

Petit guide de
l'acceptation universelle

L'« acceptation universelle », c'est quoi ?



ACCEPTATION



VALIDATION



STOCKAGE



TRAITEMENT



AFFICHAGE

Les logiciels et services en ligne soutiennent l'acceptation universelle lorsqu'ils offrent les fonctionnalités susmentionnées pour tous les noms de domaine et e-mails.

L'acceptation universelle (UI) correspond à l'état dans lequel tous les noms de domaine et adresses e-mail valides sont acceptés, validés, stockés, traités et affichés correctement et de façon cohérente par l'ensemble des applications, dispositifs et systèmes Web.

En raison de l'évolution rapide du paysage des noms de domaine, de nombreux systèmes ne reconnaissent pas ou ne traitent pas adéquatement les nouveaux noms de domaine, principalement car ils ne peuvent qu'ils aient plus de trois caractères ou se trouvent sous un format non ASCII. Il en va de même pour les adresses e-mail qui incorporent ces nouvelles extensions.

Le groupe directeur sur l'acceptation universelle (UASG), créé par la Société pour la distribution des noms de domaine et des numéros sur Internet (ICANN), constitue une initiative communautaire à l'échelle de l'industrie visant à sensibiliser et à identifier et résoudre les problèmes associés à l'acceptation universelle des noms de domaine. Le but de cette initiative est d'aider à garantir une expérience uniforme et positive pour les internautes à travers le monde.

Pour de plus amples informations sur l'UASG et ses récentes évolutions, consultez : www.uasg.tech.

ACCEPTATION



L'acceptation est le processus via lequel une adresse e-mail ou un nom de domaine est reçu en tant que chaîne de caractères d'une interface utilisateur, d'un dossier ou d'une API (interface de programmation d'application) utilisé par une application logiciel ou un service en ligne.

Recommandations de l'UASG

- Tout élément d'une interface utilisateur imposant à un utilisateur de saisir un nom de domaine ou une adresse e-mail doit prendre en charge le format Unicode et des chaînes allant jusqu'à 256 caractères.
- Les utilisateurs doivent être autorisés, mais non obligés, à saisir du texte à encodage compatible ASCII (soit « Punycode ») au lieu de son équivalent Unicode. Toutefois, Unicode doit être affiché par défaut, du texte en Punycode devant être affiché à l'utilisateur uniquement **lorsque cela est avantageux**



VALIDATION



Il s'agit du processus de vérification de la syntaxe d'une adresse e-mail ou d'un nom de domaine, reçu ou émis. De nombreux programmeurs ont été formés afin de procéder à la validation à l'aide d'heuristiques consistant à vérifier qu'un domaine de premier niveau a le « bon » nombre de lettres ou que les lettres proviennent bien de l'ensemble de caractères ASCII. Ces heuristiques ne peuvent plus être utilisées de par l'introduction de noms de domaine de plus de trois caractères et de caractères Unicode (non ASCII).

Recommandations de l'UASG


- Il ne doit être procédé aux validations que si cela est nécessaire pour l'exploitation de l'application ou du service. C'est la façon la plus simple de garantir que tous les noms de domaine valides sont acceptés dans les systèmes.
- Si la validation est requise, examiner les éléments suivants :
 - ▶ Vérifier la partie TLD d'un nom de domaine par rapport à un tableau faisant autorité :
<http://www.internic.net/domain/root.zone>
<http://www.dns.icann.org/services/authoritative-dns/index.html>
<http://data.iana.org/TLD/tlds-alpha-by-domain.txt>
Voir aussi le SAC070 : <https://tinyurl.com/sac070>.
 - ▶ Interroger le nom de domaine par rapport au système des noms de domaine (DNS).
 - ▶ Exiger qu'une adresse e-mail soit répétée plusieurs fois afin d'éviter les erreurs de frappe.
 - ▶ Valider les caractères sur les étiquettes uniquement afin de déterminer que l'étiquette U ne contient pas de points de code « NON AUTORISÉS » ou points de code qui ne sont pas attribués dans leur version d'Unicode. Consulter : <https://tools.ietf.org/html/rfc5892>.
 - ▶ Limiter la validation d'étiquettes à un nombre limité de règles s'appliquant à toutes les étiquettes définies dans les appels à commentaires (RFC). Consulter : <https://tools.ietf.org/html/rfc5894>.
 - ▶ Si une chaîne ressemblant à un nom de domaine contient le caractère point idéographique « 。

STOCKAGE



Le stockage fait référence au stockage à long terme et/ou temporaire de noms de domaine et adresses e-mail. Indépendamment de la durée de vie des données, elles doivent être stockées sous des formats définis par le RFC (de préférence) ou d'autres formats capables de se transformer en formats définis par le RFC.

