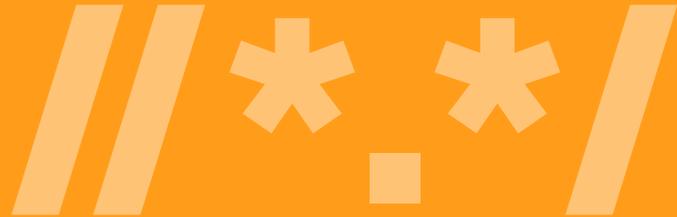


الدليل السريع
للقبول الشامل



ما الذي يعنيه "القبول الشامل"؟

القبول



التصديق



التخزين



المعالجة



العرض



تدعم البرمجيات والخدمات المقدمة عبر الإنترنت القبول الشامل عندما يقوموا بعرض القدرات المدرجة في أعلاه لكافة أسماء النطاقات والبريد الإلكتروني.

يتناول هذا الدليل السريع توصيات المجموعة التوجيهية للقبول الشامل UASG من أجل تحقيق القبول الشامل في خمسة سياقات—القبول والتصديق والتخزين والمعالجة والعرض—التي تتعامل فيها النظم مع أسماء النطاقات وعناوين البريد الإلكتروني. وهي مخصصة للمديرين التنفيذيين والمديرين المسؤولين عن تقنية المعلومات أو أنشطة هندسة المنتجات البرمجية. وهو يستعرض توصيات المجموعة التوجيهية للقبول الشامل UASG في المستوى العالي بدون الخوض ببعض التفاصيل التي يمكن أن تكون مفيدة لأي مصمم أو مهندس برمجيات. للحصول على تلك التفاصيل، راجع UASG 007، "مقدمة إلى القبول الشامل".

إن نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII هو ترميز الحروف المستخدم تاريخياً على الإنترنت، والمعروف في معيار الإنترنت RFC 20 على الرابط <https://tools.ietf.org/html/rfc20>. يُعرّف الكود الموحد (الوينيكود) من خلال اتحاد الكود الموحد (الوينيكود) (<http://unicode.org>).



توصيات المجموعة التوجيهية للقبول الشامل UASG

يجب أن تكون حقول الإدخال أكبر بما يكفي لقبول أي مدخلات صحيحة. واستناداً إلى طريقة ترميزها، قد يتطلب أي اسم نطاق عدد بايت يصل إلى 670. وقد يكون لأي عنوان بريد إلكتروني جزءاً محلياً (الجزء الذي يسبق علامة @) الذي ينطوي على ما يصل إلى 64 بايت بالإضافة إلى اسم نطاق بطول إجمالي يصل إلى 735 بايت.

ويجب أن تقبل التطبيقات والخدمات أسماء النطاقات وعناوين البريد الإلكتروني برموز نظام UTF-8²، كما يجب أن تميز بأن عدد البايت المشغولة من قبل ترميز UTF-8 قد يكون أكبر من عدد الحروف المعروضة.

ويمكن إدخال اسم نطاق منوّل وعرضه سواء بنصه الأصلي أو بإصدار ASCII المحدد للتوافق الخلفي، على سبيل المثال، 测试 وأيضاً xn--0zwm56d. ويطلق على ترميز الكود الموحد (اليونيكود) للنص الأصلي رمز-U؛ أما الترميز المتوافق مع ASCII المكافئ فيطلق عليه اسم رمز-A³. ويجب أن تقبل البرمجيات كل من رموز-A ورموز-U، ولكن يجب أن تقوم بتحويل رموز-A إلى رموز-U لأجل العرض وأيضاً لأي معالجة لا تتطلب رموز-A.

