

TYPO3.Neos - Bug # 49761

Status:	Closed	Priority:	Should have
Author:	Philipp Gampe	Category:	
Created:	2013-07-07	Assigned To:	
Updated:	2014-10-01	Due date:	
Subject:	Hostname validator produces false negatives		
Description	<p>The regex in the hostname validator is completely wrong: \$pattern = '/([a-zA-Z0-9\-_]+\.)?[a-zA-Z0-9\-_]+\.[a-zA-Z]{2,5}/';</p> <p>First of all a hostname does not need to have a dot. This is a common use case in intranets where the DNS suffix is used while finding the IP, but the actual hostname does not contain additional parts: intranet[.my-company.com] -> http://intranet/.</p> <p>Second, the new TLDs allow longer domain names than five chars, e.g. travel which is already included in the list of valid domain suffixes: http://data.iana.org/TLD/tlds-alpha-by-domain.txt http://www.iana.org/domains/root/db</p> <p>There are more limits that could be checked, e.g.</p> <ul style="list-style-type: none">- 127 level limit- 253 char limit total in text representation <p>http://tools.ietf.org/html/rfc1035</p> <p>Also internationalized domain names should be taken into account.</p>		

History

#1 - 2013-07-07 11:42 - Philipp Gampe

Short discussion about this ...:

- If there is a dot, the last segment should be split of and be checked for >2 chars and [a-zA-Z]
- Any other segments should only be checked for ascii (if neos does unicode -> punycode conversion automatically)
- Otherwise check RFC for special rules (no dash as first segment char???)

#2 - 2013-07-11 22:49 - Aske Ertmann

- Status changed from New to Accepted
- Target version set to 1.0 beta 1

Hey Philipp

We just didn't put a lot of effort into this, so that's why it only works for common domain names.

You're very welcome to find a better regular expression to use instead. Searching google finds quite some options, so if you can try to find one that follows everything you'd like please push a patch with it and explain why you chose that one..

#3 - 2013-07-13 11:34 - Philipp Gampe

According to wikipedia, we have the following rules:

- each label may contain up to 63 chars
- max 127 levels
- full domain name may not exceed 253 chars in textual representation
- root uses LDH [a-zA-Z0-9-]
- warning for labels that do not conform to LDH (but not error)
- top level domains may not be numeric (but local domains could)

<http://tools.ietf.org/html/rfc1034>

- labels must start with a letter and end with a letter or digit; they can contain hyphens (3.5)

Thus a single regexp will not do it.

Ping me next week and I might be able to compose a patch with those rules.

#4 - 2014-10-01 14:56 - Jonas Renggli

- *Status changed from Accepted to Closed*

Moved to <https://jira.typo3.org/browse/NEOS-475>