

Groupe de travail Réseau
Request for Comments : 2047
 RFC rendues obsolètes : 1521, 1522, 1590
 Catégorie : Sur la voie de la normalisation

K. Moore
 University of Tennessee
 novembre 1996
 Traduction Claude Brière de L'Isle

MIME (Extensions de messagerie Internet multi-objets) Partie trois : extensions d'en-tête de message pour texte non ASCII

Statut du présent mémoire

Le présent mémoire définit un protocole expérimental pour la communauté de l'Internet. Le présent mémoire ne spécifie aucune sorte de norme de l'Internet. La discussion et les suggestions pour son amélioration sont bienvenues. La distribution du présent mémoire n'est soumise à aucune restriction.

(La présente traduction incorpore les errata n° 504 du 11/10/2001 et n° 506 du 13/02/2003)

Résumé

Le STD 11, RFC 822, définit un protocole de représentation de message qui spécifie des détails considérables sur les en-têtes de message en US-ASCII, et laisse le contenu du message, un corps de message, en texte US-ASCII plat. Cet ensemble de documents, collectivement appelés "Extensions de messagerie Internet multi-objets", ou MIME, redéfinit le format des messages pour permettre :

- (1) des corps de message textuels dans des jeux de caractères autres que l'US-ASCII,
- (2) un ensemble extensible de formats différents pour les corps de message non textuels,
- (3) des corps de message multi-parties, et
- (4) des informations d'en-tête textuelles dans des jeux de caractères autres que l'US-ASCII.

Ces documents se fondent sur des travaux antérieurs documentés dans la RFC 934, STD 11, et dans la RFC 1049, mais les étendent et les révisent. Parce que la RFC 822 en disait si peu sur les corps de message, ces documents prennent largement le contre-pied de la RFC 822 (plutôt que d'en être une révision).

Ce document particulier est le troisième de la série. Il décrit les extensions à la RFC 822 pour permettre des données de texte non US-ASCII dans les champs d'en-tête de messagerie Internet.

Les autres documents de la série incluent :

- + la RFC 2045, qui spécifie les divers en-têtes utilisés pour décrire la structure des messages MIME.
- + la RFC 2046, qui définit la structure générale du système de classification des supports MIME et définit un ensemble initial de types de supports,
- + la RFC 2048, qui spécifie diverses procédures d'enregistrement de l'IANA pour les services en rapport avec MIME,
- + la RFC 2049, qui décrit les critères de conformité à MIME et donne quelques exemples comme illustration de formats de message MIME, les remerciements et la bibliographie.

Ces documents sont les révisions des RFC 1521, 1522 et 1590, qui étaient elles-mêmes les révisions des RFC 1341 et 1342. Un appendice de la RFC 2049 décrit les différences et changements par rapport aux versions précédentes.

Table des matières

1. Introduction.....	2
2. Syntaxe des mots codés.....	3
3. Jeux de caractères.....	3
4. Codages.....	4
4.1 Codage "B".....	4
4.2 Codage "Q".....	4
5. Utilisation des mots codés dans les en-têtes de message.....	4
6. Prise en charge des "mots codés" par les lecteurs de messagerie.....	5
6.1 Reconnaissance des "mots codés" dans les en-têtes de message.....	5
6.2 Affichage des "mots codés".....	6
6.3 Traitement par le lecteur de messagerie des "mots codés" incorrectement formés.....	6
7. Conformité.....	7
8. Exemples.....	7
9. Références.....	8

10. Considérations pour la sécurité.....	8
11. Remerciements.....	8
12. Adresse de l'auteur.....	9
Appendice - Changements depuis la RFC 1522 (sans ordre particulier).....	9

1. Introduction

La RFC 2045 décrit un mécanisme pour noter la partie textuelle du corps qui est codée en divers jeux de caractères, ainsi que les méthodes de codage de telles parties de corps comme des séquences de caractères US-ASCII imprimables. Le présent mémoire décrit des techniques similaires pour permettre le codage de texte non ASCII dans diverses portions d'un en-tête de message de la RFC 822 [2], d'une manière qui ait peu de chances de troubler les logiciels de traitement de message existants.

