

Groupe de travail Réseau
Request for Comments : 1630
 Catégorie : Information
 Traduction Claude Brière de L'Isle

T. Berners-Lee
 CERN
 juin 1994

Identifiants de ressource universels dans la Toile mondiale ; syntaxe unificatrice pour l'expression des noms et adresses des objets du réseau utilisés sur la Toile mondiale

Statut de ce mémoire

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

Note de l'IESG :

Noter que le travail contenu dans ce mémoire ne décrit pas une norme de l'Internet. Une norme Internet pour les identifiants de ressource généraux est en cours de développement au sein de l'IETF.

Note du traducteur

La numérotation des titres des sections et paragraphes ainsi que la table des matières ont été ajoutées afin de faciliter la lecture.

Table des Matières

1.	Introduction.....
2.	Besoin d'une syntaxe universelle.....
3.	Critères et choix de la conception.....
3.1	Critères de conception.....
3.2	Choix pour une syntaxe universelle.....
4.	Recommandations.....
4.1	Syntaxe d'URI.....
5.	Exemples.....
5.1	HTTP.....
5.2	FTP.....
5.3	Gopher.....
5.4	Mailto.....
5.5	News.....
5.6	URN.....
5.7	WAIS.....
5.8	file.....
5.9	Message-Id.....
6.	Schémas pour des études complémentaires.....
6.1	X500.....
6.2	WHOIS.....
6.3	NNTP.....
6.4	Prospero.....
7.	Enregistrement des schémas de dénomination.....
8.	BNF de la syntaxe générique d'URI.....
9.	BNF pour les schémas d'URL spécifiques.....
10.	Références.....
11.	Considérations pour la sécurité.....
12.	Adresse de l'auteur.....

1. Introduction

Le présent document définit la syntaxe utilisée par l'initiative pour la Toile Mondiale pour coder les noms et les adresses des objets de l'Internet. La Toile est considérée comme incluant des objets auxquels on accède en utilisant un nombre extensible de protocoles, existants, inventés pour la Toile elle-même, ou à inventer à l'avenir. Les instructions d'accès pour un objet individuel sous un certain protocole sont codés sous forme de chaînes d'adresse. D'autres protocoles permettent l'utilisation de noms d'objet de formes diverses. Afin d'abstraire les idées d'un objet générique, la Toile a besoin d'un concept d'ensemble universel d'objets, et d'un ensemble universel de noms ou d'adresses des objets.

Un identifiant de ressource universel (URI, *Universal Resource Identifier*) est un membre de cet ensemble universel de noms dans des espaces de noms et d'adresses enregistrés se référant à des protocoles ou espaces de noms enregistrés. Un localisateur de ressource uniforme (URL, *Uniform Resource Locator*) défini ailleurs, est une forme d'URI qui exprime une adresse qui se transpose en un algorithme d'accès qui utilise des protocoles réseau. Les schémas d'URI existants qui correspondent au concept (encore mouvant) d'URL de l'IETF sont énumérés ci-dessous. Le débat sur le nom de ressource uniforme (URN, *Uniform Resource Name*) tente de définir un espace de noms (et vraisemblablement des protocoles de résolution) pour des noms d'objet persistants. Ce domaine n'est pas abordé par le présent document, qui a été rédigé pour documenter les pratiques existantes et fournir un point de référence pour les discussions sur les URL et les URN.

Les protocoles de la Toile Mondiale sont discutés sur la liste de diffusion "www-talk-request@info.cern.ch" et dans le groupe de nouvelles "comp.infosystems.www" qui est préférable pour les questions des débutants. La liste de diffusion "uri-request@bunyip.com" a des discussions qui se rapportent particulièrement à la question des URI. L'auteur peut être contacté à "timbl@info.cern.ch".

Le présent document est disponible sous forme hypertext à ["http://info.cern.ch/hypertext/WWW/Addressing/URL/URI_Overview.html"](http://info.cern.ch/hypertext/WWW/Addressing/URL/URI_Overview.html).

2. Besoin d'une syntaxe universelle

Cette section décrit le concept d'URI et ne fait pas partie de la "spécification".

De nombreux protocoles et systèmes sont actuellement utilisés pour la recherche et la restitution de document, et bien plus encore de protocoles ou de raffinements des protocoles existants sont attendus dans un champ dont l'expansion est explosive.

Ces systèmes visent à réaliser une recherche et un lectorat mondial des documents à travers des plates-formes de calcul différentes, et en dépit de la pléthore des protocoles et des formats. Avec l'évolution des protocoles, les passerelles peuvent permettre que reste possible l'accès mondial. Avec l'évolution des formats de données, les programmes de conversion de format peuvent préserver l'accès global. Il y a cependant un domaine dans lequel il n'est pas praticable de faire des conversions, et c'est celui des noms et adresses utilisés pour identifier les objets. C'est parce que les noms et les adresses des objets sont transmis de façons si nombreuses, depuis le dos des enveloppes aux objets hypertexte, et qu'ils peuvent avoir une longue vie.

Une caractéristique commune à presque tous les modèles de données des systèmes passés et proposés est quelque chose qui peut être transposé en un concept "d'objet" et une forme de nom, d'adresse, ou d'identifiant pour cet objet. On peut donc définir un ensemble d'espaces de noms dans lequel ces objets peuvent être dits exister.

Des systèmes pratiques ont besoin d'accéder et de mélanger des objets qui font partie de systèmes différents existants et proposés. Donc, le concept d'ensemble universel de tous les objets, et donc, l'ensemble universel des noms et adresses, dans tous les espaces de noms, devient important. Cela permet que les noms dans des espaces différents soient traités d'une façon commune, même si les noms dans les différents espaces ont des caractéristiques différentes, comme les objets auxquels ils se réfèrent.

URI

Ce document définit un moyen pour encapsuler un nom dans tout espace de noms enregistré, et de l'étiqueter avec l'espace de noms, produisant un membre de l'ensemble universel. Un tel membre codé et étiqueté de cet ensemble est appelé un identifiant de ressource universel ou URI (*Universal Resource Identifier*).

La syntaxe universelle permet l'accès aux objets disponibles en utilisant les protocoles existants, et peut être étendue avec la technologie.

La spécification de la syntaxe d'URI n'a pas d'implication sur les propriétés des noms et adresses dans les divers espaces de noms qui sont transposés dans l'ensemble des chaînes d'URI. Les propriétés découlent de la spécifications des protocoles et des conventions d'usage associées à chaque schéma.

URL

Pour les protocoles d'accès Internet existants, il est nécessaire dans la plupart des cas de définir le codage de l'algorithme d'accès en quelque chose d'assez concis pour être appelé une adresse. Les URI qui se réfèrent aux objets auxquels on accède avec les protocoles existants sont connus comme "localisateurs de ressource uniforme" (URL, *Uniform Resource Locator*) et sont énumérés ici comme ils sont utilisés sur la Toile, mais sont définis de façon formelle dans un document distinct.

