

Groupe de travail Réseau
Request for Comments : 1737
Catégorie : Information
Traduction Claude Brière de L'Isle

K. Sollins, MIT/LCS
L. Masinter, Xerox Corporation
décembre 1994

Exigences fonctionnelles pour les noms de ressources uniformes

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.

1. Introduction

Le présent document spécifie un ensemble minimum d'exigences pour un type d'identifiant de ressource de l'Internet connu comme nom de ressource uniforme (URN, *Uniform Resource Name*). Les URN entrent dans une architecture plus large des informations de l'Internet, qui est en plus composée des caractéristiques de ressource uniforme (URC, *Uniform Resource Characteristics*) et des localisateurs de ressource uniformes (URL, *Uniform Resource Locator*). Les URN sont utilisés pour l'identification, les URC pour l'inclusion de méta informations, et les URL pour localiser ou trouver des ressources. Il est fourni comme base de l'évaluation standard des URN. La discussion de ce travail s'est faite sur la liste de diffusion uri@bunyip.com et aux sessions du groupe de travail URI de l'IETF.

Les exigences décrites ici ne sont pas nécessairement exhaustives ; par exemple, plusieurs questions en rapport avec la prise en charge de la duplication des ressources et avec la sécurité ont été discutées ; cependant, les problèmes ne sont pas suffisamment bien compris pour l'instant pour inclure ici des exigences spécifiques de ces domaines.

Au sein du domaine du concept général de systèmes d'objets répartis, il y a de nombreux concepts qui sont discutés sous le thème général de "désignation". Les exigences pour les URN exposées ici sont pour une facilité qui s'adresse à un ensemble différent (et, en général, plus contraignant) de besoins qui sont fréquemment le domaine de la désignation générale des objets.

Les exigences des noms de ressource uniformes se tiennent dans l'architecture globale de l'identification de ressource uniforme. Afin de construire des applications dans le cas le plus général, l'utilisateur doit être capable de découvrir et identifier les informations, objets, ou ce que, dans cette architecture, nous appellerons des ressources, sur lesquelles l'application va fonctionner. Au delà de cette déclaration, l'architecture d'URI ne définit pas de "ressource". Avec la croissance du réseau et de l'interconnectivité, la capacité à utiliser les ressources distantes, peut-être gérées de façon indépendante, va devenir de plus en plus importante. Cette activité de découverte et d'utilisation des ressources peut être scindée en ces activités où une des principales contraintes est l'utilité et la facilité humaine et celles dans lesquelles l'implication humaine est faible ou inexistante. La désignation humaine doit avoir des caractéristiques qui soient à la fois mémorisables et courtes. Les humains, à la différence des ordinateurs, sont doués pour lever les ambiguïtés heuristiques et pour une grande variabilité des structures. Pour que les systèmes fondés sur les ordinateurs et le réseau prennent en charge une désignation globale et l'accès à des ressources qui ont peut-être une durée de vie indéterminée, la flexibilité et la non fiabilité latente des noms faciles à retenir pour les humains devrait être traduite dans une infrastructure de dénominations plus appropriée pour le système support sous-jacent. C'est à ce système de support sous-jacent que s'adresse l'architecture d'infrastructure d'information de l'Internet (IIIA, *Internet Information Infrastructure Architecture*).

Au sein de l'IIIA, plusieurs sortes d'informations sur les ressources sont spécifiées et divisées parmi différentes sortes de structures, selon des lignes fonctionnelles. Afin d'accéder à des informations, on doit être capable de découvrir ou identifier les informations particulières désirées, déterminées à la fois par comment et par où elles peuvent être utilisées ou accédées. La partition des fonctions dans cette architecture est en noms de ressource uniformes (URN), caractéristiques de ressource uniformes (URC), et localisateurs de ressource uniformes (URL). Un URN identifie une ressource ou unité d'information. Il peut identifier, par exemple, un contenu intellectuel, une présentation particulière d'un contenu intellectuel, ou tout ce qu'une autorité d'allocation de nom a déterminé comme étant une entité nommable distincte. Un URL identifie la localisation ou un contenant d'une instance d'une ressource identifiée par un URN. La ressource identifiée par un URN peut résider dans une ou plusieurs localisations à tout moment, peut bouger, ou peut n'être pas disponible du tout. Bien sûr, toutes les ressources ne vont pas bouger pendant la durée de leur vie, et toutes les ressources, bien qu'identifiables et identifiées par un URN seront instanciées à tout moment. À ce titre, un URL identifie un lieu où peut résider une ressource, ou un contenant, distinct de la ressource elle-même identifiée par l'URN. Un URC est un ensemble d'informations de niveau inférieur sur une ressource. Des exemples de telles méta informations sont : le propriétaire, le codage, les restrictions d'accès (peut-être pour des instances particulières), le coût.

Cela étant dit, on peut faire les déclarations suivantes :

- o l'objet, ou la fonction d'un URN est de fournir un identifiant unique au monde et persistant qui est utilisé pour la reconnaissance, pour l'accès aux caractéristiques de la ressource ou pour l'accès à la ressource elle-même.