وتقريباً في أغلب الحالات يجب تحويل اسم النطاق أو عنوان البريد الإلكتروني الذي تم إدخاله إلى نموذج C لتطبيع اليونيكود⁴ (NFC) قبل إجراء المزيد من المعالجات عليه. ونظراً لأن معيار NFC ليس شامل بشكل تام، ففي ظروف نادرة قد يكون من الضروري تأجيل التطبيع إلى أن تقوم معالجة أخرى بإقرار السياقات (السياقات) النوعية والتي يجب تطبيق المعيار وفق ذلك السياقات (أو السياقات).

القبول هو العملية التي يتم من خلالها الحصول على اسم نطاق أو عنوان بريد إلكتروني من واجهة مستخدم أو ملف أو API (واجهة برمجة التطبيقات) لاستخدامه من خلال إحدى التطبيقات البرمجية أو من خدمة متوفرة على الإنترنت.

²ويقوم UTF-8 بترميز كل نقطة كود موحد (يونيكود) كمتتالية من بايت واحد إلى أربعة بايت، وهذا محدد في المعيار RFC 3629.

³والتحويل بين رموز-A ورموز-U يكون مصحوباً بخوارزمية Punycode (يونيكود) المعرفة في المعيار RFC 3492 والمعيار RFC 5891.

⁴اطلع على الملحق 15 لمعيار اليونيكود، "مادج تطبيع اليونيكود" (<https://www.unicode.org/reports/tr15/tr15-47.html>).



توصيات المجموعة التوجيهية للقبول الشامل UASG

يجب تصديق (التحقق من صحة) المدخلات بطريقة مناسبة للغرض منها. ويجب تصديق جميع أسماء النطاقات في مقابل أسماء النطاقات المدوّلة في معيار التطبيقات، وهو حالياً IDNA2008⁵. وهذا من شأنه أن يضمن أن الاسم صحيح من الناحية التركيبية.

وإذا كان من المتوقع لسلسلة إدخال أن تكون إدخالاً حالياً في نظام أسماء النطاقات، فقم بالتحقق من صحتها مستخدماً بحث DNS.

وإذا كان من المتوقع لسلسلة إدخال أن تكون عبارة عن اسم نطاق نافذ قد لا يكون في نظام أسماء النطاقات DNS (لغاية اللحظة)، فإنه قد لا يزال من الممكن التحقق من صحة جزء منها. على سبيل المثال، يمكن التحقق من اسم نطاق المستوى الأعلى (TLD) في مقابل قائمة معتمدة من أسماء TLD الصحيحة التي تحتفظ بها هيئة الإنترنت للأرقام المخصصة⁶ (IANA).

للتحقق من صحة عنوان بريد إلكتروني، يجب تصديق جزء النطاق الموصوف أعلاه. لأن الجزء المحلي لعنوان البريد الإلكتروني يُحدد فقط من خلال نظام البريد الذي يتلقى الرسائل، وعموماً ليس من الممكن تصديقه. مطالبة المستخدم بإدخال عنوان البريد الإلكتروني مرتان قد يكشف أخطاء في الكتابة.

وفي أغلب الأحوال، جميع مكونات أي اسم من أسماء النطاقات أو أي عنوان من عناوين البريد الإلكتروني (باستثناء اسم TLD إن لم يكن اسم نطاق مدوّل) يجب أن تكون برموز حروف لغة واحدة (مثل رموز حروف اللغة العربية أو لغة الهان) أو رموز حروف اللغات القريبة منها (على سبيل المثال كانجي اليابانية وكتاكانا وهيراجانا وروماجي). استخدم معيار اليونيكود الفني رقم 39، "البيات أمن اليونيكود" (https://www.unicode.org/reports/tr39/#Restriction_Level_Detection)، للتحقق من أن رموز الحروف في أي متتالية يونيكود تتبع التطبيق الصحيح.

التصديق وهو عملية فحص عنوان البريد الإلكتروني أو اسم النطاق للتأكد من التركيب الصحيح ومن أن الاسم الذي من المتوقع وجوده في نظام اسم النطاق هو موجود بالفعل، متى ما كان ذلك مناسباً. وقد تكون تقنيات التصديق بحاجة للتحديث لأجل ضمان التفاعل مع أسماء النطاقات وعناوين البريد الإلكتروني الحديثة.