URN

Il y a actuellement une piste pour définir un espace de noms plus persistants que tous les URL. Ces "Noms de ressource uniformes" font l'objet des discussions d'un groupe de travail de l'IETF. (Voir de Sollins et Masinter, "Spécifications fonctionnelles pour les URN", [RFC1737].)

La syntaxe d'URI et les formes d'URL ont été d'un large usage dans les logiciels de la Toile depuis 1990.

3. Critères et choix de la conception

La présente section n'a pas de caractère de spécification : elle est une simple explication de la façon dont la spécification a été déduite.

3.1 Critères de conception

La syntaxe a été conçue comme devant être :

- Extensible De nouveaux schémas de dénomination seront ajoutés ultérieurement.
- Complète Il est possible de coder tous les schémas de dénomination.
- Imprimable Il est possible d'exprimer tous les URI en utilisant les caractères ASCII à 7 bits de telle sorte que les URI peuvent, si nécessaire, être passés par écrit.

3.2 Choix pour une syntaxe universelle

Pour la syntaxe elle-même, il y a peu de choix, excepté pour l'ordre et la ponctuation des éléments, les caractères acceptables et les règles d'échappement.

L'exigence d'extensibilité est satisfaite en permettant qu'une chaîne arbitraire (mais enregistrée) soit utilisée comme préfixe. Un préfixe est choisi car l'analyse de gauche à droite est plus courante que celle de droite à gauche. Le choix des deux-points comme séparateur du préfixe du reste de l'URI est arbitraire.

Le décodage du reste de la chaîne est défini comme une fonction du préfixe. De nouveaux préfixes sont introduits pour de nouveaux schémas en tant que de besoin, en accord avec les autorités d'enregistrement. L'enregistrement d'un nouveau schéma exige évidemment la définition du décodage de l'URI en un espace de nom donné, et une définition des propriétés et, selon le cas, des protocoles de résolution, pour l'espace de noms.

L'exigence de complétude est aisément satisfaite en permettant que les noms particulièrement étranges ou complètement binaires soient codés en base 16 ou 64 en utilisant les caractères acceptables.

L'exigence d'être imprimable aurait pu être satisfaite en exigeant que tous les schémas codent les caractères qui ne font pas partie de l'ensemble de base. Cela a conduit à de nombreuses discussions sur ce que devrait être l'ensemble de base. Un cas difficile, par exemple, est quand une chaîne ISO latin 1 apparaît dans un URL, et au sein d'une application avec la capacité de traiter l'ISO Latin-1, elle peut être traitée telle quelle. Cependant, pour le transport en général, les caractères non US-ASCII doivent subir un échappement.

La solution à ce problème a été de spécifier un ensemble sûr de caractères, et un schéma général d'échappement qui peut être utilisé pour coder les caractères "non sûrs". Ce "sûr" convient pour l'utilisation, par exemple, de la messagerie électronique. C'est la forme canonique d'un URI.

Le choix d'un caractère d'échappement pour introduire les représentations de caractères non admis tend aussi à être une question de goût. Il existe une norme ANSI dans le langage C, qui utilise le caractère barre oblique inverse "\". L'utilisation de ce caractère sur les lignes de commandes unix peut cependant être un problème car il est interprété par de nombreux programmes incorporés, et devait lui-même subir un échappement. C'est aussi un caractère qui n'est pas disponible sur certains claviers. Le signe égal est couramment utilisé pour le codage des noms qui ont des paires attribut=valeur. Le signe pourcentage a finalement été choisi comme caractère d'échappement convenable.

Il y a un conflit entre le besoin d'être capable de représenter directement de nombreux caractères y compris les espaces au sein d'un URI, et celui d'être capable d'utiliser l'URI dans des environnements qui ont des jeux de caractères limités ou dans lesquels certains caractères sont enclins à la corruption. Ce conflit a été résolu par l'utilisation d'une méthode

d'échappement hexadécimale qui peut être appliquée à tout caractère interdit dans un contexte donné. Lorsque les URL sont déplacés d'un contexte à un autre, le jeu de caractères échappés peut être élargi ou réduit sans ambiguïté.

L'utilisation de caractères d'espace blanche est risqué dans les URI qui doivent être imprimés ou envoyés par courrier électronique, et l'usage de plusieurs caractères d'espace blanche est très risqué. Cela parce que l'introduction fréquente d'espaces blanches étrangères lorsque les lignes sont repliées par des systèmes comme la messagerie, ou la nécessité absolue de réduire la largeur des colonnes, et à cause de l'inter-conversion de diverses formes d'espaces blanches qui survient durant la conversion de code de caractères et le transfert de texte entre applications. C'est pourquoi la forme canonique des URI a toutes les espaces blanches codées.

4. Syntaxe d'URI

La présente section décrit la syntaxe pour les URI telle qu'utilisée dans l'initiative pour la Toile Mondiale. La syntaxe générique donne un cadre pour que de nouveaux schémas pour les noms soient résolus en utilisant des protocoles encore non définis pour l'instant.

4.1 URI complet

Un URI complet consiste en une spécification de schéma de dénomination suivi par une chaîne dont le format est fonction du schéma de dénomination. Pour les localisateurs des informations sur l'Internet, une syntaxe commune est utilisée pour la partie adresse IP. Une description en BNF de la syntaxe d'URL est donnée plus loin. Les composants sont donnés ci-après. Les identifiants de fragment et les URI relatifs ne sont pas impliqués dans la définition de l'URL de base.

Schéma

Au sein de l'URI d'un objet, le premier élément est le nom du schéma, séparé du reste de l'objet par deux points ":".

Chemin

Le reste de l'URI suit les deux points dans un format qui dépend du schéma. Le chemin est interprété d'une manière qui dépend du protocole utilisé. Cependant, quand il contient des barres obliques, elles doivent impliquer une structure hiérarchique.

4.1.1 Caractères réservés

Le chemin dans l'URI a une signification définie par le schéma particulier. Normalement, il est utilisé pour coder un nom dans un espace de noms particulier, ou un algorithme pour accéder à un objet. Dans l'un et l'autre cas, le codage peut utiliser les caractères admis par la syntaxe BNF, ou le codage hexadécimal des autres caractères.

Certains des caractères réservés ont l'utilisation particulière définie ci-dessous.

Signe de pourcentage

Le signe pour cent ("%") ASCII, 25 en hexadécimal) est utilisé comme caractère d'échappement dans le schéma de codage et n'est jamais admis pour autre chose.

Formes d'indication de hiérarchie

Le caractère barre oblique ("/" en ASCII, 2F en hexadécimal) est réservé pour la délimitation des sous chaînes dont la relation est hiérarchique. Cela permet des formes partielles de l'URI. Les sous chaînes qui consistent en un seul point ou en doubles points ("." ou "..") sont également réservées.

