Additional Information on Email Validation

TopDog Assignment SQL Injection 2

Joakim Argillander
joakim.argillander@liu.se
Department of Electrical Engineering
Linköping University

Thursday 26th November, 2020

Introduction

This document serves as an addendum to the lab material and intends to further explain the validation criteria that is used by TopDog after many similar questions regarding SQL Injection 2. The user-inputted email address is validated against the standard RFC 822 [1] (newer standards are RFC 5321 [2] and RFC 5322 [3]) before it is injected into the query. This means that in addition to being valid SQL syntax for the database to execute, the injection string must also pass this email validation. Below is a summary of the standards that apply to this validation.

Email Address Specification

An email address consists of a local part and a domain[1] like

`name` @ `location`
local part domain

which is often used like

`firstname.lastname` @ `liu.se`
local part domain

both of which are subject to individual validation rules.

A. *Local-part*

A standard-conforming local part may be enclosed in quotation marks or not enclosed in quotation marks [4]. If unenclosed, then the local part must only consist of one or more the following ASCII-characters

- Uppercase and lowercase Latin letters `A` to `Z` and `a` to `z`
- Digits `0` to `9`
- Printable characters `!#$%&'*+-/=?^_`{|}
- Dot (`.`), if it is not the first or last character and there is no two consecutive dots (`..`)

in addition to being less than 64 characters long [4] (64 characters for RFC 5321 [2] and 64 characters for RFC 532 [3]). If enclosed in quotes, whitespace, semicolon, brackets and at-signs are allowed.

Examples

`firstname..lastname@liu.se`	**Invalid**
`"firstname..lastname"@liu.se`	**Valid**
`firstname lastname@liu.se`	**Invalid**
`"firstname lastname"@liu.se`	**Valid**
`"firstname lastname@liu.se"`	**Invalid**
`"firstname lastname@liu.se"@liu.se`	**Valid**
`@liu.se`	**Invalid**
`" "@liu.se`	**Valid**
`'#@liu.se`	**Valid**

B. *Domain*

A standard-conforming domain must only consist of one or more the following ASCII-characters

- Uppercase and lowercase Latin letters `A` to `Z` and `a` to `z`
- Digits `0` to `9`
- Hyphen (`-`), if it is not the first or last character and there is no two consecutive hyphens (`--`)
- Dot (`.`) to denote subdomains[2] (for example `student.liu.se`) which are only valid if there is no two consecutive dots (`..`)

in addition to being less than 64 characters long [4] (255 characters for RFC 5321 [2] and 255 characters for RFC 532 [3]).

Examples

`firstname.lastname@liu.se-`	**Invalid**
`firstname.lastname@li=u.se`	**Invalid**
`firstname.lastname@li#u.se`	**Invalid**
`firstname.lastname@li'u.se`	**Invalid**
`firstname.lastname@li-u.se`	**Valid**
`firstname.lastname@domain`	**Valid**

References

[1] D. H. Crocker, "Standard for the Format of ARPA Internet Text Messages." RFC 822, Aug. 1982.
[2] J. C. Klensin, "Simple Mail Transfer Protocol." RFC 5321, Oct. 2008.
[3] P. Resnick, "Internet Message Format." RFC 5322, Oct. 2008.
[4] J. B. Postel, "Simple Mail Transfer Protocol." RFC 821, Aug. 1982.

[1] The domain may also only consist of a (generic-) TLD alone[1]. E.g. `firstname.lastname@company` (note the lack of country TLD) which is a valid domain (compare this to `localhost`)

[2] An arbitrary number of subdomains is possible, limited only by the maximum length