Recommandations de l'UASG

- Les applications et services doivent répondre aux normes Unicode en vigueur.
- Les informations doivent être stockées sous le format UTF-8 (Format de transformation Unicode) autant que possible. Il est possible que certains systèmes exigent de stocker également sous le format UTF-16, mais en règle générale le format UTF-8 est préféré. Les formats UTF-7 et UTF-32 doivent être évités.
- Envisager tous les scénarios dans leur globalité avant de convertir des étiquettes A en étiquettes U et vice versa lors du stockage. Il peut être souhaitable de ne conserver que les étiquettes U dans un dossier ou une base de données, cela simplifiant la recherche et le tri.  Parfois, la conversion peut avoir des conséquences lors de l'interopération avec des applications et services plus anciens ne respectant pas le format Unicode. Envisager de stocker sous les deux formats.
- Indiquer clairement les adresses e-mail et noms de domaine lors du stockage afin d'y accéder plus facilement. Les cas où les adresses e-mail et noms de domaine ont été déposés en utilisant le champ « auteur » d'un document ou « info contact » d'un journal ont entraîné la perte d'origine en tant qu'adresse.





Le traitement intervient lorsqu'une adresse e-mail ou un nom de domaine est utilisé par une application ou un service afin de mener une activité (par exemple la recherche ou le tri d'une liste) ou est transformé en un format distinct (par exemple le stockage ASCII en tant qu'Unicode). Une validation supplémentaire peut également être réalisée lors du traitement.

Il existe un nombre illimité de méthodes de traitement* des noms de domaine et adresses e-mail, ce qui renforce la nécessité de disposer de conventions veillant à ce que les données soient bien comprises et classées uniformément.

Recommandations de l'UASG

- Étant donné que la norme Unicode est en perpétuelle évolution, les points de code non définis lors de la création de l'application ou du service doivent être vérifiés afin de veiller à ce qu'ils ne « cassent » pas l'expérience de l'utilisateur. Les polices qui manquent dans le système d'exploitation sous-jacent peuvent conduire à des caractères non affichables (le caractère « 0 » est souvent utilisé pour les représenter) mais cette situation ne doit pas entraîner d'incident fatal.
- Utiliser les API sous format Unicode prises en charge.
- Utiliser les documents relatifs au protocole [<http://tools.ietf.org/html/rfc5891>] et aux tableaux [<http://tools.ietf.org/html/rfc5892>] portant sur les noms de domaine internationalisés dans les applications (IDNA) pour les noms de domaine internationalisés (IDN).
- Traiter autant que possible sous le format UTF-8.
- Veiller à ce que le produit ou la fonctionnalité en charge les chiffres tel que prévu. Par exemple, les chiffres ASCII et les représentations de nombres idéographiques asiatiques doivent être traités en tant que nombres. [[RFC5892](http://tools.ietf.org/html/rfc5892), lien ci-dessous]
- Mettre à jour les applications et serveurs/services ensemble. Si le serveur est Unicode et que le client ne l'est pas ou vice versa, les données devront être converties sur chaque page de code chaque fois que les données se déplacent entre le serveur et le client.
- Effectuer des révisions de code afin d'éviter les attaques liées au dépassement de la mémoire tampon. Lors de la transformation de caractères, il se peut que des chaînes de texte grandissent ou rétrécissent considérablement.

Exemples : Identifier des individus en Nouvelle-Zélande en cherchant le ccTLD .nz ; identifier des pharmaciens en cherchant des adresses e-mail utilisateur@.pharmacien.




AFFICHAGE



L'affichage intervient dès qu'une adresse e-mail ou un nom de domaine est délivré à l'intérieur d'une interface utilisateur. L'affichage des noms de domaine et adresses e-mail est normalement simple lorsque les scripts utilisés sont pris en charge dans le système d'exploitation sous-jacent et que les chaînes sont stockées sous format Unicode ; toutefois, des transformations propres aux applications peuvent être **autrement requises**.



