WEB TECHNOLOGIES 1

MIME - Multipurpose Internet Mail Extension

MULTIPURPOSE INTERNET MAIL EXTENSIONS

- MIME(Multipurpose Internet Mail Extensions) is a standard which was proposed by Bell Communications in 1991
- In order to expand upon the limited capabilities of email, and in particular to allow documents such as images, sound, and text to be inserted in a message.
- It is used to describe message content types.
- It was originally defined by RFCs 1341 and 1342 in June 1992.
- A Request for Comments (RFC) is a document published by the Internet Engineering Task Force(IETF) describing an internet standard .

MULTIPURPOSE INTERNET MAIL EXTENSIONS

- MIME messages can contain
  - Text
  - Images
  - Audio
  - Video
  - Other application-specific data (e.g. PDF Files,Microsoft Word Documents, and so on)
- MIME types are standardized by a group called the IANA (Internet Assigned Numbers Authority).

Structure of a MIME type

- The simplest MIME type consists of a *type* and a *subtype*; these are each strings which, when concatenated with a slash (/) between them, comprise a MIME type. **No whitespace is allowed in a MIME.**

  type: ***type/subtype***

  - The ***type*** represents the general category into which the data type falls, such as video or text.
  - The ***subtype*** identifies the exact kind of data of the specified type the MIME type represents.

- For example, for the MIME type text, the subtype might be **plain** (plain text), **html** (HTMLsource code), or **calendar** (for iCalendar/.ics) files.

Structure of a MIME type

- Each type has its own set of possible subtypes, and a MIME type always has both a type and a subtype, never just one or the other.
- An optional **parameter** can be added to provide additional details:

type/subtype; parameter = value

- For example, for any MIME type whose main type is text, the optional charset parameter can be used to specify the character set used for the characters in the data.
- If **no charset** is specified, the default is ASCII(US-ASCII) unless overridden by the user agent's settings.
- To specify a UTF-8 text file, the MIME type **text/plain;charset=UTF-8** is used.
- MIME types are case-insensitive but are traditionally written in lowercase, with the exception of parameter values, whose case may or may not have specific meaning.

MULTIPURPOSE INTERNET MAIL EXTENSIONS

- MIME uses special header directives to describe the format used in a message body, so that the email client can interpret it correctly
  - MIME-Version: This is the version of the MIME standard used in the message. Currently only version 1.1.
  - Content-type: Describes the data's type and subtype. It can include a "charset" parameter, separated by a semi-colon, defining which character set to use.
  - Content-Transfer-Encoding: Defines the encoding used in the message body.
  - Content-ID: Represents a unique identification for each message segment
  - Content-Description: Gives additional information about the message content like if its Image or Video.
  - Content-Disposition: Defines the attachment's settings, in particular the name associated with the file, using the attribute filename.

Content-Type Header Field

- The purpose of the Content-Type field is to describe the data contained in the body.
- Its fully enough that the receiving user agent can pick an appropriate agent or mechanism to present the data to the user, or otherwise deal with the data in an appropriate manner.
- After the type and subtype names, the remainder of the header field is simply a set of parameters, specified in an attribute/value notation.
- The set of meaningful parameters differs for the different types. The ordering of parameters is not significant .

