

مقدمة حول القبول الشامل (UA)

المجموعة التوجيهية للقبول العالمي (UASG)

2019/09/23



جدول المحتويات

4	نبذة عن هذه الوثيقة
4	الجمهور المستهدف
5	مفاهيم أساسية
5	أسماء النطاقات
5	نطاقات المستوى الأعلى لرموز البلدان (ccTLD)
5	نطاقات المستوى الأعلى العامة (gTLD)
5	تدويل أسماء النطاقات
6	الحاجة إلى القبول العالمي (UA)
6	مسميات U ومسميات A
7	تدويل عناوين البريد الإلكتروني (EAI)
8	إنشاء ارتباط ديناميكي (توفير الربط)
8	الطبيعة الديناميكية لسجل منطقة الجذر
8	القبول الشامل من حيث التنفيذ
8	خمسة معايير للقبول الشامل
10	سيناريوهات المستخدم
12	عدم التوافق مع الممارسات العالمية
12	المتطلبات الفنية لجاهزية القبول الشامل
12	المتطلبات رفيعة المستوى
13	اعتبارات المطور
13	تصميم البرامج من أجل التوافق والمرونة
14	الممارسات الجيدة لتطوير وتحديث برمجيات لتحقيق الجاهزية للقبول الشامل
20	الموارد الموثوقة لأسماء النطاقات: منطقة جذر DNS وقوائم IANA
20	البريد الإلكتروني مع نطاقات IDN ولماذا ليس هو نفس EAI
21	الربط وتحدياته
21	توصيات الممارسات الجيدة
22	الكود الموحد (اليونيكود)—الخلفية وسمات نقاط الأكواد
22	UTF8 و UTF16 وغيرها من طرق التشفير
23	أسماء النطاقات المدوّلة في التطبيقات IDNA - سرد تاريخي موجز وعرض للحالة الراهنة
23	استخدام الحالات للاختبار
23	ترقية البرامج من أجل EAI
24	موضوعات متقدمة
24	النصوص المعقدة
24	اللغات المكتوبة من اليمين إلى اليسار وتوافق الكود الموحد (اليونيكود)
24	الخوارزمية ثنائية الاتجاه
25	قاعدة الاتجاه الثنائي لأسماء النطاقات
25	وحدات الوصل



26	الرموز الصوتية والحروف المماثلة
27	التطبيع وقواعد التغاضي عن حالة الأحرف وإعداد السلسلة
28	قواعد التغاضي عن حالة الأحرف والتخطيط
29	المسرد والموارد الأخرى
29	قاموس المصطلحات
32	طلبات الحصول على الملاحظات والتعليقات والمعايير الأساسية
34	المعايير الأساسية
35	الموارد على الانترنت



نبذة عن هذه الوثيقة

تشهد تقنيات الإنترنت، بما في ذلك مكونات التسمية الخاصة بها، تطورًا وتغيرًا مستمرين. ففي السنوات الأخيرة، وافقت هيئة الإنترنت للأسماء والأرقام المخصصة (ICANN) على نطاقات المستوى الأعلى الجديدة (TLD)، والتي يأتي بعضها بأحرف ASCII التقليدية وبعضها بأحرف غير ASCII (وهي أسماء النطاقات الدولية). وتشمل أمثلة ذلك .nyc، ونطاق .भारत، ونطاق .eco، ونطاق .католик. ومع ذلك، لم يتم تحديث العديد من التطبيقات والخدمات لإدارة هذه المجموعة الموسعة من نطاقات المستوى الأعلى TLD. بالإضافة إلى ذلك، تتيح معايير البريد الإلكتروني عبر الإنترنت الآن استخدام أحرف غير ASCII في عناوين البريد الإلكتروني، لذلك إلى أن تتم ترقية البرنامج، فإنه لن يتعامل مع هذه النطاقات والعناوين بشكل صحيح. وهذا يؤثر على تجربة المستخدم بعدة طرق:

- لا يتم الاعتراف بعناوين البريد الإلكتروني الصالحة أو قبولها.
- يتم التعامل مع أسماء النطاقات عن طريق الخطأ باعتبارها مصطلحات بحث في شريط العنوان الخاص بالمتصفح.

وإلى أن يتعرف البرنامج على جميع أسماء النطاقات وعناوين البريد الإلكتروني ويعالجها - وهي حالة تعرف باسم القبول العالمي (UA) - لن يكون من الممكن توفير تجربة متسقة وإيجابية لجميع مستخدمي الإنترنت. توفر هذه الوثيقة مقدمة تعريفية واسعة حول القبول العالمي والجهود المبذولة حاليًا للمساعدة في تطوير برمجيات جاهزة للقبول الشامل.