Comme les techniques de codage décrites dans la RFC 2045, les techniques exposées ici ont été conçues pour permettre l'usage de caractères non ASCII dans les en-têtes de message d'une façon qui ait peu de chances d'être perturbée par les bizarreries des programmes de traitement de messagerie Internet existants. En particulier, certains programmes de relais de messagerie sont connus pour (a) supprimer certains champs d'en-tête de message tout en conservant d'autres, (b) réarranger l'ordre des adresses dans les champs To ou Cc, (c) réarranger l'ordre (vertical) des champs d'en-tête, et/ou (d) "envelopper" les en-têtes de message en des endroits différents de ceux du message d'origine. De plus, certains programmes de lecture de messagerie sont connus pour avoir des difficultés à analyser correctement les en-têtes de message qui, tout en étant légaux au regard de la RFC 822, font usage de l'encadrement de citations entre des barres obliques inverses pour "cacher" les caractères spéciaux tels que "<", ",", ou ":", ou qui exploitent d'autres dispositifs d'utilisation rare de cette spécification.

Bien qu'on déplore que ces programmes n'interprètent pas correctement les en-têtes de la RFC 822, "casser" ces programmes causerait de sévères problèmes opérationnels au système de messagerie de l'Internet. Les extensions décrites dans le présent mémoire ne s'appuient donc pas sur les dispositifs peu utilisés de la RFC 822.

Au lieu de cela, certaines séquences de caractères ASCII imprimables "ordinaires" (appelées "mots codés") sont réservées pour être utilisées comme données codées. La syntaxe des mots codés est telle qu'il est peu vraisemblable qu'ils apparaissent "accidentellement" comme texte normal dans les en-têtes de message. De plus, les caractères utilisés dans les mots codés se restreignent à ceux qui n'ont pas de signification particulière dans le contexte dans lequel apparaissent les mots codés.

Généralement, un "mot codé" est une séquence de caractères ASCII imprimables qui commence par "=?", se termine par "=?", et a deux "?"s entre. Cela spécifie un jeu de caractères et une méthode de codage, et cela inclut aussi le texte original codé comme des caractères ASCII graphiques, conformément aux règles pour cette méthode de codage.

Un composeur de messagerie qui met en œuvre la présente spécification fournira un moyen pour entrer du texte non ASCII dans les champs d'en-tête, mais traduira ces champs (ou les portions appropriées de ces champs) en mots codés avant de les insérer dans l'en-tête de message.

Un lecteur de messagerie qui met en œuvre la présente spécification reconnaîtra les mots codés lorsque ils apparaissent dans certaines portions de l'en-tête de message. Au lieu d'afficher le mot codé "tel qu'il est", il inversera le codage et affichera le texte original dans le jeu de caractères pour lequel il a été conçu.

Notes :

Le présent mémoire s'appuie largement sur la notation et les termes définis dans les RFC 822 et 2045. En particulier, la syntaxe de l'ABNF utilisée dans le présent mémoire est définie dans la RFC 822, tout autant que beaucoup des symboles terminaux ou non terminaux de la RFC 822 sont utilisés dans la grammaire des extensions d'en-tête définies ici. Parmi les symboles définis dans la RFC 822 et référencés dans le présent mémoire figurent : 'addr-spec', 'atom', 'CHAR', 'comment', 'CTL', 'ctext', 'linear-white-space', 'phrase', 'quoted-pair', 'quoted-string', 'SPACE', et 'word'. La réussite de la mise en œuvre de la présente extension de protocole exige le respect scrupuleux des définitions de ces termes de la RFC 822.

Lorsque le terme "ASCII" apparaît dans le présent mémoire, il se réfère à la "norme américaine de code à 7 bits pour les échanges d'information", ANSI X3.4-1986. Le nom du jeu de caractères MIME pour ce jeu de caractères est "US-ASCII". Lorsque on ne se réfère pas spécifiquement au nom du jeu de caractères MIME, le présent document utilise le terme "ASCII" à la fois pour faire court et par souci de cohérence avec la RFC 822. Cependant, les développeurs sont

avertis que le nom du jeu de caractères doit être épelé "US-ASCII" dans les messages et en-têtes de parties de corps MIME.

Le présent mémoire spécifie un protocole pour la représentation de texte non ASCII dans les en-têtes de message. Il NE DÉFINIT PAS de traduction entre les "en-têtes à 8 bits" et les purs en-têtes ASCII, pas plus qu'il ne suppose la possibilité d'une telle traduction.

