

Universal Acceptance in Brazil: Proof of Concept E-mail Address Acceptance Study

October, 2018

Team

- Leader: **Paulo Milliet Roque** – Business Constituency, ABES
- Drafter: **Mark W. Datysgeld** – Business Constituency, Governance Primer
- Tester: **Sávyo Vinícius de Moraes** – NextGen, IFRN
- Coordinator: **Daniel Fink** – ICANN LAC GSE
- Supervisor: **Don Hollander** – UASG
- Special thanks: **Rubens Kuhl** (NIC.br), **Ajay Data** (Datainfosys)

Introduction

The objective of this project is to evaluate the compliance of the 50 most popular websites in Brazil¹ in relation to the standards of internationalization and the new domain names that fall under ICANN's Universal Acceptance umbrella. This proof of concept study was carried out as an initial exercise that could lead to a more encompassing follow-up and create a pathway towards direct engagement with businesses in the country, so that more compliance can be achieved in the future.

The main method employed to perform this appraisal was testing the acceptance of e-mail addresses entered in public-facing contact forms of the selected websites, with said addresses having varied degrees of complexity. By evaluating what types of response, or lack thereof, were obtained, we were able to obtain a general impression of what the rate of acceptance looks like among leading market players.

The methodology followed the steps laid out by 2017's "*UASG017: Evaluation of Websites for Acceptance of a Variety of Email Addresses*", which conducted a study of more than 1000 websites at a global level, and concluded that compliance was quite insufficient, with all but one of the test cases achieving less than a 50% rate of success, with the most elaborate e-mail address obtaining only an 8% rate.

This project is a multistakeholder initiative encompassing ABES (the Brazilian Association of Software Companies), the Business Constituency, Fellowship/NextGen Alumni and ICANN LAC GSE; all counting with the invaluable support of the Universal Acceptance Steering Group (UASG).

¹ According to Alexa: <https://www.alexa.com/topsites/countries/BR>.

Motivation

Brazil and Latin America more broadly have been undergoing a crisis of productivity in the late 2010s, and there is a mounting necessity for actions to be taken towards internationalization and improving avenues of communication with businesses outside of the continent, from the micro to the macro level. While not enough to generate systemic changes by itself, achieving UA compliance would be a strong signal of willingness to engage with emerging foreign partners from the CIS region, Eastern Asia, and the Arab world, as well as facilitating relations with innovators that choose to use new gTLDs as part of their branding on the Internet.

Methodology

The project evaluated the 50 Brazilian top websites based on Alexa rankings data from august 4, 2018, in accordance to the following criteria:

- Brazilian Portuguese is the main language of the website;
- The website's company has a physical address/office in Brazil;
- The website has a page that allows registration of an email address.

All of the 11 addresses were then tested against the forms of the websites and the results were collected for analysis.

