Il formato MIME

Antonio Lioy
< lioy @ polito.it>

Politecnico di Torino
Dip. Automatica e Informatica

Che cosa è MIME?

- **Multipurpose Internet Mail Extensions**
- **un formato dati per trasmettere informazioni in Internet**
- **inizialmente pensato per superare alcuni limiti della posta elettronica (RFC-821/822):**
  - canale a 7 bit
  - caratteri US-ASCII
  - righe inferiori a 1000 caratteri
- **oggi usato per trasmettere / identificare dati in molti altri contesti, con caso più importante il web**

Definizione base di MIME

- **RFC-2045 “MIME part one: format of Internet message bodies”**
- **RFC-2046 “MIME part two: media types”**
- **RFC-2047 “MIME part three: message header extensions for non-ASCII text”**
- **RFC-2048 “MIME part four: registration procedures”**
- **RFC-2049 “MIME part five: conformance criteria and examples”**

© Antonio Lioy - Politecnico di Torino (2003-2009)

Estensioni di MIME

- **RFC-2183 "Communicating presentation information in Internet messages: the Content-Disposition header field"**
- **RFC-2231 "MIME parameter value and encoded word extensions: character sets, languages, and continuations"**
- **RFC-2557 "MIME encapsulation of aggregate documents, such as HTML (MHTML)"**
- **RFC-3023 "XML media types"**
- **RFC-3676 "The Text/Plain format and DelSp parameters"**
- **RFC-3798 "Message disposition notification"**

Header MIME (I)

- **Mime-Version: 1.0**
- **Content-Type: tipo/sottotipo ; parametri**
  - text, application, audio, image, message, model, video, multipart
  - con sottotipo (es. image/gif)
  - parametri opzionali (es. name=bb.gif)
  - parametri spesso nella forma "nome = valore"
- **Content-Transfer-Encoding:**
  - non codificati: 7bit, 8bit, binary
  - codificati: base64, quoted-printable

Header MIME (II)

- **Content-Id:**
  - identificativo dei dati generato dal mittente
  - opzionale e poco usato
- **Content-Description:**
  - descrizione testuale dei dati generata dal mittente
  - opzionale e poco usato
- **Content-Disposition: disposizione ; parametri**
  - disposizione: inline, attachment
  - filename=
  - creation-date= , modification-date= , read-date=
  - size=

Non-codifiche MIME

- **7bit**
  - righe < 1000 caratteri
  - caratteri ASCII (ossia con MSB=0)
  - è il default
- **8bit**
  - righe < 1000 caratteri
  - caratteri potenzialmente con MSB=1
- **binary**
  - righe di lunghezza anche > 1000 caratteri
  - caratteri potenzialmente con MSB=1

Codifica "base64"

- **6 bit = 1 carattere ASCII stampabile**
- **dimensione del messaggio: +33%**
- **genera righe di 76 caratteri ASCII stampabili**
- **utile per messaggi binari**

- **esempio: "`C'è`" diventa "`Qyfo`"**

```
C'è = 0x43       0x27       0xe8
    = 0100 0011 0010 0111 1110 1000
    = 010000 110010 011111 101000
    = 16     50     31     40
    = Q      y      f      o
```

Tabella di conversione base64

0 A	17 R	34 i	51 z
1 B	18 S	35 j	52 0
2 C	19 T	36 k	53 1
3 D	20 U	37 l	54 2
4 E	21 V	38 m	55 3
5 F	22 W	39 n	56 4
6 G	23 X	40 o	57 5
7 H	24 Y	41 p	58 6
8 I	25 Z	42 q	59 7
9 J	26 a	43 r	60 8
10 K	27 b	44 s	61 9
11 L	28 c	45 t	62 +
12 M	29 d	46 u	63 /
13 N	30 e	47 v	
14 O	31 f	48 w	(pad) =
15 P	32 g	49 x	
16 Q	33 h	50 y	

Codifica “quoted-printable”

- codificati solo i caratteri ASCII > 127
- codifica: ‘=hh’
- codifica dipendente dall’alfabeto (!)
- aumento variabile del messaggio
- righe non superiori a 76 caratteri
- righe spezzate mettendo ‘=’ come ultimo carattere
- codificato anche il carattere ‘=’ come ‘=3D’