2. Syntaxe des mots codés

Un "mot codé" est défini par la grammaire ABNF suivante. La notation de la RFC 822 est utilisée, sauf que des caractères d'espace blanche NE DOIVENT PAS apparaître entre les composants d'un "mot codé".

```

mot-codé = "=?" jeu de caractères "?" codage "?" texte codé "?="
jeu de caractères = jeton ; voir à la section 3
codage = jeton ; voir à la section 4
jeton = 1*<tout CARACTÈRE excepté ESPACE, CTL, et spéciaux>
spéciaux = "(" / ")" / "<" / ">" / "@" / "," / ";" / ":" / "\" / "<" / ">" / "[" / "]" / "?" / "." / "="
texte codé = 1*<tout caractère ASCII imprimable autre que "?" ou ESPACE>
; (mais voir "Utilisation des mots codés dans les en-têtes de message", section 5)

```

Les noms de 'codage' et 'jeu de caractères' sont indépendants de la casse. Et donc le nom de jeu de caractère "ISO-8859-1" est équivalent à "iso-8859-1", et le codage dénommé "Q" peut s'écrire "Q" ou "q".

Un "mot codé" ne peut pas être long de plus de 75 caractères, y compris 'jeu de caractères', 'codage', 'texte codé', et les délimiteurs. Si on désire coder plus de texte qu'il n'en va tenir dans un "mot codé" de 75 caractères, plusieurs "mots codés" (séparés par CRLF ESPACE) peuvent être utilisés.

Bien qu'il n'y ait aucune limite à la longueur d'un champ d'en-tête de plusieurs lignes, chaque ligne d'un champ d'en-tête qui contient un ou plusieurs "mots codés" est limitée à 76 caractères.

Les restrictions de longueur sont incluses à la fois pour faciliter l'interopérabilité à travers les passerelles de messagerie inter réseaux, et pour imposer une limite à la quantité de données exploratoires qu'un analyseur d'en-tête doit utiliser (lorsque il recherche un délimiteur final ?=) avant qu'il puisse décider si un jeton est un "mot codé" ou quelque chose d'autre.

IMPORTANT : Les "mots codés" sont conçus pour être reconnus comme des "atomes" par un analyseur de la RFC 822. Par conséquent, les caractères d'espaces blanches non codés (tels que ESPACE et HTAB) sont INTERDITS au sein d'un "mot codé". Par exemple, la séquence de caractères =?iso-8859-1?q?ceci est un texte?= serait analysée comme quatre "atomes", plutôt que comme un seul "atome" (par un analyseur de la RFC 822) ou comme "mot codé" (par un analyseur qui comprend les "mots codés"). La façon correcte de coder la chaîne "ceci est un texte" est de coder aussi les caractères ESPACE, par exemple, =?iso-8859-1?q?ceci=20est=20un=20texte?=.

Les caractères qui peuvent apparaître dans le "texte codé" sont de plus limités par les règles de la section 5.

3. Jeux de caractères

La portion "jeu de caractères" d'un "mot codé" spécifie le jeu de caractères associé au texte non codé. Un "jeu de caractères" peut être tout nom de jeu de caractères permis dans un paramètre "jeu de caractères" MIME d'une partie de corps "texte/clair", ou tout nom de jeu de caractères enregistré auprès de l'IANA pour être utilisé avec le type de contenu MIME texte/clair.

Certains jeux de caractères utilisent des techniques de changement de code pour passer du "mode ASCII" à d'autres modes. Si du texte non codé dans un "mot codé" contient une séquence qui amène l'interpréteur de jeu de caractères à sortir du mode ASCII, il DOIT contenir des codes de contrôle supplémentaires tels que le mode ASCII soit à nouveau choisi à la fin du "mot codé". (Cette règle s'applique séparément à chaque "mot codé", y compris les "mots codés" adjacents au sein d'un seul champ d'en-tête.)

Lorsque il y a une possibilité d'utiliser plus d'un jeu de caractères pour représenter le texte dans un "mot codé", et en l'absence d'accords privés entre l'envoyeur et les receveurs d'un message, il est recommandé que les membres de la série ISO-8859-* soient utilisés de préférence à tout autre jeu de caractères.

4. Codages

Au départ, les valeurs légales pour "codage" sont "Q" et "B". Ces codages sont décrits ci-dessous. L'utilisation du codage "Q" est recommandée lorsque la plupart des caractères à coder sont dans le jeu de caractères ASCII ; autrement, le codage "B" devrait être utilisé. Néanmoins, un lecteur de messagerie qui revendique la reconnaissance des "mots codés" DOIT être capable d'accepter l'un et l'autre codage pour tout jeu de caractères qu'il prend en charge.

