

Groupe de travail Réseau
Request for Comments : 2376
 Catégorie : Information
 Traduction Claude Brière de L'Isle

E. Whitehead, UC Irvine
 M. Murata, Fuji Xerox Info. Systems
 juillet 1998

Types de supports XML

Statut de ce mémoire

Le présent mémoire apporte des informations pour la communauté de l'Internet. Le présent mémoire ne spécifie aucune sorte de norme de l'Internet. La distribution du présent mémoire n'est soumise à aucune restriction.

Notice de copyright

Copyright (C) The Internet Society (1998). Tous droits réservés.

Résumé

Le présent document propose deux nouveaux sous-types de supports, `text/xml` et `application/xml`, à utiliser dans les échanges d'entités réseau qui sont conformes au langage de balisage extensible (XML, *Extensible Markup Language*). Les entités XML qui sont couramment échangées via le protocole de transfert Hypertexte sur la Toile mondiale, font partie intégrante du protocole WebDAV pour la collecte ordonnée des auteurs et des versions distribuée sur la Toile, et sont supposées avoir une utilité dans de nombreux domaines.

Table des matières

1. Introduction.....	1
2. Conventions de notation.....	2
3. Types de supports XML.....	2
3.1 Enregistrement de <code>text/xml</code>	2
3.2 Enregistrement de <code>application/xml</code>	3
4. Considérations sur la sécurité.....	5
5. Marque d'ordre des octets (BOM) et conversions de/en UTF-16.....	5
6. Exemples.....	5
6.1 <code>text/xml</code> avec jeu de caractères UTF-8.....	6
6.2 <code>text/xml</code> avec jeu de caractères UTF-16.....	6
6.3 <code>text/xml</code> avec jeu de caractères ISO-2022-KR.....	6
6.4 <code>text/xml</code> avec jeu de caractères omis.....	6
6.5 <code>application/xml</code> avec jeu de caractères UTF-16.....	6
6.6 <code>application/xml</code> avec jeu de caractères ISO-2022-KR.....	7
6.7 <code>application/xml</code> avec jeu de caractères omis et entité XML UTF-16.....	7
6.8 <code>application/xml</code> avec jeu de caractères omis et entité UTF-8.....	7
6.9 <code>application/xml</code> avec jeu de caractères omis et déclaration de codage interne.....	7
7. Références.....	8
8. Remerciements.....	8
9. Adresse des auteurs.....	8
10. Déclaration complète de droits de reproduction.....	9

1. Introduction

Le Consortium de la Toile mondiale (W3C, *World Wide Web Consortium*) a produit une recommandation [XML] qui définit la version 1 du langage de balisage extensible (XML, *Extensible Markup Language*). Pour permettre l'échange d'entités réseau XML, le présent document propose deux nouveaux types de supports, `text/xml` et `application/xml`.

Les entités XML sont actuellement échangées sur la Toile mondiale, et XML est aussi utilisé pour la distribution des valeurs et paramètres de propriétés par le protocole WebDAV pour la collecte à distance d'informations sur les auteurs de la Toile. Donc, il y a un besoin d'un type de support pour étiqueter de façon appropriée l'échange des entités réseau XML. (Noter que, comme cela arrive parfois entre deux communautés, MIME et XML ont tous deux défini le terme entité, avec des significations différentes.)

Bien que XML soit un sous-ensemble du langage de balisage standard généralisé (SGML) [ISO-8897], et qu'il ait alloué les types de supports `text/sgml` et `application/sgml`, il y a plusieurs raisons pour lesquelles l'utilisation de `text/sgml` ou

application/sgml pour étiqueter XML est inappropriée. D'abord, il existe de nombreuses applications qui peuvent traiter XML, mais qui ne peuvent pas traiter SGML, à cause du plus grand ensemble de caractéristiques de SGML. Ensuite, les applications SGML ne peuvent pas toujours traiter les entités XML, parce que XML utilise des caractéristiques de corrections techniques récentes à SGML. Enfin, la définition de text/sgml et application/sgml [RFC1874] inclut des paramètres pour le format de transformation de combinaisons de bits SGML (SGML-bctf), et l'attribut d'amorçage SGML (SGML-boot). Comme XML n'utilise pas ces paramètres, il serait ambigu que de tels paramètres soient donnés pour une entité XML. Pour ces raisons, la meilleure approche pour étiqueter les entités réseau XML est de fournir de nouveaux types de supports pour XML.