La signification de la barre oblique entre deux segments est que le segment du chemin à gauche est de plus fort poids que le segment du chemin à droite. ("Plus fort poids" dans ce cas se réfère seulement à la proximité de la racine de la structure hiérarchique et ne porte pas de jugement de valeur !)

Note : La similarité des conventions de nom de fichier des systèmes d'exploitation de disque unix et des autres systèmes doit être tenue pour une pure coïncidence, et ne devrait pas être prise comme l'indication que les URI devraient être interprétés comme des noms de fichiers.

Dièse pour les identifiants de fragment

Le caractère dièse ("#" en ASCII, 23 en hexadécimal) est réservé comme délimiteur pour séparer l'URI d'un objet d'un identifiant de fragment.

Chaînes d'interrogation

Le point d'interrogation ("?" en ASCII, 3F en hexadécimal) est utilisé pour délimiter la frontière entre l'URI d'un objet interrogeable, et un ensemble de mots utilisés pour exprimer une interrogation sur cet objet. Lorsque cette forme est utilisée, l'URI combiné représente l'objet qui résulte de l'interrogation qui est appliquée à l'objet original.

Au sein de la chaîne d'interrogation, le signe plus est réservé à la notation abrégée d'une espace. Donc, les signes plus réels doivent être codés. Cette méthode était utilisée pour rendre plus faciles les interrogations d'URI pour les passer dans des systèmes qui n'admettaient pas les espaces.

La chaîne d'interrogation représente des opérations appliquées à l'objet, mais la présente spécification ne donne pas de syntaxe ou sémantique commune pour elles. En pratique, la syntaxe et la sémantique peuvent dépendre du schéma et peuvent même dépendre de l'URI de base.

Autres caractères réservés

L'astérisque ("*" en ASCII, 2A en hexadécimal) et aussi le point d'exclamation ("!" en ASCII, 21 en hexadécimal) sont réservés pour être utilisés comme ayant une signification particulière dans certains schémas spécifiques.

Caractères dangereux

En forme canonique, certains caractères comme les espaces, les caractères de contrôle, certains caractères dont le code ASCII est utilisé différemment dans des variantes de jeux de caractères nationaux différents à 7 bits, et tous les caractères à 8 bits au delà de DEL (7F en hexadécimal) du jeu de caractères ISO Latin-1, ne devront pas être utilisés non codés. C'est une recommandation pour des échanges sans problèmes, et, comme indiqué ci-dessous, le jeu codé peut être étendu ou réduit.

Codage des caractères réservés

Lorsque un système utilise un schéma d'adressage local, il est utile de fournir une transposition des adresses locales en URI de sorte que les références aux objets au sein du schéma d'adressage puissent être référés globalement, et éventuellement accédés par des serveurs passerelles.

Pour un nouveau schéma de dénominations, tout schéma de transposition peut être défini pourvu qu'il soit sans ambiguïté, réversible, et donne des URI valides. Il est recommandé que lorsque existent des aspects hiérarchiques du schéma de dénomination local, ils soient transposés dans la syntaxe de chemin hiérarchique d'URL afin de permettre d'utiliser la forme partielle.

4.1.2 Schéma conventionnel de codage d'URI

Il est aussi recommandé que le schéma conventionnel ci-dessous soit utilisé dans tous les cas, sauf pour un schéma qui code des données binaires par opposition à du texte, auquel cas un codage plus compact comme l'hexadécimal pur ou le base 64 pourrait être plus approprié. Par exemple, la méthode conventionnelle de codage d'URI est utilisée pour transposer les adresses WAIS, FTP, Prospero et Gopher en spécifications d'URI.

Lorsque le schéma de dénomination local utilise les caractères ASCII qui ne sont pas admis dans l'URI, cela peut être représenté dans l'URL par un signe pour cent "%" immédiatement suivi par deux chiffres hexadécimaux (0-9, A-F) donnant le code ISO Latin 1 pour ce caractère. Les codes de caractères autres que ceux admis par la syntaxe ne devront pas être utilisés non codés dans un URI.

Ensembles réduits ou augmentés de jeux de caractères sûrs

La même méthode de codage peut être utilisée pour coder les caractères dont l'utilisation, bien que techniquement admise dans un URI, serait déraisonnable eu égard aux problèmes de corruption par des passerelles imparfaites ou de mauvaise représentation due à l'utilisation de jeux de caractères déviants, ou qui vont simplement être fâcheux dans un environnement particulier. Parce qu'un signe % indique toujours un caractère codé, un URI peut être "sûr" simplement en codant tout caractère considéré comme non sûr, tout en laissant codés les caractères qui le sont déjà. De même, dans les cas où un jeu de caractères plus large est acceptable, des signes % peuvent être répandus de façon sélective et réversible.

Avant que deux URI puissent être comparés, il est donc nécessaire de les amener tous deux au même niveau de codage.

Cependant, les caractères réservés mentionnés ci-dessus ont une signification assez différente quand ils sont codés, et ne peuvent JAMAIS être codés et décodés de cette façon.

Le signe pour cent avec cette signification doit toujours être codé, car sa présence indique autrement toujours un codage. Les séquences qui commencent par un signe pour cent mais ne sont pas suivies par deux caractères hexadécimaux sont réservées pour de futures extensions. (Voir l'exemple 3.)

Exemple 1

Les URI "http://info.cern.ch/albert/bertram/marie-claude" et "http://info.cern.ch/albert/bertram/marie%2Dclaude" sont identiques, car le %2D code un caractère "trait d'union".

Exemple 2

Les URI "http://info.cern.ch/albert/bertram/marie-claude" et "http://info.cern.ch/albert/bertram%2Fmarie-claude" NE SONT PAS identiques, car dans le second cas, la barre oblique codée n'a pas de signification hiérarchique.

Exemple 3

Les URI "fxqn:/us/va/reston/cnri/ietf/24/asdf%*.fred" et "news:12345667123%asdghfh@info.cern.ch" sont illégaux, car tous les caractères % impliquent des codages, et qu'il n'y a pas de décodage défini pour "%*" ou "%as" dans cette recommandation.

4.1.3 Forme partielle (relative)

Au sein d'un objet dont l'URI est bien défini, l'URI d'un autre objet peut recevoir une forme abrégée, et les parties des deux URI sont les mêmes. Cela permet que des objets au sein d'un groupe se réfèrent l'un à l'autre sans exiger l'espace d'une référence complète, et cela permet incidemment de déplacer le groupe d'objets sans changer aucune référence. On doit souligner que lorsque une référence est passée dans n'importe quel autre qu'un contexte bien contrôlé, la forme complète doit toujours être utilisée.

Dans les applications de la Toile Mondiale, le contexte de l'URI est celui du document ou objet qui contient une référence. Dans ce cas, des URI partiels peuvent être générés dans des objets virtuels ou mémorisés dans des objets réels, sans avoir besoin de changements importants si les parties de plus fort poids d'un système de dénominations hiérarchisé sont modifiées. En plus de la concision, cela donne une plus grande robustesse aux systèmes réels, en permettant de cacher des informations entre les composants d'un système.