Seul un sous-ensemble des caractères ASCII imprimables peut être utilisé dans un "texte codé". Les caractères d'espace et de tabulation ne sont pas admis, afin que le début et la fin d'un "mot codé" soient évidents. Le caractère "?" est utilisé au sein d'un "mot codé" pour séparer les diverses portions du "mot codé" les unes des autres, et ne peut donc pas apparaître dans la portion de "texte codé". D'autres caractères sont aussi illégaux dans certains contextes. Par exemple, un "mot codé" dans une "phrase" précédant une adresse dans un champ d'en-tête From ne peut contenir aucun des caractères "spéciaux" définis dans la RFC 822. Finalement, certains autres caractères ne sont pas permis dans certains contextes, pour assurer la fiabilité des messages qui passent à travers les passerelles de messagerie inter réseaux.

Le codage "B" satisfait automatiquement à ces exigences. Le codage "Q" permet d'utiliser une large gamme de caractères imprimables dans des localisations non critiques dans l'en-tête de message (par exemple, Subject) avec moins de caractères disponibles dans des utilisations sur d'autres localisations.

4.1 Codage "B"

Le codage "B" est identique au codage "BASE64" défini par la RFC 2045.

4.2 Codage "Q"

Le codage "Q" est similaire au codage de transfert de contenu "Quoted-Printable" défini dans la RFC 2045. Il est conçu pour permettre du texte contenant principalement des caractères ASCII qui vont être déchiffrables sur un terminal ASCII sans décodage.

- (1) Toute valeur de 8 bits peut être représentée par un "=" suivi par deux chiffres hexadécimaux. Par exemple, si le jeu de caractères utilisé était ISO-8859-1, le caractère "=" serait donc codé "=3D", et un ESPACE par "=20". (Les majuscules devraient être utilisées pour les chiffres hexadécimaux de "A" à "F".)
- (2) La valeur hexadécimale de 8 bits 20 (par exemple, ISO-8859-1 ESPACE) peut être représentée par "" (souligné, ASCII 95.). (Ce caractère peut ne pas passer à travers certaines passerelles de messagerie inter réseaux, mais son utilisation va considérablement améliorer la lisibilité des données codées en "Q" avec les lecteurs de messagerie qui ne prennent pas en charge ce codage.) Noter que le "" représente toujours l'hexadécimal 20, même si le caractère ESPACE occupe une position de code différente dans le jeu de caractères utilisé.
- (3) Les valeurs de 8 bits qui correspondent aux caractères ASCII imprimables autres que "=", "?", et "" (souligné), PEUVENT être représentés par ces caractères. (Mais voir les restrictions à la section 5.) En particulier, ESPACE et TAB NE DOIVENT PAS être représentés par eux-mêmes au sein des mots codés.

5. Utilisation des mots codés dans les en-têtes de message

Un "mot codé" peut apparaître dans un en-tête de message ou un en-tête de partie de corps conformément aux règles suivantes :

- (1) Un "mot codé" peut remplacer un jeton "texte" (comme défini par la RFC 822) dans tout champ d'en-tête Subject ou Comments, dans tout champ d'en-tête de message d'extension, ou dans tout champ de partie de corps MIME pour lequel le corps de champ est défini comme "**texte". Un "mot codé" peut aussi apparaître dans tout champ d'en-tête de partie de corps ou de message ("X-") défini par l'utilisateur.

Les textes ASCII et les "mots codés" ordinaires peuvent apparaître ensemble dans le même champ d'en-tête. Cependant, un "mot codé" qui apparaît dans un champ d'en-tête défini comme "*texte" DOIT être séparé de tout "mot codé" ou "texte" adjacent par une "espace blanche linéaire".

- (2) Un "mot codé" peut apparaître au sein d'un "commentaire" délimité par "(" et ")", c'est-à-dire, chaque fois qu'un "ctext" est admis. Plus précisément, la définition ABNF de la RFC 822 pour "commentaire" est amendée comme suit :

commentaire = "(" *(ctext / quoted-pair / commentaire / mot codé) ")"

Un "mot codé" codé en "Q" qui apparaît dans un "commentaire" NE DOIT PAS contenir les caractères "(", ")" ou "\". De plus, un "mot codé" qui apparaît dans un "commentaire" DOIT être séparé de tout "mot codé" ou "ctext" adjacent par une "espace blanche linéaire".

Il est important de noter que les "commentaires" ne sont reconnus qu'à l'intérieur des corps de champ "structuré". Dans les champs dont les corps sont définis comme "*texte", "(" et ")" sont traités comme des caractères ordinaires plutôt que comme délimiteurs de commentaire, et la règle (1) de ce paragraphe s'applique. (Voir aux paragraphes 3.1.2 et 3.1.3 de la RFC 822.)