Comme XML fait partie intégrante du protocole de collecte d'informations d'auteur réparties WebDAV, et comme les recommandations du Consortium de la Toile mondiale ont conventionnellement alloué des types de supports d'arborescence de l'IETF, et comme des types de supports similaires (HTML, SGML) ont été alloués à des types de supports d'arborescence de l'IETF, les types de supports XML appartiennent aussi à l'arborescence de types de supports de l'IETF.

2. Conventions de notation

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" en majuscules dans ce document sont à interpréter comme décrit dans le BCP 14, [RFC2119].

3. Types de supports XML

Le présent document introduit deux nouveaux types de supports pour les entités XML, text/xml et application/xml. Les informations d'enregistrement pour ces types de supports sont décrites dans les paragraphes qui suivent.

Chaque entité XML est utilisable sans modification avec le type de support application/xml. Mais cela n'exploite pas le fait que XML peut être traité dans de nombreux cas comme du texte pur. Les agents d'utilisateur MIME (et les agents d'utilisateur de la Toile) qui n'ont pas de prise en charge explicite pour application/xml vont le traiter comme un application/flux d'octets, par exemple, en offrant de le sauvegarder dans un fichier.

Pour indiquer qu'une entité XML devrait être traitée comme du texte pur par défaut, on utilise le type de support text/xml. Cela restreint le codage utilisé dans l'entité XML à ceux qui sont compatibles avec les exigences pour les types de supports text comme décrit dans les [RFC2045] et [RFC2046], par exemple, UTF-8, mais pas UTF-16 (sauf pour HTTP).

XML fournit un cadre général pour définir des séquences de données structurées. Dans certains cas, il peut être souhaitable de définir de nouveaux types de supports qui utilisent XML mais définissent une application spécifique de XML, peut-être à cause de considérations de sécurité spécifiques du domaine ou d'informations au moment du démarrage. Le présent document n'interdit pas de futurs types de supports dédiés à de telles applications XML. Cependant, il est recommandé aux développeurs de tels types de supports d'utiliser le présent document comme base. En particulier, le paramètre Jeux de caractères devrait être utilisé de la même manière.

Dans la spécification XML, les entités XML peuvent être classées en quatre types. Dans la terminologie XML, ils sont appelés des "entités de document", des "sous ensembles de DTD externes", des "entités analysées externes", et des "entités de paramètre externes". Les types de supports text/xml et application/xml peuvent être utilisés pour ces quatre types.

3.1 Enregistrement de text/xml

Nom de type de support MIME : text

Nom de sous type de support MIME : xml

Paramètres obligatoires : aucun

Paramètres facultatifs : charset

Bien que mentionné comme paramètre facultatif, l'utilisation du paramètre Jeux de caractères est **FORTEMENT RECOMMANDÉE**, car ces informations peuvent être utilisées par les processeurs XML pour déterminer d'autorité le codage de caractères de l'entité XML. Le paramètre Jeux de caractères peut aussi être utilisé pour fournir des opérations spécifiques du protocole, comme la négociation du contenu fondée sur le jeu de caractères dans HTTP. "UTF-8" [RFC2279] est la valeur recommandée, représentant le jeu de caractères UTF-8. UTF-8 est pris en charge par tous les

processeurs XML [XML] conformes.

Si l'entité XML est transmise via HTTP, qui utilise un mécanisme de style MIME qui est exempt des restrictions sur le type de niveau supérieur text (voir le paragraphe 19.4.1 de HTTP 1.1 [RFC2068]), "UTF-16" (Appendice C.3 de [UNICODE] et l'Amendement 1 de [ISO-10646]) sont aussi recommandés. UTF-16 est pris en charge par tous les processeurs XML [XML] conformes. Comme le traitement des CR, LF et NUL pour les types text dans la plupart des applications MIME causerait des transformations non désirées d'octets individuels dans les caractères multi octets de UTF-16, les passerelle de HTTP à ces applications MIME DOIVENT transformer l'entité XML de text/xml; charset="utf-16" en application/xml; charset="utf-16".