الجمهور المستهدف

يهدف هذا المستند إلى التعريف بالقبول العالمي لجمهور تقني (المطورين والمديرين والمشغلين) ممن قد يكونوا على دراية ببعض جوانب تكنولوجيا الإنترنت ولكن ليس بالضرورة تفاصيل كيفية تأثير نطاقات IDN الجديدة وأسماء النطاق وعناوين البريد الإلكتروني على الطريقة التي ينبغي بها قبولها والتحقق من صحتها وتخزينها ومعالجتها وعرضها. يمثل نقطة انطلاق للأشخاص ذوي الخلفيات التقنية المتنوعة لبدء استكشافهم لعملية القبول العالمي.



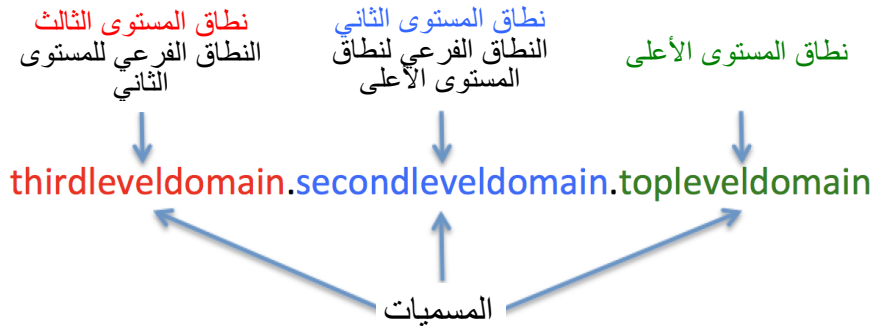
مفاهيم أساسية

أسماء النطاقات

اسم النطاق عبارة عن معرف من السهل أن تتعرف عليه أجهزة الكمبيوتر والشبكات على الإنترنت. وكثيرًا ما يتم تمثيلها في صورة سلسلة من علامات نصية تفصل بينها "نقاط" (النقطة أو علامة الوقف في آخر الجملة)؛ على سبيل المثال `www.example.tld`. وتمثل كل علامة مستوى في الترتيب الهرمي لنظام أسماء النطاقات (DNS).

ففي المستوى الأعلى أو "جذر" التسلسل الهرمي توجد تسميات نطاقات المستوى الأعلى (TLD) مثل `com` و `jp` وأيضًا، والتي تظهر في نهاية اسم النطاق. ونظرًا لأنها تظهر في النهاية، تُسمى نطاقات المستوى الأعلى TLD أحيانًا "لاحقات".

بالانتقال نزولاً في التسلسل الهرمي لنظام أسماء النطاقات من الجذر، تحدد التسمية التالية نطاقًا فرعيًا لنطاق TLD، ويُطلق عليه عادة نطاق المستوى الثاني؛ حيث تحدد التسمية بعد ذلك نطاقًا فرعيًا لنطاق المستوى الثاني، وهو ما يُطلق عليه عادةً نطاق المستوى الثالث؛ وهلم جرا، مع فصل كل علامة عن جارتها بنقطة. على سبيل المثال، قد يبدو اسم نطاق ذي ثلاثة مستويات كما يلي:



نطاقات المستوى الأعلى لرموز البلدان (ccTLD)

يتم تفويض بعض نطاقات TLD لبلدان أو أقاليم محددة. ويطلق عليها اسم نطاقات المستوى الأعلى لرموز البلدان (ccTLD). وقديمًا، كانت جميع نطاقات المستوى الأعلى لرموز البلدان ccTLD مكونة من حرفين يطابقان رمز ISO 3166 ثنائي الأحرف والذي يتم تعيينه للبلد أو الإقليم بمعرفة المنظمة الدولية للمعايير. ومنذ عام 2010، كانت هناك أيضًا نطاقات ccTLD مدوّلة تمثل اسم دولة أو إقليم بنفس النص الخاص بذلك البلد أو الإقليم.

نطاقات المستوى الأعلى العامة (gTLD)

تسمى معظم نطاقات المستوى الأعلى (TLD) التي ليست نطاقات ccTLD نطاقات مستوى أعلى عامة (gTLD)، والتي تكون إما متاحة لأي شخص أو مقيدة بأعضاء مجتمع محدد. وهي تشمل النطاقات المشهورة `com` و `net` و `org`. بالإضافة إلى بعض الإضافات الأخيرة.

فمن خلال برنامج نطاقات gTLD الجديدة – وهو عبارة عن مبادرة تنسقها مؤسسة ICANN - حقق نظام أسماء النطاقات (DNS) توسعًا كبيرًا من خلال طرح نطاقات المستوى الأعلى العامة الجديدة. ويمكن أن تمثل نطاقات gTLD الجديدة هذه علامات تجارية ومجتمعات ذات اهتمام ومناطق جغرافية (المدن والمناطق) وغير ذلك الكثير.