- (3) Comme remplacement d'une entité "mot" au sein d'une "phrase", par exemple, celle qui précède une adresse dans un champ From, To ou Cc. La définition ABNF pour "phrase" de la RFC 822 devient donc :

phrase = 1*(mot-codé / mot)

Dans ce cas, l'ensemble de caractères qui peut être utilisé dans un "mot codé" codé en "Q" est restreint à : <lettre ASCII majuscules et minuscules, chiffres décimaux, "!", "*", "+", "-", "/", "=", et "_" (souligné, ASCII 95.)>. Un "mot codé" qui apparaît dans une "phrase" DOIT être séparé de tout "mot", "texte" ou "spécial" adjacent par une "espace blanche linéaire".

Ce sont les SEULS endroits où un "mot codé" peut apparaître. En particulier :

- + Un "mot codé" NE DOIT PAS apparaître dans une portion d'une "addr-spec".
- + Un "mot codé" NE DOIT PAS apparaître au sein d'une "quoted-string".
- + Un "mot codé" NE DOIT PAS être utilisé dans un champ d'en-tête Received.
- + Un "mot codé" NE DOIT PAS être utilisé dans un paramètre d'un champ MIME Content-Type ou Content-Disposition, ni dans aucun corps de champ structuré excepté au sein d'un "commentaire" ou "phrase".

Le "texte codé" dans un "mot codé" doit être autonome ; le "texte codé" NE DOIT PAS se continuer d'un "mot codé" à un autre. Cela implique que la portion de "texte codé" d'un "mot codé" "B" sera un multiple d'une longueur de 4 caractères ; pour un "mot codé" "Q", tout caractère "=" qui apparaît dans la portion de "texte codé" sera suivi par deux caractères hexadécimaux.

Chaque "mot codé" DOIT coder un nombre entier d'octets. Le "texte codé" dans chaque "mot codé" doit être bien formé conformément au codage spécifié ; le "texte codé" peut ne pas se continuer dans le prochain "mot codé". (Par exemple, "=? jeu de caractères?Q?=?= =?jeu-de-caractères?Q?AB?=" serait illégal, parce que les deux chiffres hex "AB" doivent suivre le "=" dans le même "mot codé".)

Chaque "mot codé" DOIT représenter un nombre entier de caractères. Un caractère multi-octet ne peut pas être partagé à travers des "mots codés" adjacents.

Seuls les données de caractères imprimables et les espaces blanches devraient être codées en utilisant ce schéma. Cependant, comme ces schémas de codage permettent le codage de valeurs d'octet arbitraires, les lecteurs de messagerie qui mettent en œuvre ce décodage devraient aussi s'assurer que l'affichage des données décodées sur le terminal du receveur ne va pas causer d'effets collatéraux indésirables.

L'utilisation de ces méthodes pour coder des données non textuelles (par exemple, des images ou des sons) n'est pas définie par le présent mémoire. L'utilisation de "mots codés" pour représenter des chaînes de caractères purement ASCII est admise, mais déconseillée. Dans de rares cas, il peut être nécessaire de coder du texte ordinaire qui ressemble à un "mot codé".

6. Prise en charge des "mots codés" par les lecteurs de messagerie

6.1 Reconnaissance des "mots codés" dans les en-têtes de message

Un lecteur de messagerie doit analyser les en-têtes de message et de partie de corps conformément aux règles de la RFC 822 pour reconnaître correctement les "mots codés". La reconnaissance des "mots codés" se fait comme suit :

- (1) Tout champ d'en-tête de message ou partie de corps définie comme "*texte", ou tout champ d'en-tête défini par l'utilisateur, devrait être analysé comme suit : en commençant au début du corps du champ et en suivant immédiatement chaque occurrence "d'espace blanche linéaire", chaque séquence de jusqu'à 75 caractères imprimables (ne contenant aucune "espace blanche linéaire") devrait être examinée pour voir si elle est un "mot codé" conformément aux règles de syntaxe de la section 2. Toute autre séquence de caractères imprimables devrait être traitée comme du texte ASCII ordinaire.
- (2) Tout champ d'en-tête non défini comme "*texte" devrait être analysé conformément aux règles de syntaxe de ce champ d'en-tête. Cependant, tout "mot" qui apparaît au sein d'une "phrase" devrait être traité comme un "mot codé" si il satisfait aux règles de syntaxe de la section 2. Autrement, il devrait être traité comme un "mot" ordinaire.
- (3) Au sein d'un "commentaire", toute séquence faisant jusqu'à 75 caractères imprimables (ne contenant pas d'espace blanche linéaire), qui satisfait aux règles de syntaxe de la section 2, devrait être traitée comme un "mot codé". Autrement, elle devrait être traitée comme du texte normal de commentaire.
- (4) La présence d'un champ d'en-tête Version MIME N'EST PAS EXIGÉE pour que les "mots codés" soient interprété conformément à la présente spécification. Une raison en est que le lecteur de messagerie n'est pas supposé analyser la totalité de l'en-tête de message avant d'afficher les lignes qui peuvent contenir les "mots codés".