La forme partielle s'appuie sur une propriété de la syntaxe d'URI que certains caractères ("/") et certains éléments de chemin ("..", ".") ont une signification réservée à la représentation d'un espace hiérarchique, et doivent être reconnus comme tels à la fois par les clients et les serveurs.

Une forme partielle peut être distinguée d'une forme absolue en ce que la première doit avoir deux points et que ces deux points doivent survenir avant tout caractère barre oblique. Les systèmes qui n'exigent pas de forme partielle ne devraient pas utiliser de barre oblique non codée dans leurs schémas de désignation. Si ils le font, les URI absolus fonctionneront quand même, mais il peut en résulter une certaine confusion. (Voir la note sur Gopher un peu plus loin.)

Les règles pour l'utilisation d'un nom partiel relatif à l'URI du contexte sont :

- Si les parties du schéma sont différentes, l'URI absolu entier doit être donné. Autrement, le schéma est omis.
- Si l'URI partiel commence par un nombre positif de barres obliques consécutives, tout ce qui se trouve après l'URI de contexte jusqu'à (non inclus) la première occurrence d'exactly le même nombre de barres obliques consécutives qui n'a pas un nombre supérieur de barres obliques consécutives n'importe où à sa droite, est pris pour être le même et donc ajouté à l'URL partiel pour former l'URL complet.

Autrement :

- La dernière partie du chemin de l'URI de contexte (tout ce qui suit la barre oblique la plus à droite) est retiré, et l'URI partiel est ajouté à sa place, puis, dans le résultat, toutes les occurrences de "xxx/.." ou "/." sont retirées de façon récurrente, "xxx", ".." et "." étant des éléments de chemin complets.

Note : Barres obliques en queue. Si un chemin du localisateur de contexte se termine par une barre oblique, les URI partiels sont traités différemment de l'URI qui porte le même chemin mais sans barre oblique en queue. La barre oblique en queue indique un segment vide dans le chemin.

Note : Gopher. Le système gopher n'a pas le concept d'URI relatif, et la communauté gopher admet actuellement la barre oblique "/" comme caractères de données dans les URI gopher sans échappement en "%2F". Les formes relatives ne peuvent en général pas être utilisées pour les documents desservis par des serveurs gopher. Si ils sont utilisés, le logiciel WWW va alors supposer, normalement à juste titre, qu'il ont en fait une signification hiérarchique en dépit de leur spécification. L'utilisation du protocole HTTP plutôt que gopher est cependant recommandée.

Exemples

Dans le contexte de l'URI "magic://a/b/c//d/e/f" les URI partiels vont s'articuler comme suit :

g magic://a/b/c//d/e/g
/g magic://a/g

```
//g magic://g
../g magic://a/b/c//d/g
g:h g:h
```

Et dans le contexte de l'URI "magic://a/b/c//d/e/" le résultat serait exactement le même.

4.1.4 Identifiant de fragment

Cela représente une partie, un fragment, ou une sous fonction, au sein d'un objet. Sa syntaxe et sa sémantique sont définies par l'application responsable de l'objet, ou par la spécification du type de contenu de l'objet. La seule définition ici est celle des caractères admis par lesquels elle est représentée dans un URL.

Les syntaxes spécifiques pour la représentation des fragments dans les documents de texte par gamme de ligne et de caractères, ou dans les graphiques par des coordonnées, ou dans des documents structurés en utilisant des échelles, conviennent pour la normalisation mais ne seront pas définies ici.

L'identifiant de fragment suit l'URL de l'objet complet d'où il est séparé par un signe dièse (#). Si l'identifiant de fragment est vide, le signe dièse peut être omis. Un identifiant de fragment vide avec ou sans signe dièse signifie que l'URL se réfère à l'objet complet.

Bien que ce raccourci soit admis pour l'identification des fragments, la question de l'adressage de parties des objets, ou du groupement des objets et des relations entre des continuations d'objets et d'objets contenant, n'est pas traitée par le présent document.

Les identifiants de fragment NE TRAITENT PAS de la question des objets qui sont des versions différentes d'un objet "vivant", ni de l'expression des relations entre différentes versions et l'objet vivant.

Rien n'implique qu'un identifiant de fragment se réfère à quelque chose qui pourrait être extrait comme un objet de plein droit. Il peut, par exemple, se référer à un point indivisible au sein d'un objet.

4.1.5 Schémas spécifiques

La transposition des URI en des protocoles existants normalisés et expérimentaux est souligné dans la définition en syntaxe BNF. Des notes sur les différents protocoles suivent. On se réfère souvent à ces URI comme à des URL, bien que la définition exacte du terme d'URL soit encore en discussion (mars 1993). Les schémas couverts sont :

http	Protocole de transfert hypertexte (exemples)
ftp	Protocole de transfert de fichiers
gopher	Protocole Gopher
mailto	Adresse de messagerie électronique
news	Nouvelles Usenet
telnet, rlogin et tn3270	Référence à des sessions interactives
wais	Serveurs d'informations de zone large
file	Accès de fichier local

Les schémas suivants sont proposés comme essentiels pour l'unification de la Toile avec la messagerie électronique, mais ne sont pas encore (à la connaissance de l'auteur) mis en œuvre :

mid	Identifiants de message pour la messagerie électronique
cid	Identifiants de contenu pour partie de corps MIME

Les schémas pour X.500, base de données de gestion de réseau, et Whois++ n'ont pas été spécifiés et pourront faire l'objet d'études complémentaires. Les schémas pour Prospero, et l'utilisation restreinte de NNTP ne sont pas encore mis en œuvre actuellement à la connaissance de l'auteur.

Le préfixe "urn" est réservé à l'utilisation du codage de nom de ressource uniforme quand il aura été développé par le groupe de travail de l'IETF.

De nouveaux schémas pourront être enregistrés ultérieurement.

HTTP

Le protocole HTTP spécifie que le chemin est traité de façon transparente par ceux qui traitent les URL, sauf par les serveurs qui les déréférencent. Le chemin est passé au serveur par le client avec toute demande, mais n'est pas autrement compris par le client.

Les détails de l'hôte ne sont pas passés au client lorsque l'URL est un URL HTTP qui se réfère au serveur en question. Dans ce cas, la chaîne envoyée commence par la barre oblique qui suit les détails de l'hôte. Cependant, quand un serveur HTTP est utilisé comme passerelle (ou comme "mandataire") l'URI entier, qu'il soit HTTP ou un autre schéma, est alors passé à la ligne de commande HTTP. La partie recherche, si elle est présente, est envoyée au titre de la commande HTTP, et peut à cet égard être traitée au titre du chemin. Aucune partie fragmentée d'un URI de la Toile mondiale (le signe dièse et ce qui suit) n'est envoyée avec la demande. Les caractères d'espaces et de contrôle dans les URL doivent être échappés pour la transmission en HTTP, comme le doivent les autres caractères non admis.

