сравнению с базой данной SQL или совместно используемой электронной таблицей.

Создание главной базы в MongoDB позволяет составлять неограниченное число «наборов данных», аналогичных электронным таблицам. Наборы данных содержат все результаты, относящиеся к конкретному тесту, и после того, как им присваивается соответствующее имя, они существуют как автономные, функционально совместимые наборы данных. При соблюдении соглашения о присвоении имен «ua\_<scope>\_<year>» все они могут быть собраны в одной базе данных с удобочитаемыми для человека именами.

Например, набору данных для настоящего исследования присвоено имя «ua\_global\_2019». Это значит, что если исследованию будущего года будет присвоено имя «ua\_global\_2020», они смогут без проблем сосуществовать, а исследователи смогут ссылаться ни на них и на содержащиеся в них данные. Важно отметить, что названия наборов данных чувствительны к регистру. Например, какому-нибудь будущему исследованию в Мексике может быть присвоено имя «ua\_regional\_Mexico\_2019», и т.д.

Последним шагом стало объединение всех данных, преобразование итоговой базы данных в файл CSV и его передача на обработку и оценку для подготовки настоящего отчета. Любой файл CSV, соответствующий описанной ниже схеме метаданных, можно легко импортировать в базу данных с помощью следующей команды, предполагающей среду Ubuntu:

```
mongoimport --db ua_database --collection <scope>_<year> --  
host=<hostname> --username=<username> --password=<password> --  
drop --type csv --file <file_address>/ua_global_2019.csv --  
headerline
```

## Схема метаданных

---

<sup>4</sup> Определение: Удаление ненужных или избыточных данных без изменения способа обработки ресурса браузером; подразумевается удаление комментариев к программному коду и информация о формате, неиспользуемого кода, использование более коротких имен переменных и функций и т.п. [Wikipedia]



Выбрав для хранения данных базу MongoDB, группа предложила схему метаданных, которая, мы надеемся, будет использоваться и в будущих тестах, чтобы результаты различных выборок имели общий стандарт, а исследователи могли их сопоставлять по своему усмотрению.

Теперь мы представим данный стандарт и поясним его.

```
{
  "_id": {"$oid": "00001"}
  "domain": "test.org",
  "url":
  "https://www.test.org/signup",
  "rank": "1000",
  "testable": "Yes",
  "code": "<input
type='email'>",
  "comments": "Field triggers
a captcha",
  "mailboxes":
  {
    "mail1": "Accepted",
    "mail2": "Accepted",
    "mail3": "Accepted",
    "mail4": "Rejected",
    "mail5": "Rejected",
    "mail6": "Rejected"
  }
}
```

Каждая строка имеет свою функцию:

- { **\_id**: Автоматически сгенерированный уникальный идентификатор в базе данных.
- { **domain**: Домен без префикса «www».
- { **url**: URL-адрес, содержащий полный путь к странице с формой.
- { **rank**: Ранжирование сайта в рамках конкретного набора данных.
- { **testable**: Указывает на то, удалось ли группе протестировать сайт.
- { **code**: Содержит строку, которая подтверждает форму – при наличии формы.
- { **comments**: Любой относящийся к делу комментарий.
- { **mailboxes**: Список оценок по каждому тестовому примеру.

Для будущих тестов мы обдумываем включить поле «**eai**», указывающее на то, поддерживает ли почтовый сервер конкретного домена интернационализацию адресов электронной почты, что по сути означает возможность обмениваться почтой в UTF-8 и, соответственно, в Unicode. Это обеспечит более глубокое понимание общего уровня поддержки данной технологии с учетом того, сколько точек вероятных отказов существует на пути к успешному обмену сообщениями с более сложных адресов.

Конкретно для этого отчета по 2019 году была включена строка «**country**», чтобы получить более полное представление о географии списка. Этот подход тем не менее



не планируется сделать стандартным из-за его непрактичности в будущем в связи с изменениями, внесенными в базу данных WHOIS и не позволяющими проводить массовые просмотры, подобные тем, которые произвела группа. Если с течением времени RDAP сделает возможным выполнение подобных запросов, эту строку можно будет опять включить.



## Результаты

### Результаты теста в цифрах

Тестовый пример	принят	не принят
ascii@ascii.newshort	971	29
ascii@ascii.newlong	838	162
ascii@idn.ascii	505	495
Unicode@ascii.ascii	131	869
Unicode@idn.idn	76	924
arabic.arabic@arabic (RTL)	68	932

### Графическое отображение результатов теста





## Сравнение результатов тестов 2017 г. и 2019 г.





## Анализ и корреляция

В данном разделе мы рассмотрим некоторые результаты, полученные группой в процессе оценки программного кода и проливающие свет на неудачную практику использования форм электронной почты в сети. Мы также не упустили возможность провести небольшой эксперимент, воспользовавшись фактом сбора региональных данных.