⁵راجع معايير RFC رقم 5890 و5891 و5892 و5893 و5894 للحصول على تعريف IDNA2008.

⁶اطلع على "قائمة نطاقات المستوى الأعلى" (<https://www.icann.org/resources/pages/tlds-2012-02-25-en>).



توصيات المجموعة التوجيهية للقبول الشامل UASG

- تقريباً في أغلب الحالات يجب تطبيع أسماء النطاقات أو عناوين البريد الإلكتروني وفقاً لنموذج C لتطبيع اليونيكود (NFC) قبل التخزين. ونظراً لأن معيار NFC ليس شاملاً بشكل تام، ففي ظروف نادرة قد يكون من الضروري تأجيل التطبيع إلى أن تقوم معالجة أخرى بإقرار السياق (السياقات) النوعية والتي يجب تطبيق المعيار وفق ذلك السياق (أو السياقات).
- وفي أغلب التطبيقات، يجب تخزين أسماء النطاقات وعناوين البريد الإلكتروني في ملفات وقواعد بيانات يتم ترميزها وفق صيغة UTF-8، وهو ترميز الكود الموحد (اليونيكود) المعروف على نطاق واسع والمدعوم على نحو أفضل. وفي بعض الحالات التي يفترض فيها أن تعمل البرمجية بشكل متبادل مع قواعد البيانات القديمة، قد يكون من الأسهل استخدام نفس الترميز كقاعدة بيانات.
- وضمن كود التطبيق، يعتمد التمثيل الأنسب لليونيكود على بيئة البرمجة. وتحتوي العديد من لغات البرمجة بما في ذلك لغتي البايثون والبيرل على دعم مضمن لليونيكود والتحويل التلقائي إلى UTF-8 أو منه، عند الإدخال والإخراج.
- ويجب أن تختار التطبيقات تمثيلاً داخلياً متسقاً—سواء برموز U أو رموز A—لأسماء النطاقات المدوّلة IDN. لأن كل رمز U يمكن تحويله إلى رمز A فريد والعكس بالعكس، فأياً من الصيغتين هي مقبولة.

ويشير التخزين إلى التخزين المؤقت أو طويل الأجل لأسماء النطاقات وعناوين البريد الإلكتروني، والتي يجب تخزينها بصيغ محددة بشكل جيد بصرف النظر عن المدة المتوقعة للتخزين.



توصيات المجموعة التوجيهية للقبول الشامل UASG

مع تطور الكود الموحد (اليونيكود)، يجب ترقية (تحسين) البرمجيات عندما يكون من العملي القيام بذلك لاستخدام أحدث إصدار من المعيار وأي رسومات وخطوط متاحة. ويجب مراعاة أن أجهزة المستخدم ومكتبات البرمجيات ومعايير الويب قد لا تدعم أحدث إصدار، ومن ثم قد تعرض حروفاً خصصت حديثاً بشكل غير صحيح، في صورة مربع عام (□)، أو عدم عرضها على الإطلاق.

وعندما تتوافر واجهات برمجة التطبيقات التي تدعم مدخلات ومخرجات UTF-8، استخدمها بدلاً من واجهات برمجة التطبيقات التي لا تدعم ذلك. استخدم مكتبات قياسية منزوعة العيوب البرمجية بشكل جيد، مثل GNU libidn2 في الرابط <http://www.gnu.org/software/libidn/#libidn2>، لمعالجة أسماء النطاقات المدوّلة IDN والتحقق من صحتها؛ ولا تقم بهذا العمل بنفسك".

وتتطلب النصوص التي تكتب من اليمين إلى اليسار اعتبارات خاصة عند استخدامها في أسماء النطاقات وعناوين البريد الإلكتروني. وقد تم تناول البعض من هذه الاعتبارات في أسماء النطاقات المدوّلة في التطبيقات IDNA⁷ (لأسماء النطاقات) وملحق بمعيار اليونيكود⁸ (لعناوين البريد الإلكتروني).