Plus précisément, il y a deux sortes d'exigences pour les URN : des exigences sur les capacités fonctionnelles des URN, et des exigences sur la façon dont les URN sont codés en flux de données et en communications écrites.

2. Exigences sur les capacités fonctionnelles

Voici les exigences pour les capacités fonctionnelles des URN :

- o Une portée mondiale : un URN est un nom dont la portée mondiale n'implique pas une localisation. Il a la même signification partout.
- o Unicité mondiale : le même URN ne sera jamais alloué à deux ressources différentes.
- o Persistance : la durée de vie d'un URN est destinée à être permanente. C'est-à-dire que l'URN sera unique au monde pour toujours, et peut très bien être utilisé comme référence à une ressource bien au delà de la durée de vie de la ressource qu'il identifie ou de toute autorité de dénomination impliquée dans l'allocation de son nom.
- o Adaptabilité : les URN peuvent être alloués à toute ressource qui peut être disponible sur le réseau, pour des centaines d'années.
- o Prise en charge des systèmes traditionnels : le schéma doit permettre la prise en charge des systèmes existants de dénomination, dans la mesure où ils satisfont aux autres exigences décrites ici. Par exemple, les numéros ISBN, les identifiants publics ISO, et les codes de produit UPC semblent satisfaire aux exigences fonctionnelles, et permettre une incorporation qui satisfasse aux exigences syntaxiques décrites ici.
- o Extensibilité : tout schéma pour les URN doit permettre les extensions futures du schéma.
- o Indépendance : Il est de la seule responsabilité de l'autorité qui produit un nom de déterminer les conditions sous lesquelles elle produit un nom.
- o Résolution : un URN n'empêchera pas la résolution (traduction en un URL, à volonté). Pour être plus précis, pour les URN qui ont des URL correspondants, il doit y avoir un mécanisme possible de traduction de l'URN en URL.

3. Exigences pour le codage des URN

En plus des exigences sur les éléments fonctionnels des URN, il y a des exigences sur la façon dont ils sont codés dans une chaîne :

- o Un seul codage : le codage pour la présentation aux gens en texte en clair, en messagerie électronique et ainsi de suite, est le même que le codage dans les autres transmissions.
- o Comparaison simple : un algorithme de comparaison pour les URN est simple, local, et déterministe. C'est-à-dire qu'il y a un seul algorithme pour comparer deux URN, qui n'exige pas de contacter un serveur externe, qui est bien spécifié et simple.
- o Transcrivable en langage humain : pour que les URN soient facilement transcrivables sans erreur par l'homme, ils devraient être courts, utiliser un minimum de caractères spéciaux, et être insensibles à la casse. (Il n'y a pas d'exigence forte qu'il soit facile à l'homme de générer ou d'interpréter un URN ; que la sémantique des noms soit explicitement accessible à l'homme n'est pas une exigence.) Pour cette raison, la comparaison des URN n'est pas sensible à la casse, ni probablement aux espaces et à certaines marques de ponctuation.
- o Facilité du transport : un URN peut être transporté sans modification dans les protocoles courants de l'Internet, tels que TCP, SMTP, FTP, Telnet, etc., ainsi qu'imprimés sur papier.
- o Traitable en machine : un URN peut être analysé par un ordinateur.
- o Reconnaissance du texte : le codage d'un URN devrait améliorer la capacité à trouver et analyser les URN en texte libre.

4. Implications

Pour qu'une spécification d'URN soit acceptable, elle doit satisfaire aux exigences qui précèdent. On tire un ensemble de conclusions de ces exigences, dont la liste figure ci-dessous ; une spécification qui satisfait aux exigences sans satisfaire à ces conclusions est réputée acceptable, bien qu'il soit peu probable que cela se produise.

- o Pour satisfaire aux exigences d'unicité et d'adaptabilité, l'allocation des noms est déléguée aux autorités de désignation, qui peuvent donc allouer les noms directement ou déléguer cette autorité à des autorités subalternes. L'unicité est garantie par l'exigence que chaque autorité de dénomination garantisse l'unicité. Les noms des autorités de désignation elles-mêmes sont persistants et uniques au monde et les autorités de niveau supérieur seront enregistrées centralement.

- o Les autorités de désignation sont encouragées à prendre en charge l'adaptabilité des désignations mais elles n'y sont pas obligées. L'adaptabilité implique que le schéma pour l'attribution des noms soit adaptable à la fois pour ses terminaisons et pour sa structure ; par exemple, dans un schéma de désignation hiérarchique, une autorité de désignation peut avoir un mécanisme d'extension pour ajouter de nouveaux sous-registres.
- o Il est vivement recommandé qu'il y ait une transposition entre les noms générés par chaque autorité de désignation et les URL. Il y aura à tout moment zéro, un ou plusieurs URL dans lesquels un URN particulier peut être transposé. L'autorité de désignation n'a pas besoin de fournir elle-même la transposition d'URN en URL.
- o Pour que les URN soient transcrivables et transportés dans la messagerie, il est nécessaire de limiter le jeu de caractères utilisable dans les URN, bien qu'il n'y ait pas encore de consensus sur quelle pourrait être la limite.