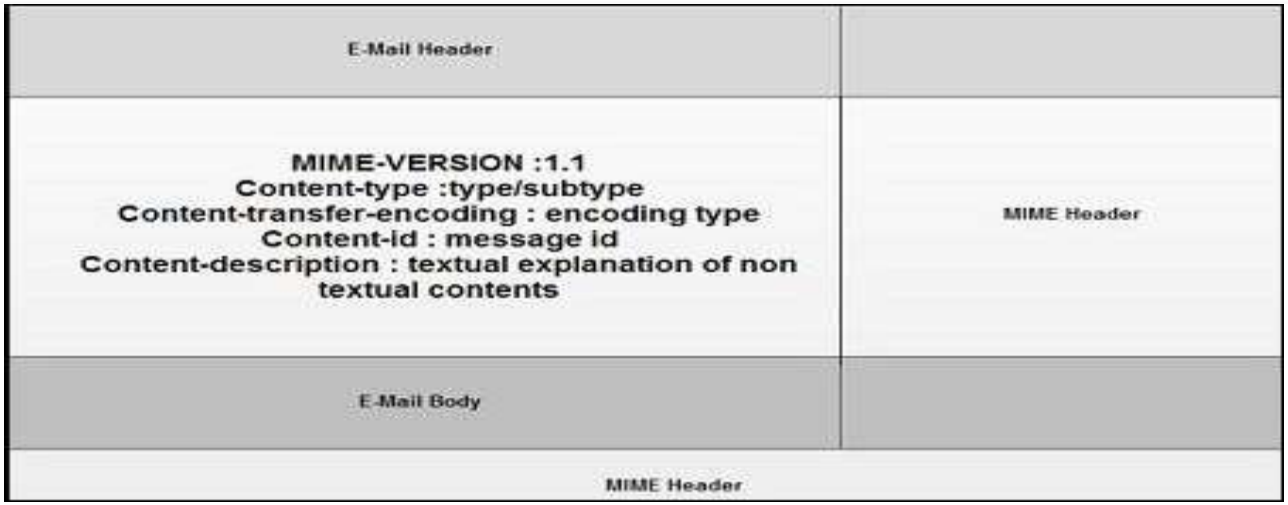
Content-Type Header Field

- Among the defined parameters is a "charset" parameter by which the character set used in the body may be declared.
- In general, the top-level *Content-Type* is used to declare the general type of data, while the *subtype* specifies a specific format for that type of data.
- Example, a Content-Type of *"image/xyz"* is enough to tell a user agent that the data is an image, even if the user agent has no knowledge of the specific image format "xyz".
- **ISO-8859-1** character standard extends the basic character set of ASCII to include many of the accented characters used in languages such as German.

Content-Type Header Field

Response headers (415 B) Raw headers

```
HTTP/2.0 204 No Content
content-type: text/html; charset=UTF-8
date: Mon, 21 Oct 2019 09:35:11 GMT
server: gws
content-length: 0
x-xss-protection: 0
x-frame-options: SAMEORIGIN
set-cookie: 1P_JAR=2019-10-21-09; expires=Wed, 20-Nov-2019 09:35:11 GMT; path=/; doma
alt-svc: h3-Q048=":443"; ma=2592000, h3-Q046=":443"; ma=2592000, h3-Q043=":443"; ma=2
X-Firefox-Spdy: h2
```



Features of MIME

- Using headers, MIME describes the type of message content and the encoding used.
- MIME adds the following features to email service:
  - Be able to send multiple attachments with a single message
  - Unlimited message length
  - Use of character sets other than ASCII code
  - Use of rich text (layouts, fonts, colors, etc)
  - Binary attachments (executables, images, audio or video files, etc.), which may be divided if needed.

Types

- There are two classes of type:
  - **Discrete Type**
  - **Multipart Type**
- **Discrete types** are types which represent a single file or medium, such as a single text or music file, or a single video.
- A **Multipart type** is one which represents a document that's comprised of multiple component parts, each of which may have its own individual MIME type; or, a multipart type may encapsulate multiple files being sent together in one transaction.
- For example, multipart MIME types are used when attaching **multiple files to an email**.

Discrete types

- Application
  - Any kind of **binary data** that doesn't fall explicitly into one of the other types; either data that will be executed or interpreted in some way or binary data that requires a specific application or category of application to use.
  - Generic binary data (or binary data whose true type is unknown) is **application/octet-stream** (unknown binary file).
  - Other common examples include

    application/pdf, and application/zip.
- Video
  - Video data or files, such as MP4 movies (video/mp4).

Discrete types

- Audio
  - Audio or music data. Examples include audio/mpeg, audio/vorbis.
- Font
  - Font/typeface data. Common examples include font/woff, font/ttf, and font/otf.
- Image
  - Image or graphical data including both bitmap and vector still images as well as animated versions of still image formats such as animated GIF or APNG. Common examples are image/jpeg, image/png, and image/svg+xml.
- Text
  - Text-only data including any human-readable content, source code, or textual data such as comma-separated value (CSV) formatted data. Examples include text/plain, text/csv, text/javascript, text/css and text/html.

Multipart types

- Multipart types indicate a category of document broken into pieces, often with different MIME types.
- They can also be used in email scenarios to represent multiple, separate files which are all part of the same transaction. They represent a **composite document**.
- There are two multipart types:
  - Message
  - Multipart

MESSAGE

- A message that encapsulates other messages.
- This can be used, for instance, to represent an email that includes a forwarded message as part of its data, or to allow sending very large messages in chunks as if it were multiple messages.
- **Examples** include message/rfc822 (for forwarded or replied-to message quoting) and message/partial to allow breaking a large message into smaller ones automatically to be reassembled by the recipient.