Test e-mails

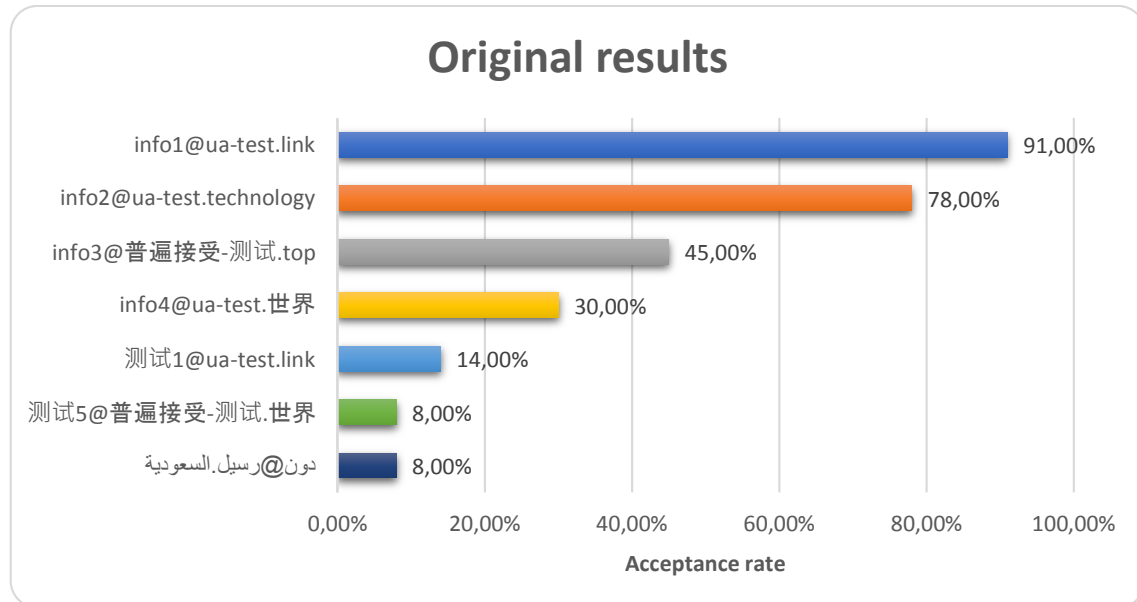
The original 2017 survey included 7 addresses of increasing complexity, which are the first entries in the table below. For the Brazilian edition, it was deemed suitable that extra addresses that reflect local realities be added, which led to the inclusion of 4 more cases. They include Latin diacritics at different levels, as well as second-level domains specific to the “.br” ccTLD.

Case	Implementation
ascii@ascii.newshort	info1@ua-test.link
ascii@ascii.newlong	info2@ua-test.technology
ascii@idn.newshort	info3@普遍接受-测试.top
ascii@ascii.idn	info4@ua-test.世界
unicode@ascii.newshort	测试1@ua-test.link
unicode@idn.idn	测试5@普遍接受-测试.世界
arabic.arabic@arabic	السعودية.رسيل@دون
ascii@unicode.gtld.cctld	info5@xçãäöy.com.br
ascii@ascii.cc2ld.cctld	info6@ua-test.campinagrande.br
ascii@unicode.cc2ld.cctld	info7@xçãäöy.campinagrande.br
unicode@unicode.cc2ld.cctld	usuário@xçãäöy.campinagrande.br