تدويل أسماء النطاقات

كانت أسماء النطاقات مقتصرة في الأصل على مجموعة فرعية من أحرف ASCII (الحروف من a-z والأرقام من 0-9 والوصلة "-"). ومنذ تسجيل نطاقات `com` الأولى، أي نطاق `symbolics.com` في عام 1985، توسع عدد وخصائص أسماء النطاقات ليعكس احتياجات الاستخدام العالمي دائمة الزيادة للإنترنت باعتباره أحد الموارد الاشتراكية. أما في الوقت الحالي، فإن غالبية مستخدمي الإنترنت من غير المتحدثين باللغة الإنجليزية؛ وعلى الرغم من ذلك، فإن اللغة السائدة المستخدمة على الإنترنت هي الإنجليزية. وللمساعدة في تدويل الإنترنت، بدأت فرقة عمل هندسة الإنترنت (IETF) في عام 2003 في إصدار معايير توفر



إرشادات تقنية لنشر وإطلاق أسماء النطاقات الدولية (IDN) من خلال آلية ترجمة لدعم تمثيلات أسماء النطاقات من غير ASCII في أي نص مدعوم بنظام يونيكود (مثل، 测试世界 - 普遍接受، و ua-test كاثوليك، إلخ).

ووافق مجلس إدارة ICANN على عملية طرح نطاقات ccTLD ذات أسماء IDN جديدة في أكتوبر/تشرين الأول 2009، مع إضافة نطاقات ccTLD ذات أسماء IDN الأولى إلى منطقة الجذر في مايو/أيار 2010. وفي يونيو/حزيران 2011، اعتمد مجلس الإدارة ووافق على إطلاق برنامج نطاقات gTLD الجديدة، والذي اشتمل على نطاقات TLD ذات أسماء IDN. وتمت إضافة الدفعة الأولى من نطاقات TLD من هذا البرنامج إلى منطقة الجذر في عام 2013.

الحاجة إلى القبول العالمي (UA)

وبعد إطلاق فريق عمل هندسة الإنترنت لإرشاداته ذات الصلة بأسماء IDN بعشر سنوات، ومن خلال برنامج نطاقات gTLD الجديدة، بات أكثر من 1,000 نطاق TLD نشطًا الآن. ومع ذلك، لا تزال بعض البرامج والتطبيقات قديمة وغير قادرة على التعامل مع نطاقات TLD الجديدة. وهذا ما يسبب مشكلات بالنسبة لمستخدمي الإنترنت، بما في ذلك من يستخدمون حروفًا ونصوصًا من غير نظام ASCII.

ويضمن القبول العالمي قبول جميع أسماء النطاقات وعناوين البريد الإلكتروني الصحيحة وتوثيقها وتخزينها ومعالجتها وعرضها بشكل صحيح ومتسق من خلال جميع التطبيقات والأجهزة والنظم ذات القدرة على الاتصال بالإنترنت. على سبيل المثال، يتم حل كل عنوان ويب صالح إلى المورد المتوقع على موقع الويب الصحيح ويسفر كل عنوان بريد إلكتروني صالح عن تسليم البريد إلى المستلم المتوقع.

المجموعة التوجيهية للقبول العالمي (UASG) عبارة عن مبادرة من مجتمع الإنترنت تدعمها ICANN، وتم تكليفها بتنفيذ أنشطة من شأنها تعزيز القبول العالمي وتساعد في ضمان خبرة متسقة وإيجابية لجميع مستخدمي الإنترنت على مستوى العالم.

مسميات U ومسميات A

يطلق على أسماء النطاقات التي تستخدم حروفًا غير ASCII لفظ أسماء النطاقات المدوّلة (IDN). ويمكن أن يكون الجزء المدوّل في أي اسم نطاق في أي مسمى—وليس فقط في TLD.

وحيث إن نظام أسماء النطاقات نفسه قد استخدم في السابق نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII فقط، كان من الضروري إنشاء ترميز إضافي من أجل تمكين نقاط رمز Unicode غير نظام ASCII من التمثيل في صورة سلاسل ASCII. والخوارزمية التي تقوم بتنفيذ هذا الترميز بمنظومة يونيكود إلى ASCII يطلق عليه اسم Punycode (بيونيوكود)؛ ويطلق على السلاسل الناتجة عن ذلك اسم مسميات A. ويمكن تنفيذ مسميات A عن مسميات نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII العادية لأنها تبدأ دائمًا بالحروف الأربعة التالية:

xn--

ويطلق على هذه الحروف اسم بادئة ACE¹.

علمًا بأن نظام تحويل Punycode (بيونيوكود) يمكن الرجوع فيه: حيث يمكن التحول من نظام الكود الموحد (اليونيوكود) إلى A-Label الترميز بمنظومة الأسكي ومن A-Label الترميز بمنظومة الأسكي مرة أخرى إلى سلسلة حروف الكود الموحد (اليونيوكود) (المعروفة باسم مسميات U).

