Default charset parameter values for text/* media types

Alexey Melnikov <alexey.melnikov@isode.com>

Julian F. Reschke <julian.reschke@greenbytes.de>

Background information: draft-melnikov-mime-default-charset-01.txt

Issues with default charsets for different text/* media types

MIME (RFC 2046, Section 4.1.2) says that the default charset is "US-ASCII"

RFC 2616 says that the default charset is "ISO-8859-1"

Confusion on which document is authoritative in case of HTTP

Additional confusion: what about structured text/* media types which include the charset in the payload itself (e.g. text/xml), e.g.

```
<?xml version="1.0" encoding="UTF-8"?>
    ?
```

Proposal (1 of 2)

- Remove the restriction from MIME that the default charset for a text/* media type is always "US-ASCII", i.e. a particular text/* media type can
  - specify a different default;
    - This option is not recommended
  - specify that the charset parameter is never used;
  - specify that the charset parameter must always be specified
- Protocols using media types MUST NOT [re-] define any default charset values for text/*

Proposal (2 of 2)

- If a text/* media type defines a default charset value, it SHOULD be "UTF-8"

Examples

- text/plain --> default charset value is "US-ASCII" (unchanged), charset parameter is allowed

- text/html or text/xml --> the charset parameter SHOULD NOT be used

- text/vcard (draft-ietf-vcarddav-vcardrev-22.txt) --> always use explicit charset="UTF-8"

Open Issues

- For media types that have internal (to their format) means of describing the charset – is use of the charset parameter **SHOULD NOT** or **MUST NOT**?

    – For existing media types – the charset parameter SHOULD NOT be used

    – For new media types the charset parameter MUST NOT be used?