6.2 Affichage des "mots codés"

Tout "mot codé" ainsi reconnu est décodé, et si possible, le texte non codé résultant est affiché dans le jeu de caractères original.

Note : Le décodage et l'affichage des mots codés survient *après* qu'un corps de champ structuré soit analysé en jetons. Il est donc possible de cacher des caractères "spéciaux" dans des mots codés qui, à l'affichage, seront indistinguables des caractères "spéciaux" du texte environnant. Pour cette raison parmi d'autres, il N'EST PAS généralement possible de traduire un en-tête de message contenant des "mots codés" dans une forme non codée qui puisse être analysée par un lecteur de messagerie de la RFC 822.

Lors de l'affichage d'un champ d'en-tête particulier qui contient plusieurs "mots codés", toute "espace blanche linéaire" qui sépare une paire de "mots codés" adjacents est ignorée. (C'est pour permettre d'utiliser plusieurs "mots codés" pour représenter de longues chaînes de texte non codé sans avoir à séparer les "mots codés" où surviennent des espaces dans le texte non codé.)

Dans le cas où d'autres codages seraient définis à l'avenir, et où le lecteur de messagerie ne prendrait pas en charge le codage utilisé, il peut soit (a) afficher le "mot codé" comme du texte ordinaire, soit (b) substituer un message approprié pour indiquer que le texte n'a pas pu être décodé.

Si le lecteur de messagerie ne prend pas en charge le jeu de caractères utilisé, il peut (a) afficher le "mot codé" comme du texte ordinaire (c'est-à-dire, comme il apparaît dans l'en-tête), (b) faire "au mieux" pour afficher en utilisant les caractères tels qu'ils sont disponibles, ou (c) substituer un message approprié qui indique que le texte décodé n'a pas pu être affiché.

Si le jeu de caractères utilisé emploie des techniques de changement de code, l'affichage du texte codé commence implicitement en "mode ASCII". De plus, le lecteur de messagerie doit s'assurer que l'appareil de sortie est bien de nouveau en "mode ASCII" après l'affichage du "mot codé".

6.3 Traitement par le lecteur de messagerie des "mots codés" incorrectement formés

Il est possible qu'un "mot codé" légal suivant la syntaxe définie à la section 2, soit incorrectement formé selon les règles de codage utilisées. Par exemple :

- (1) Un "mot codé" qui contient des caractères qui ne sont pas légaux pour un codage particulier (par exemple, un "-" dans le codage "B", ou un ESPACE ou HTAB dans les codages "B" ou "Q") est incorrectement formé.
- (2) Tout "mot codé" qui code un nombre non entier de caractères ou d'octets est incorrectement formé.

Un lecteur de messagerie n'a pas besoin d'essayer d'afficher le texte associé à un "mot codé" incorrectement formé. Cependant, un lecteur de messagerie NE DOIT PAS empêcher l'affichage ou le traitement d'un message parce qu'un "mot codé" est incorrectement formé.

7. Conformité

Un programme de composition de messagerie qui revendique la conformité à la présente spécification DOIT s'assurer que tout chaîne de caractères ASCII imprimables autre que des espaces blanches au sein d'un "*texte" ou "*ctext" qui commence par "=?" et se termine par "?=" soit un "mot codé" valide. ("commence" signifie : au début du corps de champ, immédiatement à la suite d'une "espace blanche linéaire", ou immédiatement à la suite d'une "(" pour un "mot codé" au sein d'un "*ctext" ; "se termine" signifie : à la fin du corps de champ, précédant immédiatement une "espace blanche linéaire", ou précédant immédiatement une ")" pour un "mot codé" au sein de "*ctext".) De plus, tout "mot" au sein d'une "phrase" qui commence par "=?" et se termine par "?=" doit être un "mot codé" valide.

Un programme de lecture de messagerie qui revendique la conformité à la présente spécification doit être capable de distinguer les "mots codés" du "texte", "ctexte", ou "mots", conformément aux règles de la section 6, à chaque fois qu'ils apparaissent aux endroits appropriés dans les en-têtes de message. Il doit prendre en charge les deux codages "B" et "Q" pour tout jeu de caractères qu'il accepte. Le programme doit être capable d'afficher le texte non codé si le jeu de caractères est "US-ASCII". Pour les jeux de caractères ISO-8859-*, le programme de lecture de messagerie doit au moins être capable d'afficher les caractères qui sont aussi dans le jeu ASCII.