Conformément à la [RFC2046], si une entité text/xml est reçue avec le paramètre Jeux de caractères omis, les processeurs MIME et XML DOIVENT utiliser la valeur de charset par défaut de "us-ascii". Dans les cas où l'entité XML est transmise via HTTP, la valeur de charset par défaut est encore "us-ascii".

Comme le paramètre Jeux de caractères est d'autorité, le jeu de caractères n'est pas toujours déclaré dans une déclaration de codage XML. Donc, une attention particulière est nécessaire quand le receveur supprime l'en-tête MIME et fournit une mémorisation persistante de l'entité XML reçue (par exemple, dans un système de fichiers). Sauf si le jeu de caractères est UTF-8 ou UTF-16, le receveur DEVRAIT aussi mémoriser de façon persistante les informations sur le jeu de caractères, peut-être en incorporant une déclaration de codage XML correcte dans l'entité XML.

Considérations de codage : ce type de support PEUT être codé comme approprié pour le jeu de caractères et les capacités du transport MIME sous-jacent. Pour les transports sur 7 bits, les données en UTF-8 et UTF-16 sont codées en quoted-printable ou base64. Pour un transport sur 8 bits pur (par exemple, ESMTP, 8BITMIME, ou NNTP) UTF-8 n'est pas codé, mais UTF-16 est codé en base64. Pour les transports en binaire pur (par exemple, HTTP) aucun codage de transfert de contenu n'est nécessaire.

Considérations de sécurité : Voir la Section 4.

Considérations d'interopérabilité : il a été prouvé que XML est interopérable à travers les clients et serveurs WebDAV, et pour importer et exporter à partir de plusieurs outils XML de collecte d'auteurs.

Spécification publiée : voir [XML]

Applications qui utilisent ce type de support : XML est neutre à l'égard de l'appareil, de la plateforme, et du fabricant et est pris en charge par une large gamme d'agents d'utilisateur de la Toile, de clients et serveurs WebDAV, ainsi que d'outils XML de collecte d'auteurs.

Informations supplémentaires :

Numéro magique : aucun. Bien qu'aucune séquence d'octets ne puisse être réputée être toujours présente, les entités XML dans les jeux de caractères compatibles ASCII (incluant UTF-8) commencent souvent par l'hexadécimal 3C 3F 78 6D 6C ("`<?xml`"). Pour plus d'informations, voir l'Appendice F de [XML].

Extensions de fichier : .xml, .dtd

Code de type de fichier Macintosh : "TEXT"

Adresse de messagerie des personnes à contacter pour plus d'informations :

Dan Connolly <connolly@w3.org>

Murata Makoto (Family Given) <murata@fxis.fujixerox.co.jp>

Utilisation prévue : COMMUNE

Auteur/Contrôleur des changements : la spécification XML est le produit du travail du groupe de travail XML du World Wide Web Consortium, et a été édité par :

Tim Bray <tbray@textuality.com>

Jean Paoli <jeanpa@microsoft.com>

C. M. Sperberg-McQueen <cmsmq@uic.edu>

Le W3C, et le groupe de travail W3C XML a le contrôle des changement sur la spécification XML.

3.2 Enregistrement de application/xml

Nom de type de support MIME : application

Nom de sous type de support MIME : xml

Paramètres obligatoires : aucun

Paramètres facultatifs : charset

Bien que mentionné comme paramètre facultatif, l'utilisation du paramètre Jeux de caractères est **FORTEMENT RECOMMANDÉE**, car ces informations peuvent être utilisées par les processeurs XML pour déterminer d'autorité le codage de caractères de l'entité XML. Le paramètre Jeux de caractères peut aussi être utilisé pour fournir des opérations spécifiques du protocole, comme la négociation du contenu fondée sur le jeu de caractères dans HTTP. "UTF-8" [RFC2279] est la valeur recommandée, représentant le jeu de caractères UTF-8. UTF-8 est pris en charge par tous les processeurs XML [XML] conformes.