وعند إنشاء السجلات أو هياكل البيانات الأخرى التي تشمل معلومات النصوص أو اللغات، يجب السماح لأكبر عدد ممكن، وفي المعتاد جميع ما يدعمه معيار اليونيكود⁹. كن على علم من أن بعض اللغات يمكن كتابتها باستخدام رموز نصوص مختلفة، وأن بعض النصوص يمكن استخدامها لكتابة العديد من اللغات المختلفة.

تحدث **المعالجة** كلما تم استخدام عنوان بريد أو اسم نطاق من خلال تطبيق أو خدمة لأداء نشاط (مثل، إجراء بحث أو تصنيف قائمة)، أو كلما تم تغييره إلى صيغة بديلة (مثلاً، من صيغة ترميز قديمة إلى صيغة UTF-8). وقد يحدث تحقق اضافي من الصحة (تصديق آخر) خلال عملية المعالجة.

⁷اطلع على المعيار RFC 5893، "النصوص المكتوبة من اليمين إلى اليسار لأسماء النطاقات المدوّلة في التطبيقات (IDNA)" على <https://tools.ietf.org/html/rfc5893>.

⁸انظر UAX رقم 9، "خوارزمية ثنائية الاتجاه لليونيكود" <http://unicode.org/reports/tr9>.

⁹اطلع على "النصوص المدعومة" لليونيكود <http://unicode.org/standard/supported.html>.



توصيات المجموعة التوجيهية للقبول الشامل UASG

اعتبر أنه على الرغم من قدرة البرمجيات والأجهزة الحديثة على عرض غالبية نقاط رموز الكود الموحد (اليونيكود)، قد تحتوي الأنظمة الأقدم على دعم محدود، وتتطلب قيام التطبيقات بإدارة بعض من خطوطها القديمة. وأيضاً عندما يضيف الكود الموحد (اليونيكود) نقاط رموز جديدة، فلن تعرضها الأجهزة والتطبيقات إلى أن يتم تحديث مستودع الخطوط الخاصة بها.

اعرض أسماء النطاقات المدوّلة IDN بشكل حروف لغتهم الأصلية ما لم يكن هناك شرط خاص بعرضها كرموز-A.

ويمكن عرض أسماء النطاقات وعناوين البريد الإلكتروني في نص متجه من اليسار إلى اليمين LTR، كما في الإنجليزية والروسية، أو نص من اليمين إلى اليسار RTL، كما في العربية والعبرية. ولأن الكود الموحد Unicode يعيّن سمات الاتجاه لنقاط رمز منفردة - وليس لتسلسلات نقطة الكود - فإنه من المعقول لدى المستخدمين فيما يخص بعض النصوص المختلطة LTR و RTL ("ثنائية الاتجاه")، بينما قد لا يكون البعض الآخر كذلك. استخدم معيار مستويات قيود الكود الموحد (اليونيكود)¹⁰ لتميز السلاسل التي قد تكون مثيرة للبس.

يقراً مستخدمو الإنترنت ويتحدثون العديد من اللغات المختلفة. وفي بعض الحالات، قد يكون من الضروري تصميم تطبيقات بشكل مختلف من أجل لغات أو مجموعات مختلفة من اللغات.

يحدث العرض كلما تم تقديم عنوان بريد إلكتروني أو اسم نطاق بشكل مرئي من خلال واجهة المستخدم. عرض أسماء النطاقات وعناوين البريد الإلكتروني غالباً ما يكون بشكل مباشر عندما يتم استخدام نصوص وأي آليات إخراج مطلوبة فتكون مدعومة في نظام التشغيل الأساسي ويتم تخزين السلاسل في ترميز يحدده معيار اليونيكود. وقد تكون عمليات التحويل المقتصرة على التطبيقات ملزمة خلافاً لذلك.