8. Exemples

Ci-après figurent des exemples d'en-têtes de message qui contiennent des "mots codés" :

```
From: =?US-ASCII?Q?Keith_Moore?= <moore@cs.utk.edu>
To: =?ISO-8859-1?Q?Keld_J=F8rn_Simonsen?= <keld@dikuug.dk>
Cc: =?ISO-8859-1?Q?Andr=E9?= Pirard <PIRARD@vm1.ulg.ac.be>
Subject: =?ISO-8859-1?B?SWYgeW91IGNhbiByZWFKIHRoaXMgeW8=?=
=?ISO-8859-2?B?dSB1bmRlc nN0YW5kIHRoZSBleGFtcGxlLg==?=
```

Note : Dans le premier "mot codé" du champ Subject ci-dessus, le dernier "=" à la fin du "texte codé" est nécessaire parce que chaque "mot codé" doit être autonome (le caractère "=" complète un groupe de 4 caractères base64 représentant 2 octets). Un octet supplémentaire aurait pu être codé dans le premier "mot codé" (de sorte que le mot codé contienne un multiple exact de 3 octets codés) sauf que le second "mot codé" utilise un "jeu de caractères" différent du premier.

```
From: =?ISO-8859-1?Q?Olle_J=E4rnefors?= <ojarnef@admin.kth.se>
To: ietf-822@dimacs.rutgers.edu, ojarnef@admin.kth.se
Subject: Time for ISO 10646?
```

```
To: Dave Crocker <dcrocker@mordor.stanford.edu>
Cc: ietf-822@dimacs.rutgers.edu, paf@comsol.se
From: =?ISO-8859-1?Q?Patrik_F=E4ltstr=F6m?= <paf@nada.kth.se>
Subject: Re: RFC-HDR care and feeding
```

```
From: Nathaniel Borenstein <nsb@thumper.bellcore.com>(=?iso-8859-8?b?7eXs+SDv4SDp7Oj08A==?)
To: Greg Vaudreuil <gvaudre@NRI.Reston.VA.US>, Ned Freed<ned@innosoft.com>, Keith Moore <moore@cs.utk.edu>
Subject: Essai d'un nouveau générateur d'en-tête
MIME-Version: 1.0
Content-type: texte/clair ; jeu de caractères=ISO-8859-1
```

Les exemples qui suivent illustrent comment le texte qui contient des "mots codés" apparaît dans un corps de champ structuré. Les règles sont légèrement différentes pour les champs définis comme "*texte" parce que "(" et ")" ne sont pas reconnus comme délimiteurs de "commentaire". [Section 5, alinéa (1)].

Dans chacun des exemples suivants, si la même séquence devait survenir dans un champ "*texte", la forme "affiché comme" NE SERAIT PAS traitée comme mot codé, mais serait identique à la "forme codée". Cela parce que chacun des mots codés des exemples suivants est adjacent à un caractère "(" ou ")".

forme codée	affichée comme
(=?ISO-8859-1?Q?a?=)	(a)
(=?ISO-8859-1?Q?a?= b)	(a b)

Au sein d'un "commentaire", les espaces DOIVENT apparaître entre un "mot codé" et le texte environnant. [Section 5, alinéa (2)]. Cependant, une espace n'est pas nécessaire entre la "(" initiale qui débute le "commentaire" et le "mot codé".

(=?ISO-8859-1?Q?a?= =?ISO-8859-1?Q?b?=)	(ab)
---	------

Les espaces entre les "mots codés" adjacents ne sont pas affichées.

(=?ISO-8859-1?Q?a?= =?ISO-8859-1?Q?b?=)	(ab)
---	------

Même plusieurs ESPACE entre les "mots codés" sont ignorés pour les besoins de l'affichage.

(=?ISO-8859-1?Q?a?= =?ISO-8859-1?Q?b?=)	(ab)
---	------

Toute quantité d'espace blanche linéaire entre des "mots codés", même si elle inclut un CRLF suivi par un ou plusieurs ESPACE, est ignorée pour les besoins de l'affichage.

(=?ISO-8859-1?Q?a_b?=)	(a b)
------------------------	-------

Pour causer l'affichage d'un ESPACE au sein d'une portion de texte codé, l'ESPACE DOIT être codé au titre du "mot codé".