5. Exemples

Ces exemples ne font pas partie de la "spécification" : ils ne sont donnés que comme illustrations.

5.1 HTTP

L'URI de la page "d'accueil" d'un serveur est par convention :

`http://www.my.work.com/`

Comme le reste de l'URL (après le nom d'hôte et l'accès) est opaque pour le client, il présente une grande variété mais ce qui suit est assez typique.

`http://www.my.uni.edu/info/matriculation/enroling.html`

`http://info.my.org/AboutUs/Phonebook`

`http://www.library.my.town.va.us/Catalogue/76523471236%2Fwen44--4.98`

`http://www.my.org/462F4F2D4241522A314159265358979323846`

Un URL pour un serveur sur un accès différent pour 80 ressemble à "`http://info.cern.ch:8000/imaginary/test`".

Une référence à une partie particulière d'un document peut, y compris l'identifiant de fragment, ressembler à :

`http://www.myu.edu/org/admin/people#andy`

auquel cas la chaîne "#andy" n'est pas envoyée au serveur, mais est conservée par le client et utilisée lorsque la totalité de l'objet a été restituée.

Une recherche dans une base donnée de texte peut ressembler à :

`http://info.my.org/AboutUs/Index/Phonebook?dobbins`

et dans une autre base de données :

`http://info.cern.ch/RDB/EMP?*%20where%20name%%3Ddobbins`

Dans tous les cas, le client passe la chaîne de chemin au serveur non interprétée, et il appartient au client d'en déduire quelque chose.

5.2 FTP

Le préfixe ftp: indique que le protocole FTP est utilisé, comme défini dans la [RFC0959], STD 9 ou toute RFC qui lui succède. Le numéro d'accès, si il est présent, donne l'accès du serveur FTP si ce n'est pas le FTP par défaut.

5.2.1 Nom d'utilisateur et mot de passe

La syntaxe permet l'inclusion d'un nom d'utilisateur et même d'un mot de passe pour les systèmes qui n'utilisent pas la convention de FTP anonyme. Cette convention sera utilisée par défaut, cependant, si aucun nom d'utilisateur ou mot de passe n'est fourni, on supposera que le nom d'utilisateur est "anonymous" et que le mot de passe est l'adresse de messagerie de styme Internet de l'utilisateur.

Lorsque possible, cette adresse de messagerie devrait correspondre à une adresse de messagerie utilisable pour l'utilisateur, et devrait de préférence donner un nom d'hôte du DNS qui résolve l'adresse IP du client. Noter qu'actuellement, le traitement du mot de passe anonyme par les serveurs est assez variable .

5.2.2 Chemin

Le protocole FTP permet une séquence de commandes CWD (changer le répertoire de travail, *change working directory*) et une commande TYPE avant les commandes de service telles que RETR (*retrieve*, restituer) ou NLIST (etc.) qui permettent en fait d'accéder à un fichier.

Les arguments de toute commande CWD sont des parties de l'URL en segments successifs délimités par une barre oblique, et le segment final convient comme argument de nom de fichier pour la commande RETR pour la restitution, ou comme argument de répertoire pour NLIST.

Pour certains systèmes de fichiers (Unix en particulier), le caractère "/" utilisé pour noter la structure hiérarchique de l'URL correspond au délimiteur utilisé pour construire une hiérarchie de noms de fichiers, et donc, le nom de fichier va ressembler à un chemin d'URL. Cela NE SIGNIFIE PAS que l'URL est un nom de fichier Unix.

Note : Restitution d'URL suivants à partir du même hôte : Il n'y a pas de modèle commun de hiérarchie dans le protocole FTP, de sorte que si une commande de changement de répertoire a été donnée, il est en général impossible d'en déduire quelle séquence devrait être donnée pour naviguer sur un autre répertoire pour une seconde restitution, si les chemins sont différents. Le seul algorithme fiable est de se déconnecter et de rétablir la connexion de commande.

5.2.3 Type de données

Le type de contenu de données d'un fichier peut seulement, dans le cas général de FTP, être déduit du nom, normalement, du suffixe du nom. Ceci n'est pas normalisé. Une solution de remplacement est qu'il soit transféré dans des informations en dehors de l'URL. Un type de transfert FTP convenable (par exemple binaire "I" ou text "A") doit à son tour être déduit du type de contenu de données. Il est recommandé que des conventions pour les suffixes des archives publiques soient établies, mais ceci sort du domaine d'application de ce document.

Un URI FTP peut facultativement spécifier le type de transfert de données FTP par lequel un objet va être restitué. La plupart des méthodes correspondent aux "types de données" FTP ASCII et IMAGE pour la restitution d'un document, comme spécifié dans FTP par la commande TYPE. Une méthode indique l'accès de répertoire.

Le type de données est spécifié par un suffixe de l'URL. Les suffixes possibles sont :

;type = <type-code> Utilise le type FTP tel que donné pour effectuer le transfert de données.
/ Utilise les commandes de liste de répertoires FTP pour lire les répertoires.

Le code de type est dans le format défini dans la [RFC0959], excepté que l'espace est omise de l'URL.

Mode de transfert : Le mode flux est toujours utilisé.

5.3 Gopher

L'URL gopher spécifie l'hôte et facultativement l'accès auquel le client devrait se connecter. Il est suivi d'une barre oblique et d'un seul code de type gopher. Ce code de type est utilisé par le client pour déterminer comment interpréter la réponse du serveur et n'est pas à envoyer au serveur. La chaîne de commande à envoyer au serveur suit immédiatement le caractère de type gopher. Il consiste en la chaîne de sélecteur gopher suivie par toute syntaxe "Gopher plus", mais toujours en omettant la paire CR LF en queue.

Lorsque la chaîne de commande gopher contient des caractères (comme des caractères CR LF et HT incorporés) non admis dans un URL, ils sont codés en utilisant le codage conventionnel.

Note : Certaines chaînes de sélecteur gopher commencent par une copie du caractère de type gopher, auquel cas ce caractère va survenir deux fois consécutivement. Noter aussi que la chaîne de sélecteur gopher peut être une chaîne vide car c'est ainsi que les clients gopher se réfèrent au répertoire de niveau supérieur sur un serveur gopher.

Si la chaîne de commande codée (avec la suppression du CR LF en queue) était vide, le caractère de type gopher pourrait être omis et on supposerait le type "1" (ASCII, 31 en hexadécimal).

Noter que la barre oblique "/" dans les chaînes de sélecteur gopher peut ne pas correspondre à un niveau dans la structure hiérarchique.

5.4 Mailto

Cela permet à un URL de spécifier une adresse de messagerie "addr-spec" de la [RFC0822]. Noter que l'utilisation de %, par exemple comme utilisé dans la formation d'une adresse de messagerie par l'intermédiaire d'une passerelle, exige la conversion en %25 dans un URL.