"UTF-8" [RFC2279] et "UTF-16" (Appendice C.3 de [UNICODE] et Amendement 1 de [ISO-10646]) sont les valeurs recommandées, représentant respectivement les jeux de caractères UTF-8 et UTF-16. Ces jeux de caractères sont préférés car ils sont pris en charge par tous les processeurs [XML] conformes.

Si une entité application/xml où le paramètre Jeux de caractères est omis est reçue, aucune information n'est fournie sur le jeu de caractères par l'en-tête MIME Content-Type. Les processeurs XML conformes **DOIVENT** respecter les exigences du paragraphe 4.3.3 de [XML] qui traitent directement de ce problème. Cependant, les processeurs MIME qui ne sont pas des processeurs XML ne devraient pas supposer un jeu de caractères par défaut si le paramètre Jeux de caractères est omis dans une entité application/xml.

Comme le paramètre Jeux de caractères est d'autorité, le jeu de caractères n'est pas toujours déclaré dans une déclaration de codage XML. Donc, une attention particulière est nécessaire quand le receveur supprime l'en-tête MIME et fournit une mémorisation persistante de l'entité XML reçue (par exemple, dans un système de fichiers). Sauf si le jeu de caractères est UTF-8 ou UTF-16, le receveur **DEVRAIT** aussi mémoriser de façon persistante les informations sur le jeu de caractères, peut-être en incorporant une déclaration de codage XML correcte dans l'entité XML.

Considérations de codage : ce type de support **PEUT** être codé comme approprié pour le jeu de caractères et les capacités du transport MIME sous-jacent. Pour les transports sur 7 bits, les données en UTF-8 et UTF-16 sont codées en quoted-printable ou base64. Pour un transport sur 8 bits pur (par exemple, ESMTP, 8BITMIME, ou NNTP) UTF-8 n'est pas codé, mais UTF-16 est codé en base64. Pour les transports en binaire pur (par exemple, HTTP) aucun codage de transfert de contenu n'est nécessaire.

Considérations de sécurité : Voir la Section 4.

Considérations d'interopérabilité : il a été prouvé que XML est interopérable à travers les clients et serveurs WebDAV, et pour importer et exporter à partir de plusieurs outils XML de collecte d'auteurs.

Spécification publiée : voir [XML]

Applications qui utilisent ce type de support : XML est neutre à l'égard de l'appareil, de la plateforme, et du fabricant et est pris en charge par une large gamme d'agents d'utilisateur de la Toile, de clients et serveurs WebDAV, ainsi que d'outils XML de collecte d'auteurs.

Informations supplémentaires :

Numéro magique : aucun. Bien qu'aucune séquence d'octets ne puisse être réputée être toujours présente, les entités XML dans les jeux de caractères compatibles ASCII (incluant UTF-8) commencent souvent par l'hexadécimal 3C 3F 78 6D 6C ("<?xml"). Pour plus d'informations, voir l'Appendice F de [XML].

Extensions de fichier : .xml, .dtd

Code de type de fichier Macintosh : "TEXT"

Adresse de messagerie des personnes à contacter pour plus d'informations :

Dan Connolly <connolly@w3.org>

Murata Makoto (Family Given) <murata@fxis.fujixerox.co.jp>

Utilisation prévue : COMMUNE

Auteur/Contrôleur des changements : la spécification XML est le produit du travail du groupe de travail XML du World

Wide Web Consortium, et a été édité par :
Tim Bray <tbray@textuality.com>
Jean Paoli <jeanpa@microsoft.com>
C. M. Sperberg-McQueen <cmsmcq@uiuc.edu>

Le W3C, et le groupe de travail W3C XML a le contrôle des changements sur la spécification XML.

4. Considérations sur la sécurité

XML, comme sous-ensemble de SGML, a les mêmes considérations de sécurité que spécifié dans la [RFC1874].