كما أن الاستخدام القياسي الوحيد لخوارزمية Punycode (بيونيوكود) مخصصة للتعبير عن النطاقات المدوّلة. وفي حين يمكن لأي أحد من الناحية الافتراضية ترميز سلاسل UTF-8 الأخرى مستخدمًا Punycode (بيونيوكود)، فلن يكون هذا الأمر قياسيًا ولن تكون له القدرة على التشغيل البيئي مع الأنظمة الأخرى.

¹بادئة الترميز المتوافق لمنظومة الأسكي (ACE) التي تميز المسميات المشفرة بنظام Punycode (بيونيوكود) عن المسميات الأخرى بنظام ASCII.



أمثلة على نطاقات IDN (التخيلية)

إصدار مسميات A	إصدار مسميات U
example.xn--q9jyb4c	example.みんな
xn--uesx7b.info	大坂.info
xn--q9jyb4c.xn--uesx7b	みんな.大坂

تحويل عناوين البريد الإلكتروني (EAI)

تحتوي عناوين البريد الإلكتروني على جزئين:

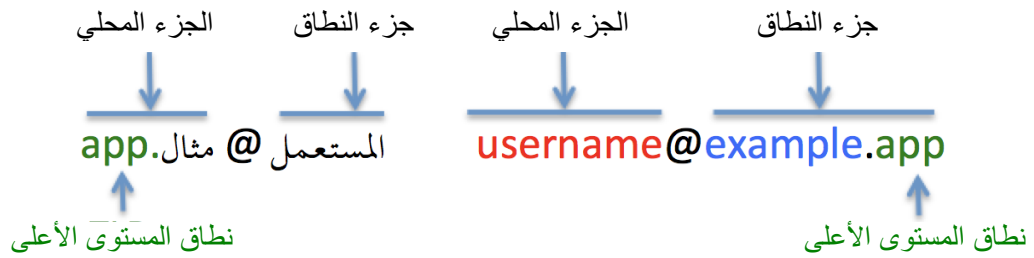
- جزء محلي (قبل الحرف "@").
- جزء نطاق (بعد الحرف "@").

حيث إن كل من نظام الكتابة من اليسار إلى اليمين (LTR) ومن اليمين إلى اليسار (RTL) في عناوين البريد الإلكتروني وأسماء النطاقات، يجب فهم لفظ "قبل" و"بعد" فيما يخص اتجاه النص وطريقة الكتابة.

أمثلة على عناوين EAI (التخيلية)

يستخدم نطاق مستوى أعلى باسم نطاق مدوّل	user@example.みんな
يستخدم نطاقًا من المستوى الثاني باسم نطاق مدوّل	user@大坂.info
يستخدم جزءًا محليًا من الكود الموحد (اليونيكود) ونطاق gTLD جديدًا	用戶@example.lawyer

أمثلة على النص من اليمين إلى اليسار في عنوان EAI



من اليمين إلى اليسار

من اليسار إلى اليمين

وفي أي عنوان بريد إلكتروني مدوّل، يمكن أن يحتوي جزء النطاق على أي اسم نطاق، بما في ذلك أسماء النطاقات المستخدمة نطاقات TLD الجديدة، ويمكن أن تحتوي على مسميات U بالكود الموحد (اليونيكود). علمًا بأن الجزء المحلي ليس اسم نطاق ويمكن أن يحتوي من حيث المبدأ على أي حرف بنظام الكود الموحد (اليونيكود) تقريبًا، على الرغم من أنه من الناحية العملية سوف تقوم أنظمة البريد بتقييد الحروف المستخدمة في أسماء صندوق البريد الخاص بها.

غالبًا ما يتم استخدام تحويل عناوين البريد الإلكتروني (EAI) لوصف استخدام العناوين المدوّلة في البريد الإلكتروني.



إنشاء ارتباط ديناميكي (توفير الربط)

تسمح البرامج الحديثة—مثل التطبيقات الشائعة في مجال معالجة الكلمات أو جداول البيانات—في بعض الأحيان للمستخدم إنشاء ارتباط تشعبي فقط عن طريق كتابة سلسلة تبدو وكأنها عنوان ويب أو عنوان بريد إلكتروني أو مسار شبكة. على سبيل المثال، كتابة "www.icann.org" في رسالة بالبريد الإلكتروني قد تسفر تلقائيًا عن رابط قابل للنقر عليه يوصل إلى <http://www.icann.org> إذا تعرف التطبيق على "www." باعتبارها بادئة خاصة أو ".org" باعتبارها لاحقة خاصة.

ويجب أن يعمل توفير الروابط في حالة استخدام توفيره باتساق لجميع عناوين الويب أو عناوين البريد الإلكتروني أو مسارات الشبكات المصاغة جيدًا، وليس البعض منها فقط. علمًا بأن عملية توفير الروابط مسألة صعبة وتعتمد على سياق النص (على سبيل المثال، في بعض اللغات، "www" لا تشير إلى عنوان ويب)، ومن ثم لم يتم التعامل معها أكثر من ذلك هنا.