5.5 News

Les localisateurs "news" se réfèrent soit à des noms de groupe de nouvelles, soit à des identifiants de messages d'articles qui doivent se conformer aux règles pour un identifiant de message de la [RFC1036] (Horton 1987). Un identifiant de message peut être distingué d'un nom de groupe de nouvelles par la présence du caractère arobase "@". Ces règles impliquent que, au sein d'un article, une référence à un groupe de nouvelles ou à un autre article sera un URL valide (en forme partielle).

Un URI "news" peut être déréférencé en utilisant NNTP [RFC0977] (Kantor 1986) (la commande ARTICLE par message-id) ou en utilisant tout autre protocole pour convoyer les articles de nouvelles usenet, ou par référence à un corps d'articles de nouvelles déjà reçu.

Note 1 : Parmi les URL, les URL "news" sont anormaux en ce qu'ils sont indépendants de la localisation. Ils ne conviennent pas comme candidats URN parce que l'architecture NNTP s'appuie sur l'arrivée à expiration des articles et donc seul un petit nombre d'articles est disponible à tout moment. Lorsque un URL "news:" est cité, l'hypothèse est que le lecteur va aller chercher l'article ou le groupe à partir de son hôte de nouvelles local. Les noms des hôtes de nouvelles NE FONT PAS partie des URL news.

Note 2 : Un problème actuel est que l'identifiant de message est insuffisant pour permettre la restitution d'un article arrivé à expiration, car aucun algorithme n'existe pour déduire un site d'archive et un nom de fichier. L'ajout de la date et du groupe de nouvelles à l'URL de l'article permettrait cela si il existait un répertoire des sites d'archive par groupe de nouvelles.

Un sujet d'étude suggéré en rapport avec le groupe de travail NNTP serait l'extension possible qui permettrait la désignation d'une filière de sujets comme objets adressables.

Telnet, rlogin, tn3270

L'utilisation des URL pour représenter des sessions interactives est une extension souhaitable à leur utilisation pour les objets. Cela permet l'accès aux systèmes d'information qui ne fournissent qu'un service interactif et pas de serveur d'informations. Comme les informations au sein du service ne peuvent pas être adressées individuellement ou, en général, restituées automatiquement, ceci est une solution moins désirable, bien que commune actuellement.

5.6. URN

Le "nom des ressource universel" est actuellement (mars 1993) en cours de développement dans l'IETF. Une spécification des exigences est en préparation. Elle ressemble actuellement à une courte chaîne qui conviendrait au codage dans la syntaxe d'URI, cas auquel le préfixe "urn:" est réservé. L'URN devra être codé précisément comme défini dans la (future) norme d'URN, sauf :

- si la description officielle de la syntaxe d'URN comporte des caractères d'enveloppe de constante, ils ne devront alors pas être omis dans le codage d'URI de l'URN ;
- si l'URN a une nature hiérarchique, le délimiteur "barre oblique" devra être utilisé dans le codage d'URI ;
- si l'URN a une nature hiérarchique, la partie de plus fort poids devra être codée sur la gauche du codage d'URI ;
- tout caractère de signification réservée dans la syntaxe d'URI devra être codé en échappement (en pourcentage).

Ces règles s'appliquent, bien sûr, à tout schéma d'URI. Il est sans doute possible que la syntaxe d'URN soit choisie de telle façon que le codage d'URI soit une transcription biunivoque.

Un exemple pourrait être celui d'un nom tel que "urn:iana/dns/ch/cern/cn/techdoc/94/1642-3" mais le lecteur devrait se référer aux dernières spécifications ou projets d'URN.

5.7. WAIS

La mise en œuvre actuelle de WAIS dans le domaine public exige qu'un client connaisse le "type" d'un objet avant sa restitution. Cette valeur est retournée avec un identifiant d'objet interne dans la réponse à la recherche. Elle a été codée dans la partie chemin de l'URL afin de rendre l'URL suffisant pour la restitution de l'objet.

Dans le monde WAIS, les noms n'ont bien sûr pas besoin d'être munis d'un préfixe "wais:" (par les règles de forme partielle).

Le chemin wpath d'un URL WAIS consiste en des champs codés de l'identifiant WAIS, dans le même ordre que dans l'identifiant WAIS. Pour chaque champ, le numéro de champ d'identifiant est constitué des chiffres qui se trouvent avant le signe égal, et le contenu du champ suit, codé selon le codage conventionnel, terminé par ";".

5.8. file

Les autres schéma d'URI (sauf nntp) partagent la propriété d'être également valides en tout lieu géographique.

Il y a cependant une réelle exigence pratique d'être capable de générer un URL pour un objet dans le système de fichiers local d'une machine.

La syntaxe est similaire à celle de ftp, mais dans ce cas, la barre oblique est utilisée pour établir des limites entre les niveaux de répertoire d'un système de fichiers hiérarchisé. Le logiciel "client" convertit l'URL de fichier en un nom de fichier dans les conventions de désignation de fichier local. Cela permet aux fichiers locaux d'être traités comme des objets du réseau sans qu'il soit besoin d'utiliser un serveur réseau pour y accéder. Cela peut être utilisé par exemple pour définir un document "en local" de l'utilisateur dans la Toile Mondiale.

Il y a évidemment un danger de confusion lorsque un lien fait pour un fichier local devrait être suivi par quelqu'un qui est sur un système différent, cela peut avoir des résultats inattendus et éventuellement dommageables. Donc, la convention est que même un URL "file" est fourni avec une partie hôte. Cela permet à un client qui est sur un autre système de savoir qu'il ne peut pas accéder au système de fichiers, ou peut-être d'utiliser quelque autre mécanisme local pour accéder au fichier.

La valeur particulière "localhost" est utilisés dans le champ Hôte pour indiquer que le nom de fichier devrait réellement être utilisé sur tout hôte quel qu'il soit. Cela permet, par exemple, les liens avec des fichiers qui sont répartis sur de nombreuses machines, ou que votre sujet de cours sur "votre fichier de mot de passe local unix" soit cohérent parmi les utilisateurs des données.

Un champ Hôte vide est équivalent à "hôte local".

5.9. Message-Id

Pour les systèmes qui comportent des informations transférées en utilisant des protocoles de messagerie, il faut être capable de faire des références croisées entre différents éléments d'informations, même si, de par la nature de la messagerie, ces éléments ne sont disponibles que pour un ensemble restreint de personnes.

Deux schémas sont définis.

5.9.1 MID

Le premier, "mid:", se réfère à l'identifiant de message de la [RFC0822], (STD 11) d'un message électronique. Cet identifiant est déjà utilisé dans la RFC0822 dans les champs Références et En réponse à, par exemple.