Pour paraphraser la Section 3 de la [RFC1874], les entités XML contiennent des informations à analyser et traiter par le système XML du receveur. Ces entités peuvent contenir, et de tels systèmes peuvent permettre, des commandes explicites de niveau système, à exécuter tout en traitant les données. Dans la mesure où un système XML va exécuter des chaînes de commandes arbitraires, les receveurs des entités XML peuvent être en danger. En général, il est possible de spécifier des commandes qui effectuent des opérations de fichier non autorisées ou font des changements à l'affichage de l'environnement du processeur qui affectent les opérations suivantes.

L'utilisation de XML est supposée être variée, et largement répandue. XML est examiné par de larges gammes de communautés pour être utilisé comme syntaxe commune pour des métadonnées spécifiques de la communauté. Par exemple, le groupe Dublin Core utilise XML pour des métadonnées de document, et un nouveau travail a commencé qui examine l'utilisation de XML pour des informations médicales. D'autres groupes voient XML comme un mécanisme pour disposer les paramètres des appels de procédure distante. D'autres utilisations de XML vont sans doute se faire jour.

Les considérations de sécurité vont varier selon le domaine d'utilisation. Par exemple, les enregistrements médicaux XML auront des considérations de confidentialité et de sécurité plus strictes que les métadonnées de bibliothèque XML. De même, l'utilisation de XML comme syntaxe de disposition des paramètres exige un examen de sécurité au cas par cas.

XML peut aussi avoir certains des mêmes soucis de sécurité que le texte pur. Comme le texte pur, XML peut contenir des séquences d'échappement qui, à l'affichage, ont le potentiel de changer l'environnement de traitement de l'affichage d'une façon qui contrarie les opérations suivantes. Les effets possibles incluent, mais ne se limitent pas, de verrouiller le clavier, de changer les paramètres d'affichage de façon que la suite du texte à afficher soit illisible, ou même de changer les paramètres d'affichage pour obscurcir ou fausser délibérément le matériel affiché suivant de telle sorte que sa signification soit perdue ou altérée. Les processeurs d'affichage devraient soit filtrer un tel matériel du texte affiché, soit s'assurer de rétablir tous les réglages importants après l'achèvement d'une certaine opération d'affichage.

Certains terminaux ont des clés dont le résultat, à la frappe, peut être changé en envoyant au processeur d'affichage une séquence de caractères. Si cela est possible, l'affichage d'un objet text contenant de telles séquences de caractères pourrait reprogrammer les clés pour effectuer une action illicite ou dangereuse quand la clé est ultérieurement pressée par l'utilisateur. Dans certains cas, non seulement des clés peuvent être programmées, mais elles peuvent être déclenchées à distance, rendant possible qu'une opération d'affichage de texte effectuée directement une action non voulue. À ce titre, la capacité de programmer les clés devrait être bloquée soit par filtrage, soit en désactivant entièrement la capacité de programmer les clés.

Noter qu'il est aussi possible de construire des documents XML qui font usage de ce que XML appelle des "références d'entité" (utilisant la signification XML du terme "entité", qui diffère de la définition MIME de ce terme) pour construire des expansions répétées de texte. Les expansions récurrentes sont interdites [XML] et il est exigé des processeurs XML qu'ils les détectent. Cependant, même des expansions non récurrentes peuvent causer des problèmes avec les ressources de calcul finies des ordinateurs, si elles sont effectuées de nombreuses fois.

5. Marque d'ordre des octets (BOM) et conversions de/en UTF-16

La recommandation XML, au paragraphe 4.3.3, spécifie que les entités XML UTF-16 doivent commencer par une marque d'ordre des octets (BOM, *byte order mark*) qui est le caractère ZERO WIDTH NO-BREAK SPACE (*espace largeur zéro sans coupure*), séquence hexadécimale 0xFEFF (ou 0xFFFE, selon l'ordre des octets). La recommandation XML déclare de plus que la BOM est une signature de codage, et ne fait partie ni du balisage ni des données de caractères du document XML.