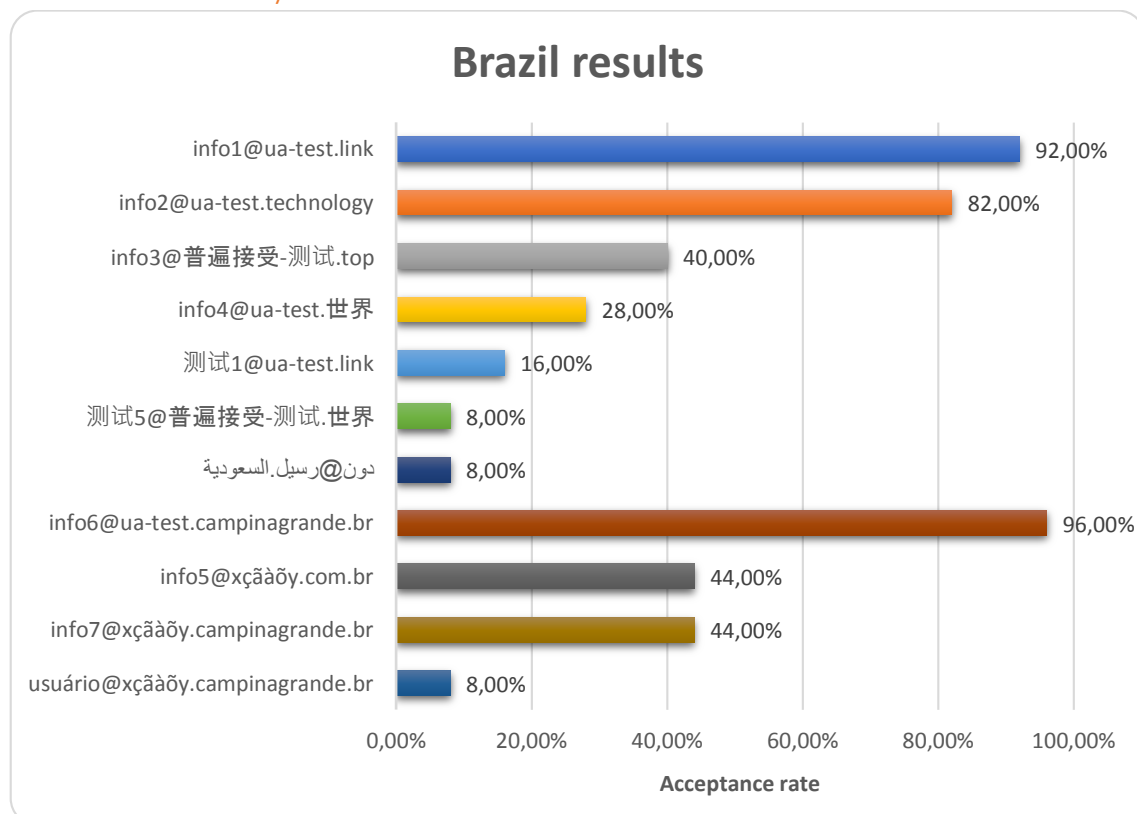
Results

First, we present the results of the original set of tests in order to provide a basis of comparison. In sequence, we present the results of the Brazilian tests following the same standard of analysis with the added cases previously described.

Results: Original study



Results: Brazil study



Result analysis

It is interesting to note that, on the overall, the results from our limited sample were very close to those of the original global research, in spite of the difference in scope between them. The deviation in acceptance falls within standard margins of error, and the results can be considered to be virtually identical. While not conclusive, this does point to a general pattern in rates of acceptance across the Internet.

Out of the 50 Brazilian websites, only 3 achieved total UA compliance, with those being that of content portal *IG* (ig.com.br), download website *Baixaki* (baixaki.com.br), and the government's Ministry of Education (mec.gov.br); a close contender was the film-oriented *Adoro Cinema* (adorocinema.com), which despite accepting all addresses, resulted in a failure when ultimately processing some of them.

Success cases

It is interesting to note that the methods used to achieve success by the compliant websites falls outside of the realm of the more standard approach of using JavaScript embedded on the page, but rather rely on third-party specialized service providers or use server-side solutions that are not apparent upon casual inspection. It might be worth engaging in further discussion with the administrators of those systems to understand the philosophy behind their design.

This is synthesized in the table below:

Website	Method
ig.com.br	The data is processed by means of a third-party solution provided by the <i>Movidesk</i> platform. The validation process could not be identified because it is a server-side solution.
baixaki.com.br	Unidentified server-side solution.
mec.gov.br	The data is processed by means of a third-party solution provided by the <i>Call Tecnologia</i> platform. The expression found on the script used by the platform is as follows: <pre>var validarEmail = function(\$email) { var emailReg = /^[^<>()\\[\]\\.,:;\s@\"']+(\.[^<>()\\[\]\\.,:;\s@\"']+)*)(\.[^<>()\\[\]\\.,:;\s@\"']+)+[\^<>()\\[\]\\.,:;\s@\"']{2,})\$/i; if(\$email === undefined) return true; if(\$email.length < 7) { mail_err.append('Campo email não preenchido corretamente.
'); mail_err.show(); return false; } else { if (!emailReg.test(\$email)) { mail_err.append('O e-mail digitado não é válido.
'); mail_err.show(); return false; } else { return true; } };</pre>

Failure cases

Problems encountered came from different vectors, and several different implementations of code ended up proving problematic, such as was the case with the popular library AngularJS, native HTML5 forms, and some server-side solutions. As would be expected, the most common issue met was that of codes with insufficient amplitude to cover non-ASCII characters, that was never intended to handle anything other than a limited set of A-Z, 0-9 and very few special symbols.

