

EAI

Email Address Internationalization

Marcos Sanz
DENIC eG
sanz@denic.de

Abstract

- Introduction
- Overview
- Components:
 - UTF-8 Headers
 - SMTP Service Extension
 - Downgrading mechanisms
 - POP3 Extensions
- Outlook and Open Issues
- References

Introduction

- ASCII art:
$$\text{"ø"} = \text{"o"} + \text{Backspace} + \text{"/"}$$
- First MIME (Multipurpose Internet Mail Extensions) back in 1992:
 - Allowing different charsets in mail body (RFC 2045 and 2046)
 - Some header field bodies (RFC 2047)
- IDNA (RFC 3490) internationalizes domain part
- First email address internationalization efforts in the IETF back to 2003

Introduction (II)

Current mail system allows for

Marcos Sanz Grossón <sanz@denic.de>

=?ISO-8859-1?Q?Marcos_Sanz_Gross=F3n?= <sanz@denic.de>

but not for

マルコス サンズ <サンズ@denic.de>

or

マルコス サンズ <マルコス@サンズ.de>

Introduction (III)

- The "display name" solution is not perceived as sufficient:
 - No global directories
 - Protocol elements displayed to the user
 - "I want an address with my proper name"
- Solution alternatives:
 - Client-side presentation layer?
 - Mail protocol?
 - New layer on top?

Overview

- Solution in the mail architecture:
 - Update of Internet message format to allow for Unicode header field bodies
 - Update of SMTP to allow for Unicode addresses in the envelope
 - Backward-compatibility mechanisms
 - Update of other mail-related protocols: POP, IMAP...

UTF-8 Headers

- Redefinition of **local-part** and **domain** in RFC 2822 to accept Unicode characters beyond ASCII:

addr-spec = local-part "@" domain

- Domain part compliant to IDNA
- Originating MUA includes a presence header:

i18n-email: 1.0

UTF-8 Headers (II)

- Header field *bodies* affected:
from, sender, reply-to, to, cc, bcc, resent-*
and **unstructured, comment** and **phrase** like
subject, comments, keywords
- Header field *bodies not* affected:
date, message-id, in-reply-to, references,
resent-*
- **received** and **oder** trace fields are under discussion

SMTP Service Extension

- EHLO keyword value without parameters:
IEmail
- SMTP server must support 8BITMIME (RFC 1652)
- Commands MAIL and RCPT become two optional parameters:
 - **ALT-ADDRESS**
 - **ATOMIC**

SMTP Service Extension (II)

- **IMail** = server accepts UTF-8 headers and a UTF-8 string in any instance of **mailbox** in the envelope
 - **local-part** must be treated transparently (unless MDA)
 - **domain** must be dealt according to IDNA
- **ALT-ADDRESS** = fallback ASCII address
- **ATOMIC** = boolean to allow for conversion
ALT-ADDRESS = ACE(mailbox)

SMTP Service Extension (III)

müller@example.org

- If SMTP server supports **IMail**, no problem
- If SMTP server does not support **IMail**:
 - If **ALT-ADDRESS** set by client, use it
mueller@example.org
 - If **ALT-ADDRESS** not set
 - If **ATOMIC** true, apply ACE conversion
bq--mller-kva@example.org
 - If **ATOMIC** false, reject & bounce

Downgrading

- If any link in the chain does not support EAI, downgrading takes place for a best-effort delivery:
 - SMTP session downgrade
 - Header downgrade
- Requirements:
 - unique
 - automated
 - lightweight
 - header-preserving
 - recorded

Downgrading (II)

- SMTP session downgrade:
 - **MAIL FROM** and **RCPT TO** use **ALT-ADDR** and **ATOMIC**
 - Tracking by including new header fields in the Message itself:
 - **IMA-Downgraded-From**
 - **IMA-Downgraded-To**

IMA-Downgraded-To: <müller@example.org>
<bq--mller-kva@example.org>

Downgrading (III)

- Header downgrade (1st option)
 - MIME encapsulation, Content-Type: **Message/EAI**
 - Message is MIME-encoded and becomes a new MIME part
 - Reuse as many headers as possible
 - Pros: Preserves headers, easy for MTAs
 - Cons: MUAs must be updated