Du fait de la BOM, les applications qui convertissent XML à partir d'un codage UTF-16 en un autre codage DEVRAIENT

supprimer la BOM avant conversion. De même, lors de la conversion d'un autre codage en UTF-16, la BOM DEVRAIT être ajoutée après l'achèvement de la conversion.

6. Exemples

Les exemples ci-dessous donnent la valeur de l'en-tête MIME Content-type et la déclaration XML (qui inclut la déclaration de codage) à l'intérieur de l'entité XML. Pour les exemples de UTF-16, le caractère de marque d'ordre des octets est noté par "{BOM}", et la déclaration XML est supposée venir au début de l'entité XML, suivant immédiatement la BOM. Noter que d'autres en-têtes MIME peuvent être présents, et que l'entité XML peut contenir d'autres données en plus de la déclaration XML ; les exemples se concentrent sur l'en-tête Content-type et la déclaration de codage pour la clarté du texte.

6. text/xml avec jeu de caractères UTF-8

Content-type: text/xml; charset="utf-8"

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

C'est la valeur de jeu de caractères recommandée pour utiliser avec text/xml. Comme le paramètre Jeux de caractères est fourni, les processeurs MIME et XML doivent traiter l'entité enclose comme codée en UTF-8.

Si il est envoyé en utilisant un transport sur 7 bits (par exemple SMTP) l'entité XML doit utiliser un codage de transfert de contenu de quoted-printable ou base64. Pour un transport sur 8 bits pur (par exemple, ESMTP, 8BITMIME, ou NNTP) ou un transport binaire pur (par exemple, HTTP) aucun codage de transfert de contenu n'est nécessaire.

6.2 text/xml avec jeu de caractères UTF-16

Content-type: text/xml; charset="utf-16"

```
{BOM}<?xml version='1.0' encoding='utf-16'?>
```

Ceci n'est possible que quand l'entité XML est transmise via HTTP, qui utilise un mécanisme de type MIME et est un protocole binaire pur, donc n'effectue pas les transformations CR et LF et permet les octets NUL. Cela diffère du traitement normal de type text MIME (voir le paragraphe 19.4.1 de HTTP 1.1 [RFC2068] pour les détails).

Comme HTTP est binaire pur, aucun codage de transfert de contenu n'est nécessaire.

6.3 text/xml avec jeu de caractères ISO-2022-KR

Content-type: text/xml; charset="iso-2022-kr"

```
<?xml version="1.0" encoding='iso-2022-kr'?>
```

Cet exemple montre text/xml avec un jeu de caractère coréen (par exemple, Hangul) codé suivant la spécification de la [RFC1557]. Comme le paramètre Jeux de caractères est fourni, les processeurs MIME et XML doivent traiter l'entité enclose comme codée selon la [RFC1557].

Comme ISO-2022-KR a été défini pour être seulement utilisé avec des données de 7 bits, aucun codage de transfert de contenu n'est nécessaire avec tout transport.

6.4 text/xml avec jeu de caractères omis

Content-type: text/xml

```
{BOM}<?xml version="1.0" encoding="utf-16"?>
```

Cet exemple montre text/xml sans le paramètre Jeux de caractères. Dans ce cas, les processeurs MIME et XML doivent supposer que le jeu de caractères est "us-ascii", la valeur de jeu de caractères par défaut pour les types de support text

spécifiée dans la [RFC2046]. La valeur par défaut de "us-ascii" tient même si l'entité text/xml est transportée avec HTTP.

Omettre le paramètre Jeux de caractères N'EST PAS RECOMMANDÉ pour text/xml. Par exemple, même si le contenu de l'entité XML est UTF-16 ou UTF-8, ou si l'entité XML a une déclaration explicite de codage, les processeurs XML et MIME doivent supposer que le jeu de caractères est "us-ascii".

6.5 application/xml avec jeu de caractères UTF-16

Content-type: application/xml; charset="utf-16"

```
{BOM}<?xml version="1.0"?>
```

C'est une valeur de jeu de caractères recommandée pour utiliser avec application/xml. Comme le paramètre Jeux de caractères est fourni, les processeurs MIME et XML doivent traiter l'entité enclose comme codée en UTF-16.