الطبيعة الديناميكية لسجل منطقة الجذر

إن نظام أسماء النطاقات DNS عبارة عن قاعدة بيانات ضخمة موزعة ومقسمة إلى قطاعات يطلق عليها اسم المناطق. والقطاع الذي يحتوي على جميع نطاقات TLD يطلق عليه اسم منطقة الجذر لأنه من ناحية المفهوم في جذر أسماء DNS الثلاثة. ويجري تحديث جميع مناطق نظام أسماء النطاقات DNS، بما في ذلك منطقة الجذر حسب الما يلزم. ومع إضافة نطاقات TLD جديدة أو حذف نطاقات TLD قديمة، تتم إضافة أسمائها أو حذفها من منطقة الجذر.

وهذا يعني أن أي قائمة محددة بنطاقات TLD مثل أي قائمة مخزنة في أي تطبيق أو في أي ملف، سوف تصبح عتيقة الطراز في حتميًا وفي نهاية المطاف. ولتوثيق نطاق المستوى الأعلى بموثوقية في أي اسم نطاق، يمكن لبرمجيات فحصه عبر الإنترنت باستخدام استعمال DNS، أو في حالة استخدام ملف، أن تقوم بتحديث الملف بانتظام. وسوف يتم وصف هذين الأمرين بمزيد من التفصيل لاحقًا.

القبول الشامل من حيث التنفيذ

خمس معايير للقبول الشامل

القبول العالمي هو الحالة التي يتم فيها قبول جميع أسماء النطاقات وعناوين البريد الإلكتروني الصحيحة وتوثيقها وتخزينها ومعالجتها وعرضها بشكل صحيح ومتسق من خلال جميع التطبيقات والأجهزة والنظم ذات القدرة على الاتصال بالإنترنت. وفيما يلي وصف لمعايير القبول الشامل الخمسة.

<p>يحدث القبول عندما يتم استلام أي عنوان بريد إلكتروني أو اسم نطاق في صورة سلسلة من الحروف من واجهة مستخدم أو ملف أو واجهة برمجة التطبيقات مما يستخدمه أحد التطبيقات البرمجية أو خدمة متوفرة على الإنترنت.</p> <p>وتتيح التطبيقات والخدمات لأسماء النطاقات وعناوين البريد الإلكتروني:</p> <ul style="list-style-type: none">■ أن يتم إدخالها عن طريق واجهات المستخدم■ أو أن يتم تلقيها من تطبيقات وخدمات أخرى عن طريق واجهات معاملة التطبيقات.	<p>1. القبول²</p>
---	------------------------------

² يتم التعامل مع القبول بمنأى عن التوثيق في هذه الوثيقة. ومن الناحية العملية، ربما تتداخل الإجراءات.



<p>2. التوثيق³</p> <p>التوثيق قد يحدث في العديد من الأماكن في أي مكان يتم فيه تلقي أي عنوان بريد إلكتروني أو اسم نطاق أو إطلاقه كسلسلة من الحروف من خلال تطبيق أو خدمة على الإنترنت.</p> <p>ويتمثل الهدف من التوثيق في ضمان أن ما يتم إدخاله من معلومات إما صحيح أو عبر الأقل غير زائف بالتأكد. يضمن التوثيق احتواء المعلومات على السياق الصحيح ويمكنه إجراء اختبارات أخرى.</p> <p>بالنسبة لأسماء النطاقات وعناوين البريد الإلكتروني، فقد اعتمد العديد من المبرمجين من الناحية التقليدية على طرق توثيق مخصصة مثل التحقق من أن نطاق TLD يقع ضمن حدود الطول، أو أن الحروف المستخدمة من مجموعة حروف ASCII. وعلى الرغم من ذلك، تعتمد هذه الطرق على افتراضات لم تعقد منطبقة لأن الإنترنت دائم التغيير:</p> <ul style="list-style-type: none">▪ حيث يمكن لأسماء النطاقات وعناوين البريد الإلكتروني الآن أن تحتوي على حروف يونيكود من غير ASCII.▪ وقائمة نطاقات TLD في تغيير.▪ وأي تسمية في اسم نطاق، بما في ذلك تسمية TLD، يمكن أن تصل إلى 63 حرفاً من حيث الطول.⁴ <p>ولا يزال من الممكن توثيق نطاقات TLD من خلال استخدام تقنيات أخرى، وفقاً لما هو مذكور لاحقاً.</p>	
<p>3. التخزين</p> <p>التخزين عندما يتم تخزين عنوان بريد إلكتروني أو اسم نطاق في صورة سلسلة من الحروف في قاعدة بيانات أو ملف يستخدم من خلال تطبيق برمجي أو خدمة عبر الإنترنت ويتم استعادتها لاحقاً من خلال نفس التطبيقات البرمجية أو تطبيقات غيرها.</p> <p>قد تتطلب التطبيقات والخدمات تخزيناً طويلاً الأجل و/أو عابراً لأسماء النطاقات وعناوين البريد الإلكتروني. وبصرف النظر عن المدى العمري للبيانات، فيجب تخزينها في:</p> <ul style="list-style-type: none">▪ تنسيقات محددة من خلال طلب تعليقات قياسي (RFC) حسب الإنترنت (أو أقل استجابياً)▪ التنسيقات البديلة التي يمكن ترجمتها من التنسيقات المحددة بطلبات الحصول على التعليقات وإلها. <p>على الرغم من أن يونيكود في أسماء DNS وعناوين البريد الإلكتروني تُخزن في نسق UTF-8، إلا أنه يمكن مقابلة تنسيقات أخرى في الأكواد القديمة. انظر قسم "الممارسات الجيدة" أدناه.</p>	