¹⁰اطلع على معيار اليونيكود الفني رقم 93، "آليات أمن اليونيكود" (https://www.unicode.org/reports/tr39/#Restriction_Level_Detection)، للتعرف على مستويات الحدود المقيدة بشكل والمقيدة بشكل كبير للتحقق من أن النصوص في أي متتالية يونيكود تتبع أفضل التطبيقات.

كن جاهزاً للقبول الشامل

مراجعات الرمز الأصلي واختبار الوحدات

يجب مراجعة البرمجيات والأنظمة التي تم وضعها أو ترقيتها لدعم القبول الشامل واختبارها للتحقق من صحتها ولاكتشاف العيوب وتصحيحها. وكجزء من جهود التوعية التي تقوم بها المجموعة التوجيهية للقبول الشامل UASG، تتواصل المجموعة بشكل مباشر مع مطوري التطبيقات وموفري الخدمات على الإنترنت من أجل تشجيعهم على أداء مراجعات لأكواد مصادر القبول الشامل واختبار ومشاركة قائمة المعايير يمكن استخدامها في تطوير حالات الاختبار المعيارية.

الاختبارات

كما أن المجموعة التوجيهية للقبول الشامل تعمل على وضع قائمة بمواقع الويب والتطبيقات وعناوين البريد الإلكتروني وأسماء النطاقات المناسبة لإجراء الاختبارات. وفي بعض الحالات، يمكن جعل الاختبارات تلقائية وإدارتها بدون تدخل يدوي. مثال واقعي يتمثل في التحري الأخير عن نطاق gTLD يجريه مركز APNIC Labs بالنيابة عن ICANN: <https://tinyurl.com/new-gtld-ua>. وتتحرى المجموعة للقبول العالمي UASG عن طرق الفحص التلقائية بالنسبة للقبول العالمي وسوف تعلن عن النتائج حال توفرها.

لمزيد من القراءة

تحتوي هذه المستندات على معلومات أخرى حول القبول الشامل والكود الموحد (اليونيكود) وأسماء النطاقات المدوّلة.

UASG 007، "مقدمة إلى القبول الشامل" [. \(https://uasg.tech/documents\)](https://uasg.tech/documents)

معياري RFC 5894 "أسماء النطاقات المدوّلة في التطبيقات (IDNA): الخلفية والشرح والحيثيات" [. \(https://www.rfc-editor.org/info/rfc5894\)](https://www.rfc-editor.org/info/rfc5894)

"الطوبوغرافية الدولية على الويب"، ملخص تخطيطي يوضح المشكلات والقضايا التي تتناول مختلف اللغات على الويب [. \(https://w3c.github.io/typography/gap-analysis/language-matrix.html\)](https://w3c.github.io/typography/gap-analysis/language-matrix.html)

ملاحظة حول المصطلحات

من بين الصعوبات المصاحبة لتحقيق القبول الشامل هو أن العديد من المصطلحات والمفاهيم المعروفة لمن هم معادون على النصوص البسيطة ذات العدد الصغير من الحروف "الأبجدية" المتميزة، مثل النصوص اللاتينية، قد تؤدي إلى الكثير من حالات اللبس عند تطبيقها على أنظمة الكتابة التي تستخدم أسس مختلفة. ولقد تطلّب ضم مجموعة واسعة من أنظمة الكتابة في مجال أسماء النطاقات المدوّلة (IDN) ابتكار مصطلحات جديدة واستخدام المصطلحات الشائعة (مثل "الحروف") بطريقة خاصة ومحددة للغاية. يحاول هذا الدليل السريع تجنب مثل تلك المصطلحات أو تعريفها عند استخدامها، لكن من المرجح أن البحث في مصادر أخرى بما في ذلك بعض المستندات المشار إليها هنا، يتطلّب فهماً أعمق بالمصطلحات.