MULTIPART

- Data that is comprised of multiple components which may individually have different MIME types.
- Examples include
  - **multipart/form-data:** for data produced using the Formdata API .
  - **multipart/byteranges:** defined in RFC 7233: 5.4.1 and used with HTTP's 206 (Status code) "Partial Content" response returned when the fetched data is only part of the content, such as is delivered using the Range header.

List of MIME types

MIME Type	Type of file	Associated extension
application/atom+xml	Files in ATOM format	atom
application/iges	CAS files	iges
application/javascript	JavaScript files	js
application/dxf	AutoCAD files	dxf
application/mp4	MPEG4 files	mp4
application/iges	IGES CAD exchange format	igs,iges
application/octet-stream	Non-interpreted binary files	bin
application/msword	Microsoft Word document files	doc
application/pdf	Adobe Acrobat files	pdf
application/postscript	PostScript files	ai,eps,ps
application/rtf	Rich text format	rtf
application/sgml	SGML files	sgml
application/vnd.ms-excel	Microsoft Excel spreadsheet files	xls
application/vnd.ms-powerpoint	Microsoft Powerpoint presentation files	ppt

CONTEXT-TYPE: <TYPE / SUBTYPE>

Type	Sub type	Description
Text	Plain	Unformatted text in US ASCII ISO 8859.
Image	jpeg	Image in JPEG Format.
	gif	Image in GIF format.
Video	mpeg	MPEG format.
Audio	Basic	Single- channel encoding of voice at 8kHz.
Message	rfc 882	The boby is an encapsulated message that confirms to RFC 822.
	partial	Large mail is fragmented.
	External Body	contains pointer to an object that exists elsewhere and is accessible via FTP. TFTP etc.
Multipart	Mixed	The different parts are independent but are to be transmitted together. They should be presented to receiver in the appear in mail message.
	Parallel	same as mixed but order not defined.
	Alternate	The different parts are alternate versions of the same information
	Digest	similar to mixed, but the default type/subtype of each part is message/rfc 822.
Applciation	Postscript	Adobe postscript .
	Octet-stream	General binary data consisting of 8-bit bytes (Octets).

Header encoding

- The transfer-encoding header is used to specify an encoding format for the message body, but it doesn't solve the problem of encoding headers themselves (such as the message subject).
- To encode headers with character sets which use more than 7 bits, such as for including accented letters in an email's subject, the MIME standard offers the following format:

  =?charset?encoding?result?=

- Charset represents the character set used,
- Encoding defines the encoding desired with two possible values:
  - Q for quoted-printable
  - B for base64
- Result: text encoded using the method specified.
- Below is an example of Quoted-Printable encoding with "Building faÃ§ade" as the email's subject.

  Subject: Building fa=?ISO-8859-1?Q?=E7ade?=

PRIMARY MIME TYPES

- MIME types, used in the Content-Typeheader, are used to classify documents attached to an email.
- A MIME type is comprised as follows:
  - Content-type:main_mime_type/mime_subtype
  - For example, A GIF image has the following MIME type:
    - **Content-type:image/gif**
- The primary data types, sometimes called "discrete data types," are:
  - text: readable text data text/rfc822 [RFC822]
  - text/plain [RFC2646]; text/html [RFC2854]
  - image: binary data representing digital images
  - image/jpeg; image/gif; image/png.

Primary MIME types

- The primary data types, sometimes called "discrete data types," are:
  - audio: digital sound data: audio/basic; audio/wav
  - video: video data: video/mpeg
  - application:Other binary data: application/octet-stream; application/pdf.
- MIME types are also used on the Web to classify documents transferred using the protocol HTTP.
- Thus during a transaction between a web server and a browser, the first thing the web server does is send the MIME type of the file to the browser, so that the browser knows how to display the document.

ENCODING FORMATS

- To transfer binary data, MIME offers five encoding formats which canbe used in the header transfer-encoding:
  - 7bit: 7-bit text format (for messages without accented characters);
  - 8bit: 8-bit text format;
  - quoted-printable: Quoted-Printableformat, recommended for messages which use a 7-bit alphabet (such as when there are accent marks);
  - base64:Base 64, recommended for sending binary files as attachments;
  - binary: binary format; not recommended.Since
- MIME is very open, it can use third-party encoding formats such as:
  - BinHex(a proprietary format belonging to Apple),
  - uuencode
  - xxencode