Важно отметить, что, аналогично тому, что было отмечено по результатам исследования 2017 года, не существует унифицированного подхода к кодированию функций подтверждения, кроме как в случаях использования HTML5 и ряда стандартизированных наборов символов, таких как «jquery-validation.js». Значительное число сайтов опирается на регулярные выражения (REgEx) в JavaScript с разной степенью успеха при их использовании. Но, как показывает опыт, это редко оказывается хорошим решением.

### Случаи неподтверждения выборки

На сайтах с более качественным кодом в целом, но низкими показателями поддержки универсального принятия, адрес электронной почты не принимался непосредственно в момент ввода адреса в форму. На сайтах с менее качественным кодом в целом, адрес сначала принимался, но затем выдавалось сообщение «произошла ошибка». Иногда более сложные тестовые примеры определялись сайтами как вредоносные и классифицировались как атаки или попытки взлома.

Ниже приведены некоторые из часто встречающихся сообщений, с небольшими вариантами, т.е. со случаями непринятия тестовых примеров:

Введите действительный адрес электронной почты.
Неправильный формат электронной почты.
Извините, указан недействительный адрес электронной почты. Пожалуйста, убедитесь в правильности адреса.
Текст перед значком @ не должен содержать символов.
Произошла ошибка.

Один тип сообщений особенно отличается от остальных — когда наличие символов перед значком @ воспринимается как ошибка. Эта ситуация встречалась достаточно систематически в наших наблюдениях и, похоже, в этом состоит главный ошибочный вывод, вызывающий проблемы с подтверждением.

### Непринятие всех адресов электронной почты

Группа столкнулась с рядом особенно неудачных примеров написания программного кода среди 1000 ведущих сайтов мира. Ниже мы выделили один пример в JavaScript, выдававший отказ во всех тестовых случаях при работе с одним вьетнамским сайтом, ни много ни мало. Это не единичный случай, другие сайты из списка выдавали самые разные форматы сообщения о такой же ошибке.



RegEx:

```
/^[ -a-z0-9~!$%^&* _+=}{\'}?]+\(\. [ -a-z0-9~!$%^&* _+=}{\'}?+\)*@([a-z0-9_][ -a-z0-9_]*\(\. [ -a-z09_]+\)*\.(aero|arpa|biz|com|coop|edu|gov|info|int|mi1|museum|name|net|org|pro|travel|mobi|[a-z][a-z])|([0-9]{1,3}\. [0-9]{1,3}\. [0-9]{1,3}\. [0-9]{1,3}))(: [0-9]{1,5})?$/i
```

Этот фрагмент кода выполняет строгую проверку некоторых конкретных параметров, определяющих структуру исторического адреса электронной почты, и уделяет особое внимание занесению в белый список некоторых спонсированных TLD, а также повторным образом заносит в белый список такие исторические TLD, как «.edu» и «.org». В этом нет особого смысла, но варианты этого поведения были найдены присутствовали и в других случаях.

### Тестирование региональной корреляции

Хотя это и не было основной целью исследования, группа заинтересовалась возможностью определить показатели соответствия принципам универсального принятия на ведущих сайтах стран, использующих отличные от латинского алфавиты, предполагая, что эти показатели могут быть выше. Проанализировав кириллицу и хань, мы пришли к выводу, что на этом уровне **эта корреляция не прослеживается**. Возможно, в будущих локальных тестах удастся точнее оценить, намного ли лучше аналогичные показатели у 100 ведущих региональных сайтов, что в перспективе представляет собой важный вопрос.

С результатами можно ознакомиться ниже:

- **Хань:** пул состоял из 47 сайтов из Китая, 8 из Гонконга, 2 из Сингапура и 12 из Тайваня; в итоге только 4 сайта из Китая смогли обработать умеренно сложный тестовый пример «unicode@ascii.ascii».
- **Кириллица:** пул состоял из 37 сайтов из России и 2 из Украины; в итоге только 5 российских сайтов смогли обработать умеренно сложный тестовый пример «unicode@ascii.ascii».



## HTML5: препятствие на пути к универсальному принятию

Результаты исследования «ua\_regional\_Brazil\_2018»<sup>5</sup> показали, что 30 процентов 50 ведущих сайтов страны используют строку `<input type="email">` в качестве решения для подтверждения адресов электронной почты в части браузера. Группа предполагала обнаружить аналогичную картину в глобальном исследовании, и эти ожидания подтвердились. Согласно полученным нами данным, интенсивность использования этой строки находится в диапазоне между 20 и 30 процентами.

Проблема заключается в том, что в середине 2019 года стандарт HTML5 все еще не поддерживает принципы универсального принятия. Это вызывает беспокойство, поскольку использование стандарта продолжает расти, и разработчики все чаще опираются на его функционал для того, чтобы их код работал в стольких сценариях, скольких требует текущее состояние Интернета.

Ниже представлено текущее положение дел по принятию строки `<input type="email">`:

Тестовый пример	Результат
ascii@ascii.newshort	принят
ascii@ascii.newlong	принят
ascii@idn.ascii	принят
unicode@ascii.ascii	непринят
unicode@idn.idn	непринят
arabic.arabic@arabic (RTL)	непринят

Данный результат только подтверждает наше прежнее утверждение о том, что главной ошибкой, очевидно, является представление о том, что перед знаком @ не должно быть символов.

