

python / cpython Public

<> Code Issues 5k+ Pull requests 2.2k Actions Projects Security and q

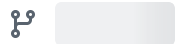
# Commit 09fab93



t-8ch authored on Feb 17, 2024 Verified

gh-100884: email/\_header\_value\_parser: don't encode list separators (GH-100885)

ListSeparator should not be encoded. This could happen when a long line pushes its separator to the next line, which would have been encoded.



1 parent [465db27](#) commit 09fab93

3 files changed +9 -1 lines changed

↑ Top ⚙️

Filter files...

- Lib
  - email
    - \_header\_value\_parser.py
  - test/test\_email
    - test\_header\_value\_parser.py
- Misc/NEWS.d/next/Library
  - 2023-01-09-14-08-02.gh-issue-100884.DcmdLI.rst

3 files changed +9 -1 lines changed

Search within code



Lib/email/\_header\_value\_parser.py



```

@@ -949,6 +949,7 @@ class _InvalidEwError(errors.HeaderParseError):
949 949 # up other parse trees. Maybe should have tests for that, too.
950 950 DOT = ValueTerminal('.', 'dot')
951 951 ListSeparator = ValueTerminal(',', 'list-separator')
952 + ListSeparator.as_ew_allowed = False

```

```

952 953 RouteComponentMarker = ValueTerminal('@', 'route-component-marker')
953 954
954 955 #
@@ -2022,7 +2023,7 @@ def get_address_list(value):
    address_list.defects.append(errors.InvalidHeaderDefect(
        "invalid address in address-list"))
    if value: # Must be a , at this point.
-   address_list.append(ValueTerminal(',', 'list-separator'))
+   address_list.append(ListSeparator)
    value = value[1:]
    return address_list, value

```

...est/test\_email/test\_\_header\_value\_parser.py

```

@@ -2985,6 +2985,11 @@ def
test_address_list_with_unicode_names_in_quotes(self):
    '?utf-8?q?H=C3=BCbsch?= Kaktus <beautiful@example.com>,\n'
    '=?utf-8?q?bei=C3=9Ft_bei=C3=9Ft?= <biter@example.com>\n')
+   def test_address_list_with_list_separator_after_fold(self):
+       to = '0123456789' * 8 + '@foo, ä <foo@bar>'
+       self._test(parser.get_address_list(to)[0],
+           '0123456789' * 8 + '@foo,\n =?utf-8?q?=C3=A4?=
+       <foo@bar>\n')
+
2988 2993 # XXX Need tests with comments on various sides of a unicode token,
2989 2994 # and with unicode tokens in the comments. Spaces inside the quotes
2990 2995 # currently don't do the right thing.

```

...3-01-09-14-08-02.gh-issue-100884.DcmdL1.rst

```

@@ -0,0 +1,2 @@
1 + email: fix misfolding of comma in address-lists over multiple lines in
2 + combination with unicode encoding.

```

### Comments 0