(=?ISO-8859-1?Q?a?= =?ISO-8859-2?Q?_b?=)	(a b)
--	-------

Pour causer l'affichage d'un ESPACE entre deux chaîne de texte codé, l'ESPACE PEUT être codé au titre de l'un des "mots codés".

9. Références

- [RFC0822] D. Crocker, "Norme pour le [format des messages de texte](#) de l'ARPA-Internet", STD 11, août 1982. (*Obsolète, voir RFC5322*)
- [RFC2045] N. Freed et N. Borenstein, "[Extensions de messagerie Internet](#) multi-objets (MIME) Partie 1 : Format des corps de message Internet", novembre 1996. (*D. S., MàJ par 2184, 2231, 5335.*)
- [RFC2046] N. Freed et N. Borenstein, "[Extensions de messagerie Internet](#) multi-objets (MIME) Partie 2 : Types de support", novembre 1996. (*D. S., MàJ par 2646, 3798, 5147, 6657, 8098*)
- [RFC2048] N. Freed, J. Klensin et J. Postel, "Extensions multi-objets de la messagerie Internet (MIME) Partie 4 : Procédures d'enregistrement", BCP 13, novembre 1996. (*Rendue obsolète par les RFC 4288-4289*)
- [RFC2049] N. Freed, N. Borenstein, "[Extensions multi-objets de la messagerie](#) Internet (MIME) Partie cinq : critères de conformité et exemples", novembre 1996. (*Remplace RFC1521, RFC1522, RFC1590*) (*D.S.*)

10. Considérations pour la sécurité

Les questions de sécurité ne sont pas abordées dans le présent mémoire.

11. Remerciements

L'auteur tient à remercier Nathaniel Borenstein, Issac Chan, Lutz Donnerhacke, Paul Eggert, Ned Freed, Andreas M. Kirchwitz, Olle Jarnefors, Mike Rosin, Yutaka Sato, Bart Schaefer et Kazuhiko Yamamoto de leurs utiles conseils, de leurs commentaires pertinents, et des questions éclairantes en réponse à des versions précédentes de cette spécification.

12. Adresse de l'auteur

Keith Moore
University of Tennessee
107 Ayres Hall
Knoxville TN 37996-1301
USA
mél : moore@cs.utk.edu

Appendice - Changements depuis la RFC 1522 (sans ordre particulier)

- + déclare explicitement que la version MIME n'est pas obligée d'utiliser les "mots codés".
- + ajoute dans une note explicite que les SPACE et TAB ne sont pas admis au sein des "mots codés", qui explique qu'un "mot codé" doit ressembler à un "atome", à une valeur parser.values de la RFC822, pour être précis.
- + ajout des exemples de Olle Jarnefors (merci !) qui illustrent comment sont affichés les mots codés avec des espaces blanches linéaires adjacentes.
- + fait une liste explicite des termes définis dans la RFC822 et référencés dans ce mémoire.
- + corrige les erreurs de transcription qui ont causé la disparition d'une ou deux lignes et de quelques caractères dans le texte final, du fait de bizarreries de nroff.
- + précise que les mots codés sont admis dans les champs "*texte" des en-têtes de la RFC 822 et dans les en-têtes de partie corps de MIME, mais PAS comme valeurs de paramètres.
- + précise les exigences de retour à l'ASCII au sein de la portion codée d'un "mot codé", pour tout jeu de caractères qui utilise des séquences de changement de code.
- + ajout d'une note sur la délimitation des "mots codés" par "(" et ")" au sein d'un commentaire, mais pas dans un *texte (comme c'est bizarre !).
- + correction de l'exemple de André Pirard pour se débarrasser du "_" en queue après le =E9. (plus nécessaire post-1342).
- + précision : un "mot codé" peut apparaître immédiatement après le "(" initial ou immédiatement avant le ")" final qui délimite un commentaire, pas simplement adjacent à "(" et ")" *au sein de* *ctext.
- + ajout d'une note pour expliquer qu'un "mot codé" "B" va toujours avoir un multiple de 4 caractères dans la portion de "texte codé".
- + ajout d'une note sur le "=" dans les exemples
- + note que le traitement des "mots codés" survient *après* l'analyse, et certaines des implications qui en découlent.
- + déclare explicitement qu'on ne peut traduire du texte de la RFC 1522 en du texte soit de la RFC 822, soit de ce qu'on appelle les "en-têtes de 8 bits".
- + déclare explicitement que les "mots codés" ne sont pas valides au sein d'une "chaîne entre guillemets".