³ تتم معاملة القبول والمعالجة بمنأى عن التوثيق في هذه الوثيقة. ومن الناحية العملية، ربما تتداخل الإجراءات.

⁴ وينطبق حد الطول عند 63 حرفاً على التسمية نفسها إذا كانت تسمية بنظام الترميز ASCII، أو على نموذج تسمية A للتسمية إذا كانت اسم نطاق مدوّل.



<p>4. المعالجة⁵</p> <p>المعالجة تتم عندما يتم استخدام عنوان بريد أو اسم نطاق من خلال تطبيق أو خدمة من أجل أداء نشاط مثل بحث أو تصنيف قائمة، أو تحويلها إلى تنسيق بديل (على سبيل المثال تحويل مسميات A إلى مسميات U).</p> <p>وقد يحدث تحقق اضافي من الصحة (تصديق آخر) خلال عملية المعالجة. علمًا بأن الطرق التي يمكن بها معالجة عناوين البريد الإلكتروني وأسماء النطاقات مقتصرة فقط على تصور مطوري التطبيقات، لكن من الضروري عدم القيام بأي افتراضات (على سبيل المثال، في حالة توجيه البريد الإلكتروني إلى pākehā@tetaurawhiri.govt.nz والهدف إرساله إلى شخص في نيوزيلندا) فهذا يعتمد على السياسات المستخدمة خارج DNS.</p>	
<p>5. العرض</p> <p>يحدث إجراء العرض عندما يتم تقديم عنوان بريد إلكتروني أو اسم نطاق داخل وجهة مستخدم.</p> <p>عرض أسماء النطاقات وعناوين البريد الإلكتروني غالبًا -وليس دائمًا- ما يكون بشكل مباشر عندما يتم استخدام نصوص في الاسم أو العنوان المدعوم في نظام التشغيل الأساسي وعندما يتم تخزين السلاسل في ترميز يحدده معيار اليونيكود.⁶ وفي حالة عدم استيفاء هذه الشروط، فقد يكون من المطلوب توفير عمليات نقل خاصة بالتطبيقات. وعلاوة على ذلك، حتى وإن كان نظام التشغيل الأساسي لا يدعم السلاسل، فقد يكون العرض معقدًا إذا ما تم الخلط بين النصوص التي تكتب من اليمين إلى اليسار والنصوص التي تكتب من اليسار إلى اليمين على سبيل المثال، أو أن يكون الاتجاه العام للنص غير واضح.</p>	

سيناريوهات المستخدم

قد توفر الأمثلة والتعريفات السابقة انطباعًا بأن القبول الشامل يتعلق فقط بأنظمة الكمبيوتر والخدمات المتوفرة على الإنترنت. لكن واقع الأمر مع ذلك، أنه يتعلق أيضًا بالأشخاص الذين يستخدمون تلك الأنظمة والخدمات.

فيما يلي بعض الأمثلة على الأنشطة التي تتطلب توافر القبول الشامل:

<p>أي منظمة تعتمد نطاق TLD يحمل "علامة تجارية" لكي توفر لعملائها خبرة عملاء متميزة من خلال توفير عناوين بريد إلكتروني بالتنسيق: <code>customername@example.brand</code></p> <p>يعني القبول الشامل أن:</p> <ul style="list-style-type: none">مواقع وتطبيقات الويب تقبل عناوين البريد الإلكتروني "@example.brand" هذه كما هو الحال مع نطاقات TLD الأقدم مثل <code>.com</code> و <code>.net</code> و <code>.org</code>	<p>تسجيل نطاق TLD جديدًا</p>
---	------------------------------

⁵ يتم التعامل مع المعالجة بمنأى عن التوثيق في هذه الوثيقة. ومن الناحية العملية، ربما تتداخل الإجراءات.

⁶ ومن المهم معرفة أن العرض غير مباشر ومستقيم، حتى عند استيفاء هذه الشروط لبعض النصوص المعقدة.