Le reste de l'URL après le "mid:" est l'identifiant de message de la RFC822 dont l'enveloppe \diamond de constante a été retiré, laissant un identifiant dont le format se trouve être en fait le même que le format de "addr-spec" pour les boîtes aux lettres (bien que sa sémantique soit différente).

L'utilisation d'un URL "mid" implique l'accès à un corps de messagerie déjà reçu. Si un message a été distribué en utilisant NNTP ou un autre protocole usenet sur le système "news", la forme "news:" devrait alors être utilisée.

5.9.2 Content-Id

Le second schéma, "cid:", est similaire à "mid:", mais fait référence à une partie de corps de message MIME par la valeur de son champ Identifiant de contenu. Cela permet, par exemple, qu'un document maître soit la première partie d'un message multipart/relatif MIME et se réfère à des parties de composants qui sont transférées dans le même message.

Note : Attention cependant, qu'il n'est exigé des identifiants de contenu qu'ils soient uniques qu'au sein du contexte d'un message MIME donné, et donc l'URL cid: n'a de signification que dans le contexte du même message MIME. Pour une référence en dehors du message, il faudrait ajouter qu'il soit ajouté à l'identifiant de message du message entier. La syntaxe de cela n'a pas encore été définie.

6. Schémas pour des études complémentaires

6.1 X500

La transposition des noms x500 en URL n'est pas définie ici. Il faut prendre une décision pour savoir s'il faut admettre les "noms distinctifs" ou les "noms faciles à lire" (ufn, *user friendly name*), ou les deux. Si des conversions de ponctuation sont nécessaires sur les représentations x500 adoptées (comme l'utilisation des barres obliques entre les parties d'un ufn) elles doivent être définies. Ceci est encore un sujet d'études.

6.2 WHOIS

Ce préfixe décrit l'accès qui utilise le schéma "whois++" dans le processus de définition. La partie nom d'hôte est la même que pour les autres schémas fondés sur IP. La partie chemin peut être soit une bretelle whois pour un objet whois, soit une chaîne d'interrogation whois valide. Ceci est encore un sujet d'études.

Base de données de gestion de réseau
Ceci est encore un sujet d'études.

6.3 NNTP

C'est une forme de référence de remplacement pour les articles de nouvelles, précisément, à utiliser avec les serveurs NNTP, et particulièrement avec les mises en œuvre incomplètes de serveur qui ne permettent pas de restitution par un identifiant de message. Dans tous les autres cas, le schéma "news" devrait être utilisé.

Il donne le nom de serveur de nouvelles, le nom du groupe de nouvelles, et le numéro d'indice d'un article au sein du groupe de nouvelles sur ce serveur particulier. Le protocole NNTP doit être utilisé.

Note 1 : Cette forme d'URL n'est pas pour une accessibilité mondiale, car normalement les serveurs NNTP ne permettent l'accès qu'à partir des clients locaux. Noter que les numéros d'article au sein des groupes varient d'un serveur à l'autre.

Cette forme d'URL ne devrait pas être citée en dehors de sa zone locale. Elle ne devrait pas être utilisée au sein d'articles de nouvelles pour une circulation plus large que le seul serveur. C'est un identifiant local pour une ressource qui est souvent disponible mondialement, et il n'est donc pas recommandé sauf dans le cas où une mise en œuvre NNTP incomplète sur le serveur local force son adoption.

6.4 Prospero

Le service de répertoire Prospero (Neuman, 1991) est utilisé pour résoudre les URL. Il donne une méthode d'accès pour les objets (qui peuvent alors être eux-mêmes représentés en URL si ils sont traduits). La partie hôte contient un nom d'hôte ou une adresse Internet. La partie accès est facultative.

La partie chemin contient un nom d'objet spécifique de l'hôte et un numéro de version facultatif. Si il est présent, le numéro de version est séparé du nom d'objet spécifique de l'hôte par les caractères "%00" (pour cent zéro zéro) ceci étant une terminaison de chaîne d'échappement (nulle). Les liens Prospero externes sont représentés comme des URL de la méthode d'accès sous-jacente et ne sont pas représentés comme des URL Prospero.

7. Enregistrement des schémas de dénomination

Un nouveau schéma de dénomination peut être introduit en définissant une transposition dans la syntaxe d'URL conforme en utilisant un nouveau préfixe. Des préfixes expérimentaux peuvent être utilisés par accord mutuel entre les parties, et doivent commencer par les caractères "x-". Le nom de schéma "urn:" est réservé pour le travail en cours sur un schéma pour des noms plus persistants.

Il est proposé que l'Autorité d'allocation des numéros de l'Internet (IANA) assume la fonction d'enregistrement des

nouveaux schémas. Toute soumission d'un nouveau schéma d'URI doit comporter une définition d'un algorithme pour la restitution de tout objet au sein de ce schéma. L'algorithme doit prendre l'URI et produire soit un ensemble d'URL qui va conduire à l'objet désiré, soit l'objet lui-même, dans un format bien défini ou déterminable.

Il est recommandé que ceux qui proposent un nouveau schéma démontrent son utilité et sa capacité de fonctionnement par la fourniture d'une passerelle qui produira des images des objets dans le nouveau schéma pour les clients qui utilisent un protocole existant. Si le nouveau schéma n'est pas un schéma de localisateur, les propriétés des noms dans le nouvel espace devraient alors être clairement définies. Il est de même recommandé que, lorsque un protocole permet la restitution par l'URL, que le logiciel client ait la possibilité d'être configuré pour utiliser des localisateurs de passerelles spécifiques pour l'accès indirect à travers de nouveaux schémas de dénomination.

8. BNF de la syntaxe générique d'URI

Voici une description dans le style BNF de la syntaxe d'URI. au niveau où les schémas spécifiques ne sont pas pris en compte. Une ligne verticale "|" indique les alternatives, et les [crochets] indiquent les parties facultatives. Les espaces sont représentés par le mot "space", et le caractère ligne verticale par "vline". Les lettres seules sont des lettres seules. Tous les mots de plus d'une lettre ci-dessous sont des entités décrites quelque part dans cette description. La production "generic" donne une analyse de niveau supérieur des mêmes URI que les autres productions. Les caractères "national" et de "punctuation" n'apparaissent dans aucune production et ne peuvent donc pas apparaître dans les URI.

```

fragmentaddress uri [ # fragmentid ]
uri              scheme : path [ ? search ]
scheme          ialpha
path            void | xpalphas [ / path ]
search         xalphas [ + search ]
fragmentid     xalphas
xalpha         alpha | digit | safe | extra | escape
xalphas       xalpha [ xalphas ]
xpalpha       xalpha | +
xpalphas     xpalpha [ xpalpha ]
ialpha       alpha [ xalphas ]
alpha        a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | A | B | C | D | E | F | G | H | I
              | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z
digit        0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
safe         $ | - | _ | @ | . | &
extra       ! | * | " | ' | ( | ) | ,
reserved    = | ; | / | # | ? | : | space
escape      % hex hex
hex         digit | a | b | c | d | e | f | A | B | C | D | E | F
national    { | } | vline | [ | ] | \ | ^ | ~
punctuation < | >
void
(fin du BNF d'URI)