Recommandations de l'UASG

- Afficher tous les points de code Unicode qui sont pris en charge par le système d'exploitation sous-jacent. Si une application conserve ses propres ensembles de polices, une prise en charge Unicode complète doit être offerte à l'ensemble de polices proposé par le système d'exploitation.
- Lors du développement d'une application ou d'un service, ou lors de l'exploitation d'un registre, passer en revue les langues prises en charge et s'assurer que le système d'exploitation et les applications prennent en charge ces langues.
- Convertir les données non Unicode en données Unicode avant l'affichage. Par exemple, l'utilisateur final doit voir « everyone.みんな » et non pas « everyone.xn--q9jyb4c ». (Cette conversion est un exemple de traitement prenant en compte l'UA).
- Afficher Unicode par défaut. Utilisez  texte en Punycode pour l'utilisateur uniquement lorsque cela **est avantageux**. Augmenter l'affichage Unicode avec du texte infobulle en Punycode en tant que réduction.
- Garder à l'esprit que les adresses à plusieurs scripts deviendront plus courantes. Il se peut que certains caractères Unicode se ressemblent à l'œil humain mais soient différents pour les ordinateurs. Ne pas présumer que les chaînes à plusieurs scripts sont utilisées à des fins malveillantes, par exemple pour le hameçonnage, et si l'interface utilisateur attire l'attention de l'utilisateur sur les chaînes, s'assurer qu'elle le fait sans porter préjudice aux utilisateurs de scripts non latins. **En savoir  davantage** sur les impératifs de sécurité Unicode : <http://unicode.org/reports/tr36/>.
- Utiliser un traitement de compatibilité IDNA Unicode afin de satisfaire les attentes de l'utilisateur. Pour en savoir plus, consultez l'adresse suivante : <http://unicode.org/reports/tr46/>.
- Prendre connaissance des caractères non attribués et non autorisés. **En savoir  davantage** sur le RFC 5892 ici : <https://tools.ietf.org/rfc/rfc5892.txt>.




Devenir un pro de l'acceptation universelle


Révision du code source et test de l'unité


Il s'agit du processus consistant à contrôler le code source et à vérifier que seules les techniques de programmation, les bibliothèques logicielles et les interfaces (les « API ») adéquates ont été utilisées. Après cela, l'administrateur peut vérifier que l'application ou le service fonctionne correctement en testant les fonctionnalités spécifiques (acceptation, validation, etc.) susmentionnées. Cette méthode est en règle générale uniquement utilisée par des développeurs d'applications et des fournisseurs de services en ligne.

Dans le cadre des initiatives de sensibilisation de l'UASG, le groupe cible directement les développeurs d'applications et les plus gros fournisseurs de services en ligne afin de les encourager à effectuer des révisions du code source d'acceptation

universelle et un test et à partager une liste de critères pouvant être utilisés afin de développer le recours au test. 

Test manuel

Il convient d'effectuer de multiples tests par rapport aux nouveaux domaines et aux domaines non ASCII, tels que soumettre une adresse e-mail lors de l'enregistrement d'un service en ligne et vérifier qu'elle a été acceptée. Étant donné qu'il existe un nombre incalculable de services en ligne potentiels auxquels il est possible de souscrire ainsi que de potentielles nouvelles combinaisons d'adresses e-mail, cette méthode implique d'essayer différentes combinaisons d'applications, de services, d'adresses e-mail et/ou de noms de domaine afin de couvrir une large gamme de cas d'utilisation. Cette méthode peut être suivie par quiconque mais il s'agit de la méthode à plus forte intensité en main-d'œuvre. 

L'UASG aide également à assurer la promotion de cette méthode en développant une liste de sites Web, applications, adresses e-mail et noms de domaine particulièrement adaptés pour le test. 

Test automatisé

Il convient d'utiliser des scripts ou des directives automatisés afin de tester une variété d'URL. Cette méthode implique qu'un gros travail technique soit mené en amont mais elle est plus susceptible d'être reproduite par d'importantes initiatives en termes de mesure ou de suivi. Voici un exemple concret d'enquête sur les gTLD récemment menée par l'APNIC au nom de l'ICANN : <https://tinyurl.com/new-gtld-ua>.

L'UASG cherche actuellement des méthodes de test automatisé pour l'acceptation universelle et partagera ses conclusions dès qu'elles seront disponibles.