Si il est envoyé en utilisant un transport sur 7 bits (par exemple, SMTP) ou un transport sur 8 bits pur (par exemple, ESMTP, 8BITMIME, ou NNTP) l'entité XML doit être codée en quoted-printable ou base64. Pour un transport binaire pur (par exemple, HTTP) aucun codage de transfert de contenu n'est nécessaire.

6.6 application/xml avec jeu de caractères ISO-2022-KR

Content-type: application/xml; charset="iso-2022-kr"

```
<?xml version="1.0" encoding="iso-2022-kr"?>
```

L'entité XML montre application/xml avec un jeu de caractères coréen (par exemple, Hangul) codé suivant la spécification de la [RFC1557]. Comme le paramètre Jeux de caractères est fourni, les processeurs MIME et XML doivent traiter l'entité enclose comme codée selon la [RFC1557], indépendamment de si l'entité XML a une déclaration de codage interne (cet exemple montre une telle déclaration, qui est en accord avec le paramètre Jeux de caractères).

Comme ISO-2022-KR a été défini pour utiliser seulement des données de 7 bits, aucun codage de transfert de contenu n'est nécessaire avec tout transport.

6.7 application/xml avec jeu de caractères omis et entité XML UTF-16

Content-type: application/xml

```
{BOM}<?xml version="1.0"?>
```

Pour cet exemple, l'entité XML commence par une BOM. Comme le jeu de caractères a été omis, un processeur XML conforme suit les exigences de [XML], paragraphe 4.3.3. Précisément, le processeur XML lit la BOM, et donc sait de façon déterministe que le codage du jeu de caractères est UTF-16.

Un processeur MIME sans capacité XML ne devrait faire aucune hypothèse sur le jeu de caractères de l'entité XML.

6.8 application/xml avec jeu de caractères omis et entité UTF-8

Content-type: application/xml

```
<?xml version="1.0"?>
```

Dans cet exemple, le paramètre Jeux de caractères a été omis, et il n'y a pas de BOM. Comme il n'y a pas de BOM, le processeur XML suit les exigences du paragraphe 4.3.3, et applique facultativement le mécanisme décrit à l'Appendice F (qui est non normatif) de [XML] pour déterminer le codage du jeu de caractères de UTF-8. Cet exemple ne contient pas de déclaration de codage, mais comme le codage est UTF-8, ceci est quand même un exemple conforme.

Un processeur MIME sans capacité XML ne devrait faire aucune hypothèse sur le jeu de caractères de cet exemple.

6.9 application/xml avec jeu de caractères omis et déclaration de codage interne

Content-type: application/xml

```
<?xml version='1.0' encoding="ISO-10646-UCS-4"?>
```

Dans cet exemple, le paramètre Jeux de caractères a été omis, et il n'y a pas de BOM. Cependant, l'entité XML a une déclaration de codage dans l'entité XML qui spécifie le jeu de caractères de l'entité. Suivant les exigences du paragraphe 4.3.3, et en appliquant facultativement le mécanisme décrit à l'Appendice F (non normatif) de [XML], le processeur XML détermine le codage du jeu de caractères de l'entité XML (dans cet exemple, UCS-4).

Un processeur MIME sans capacité XML ne devrait faire aucune hypothèse sur le jeu de caractères de l'entité XML.