Downgrading (IV)

- Header downgrade (2nd option)
 - Each header has its own downgrading method: Basically MIME-encoding RFC 2047
 - Origin/Destination addresses converted to ACE with the same method as SMTP downgrading.
 - Pros: Old MUAs can deal with it (except for original origin/destination addresses)
 - Cons: Difficult to implement, header by header

POP3 Extensions

- RFC 2449 discourages POP3 extensions
- "Just send UTF-8":
 - deploys faster
 - creates interoperability problems
- Capability Tag: **UTF8**
- New commands: **RET8, LST8**
 - If **USER** (optional argument): UTF-8 in user names and passwords and applies SASLprep to the args of **USER, PASS** and **APOP**
 - If **LST8** (optional argument): Same as **LIST** but with octet counts as of **RET8**

POP3 Extensions (II)

- Command **NO-RETR** for future **RETR** deprecation
- Up-conversion:
 - POP3 clients best when simple
 - Mandatory for main headers
 - Nice-to-have for MIME headers
- **TOP8** needed?
 - **TOP** is an optional command
 - Partial fetch, use IMAP instead?

Outlook and Open Issues

- Milestones of the IETF wg:
 - All documents ready by the end of 2006
 - Start with Experimental Status
 - Diagnose potential fragmentation resulting from experiment
 - Recharter to Standards Track in March 2007, if experiments successful

Outlook and Open Issues (II)

- Better separation of envelope vs data
- Better support/definition of upgrading
- Identify scenarios with special attention to interactions in mailing lists
- Interaction with DNS-based mail authentication mechanisms (SPF, DomainKeys/DKIM...)
- Explore right-to-left writing systems
- Revision of collateral damages:
 - mailto URL scheme, IRI, IMAP, LDAP, vCARD, ACAP, S/MIME...

Outlook and Open Issues (III)

- Local part syntax definition
 - All of the Unicode range? Cf. existing namespace:
 A-Z, a-z, 0-9, !, #, \$, %, &, ', *, +, -,
 /, =, ?, ^, _, ` , {, |, }, ~
 - Security issues. Remember?
 ZZz@PUBLIKUM.DE $\neq$ ZZz@PUBLIKUM.DE
 Now it's the problem of *mail providers*!
 - Identifier and Pattern Syntax (Unicode UAX#31)
 - Maybe even define a profile of UAX #31:
 - Security Profile for Identifiers (Unicode UTR#36)

- The question of about 11,000 DE domain holders is obviously:

"Will the 'ß' in my name be allowed in my email address?"

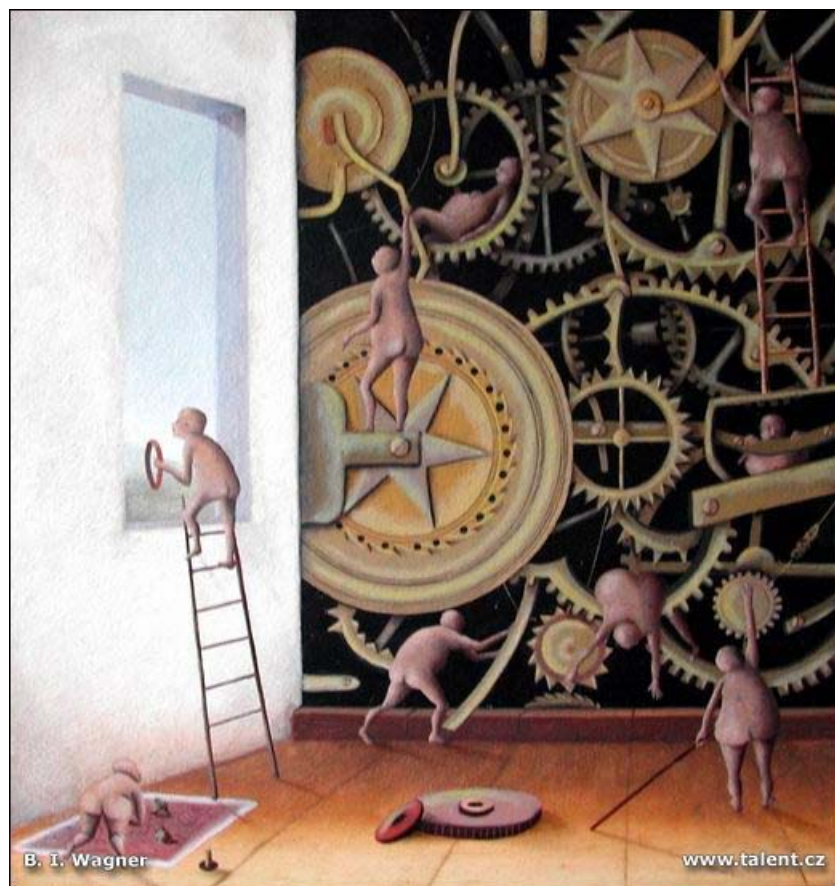
- The answer so far: *"It depends"*
 - If local-part remains UTF-8 unnormalized or applies a normalization that does not require case roundtrip, it will.

