



IETF 81 - Thoughts on HTTP Header Field Parsing

[Julian Reschke](#), greenbytes

Background

- HTTPbis WG Work on Content-Disposition (now [RFC 6266](#))
- Various HTTPbis WG issues, such as [231: Considerations for new headers](#)

Problem Statement

- The parsing of many HTTP header fields is *hard*!
- Implementations *do* get it wrong.
- Extension points not well understood.
- I18N not well understood and frequently considered too late.
- We can't fix the past, but we can try to do better.

Example: the List Production and repeating Header Field instances

Foo: a

Foo: b

is equivalent to

Foo: a, b

- This is fine for simple stuff like method names.
- It falls apart when people who define new header fields do not get it (Example: Set-Cookie).
- It helps for folding multiple instances into one, but *not* for parsing.

If-Match: "strong",W/"weak", "oops, we should have thought of that"

Example: the List Production and repeating Header Field instances

Combining list production with structured field syntax:

```
WWW-Authenticate = 1#challenge
challenge        = auth-scheme 1*SP 1#auth-param
auth-param       = token "=" ( token / quoted-string )
```

Example:

```
WWW-Authenticate: Newauth realm="newauth";
  test="oh, a \"comma\""; foo=a'b'c, Basic realm="basic"
```

Example: Parameters - Whitespace, Quoting

```
param = token "=" ( token / quoted-string )
```

```
foo=bar; foo='bar'; foo="bar"; foo = "bar"
```

- Whitespace sometimes allowed, sometimes not.
- Lots of confused parsers.
- Single quote *is* used in `token` values.
- Definitions special-case the right hand side for individual parameter names, generic parsers can't do that (example: [RFC 5988](#) disallows token form for `title`, uses double quotes for `quoted-mt` without making it a `quoted-string`)

Proposals

- Test Cases. Examples. Lots.
- Make existing syntax more consistent where we can (fix mistakes where possible, discourage generating useless whitespace, require recipients to deal with it nevertheless)
- Encourage authors of new header fields to re-use existing syntax.
- *Force* them to think about extensibility.

Links

My tests:

- Content-Disposition - <http://greenbytes.de/tech/tc2231/>
- Content-Type - <http://greenbytes.de/tech/tc/httpcontenttype/>
- Link - <http://greenbytes.de/tech/tc/httplink/>
- WWW-Authenticate - <http://greenbytes.de/tech/tc/httpauth/>

...and then there's also <http://redbot.org/>.