Dans l'allocation des noms, une autorité d'allocation des noms doit respecter les contraintes ci-dessus, ainsi que définir ses propres critères pour déterminer la nécessité ou l'indication d'une nouvelle allocation de nom.

5. Autres considérations

Il y a encore trois autres questions sur lesquelles le présent document n'a intentionnellement pas pris position, parce que on estime que ces questions doivent être réglées par une décision locale ou par d'autres services au sein d'une infrastructure d'informations. Ces questions sont l'égalité des ressources, le reflet de la sémantique visible dans un URN, et la résolution des noms.

Une des façons dont les autorités de désignation, qui allouent les noms, peuvent choisir de se distinguer elles-mêmes, est par les algorithmes par lesquels elles distinguent ou non des ressources les unes des autres. Par exemple, un éditeur peut choisir de distinguer parmi plusieurs éditions d'un ouvrage, dans lesquelles des erreurs typographiques et d'orthographe mineures ont été commises, mais une bibliothèque peut préférer ne pas faire cette distinction. De plus, aucun algorithme pour vérifier l'égalité ne sera vraisemblablement applicable à toutes ces sortes d'informations. Par exemple, un algorithme fondé sur la vérification d'égalité de deux livres ne sera sans doute d'aucune utilité pour vérifier l'égalité de deux tableaux. Donc, bien que le présent document exige que toute autorité de désignation particulière utilise un algorithme pour déterminer si deux ressources sont les mêmes ou sont différentes, chaque autorité de désignation peut utiliser un algorithme différent et une autorité de désignation peut restreindre l'ensemble de ressources qu'elle choisit d'identifier.

Une autorité de désignation aura aussi un algorithme pour en fait choisir un nom au sein de son espace de noms. Elle peut avoir un algorithme qui incorpore en fait certaines informations sur la ressource. Par ailleurs, cette incorporation peut être rendue publique ou non, et peut être ou non visible aux clients potentiels. Par exemple, un URN non réfléchi va simplement donner des numéros de série à accroissement monotone aux ressources. Cela ne porte rien d'autre que l'identité déterminée par l'algorithme de vérification d'égalité et un ordre d'allocation des noms par ce serveur. Cela ne porte pas d'information sur la ressource elle-même.

Un algorithme de résumé de message MD5 de la ressource peut dans une certaine mesure, dans et par lui-même refléter son contenu, et en fait, l'autorité de désignation peut être parfaitement désireuse de publier le fait qu'elle utilise MD5, mais si la ressource est susceptible d'être déplacée, il sera toujours vrai que tout client potentiel ne pourra pas faire beaucoup plus avec l'URN que de vérifier son égalité. Si, par contre, un schéma d'URN a beaucoup en commun avec l'allocation des numéros ISBN, l'algorithme pour les affecter est public et en le connaissant, sachant un numéro ISBN particulier, on peut apprendre quelque chose de plus sur la ressource en question. Cette gamme complète de possibilités est permise selon le présent document sur les exigences, bien qu'il soit déconseillé aux autorités de désignation de rendre accessible aux clients les informations de sémantique sur la ressource, dans l'hypothèse qu'elle pourrait changer avec le temps et donc qu'il n'est pas sage d'encourager de quelque façon que ce soit les gens à croire que cette sémantique serait toujours valide.

Enfin, le présent document ne traite intentionnellement pas du problème de la résolution des noms, autrement que pour recommander qu'il existe un mécanisme de traduction de nom pour chaque autorité de désignation. Les autorités de désignation allouent des noms, alors que les résolveurs ou les services de localisation aident d'une certaine façon ou fournissent la transposition des URN en URL. Il peut y avoir un ou plusieurs de ces services pour la ressource désignée par une autorité de désignation particulière. Il peut aussi se trouver qu'il y en a de génériques qui fournissent le service pour de nombreuses ressources de différentes autorités de désignation. Certaines peuvent être d'autorité et d'autres non. Certaines peuvent être très fiables ou très disponibles ou très rapides à se mettre à jour, ou très centrées sur d'autres critères tels que les sujets concernés. Bien sûr, il est aussi possible que certaines autorités de désignation agissent aussi comme résolveurs pour les ressources qu'elles ont désignées. Le présent document soutient et encourage les services de tiers et les services distribués dans ce domaine, et ne fait donc intentionnellement aucune déclaration sur les exigences pour les URN ou les autorités de désignation en matière de résolution de noms.

Considérations pour la sécurité

Les applications qui requièrent la traduction des noms en localisations, et les ressources elles-mêmes, peuvent exiger que les ressources soient authentifiées. Il semble généralement que les informations sur l'authentification du nom ou de la ressource à laquelle il se réfère devraient être portées dans des informations séparées passées avec l'URN plutôt que dans l'URN lui-même.

Adresse des auteurs

Larry Masinter
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
téléphone : (415) 812-4365
Fax: (415) 812-4333
mél : masinter@parc.xerox.com

Karen Sollins
MIT Laboratory for Computer Science
545 Technology Square
Cambridge, MA 02139
téléphone : (617) 253-6006
Fax : (617) 253-2673
mél : sollins@lcs.mit.edu