References

- RFC 1652 SMTP Service Extension for 8bit-MIME transport. J. Klensin, N. Freed, M. Rose, E. Stefferud, D. Crocker. July 1994
- RFC 1939 Post Office Protocol - Version 3. J. Myers, M. Rose. May 1996
- RFC 2045 Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies. N. Freed, N. Borenstein. November 1996
- RFC 2046 Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types. N. Freed, N. Borenstein. November 1996
- RFC 2047 MIME (Multipurpose Internet Mail Extensions) Part Three: Message Header Extensions for Non-ASCII Text. K. Moore. November 1996
- RFC 2244 ACAP -- Application Configuration Access Protocol. C. Newman, J. G. Myers. November 1997
- RFC 2368 The mailto URL scheme. P. Hoffman, L. Masinter, J. Zawinski. July 1998
- RFC 2449 POP3 Extension Mechanism. R. Gellens, C. Newman, L. Lundblade. November 1998
- RFC 2821 Simple Mail Transfer Protocol. J. Klensin, Ed.. April 2001
- RFC 2822 Internet Message Format. P. Resnick, Ed.. April 2001
- RFC 3490 Internationalizing Domain Names in Applications (IDNA). P. Faltstrom, P. Hoffman, A. Costello. March 2003
- RFC 3501 INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1. M. Crispin. March 2003
- RFC 3629 UTF-8, a transformation format of ISO 10646. F. Yergeau. November 2003
- RFC 3987 Internationalized Resource Identifiers (IRIs). M. Duerst, M. Suignard. January 2005
- RFC 4013 SASLprep: Stringprep Profile for User Names and Passwords. K. Zeilenga. February 2005

References

- Email Address Internationalization, IETF working group, <http://www.ietf.org/html.charters/eai-charter.html>
<https://www1.ietf.org/mailman/listinfo/ima>
- Internationalized Email Addresses: Scenarios, <http://www.ietf.org/internet-drafts/draft-alvestrand-i18mail-scenarios-00.txt>
- Overview and Framework for Internationalized Email, <http://www.ietf.org/internet-drafts/draft-klensin-ima-framework-01.txt>
- Internationalization in Internet Applications: Issues, Tradeoffs, and Email Addresses, <http://www.ietf.org/internet-drafts/draft-klensin-ima-constraints-00.txt>
- Internationalized Email Headers, <http://www.ietf.org/internet-drafts/draft-yeh-ima-utf8headers-01.txt>
- SMTP extension for internationalized email address, <http://www.ietf.org/internet-drafts/draft-yao-ima-smtpext-02.txt>
- POP3 Support for UTF-8, <http://www.ietf.org/internet-drafts/draft-newman-ima-pop-00.txt>
- Downgrading mechanism for Internationalized eMail Address (IMA), <http://www.ietf.org/internet-drafts/draft-yoneya-ima-downgrade-01.txt>
- UAX #31, Identifier and Pattern Syntax, Unicode Standard Annex #31, <http://www.unicode.org/reports/tr31/>
- UTR #36, Unicode Security Considerations, Unicode Technical Report #36, <http://www.unicode.org/reports/tr36/>
- Sicherheitsaspekte bei IDN, M. Sanz, https://member.secure.denic.de/media/pdf/technischesmeeting/20050928/TM_SANZ_IDN_Security.pdf



sanz@denic.de