- esempio: “`c'è`” diventa “`c'=e8`” (in ISO-8859-1)

Alfabeti MIME

Specificati tramite il parametro **charset=**

- US-ASCII
- ISO-8859-1 (Latin-1)
- ISO-8859-2 (Latin-2)
- ISO-8859-3 (Latin-3)
- ISO-8859-4 (Latin-4)
- ISO-8859-5 (Latin/Cyrillic)
- ISO-8859-6 (Latin/Arabic)
- ISO-8859-7 (Latin/Greek)
- ISO-8859-8 (Latin/Hebrew)
- ISO-8859-9 (Latin-15)

ISO-8859-15

- anche detto Latin-9
- adattamento di ISO-8859-1 ai seguenti simboli:
  - Euro
  - caratteri per la traslitterazione di parole russe
  - alcune legature Francesi
- al costo di perdere il simbolo di pipe (|) ed alcuni simboli diacritici isolati (umluat, cedilla, …) ma non quando usati sulle lettere

- nota: windows-1252 ha tutti i caratteri stampabili di iso-8859-1/-15 ma alcuni con codici numerici diversi

MIME content-type – esempi (I)

- **text/plain**
  - normale testo ASCII (7 bit, alfabeto US-ASCII)
  - parametro opzionale “charset=…” richiede tipicamente anche “encoding=…”
- **text/enriched**
  testo con speciali sequenze di formattazione (es. <bold> ... </bold>)
- **text/html, text/xml, text/css, text/csv, ...**
- **video/mpeg, video/quicktime, ...**
- **audio/basic**
  mono, 8 bit, codifica mu-law ISDN, campionamento a 8 KHz

MIME content-type – esempi (II)

- **multipart/mixed**
  contiene parti diverse, da visualizzare secondo l’ordine di ricezione
- **multipart/parallel**
  contiene parti diverse, da visualizzare senza un ordine prefissato
- **multipart/alternative**
  contiene gli stessi dati in formato diverso, in ordine crescente di fedeltà
- **multipart/digest**
  composizione di più messaggi di posta (parti tutte message/rfc822)

MIME content-type – esempi (III)

- **message/rfc822**
  contiene un mail (es. per il forward)
- **message/partial**
  parte di un mail che è stato spezzato
- **message/external-body**
  riferimento a dati esterni (Web, FTP, ...)
- **image/gif, image/jpeg**
  immagine in formato GIF o JPEG con codifica JFIF
- **application/postscript, application/pdf**
  documento Postscript, PDF
- **application/octet-stream**
  formato applicativo generico

© *Antonio Lioy - Politecnico di Torino (2003-2009)*

Estendibilità di MIME

- **nuove definizioni possono essere aggiunte in futuro per:**
  - content-type
  - content-transfer-enconding
  - content-disposition
- **fino a quando un'estensione non è documentata in un RFC è da considerarsi "privata" e deve iniziare con x- o X-**
- **elenco delle definizioni esistenti (gestito da IANA):**

  http://www.iana.org/assignments/media-types/

Esempio: MIME in e-mail

```
From: Antonio Lioy <lioy@polito.it>
To: Antonio Lioy <lioy@polito.it>
MIME-Version: 1.0
Subject: test di MIME
Content-Type: multipart/mixed; boundary="0107040803040507"

This is a multi-part message in MIME format.

--0107040803040507
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
Content-Transfer-Encoding: quoted-printable

Questa =E8 la parte di testo.
```

```
--0107040803040507
Content-Type: application/x-zip-compressed; name="tesi.zip"
Content-Transfer-Encoding: base64
Content-Disposition: inline; filename="tesi.zip"

dwIAAEQFAAAIAAAAdGVzaS50eHRtVMuO1DAQvEfK/as453ggfff
. . .
kudHh0UEsFBgAAAAABAAEANgAAAJ0CAAAAAA==
--0107040803040507
Content-Type: application/vnd.ms-excel; name="voti.xls"
Content-Transfer-Encoding: base64
Content-Disposition: inline; filename="voti.xls"

0M8R4KGxGuEAAAAAAAAAAAAAAAAAAAAAAPgADAP7/CFaggTTTTT
. . .
OTA1OTJjPl0NPj4Nc3RhcnR4cmVmDTE3Mw0lJUVPRg0=
--0107040803040507--
```