```

9. BNF pour les schémas d'URL spécifiques

Ceci est une description de style BNF de la syntaxe de localisateur de ressource uniforme. Une ligne verticale "|" indique les alternatives, et les [crochets] indiquent les parties facultatives. Les espaces sont représentés par le mot "space", et le caractère ligne verticale par "vline". Les lettres seules sont des lettres seules. Tous les mots de plus d'une lettre ci-dessous sont des entités décrites quelque part dans cette description. La préférence actuelle du groupe de travail URI de l'IETF est en faveur de la production prefixedurl. (novembre 1993. juillet 93: url). Les caractères "national" et de "punctuation" n'apparaissent dans aucune production et ne peuvent donc pas apparaître dans les URL. La production "afsaddress" subsiste à titre historique, mais n'est pas une production d'url.

```

refixedurl    u r l : url
url           httpaddress | ftpaddress | newsaddress | nntpaddress | prosperoaddress | telnetaddress | gopheraddress |
              waisaddress | mailtoaddress | midaddress | cidaddress
scheme       ialpha
httpaddress  h t t p : // hostport [ / path ] [ ? search ]
ftpaddress   f t p : // login / path [ filetype ]

```

afsaddress	a f s : // cellname / path
newsaddress	n e w s : groupart
nntpaddress	n n t p : group / digits
midaddress	m i d : addr-spec
cidaddress	c i d : content-identifier
mailtoaddress	m a i l t o : xalphas @ hostname
waisaddress	waisindex waisdoc
waisindex	w a i s : // hostport / database [? search]
waisdoc	w a i s : // hostport / database / wtype / wpath
wpath	digits = path ; [wpath]
groupart	* group article
group	ialpha [. group]
article	xalphas @ host
database	xalphas
wtype	xalphas
prosperoaddress	prosperolink
prosperolink	p r o s p e r o : // hostport / hsoname [% 0 0 version [attributes]]
hsoname	path
version	digits
attributes	attribute [attributes]
attribute	alphanums
telnetaddress	t e l n e t : // login
gopheraddress	g o p h e r : // hostport [/ gtype [gcommand]]
login	[user [: password] @] hostport
hostport	host [: port]
host	hostname hostnumber
ftptype	A formcode E formcode I L digits
formcode	N T C
cellname	hostname
hostname	ialpha [. hostname]
hostnumber	digits . digits . digits . digits
port	digits
gcommand	path
path	void segment [/ path]
segment	xpalphas
search	xalphas [+ search]
user	alphanum2 [user]
password	alphanum2 [password]
fragmentid	xalphas
gtype	xalpha
alphanum2	alpha digit - _ . +
xalpha	alpha digit safe extra escape
xalphas	xalpha [xalphas]
xpalpha	xalpha +
xpalphas	xpalpha [xpalphas]
ialpha	alpha [xalphas]
alpha	a b c d e f g h i j k l m n o p q r s t u v w x y z A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
digit	0 1 2 3 4 5 6 7 8 9
safe	\$ - _ @ . & + -
extra	! * " ' () ,
reserved	= ; / # ? : space
escape	% hex hex
hex	digit a b c d e f A B C D E F
national	{ } vline [] \ ^ ~
punctuation	< >
digits	digit [digits]
alphanum	alpha digit
alphanums	alphanum [alphanums]
void	

(fin du BNF d'URL)

10. Références

- Iberty, R., et.al., "Notes on the Internet Gopher Protocol", University of Minnesota, décembre 1991, <ftp://boombbox.micro.umn.edu/pub/gopher/gopher_protocol>. Voir aussi à <<gopher://gopher.micro.umn.edu/00/InformationAboutGopher/AboutGopher>>
- Berners-Lee, T., "Hypertext Transfer Protocol (HTTP)", CERN, décembre 1991, mis à jour de temps en temps à <<ftp://info.cern.ch/pub/www/doc/http-spec.txt>>
- Davis, F, et al., "WAIS Interface Protocol: Prototype Functional Specification", Thinking Machines Corporation, 23 avril 1990. <<ftp://quake.think.com/pub/wais/doc/protospec.txt>>
- International Standards Organization, "Information and Documentation - Search and Retrieve Application Protocol Specification for open Systems Interconnection", ISO-10163.
- Huitema, C., "Naming: strategies and techniques", Computer Networks and ISDN Systems 23 (1991) 107-110.
- Kahle, B., "Document Identifiers, or International Standard Book Numbers for the Electronic Age", <<ftp://quake.think.com/pub/wais/doc/doc-ids.txt>>
- Lynch, C., Coalition for Networked Information: "Workshop on ID and Reference Structures for Networked Information", novembre 1991. See <<wais://quake.think.com/wais-discussion-archives?lynch>>
- Neuman, B. Clifford, "Prospero: A Tool for Organizing Internet Resources", Electronic Networking: Research, Applications and Policy, Vol 1 n° 2, Meckler Westport CT USA, 1992. Voir aussi à <<ftp://prospero.isi.edu/pub/prospero/oir.ps>>
- [RFC0822] D. Crocker, "Norme pour le format des messages de texte de l'ARPA-Internet", STD 11, août 1982. (*Obsolète, remplacée par la RFC5322*)
- [RFC0959] J. Postel et J. Reynolds, "Protocole de transfert de fichiers (FTP)", STD 9, octobre 1985.
- [RFC0977] B. Kantor et P. Lapsley, "Protocole de transfert des nouvelles du réseau", février 1986. (*Obsolète, voirRFC3977*)
- [RFC1034] P. Mockapetris, "Noms de domaines - Concepts et facilités", STD 13, novembre 1987.
- [RFC1036] M. Horton et R. Adams, "Norme pour l'échange de messages USENET", décembre 1987. (*Remplacée par RFC5536*)
- [RFC1736] J. Kunze, "Recommandations fonctionnelles pour les localisateurs de ressource Internet", février 1995. (*Info*)
- [RFC1737] K. Sollins et L. Masinter, "Exigences fonctionnelles pour les noms de ressource uniformes", décembre 1994.
- Yeong, W., "Towards Networked Information Retrieval", Technical report 91-06-25-01, juin 1991, Performance Systems International, Inc. <<ftp://uu.psi.com/wp/nir.txt>>
- Yeong, W., "Representing Public Archives in the Directory", Travail en cours, novembre 1991, maintenant expiré.

11. Considérations pour la sécurité

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

12. Adresse de l'auteur

Tim Berners-Lee
World-Wide Web project
CERN
1211 Genève 23,
Confédération Helvétique
téléphone : +41 (22)767 3755
fax : +41 (22)767 7155
mél : timbl@info.cern.ch