One example that struck to us as particularly relevant was the case of the built-in validation feature of HTML5, which was found in 30% of the tested forms and fails for some of the e-mail addresses. The regular expressions that perform the validation in these cases are defined by the browser in use, meaning that achieving better collaboration with developers of these software carries great value.

Some relevant examples follow:

Website	Method
vagalume.com.br	email:function(email) {return /^[\\w.+\\-]+@[a-zA-Z0-9-]+\\.([a-zA-Z0-9-]+)+\$/i.test(email);}
uol.com.br	<input type="email"></input>
oi.com.br	s.emailValido=function(A){return!! /^[\\w+.-]+@[\\w+\\.\\w{2}]{?\\.\\w{2}}?\$/i.test(A)}
netshoes.com.br	valid_email:function(e){return /^[\\w+([\\-]?\\w+)*@[\\w+([\\-]?\\w+)*\\.\\w{2,3}]+\$/i.test(e.value)}
dafiti.com.br	String.prototype.isEmail=function(){var e=/^((([<>()[]\\.,;:~@'"]+) \\.[^<>()[]\\.,;:~@'"]+)*)(\\.([<>()[]\\.,;:~@'"]+) \\.[^<>()[]\\.,;:~@'"]+))*\$/;return e.test(this)}

Further studies into HTML5 compatibility

As a consequence of analyzing the results aggregated in this document, our interest was piqued in relation to the cases in which errors originate from the use of standard HTML5 fields. The most immediate questions that demanded an answer were whether there was any variation in the way each browser handles native e-mail input, and if any of them could be considered fully UA compliant.

To that end, we picked the tricky Brazilian address “usuário@xçããõy.campinagrande.br” as our test case, seeing as it is the most complex address available from our list other than the Arabic one, with an 8% acceptance rate in the standard test. The websites chosen for this test were the online retailer *Americanas* (americanas.com), the content portal UOL (uol.com.br), and music tablature website *Cifra Club* (cifraclub.com.br). The browsers used were Chrome, Edge, Firefox, Internet Explorer, Opera, and Safari.

Our results were consistent: **failure** across the board. No browser could handle correctly the input of the address. Subsequent control tests with more standard addresses demonstrated that this was indeed an issue specific to our chosen complex address.

Moving forward

The team was surprised that the results aligned so closely to that of the global study, given that original expectations were that the country, being a largely mono-linguistic and surrounded mostly by two languages that share the same character set as Brazil, might have overall poorer results.

The feasibility of performing this test on a larger scale seems very high, considering that the project, from conception to the elaboration of this document took around 3 months to be finalized. A deeper study ranging from 6 months to 1 year of duration would be able to perform a much more substantial analysis at a relatively low cost.

A factor that seems clear to us is that there are avenues to help websites become compliant with UA, and more analysis is necessary for the development of methodologies that would allow us as a group to approach companies and propose that, following certain orientations, their coding teams could improve the process by investing “x” amount of hours into improving their existing solution.

A viable next step is to start isolating interesting cases and reaching out to those developers in order to get a better understanding of how their codebases and server structure interact, so that this realistic set of solutions can be devised based on the reality of what is observed in the wild.

There are also other ideas that came up during the conception of the project, such as the possibility of coding bots with the ability of identifying fields and filling them correctly, in such a way that would allow for a truly massive study of UA across the Web and give a more definitive radiography of the problem at hand.

On the overall, we would like to signal that the project has a clear path forward that is desirable to follow, and express sustained interest of the Brazilian community in terms of continuing to look at this from the perspective of Latin America and the websites of the region.