Если эта строка используется 20-30 процентами сайтов, то, наоборот, это значит, что аналогичное число сайтов не проходит эту стадию подтверждения из-за отсутствия надлежащих стандартов в HTML5. Иными словами, в том, что касается браузеров, это наиболее действенная мера, которую можно предпринять для продвижения универсального принятия на данный момент.

Изучив ситуацию с этой проблемой, мы обнаружили, что в проектной версии 5.3 спецификации, в разделе о подтверждении адреса электронной почты<sup>6</sup>, представлен прогноз соответствия положениям RFC 6531 и RFC 5890, что обеспечивает соответствие стандарта требованиям универсального принятия. При этом эта редакция существует в качестве «первой версии проекта» с 2017 года, без прогнозов относительно даты перехода в статус «потенциальной рекомендации» или окончательной «рекомендации».

<sup>5</sup>

<http://central.abesssoftware.com.br/Content/UploadedFiles/Arquivos/Estudos,%20Pesquisas%20e%20Pareceres/Universal-Acceptance-in-Brazil-2018.pdf>

<sup>6</sup> <https://www.w3.org/TR/html53/sec-forms.html#email-state-typeemail>



У нас это вызывает беспокойство, учитывая важность вопроса для обеспечения поддержки принципов универсального принятия, в частности, в отношении интернационализированных имен. Сообщество не может себе позволить игнорировать этот вопрос, и необходимо предпринимать разнообразные шаги в наших взаимоотношениях с W3C, WHATWG и поставщиками браузеров, чтобы ускорить процесс реализации данного аспекта спецификации.

Это особенно актуально, поскольку W3C и WHATWG недавно договорились применять совместный подход к разработке стандартов<sup>7</sup>. Обращаться к соответствующим сторонам и требовать добавления опции «email» на сайтах, поддерживающих универсальное принятие, так же важно, как и оказывать давление на основных поставщиков браузеров, чтобы те сигнализировали о желании обеспечивать эту опцию.

Исследовательская группа рекомендует сделать решение этого вопроса приоритетом.

---

<sup>7</sup> <https://www.w3.org/blog/2019/05/w3c-and-whatwg-to-work-together-to-advance-the-open-web-platform/>



## Можно ли автоматизировать эти отчеты?

При подготовке данного исследования на конференциях ICANN и во время заседаний по планированию группа UASG задавала вопросы о целесообразности использования автоматизированного ПО для подготовки отчетов, аналогичных этому. Группа попыталась найти практический ответ на этот вопрос, однако пришла к выводу, что в настоящее время автоматизировать эти тесты невозможно.

Основные моменты, вызывающие у нас беспокойство:

- Слишком большое разнообразие в том, что касается программного кода сайтов, используемых в них наборов символов и организации форм. В целом, учесть все варианты практически невозможно.
- Широко используется минификация с целью сократить время загрузки сайтов и количество наборов символов. Иногда необходимо, чтобы результаты проверял специалист с опытом программирования.
- Значительное число сайтов требуют решения задачи Captcha в той или иной форме или имеют в формах некие нюансы, что делает автоматизацию бесполезной.

Некоторые из этих же обоснований можно применить при решении вопроса передачи выполнения исследования неспециалистам, например, Amazon Mechanical Turk. Подобный подход может сработать при выполнении самого тестирования, но при этом все равно будет необходимо координировать работу при подготовке списка доменов, извлечении кода и подготовке информативного отчета, содержащего не только статистику. И тем не менее, это гораздо более эффективный подход, чем автоматизация.



## Заключение

Если принять за точку отсчета 100 ведущих сайтов в мире, только пять из них приняли все тестовые примеры: «quora.com», «espn.com», «spotify.com», «txxx.com» [NSFW] и «godaddy.com». Предположим, что мы не использовали сложные тестовые примеры «unicode@idn.idn» и арабский (запись справа налево); в этом случае первые четыре тестовых примера будут все равно приняты только девятью сайтами. Поскольку мы говорим о наиболее посещаемых страницах в мире, это вызывает большую обеспокоенность.

Хотя мы с оптимизмом смотрим на перспективу того, что в скором будущем новые короткие адреса будут полностью соответствовать принципам универсального принятия и считаем реалистичным предположение о том, что новые длинные адреса движутся в том же направлении, очевидно, что интернационализированные доменные имена нуждаются в некотором толчке, и что необходимо действовать на упреждение, чтобы они успешно развивались. Необходимо не просто разработать новые, более строгие правила, а скорее надо сделать так, чтобы в интернете было больше осведомленности о существовании подобных адресов электронной почты и доменов, а этого можно достичь только путем активного информирования общества.

Наконец, мы считаем, что следующим важным шагом должна стать разработка панели управления для оценки всех отчетов по универсальному принятию, создаваемых при помощи схемы метаданных, предложенной в настоящем документе. Это поможет обеспечить создание долгосрочного репозитория, который позволит любому желающему выполнять реалистичные измерения и будет служить некоторым ориентиром для всего сообщества.