7. Références

- [ISO-10646] Norme ISO/CEI 10646, "Technologie de l'information – Jeu de caractères universel codé sur plusieurs octets (UCS) - Partie 1 : Architecture et plan multilingue de base", mai 1993.
- [ISO-8897] Norme internationale ISO 8897, "Traitement de l'information – Systèmes de textes et de bureau -- Langage de balisage standard généralisé (SGML)". Première édition -- 1986.
- [XML] Bray, T., Paoli, J. and C. Sperberg-McQueen, "Extensible Markup Language (XML) 1.0", février 1998.
- [RFC1557] U. Choi, K. Chon et H. Park, "Codage des caractères coréens pour les messages Internet", décembre 1993. (*Info.*)
- [RFC1874] E. Levinson, "Types de supports SGML", décembre 1995. (*Expérimentale*)
- [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*)
- [RFC2068] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, T. Berners-Lee, "Protocole de transfert Hypertext -- HTTP/1.1", janvier 1997. (*Obsolète, voir RFC2616*) (*P.S.*)
- [RFC2119] S. Bradner, "[Mots clés à utiliser](#) dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997. (*MàJ par RFC8174*)
- [RFC2279] F. Yergeau, "UTF-8, un format de transformation de la norme ISO 10646", janvier 1998. (*Obsolète, voir RFC3629*) (*D.S.*)
- [UNICODE] The Unicode Consortium, "The Unicode Standard -- Version 2.0", Addison-Wesley, 1996.

8. Remerciements

Chris Newman et Yaron Y. Goland ont tous deux contribué au contenu de la section des considérations sur la sécurité du présent document. En particulier, du texte de la section des considérations sur la sécurité est copié mot à mot du travail en cours, draft-newman-mime-textpara-00, avec la permission de l'auteur. Chris Newman a de plus contribué au contenu des paragraphes sur les considérations de codage. Dan Connolly a contribué à la discussion du contenu de quand utiliser text/xml. Des discussions avec Ned Freed et Dan Connolly ont aidé à préciser la compréhension des auteurs du type de support text ; des retours de Larry Masinter ont aussi été très utiles pour comprendre les questions d'enregistrement de type de support.

Les membres du groupe de travail W3C XML et du groupe d'intérêt spécial XML ont fait des contributions significatives au présent document, et les auteurs tiennent à remercier particulièrement James Clark, Martin Duerst, Rick Jelliffe, Gavin Nicol de leurs nombreux commentaires utiles.

9. Adresse des auteurs

E. James Whitehead, Jr.
Dept. of Information et Computer Science
University of California, Irvine
Irvine, CA 92697-3425
USA
mél : ejw@ics.uci.edu

Murata Makoto (Family Given)
Fuji Xerox Information Systems,
KSP 9A7, 2-1, Sakado 3-chome, Takatsu-ku,
Kawasaki-shi, Kanagawa-ken,
213 Japan
mél : murata@fxis.fujixerox.co.jp

10. Déclaration complète de droits de reproduction

Copyright (C) The Internet Society (2003). Tous droits réservés.

Le présent document et ses traductions peuvent être copiés et fournis aux tiers, et les travaux dérivés qui les commentent ou les expliquent ou aident à leur mise en œuvre peuvent être préparés, copiés, publiés et distribués, en tout ou partie, sans restriction d'aucune sorte, pourvu que la déclaration de droits de reproduction ci-dessus et le présent paragraphe soient inclus dans toutes telles copies et travaux dérivés. Cependant, le présent document lui-même ne peut être modifié d'aucune façon, en particulier en retirant la notice de droits de reproduction ou les références à la Internet Society ou aux autres organisations Internet, excepté autant qu'il est nécessaire pour le besoin du développement des normes Internet, auquel cas les procédures de droits de reproduction définies dans les procédures des normes Internet doivent être suivies, ou pour les besoins de la traduction dans d'autres langues que l'anglais.

Les permissions limitées accordées ci-dessus sont perpétuelles et ne seront pas révoquées par la Internet Society ou ses successeurs ou ayant droits.

Le présent document et les informations contenues sont fournis sur une base "EN L'ÉTAT" et le contributeur, l'organisation qu'il ou elle représente ou qui le/la finance (s'il en est), la INTERNET SOCIETY et la INTERNET ENGINEERING TASK FORCE déclinent toutes garanties, exprimées ou implicites, y compris mais non limitées à toute garantie que l'utilisation des informations encloses ne viole aucun droit ou aucune garantie implicite de commercialisation ou d'aptitude à un objet particulier.

Remerciement

Le financement de la fonction d'édition des RFC est actuellement fourni par l'Internet Society.