<p>يصل المستخدم إلى موقع ويب يحتوي اسم نطاقه على TLD جديد عن طريق كتابة عنوان في متصفح أو النقر فوق رابط في مستند.</p> <p>يعني القبول الشامل أن:</p> <ul style="list-style-type: none">▪ على الرغم من أن TLD جديد، يعرض متصفح المستخدم عنوان الويب في شكله الأصلي ويصل إلى الموقع كما يتوقع المستخدم. ولا يعرض المتصفح أسماء النطاقات في صورة مسميات A للمستخدم إلى أن يحقق استفادة للمستخدم بطريقة ما.	<p>قبول نطاقات gTLD</p>
<p>يحصل المستخدم على عنوان بريد إلكتروني مع حصة النطاق التي تستخدم نطاق gTLD جديد ويستخدم عنوان البريد الإلكتروني هذا ليكون هوية له من أجل الوصول إلى حساباته البنكية وحساب برنامج ولاء الخطوط الجوية.</p> <p>يعني القبول الشامل أن:</p> <ul style="list-style-type: none">▪ على الرغم من أن النطاق المستخدم في عنوان البريد الإلكتروني جديد، فإن البنك أو موقع الخطوط الجوية تقبل العنوان بنفس الطريقة التي تقبل بها أي عنوان في أي نطاق TLD قديم مثل biz أو eu.	<p>استخدام عنوان بريد إلكتروني يحتوي على نطاق gTLD جديدًا في صورة هوية على الإنترنت</p>
<p>يقوم مستخدم بالوصول إلى رابط URL خاص باسم نطاق مدول IDN عن طريق كتابة رابط URL في أي متصفح أو النقر فوق رابط في مستند.</p> <p>يعني القبول الشامل أن:</p> <ul style="list-style-type: none">▪ حتى وإن كان اسم النطاق يحتوي على حروف مختلفة عن إعدادات اللغة في كمبيوتر المستخدم، فسوف يعرض أي متصفح يرغب المستخدم في استخدامه عنوان الويب كما هو متوقع ويصل إلى الموقع بنجاح.	<p>الوصول إلى اسم نطاق مدول</p>
<p>حصل أحد المستخدمين على عنوان بريد إلكتروني جديد يحتوي على حروف غير ASCII في اسم النطاق (على سبيل المثال؛ 测试.世界-接受@Info).</p> <p>يعني القبول الشامل أن:</p> <ul style="list-style-type: none">▪ يمكن للمستخدم الإرسال والاستقبال من أي عنوان بريد إلكتروني، من خلال استخدام أي وكيل بريد إلكتروني.	<p>استخدام عنوان بريد إلكتروني مدول ليكون بريدًا إلكترونيًا</p>
<p>يحصل المستخدم على عنوان بريد إلكتروني من غير نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII ويستخدم عنوان البريد الإلكتروني هذا ليكون هوية له من أجل الوصول إلى حساباته البنكية وحساب برنامج ولاء الخطوط الجوية.</p> <p>يعني القبول الشامل أن:</p> <ul style="list-style-type: none">▪ يقبل البنك أو موقع شركة الطيران الهوية الجديدة تمامًا كما لو كانت أي هوية بريد إلكتروني أخرى.	<p>استخدام عنوان بريد إلكتروني مدول كهوية على الإنترنت</p>
<p>يقوم المستخدم بكتابة عنوان ويب في مستند أو رسالة بريد إلكتروني.</p> <p>يعني القبول الشامل أن:</p> <ul style="list-style-type: none">▪ القواعد التي يستخدمها التطبيق لإنشاء ارتباط تشعبي تلقائيًا هي نفس القواعد المستخدمة إذا كان العنوان غير ASCII أو يحتوي على نطاق TLD جديد.	<p>إنشاء تلقائي لرابط تشعبي في أي تطبيق</p>



تطوير تطبيق	يقوم أحد المطورين بكتابة تطبيق له القدرة على الوصول إلى موارد الويب. يعني القبول الشامل أن: ▪ تشمل الأدوات المستخدمة من جانب المطورين المكتبات التي تمكن القبول الشامل من خلال دعم نطاقات TLD وأسماء النطاقات المدوّلة IDN الجديدة.
--------------------	---

عدم التوافق مع الممارسات العالمية

تعتبر كل مما يلي ممارسة ضعيفة:

عرض مسميات A على المستخدم بدون فائدة مقابلة للمستخدم، مثل عرض المخططات بين مسميات U ومسميات A.	✗
مطالبة مستخدم بإدخال مسميات A عند الاشتراك من أجل الحصول على عنوان بريد إلكتروني جديد أو مطالبة مستخدم بإدخال مسميات A عند الاشتراك من أجل الحصول على نطاق جديد قيد الاستضافة.	✗
توثيق تركيب وصياغة اسم النطاق أو عنوان البريد الإلكتروني من خلال استخدام معايير عتيقة أو موارد غير موثقة لأسماء النطاقات على الإنترنت.	✗
استخدام قائمة قديمة بنطاقات TLD حتى وإن كانت نطاقات TLD يجري إضافتها أو حذفها بشكل منتظم.	✗
عرض الاستخدام الداخلي لمسميات A على المستخدمين.	✗
على سبيل المثال، تحويل النطاقات في عناوين تدويل عناوين البريد الإلكتروني EAI إلى مسميات A عند الرد على مستخدم EAI.	✗
معالجة بعض أسماء النطاقات مثل مصطلحات البحث وليس كأسماء نطاقات لأن التطبيق لا يتعرف عليها على هذا النحو.	✗
تعيين برامج حجب البريد العشوائي من أجل الحظر التلقائي لنطاقات المستوى الأعلى (الجديدة) بأكملها دون دليل على إساءة الاستخدام.	✗

المتطلبات الفنية لجاهزية القبول الشامل

لكي يكون أي تطبيق أو موقع على الويب جاهزًا للقبول الشامل، يجب أن يفي بمجموعة متنوعة من المتطلبات.

المتطلبات رفيعة المستوى

أي تطبيق أو خدمة تدعم القبول الشامل (UA):

1. يدعم جميع أسماء النطاقات بصرف النظر عن طرق أو مجموعة الحروف.
انظر [طلب الحصول على الملاحظات والتعقيبات رقم 5892](#).

2. يتيح جميع مجموعات الحروف الصالحة لأسماء النطاقات وعناوين البريد الإلكتروني.

يقبل نقاط رمز Unicode بالإضافة إلى نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII.



3. يمكن أن يقدم بشكل صحيح جميع نقاط الرموز في سلاسل رمز Unicode.

راجع طلب الحصول على الملاحظات والتعليقات رقم 3490. لاحظ أن الكود الموحد (اليونيكود) يضيف نقاط أكواد جديدة بانتظام، لذلك فإنه هدف متحرك.

4. يمكن أن يقدم سلاسل من اليمين إلى اليسار (RTL) مثل السلاسل المكتوبة بالعربية أو العبرية.

للمزيد من المعلومات حول النصوص المكتوبة من اليمين إلى اليسار RTL، راجع طلب الحصول على الملاحظات والتعليقات رقم 5893.

5. يمكن أن يرسل البيانات بين التطبيقات والخدمات بلغة UTF-8، وأيضًا نظم ترميز أخرى - عند الحاجة- مما يمكن تحويله من وإلى UTF-8.

للحصول على معلومات حول نظام UTF-8، راجع طلب الحصول على الملاحظات والتعليقات رقم 3629.

6. يوفر واجهات معاملة تطبيقات عامة وخاصة تدعم نظام UTF-8.

تنطبق واجهات معاملة التطبيقات الخاصة فقط على المكالمات بين الجهات المعنية من خلال نفس موفر الخدمة.

7. يتعامل مع عناوين EAI بشكل صحيح.

وعلى وجه الخصوص، لا يحول أسماء النطاقات المدوّلة IDN في العناوين إلى مسميات A.

8. يمكنه إرسال البريد الإلكتروني واستلام البريد الإلكتروني من المستقبلين بصرف النظر عن اسم النطاق أو مجموعة الحروف.

راجع طلب الحصول على الملاحظات والتعليقات رقم 6530.

9. يخزن بيانات المستخدمين في تنسيقات تدعم يونيكود وقابل للتحويل من وإلى نظام UTF-8.

ولا تكون هذه التحويلات واضحة إلا لمشغل المنتج أو الخدمة.

10. يدعم جميع أسماء نطاقات المستوى الأعلى في قائمة TLD الموثقة من ICANN بصرف النظر عن الطول أو مجموعة الحروف.

راجع القائمة الموثقة على <https://data.iana.org/TLD/>.

اعتبارات المطور

نظرًا لأن العديد من الأنظمة البرمجية الحالية تحتوي على افتراضات محدد بتعليمات برمجية ثابتة حول النطاقات وعناوين البريد الإلكتروني، قد تكون تغييرات الرموز أمرًا مطلوبًا للتعرف على أسماء النطاقات المدوّلة IDN ونطاقات المستوى الأعلى الجديدة وعناوين البريد الإلكتروني بنظام EAI. يناقش هذا القسم الطريقة التي يمكن للمطورين بها دمج تغييرات الأكواد التي ستمكّن القبول العالمي.

تصميم البرامج من أجل التوافق والمرونة

إن مبدأ المتانة والقوة الذي عبر عنه جون بوستل طلب الحصول على الملاحظات والتعليقات رقم 793، عبارة عن إرشاد تصميمي عام لتطوير البرمجيات:

"تحلى بالتحفظ فيما تقوم به؛ وكن متحررًا فيما تقبه من الآخرين".



أي كن متحفظاً فيما ترسله: في أي منطقة قد تكون فيها المواصفات غامضة أو غير واضحة، تجنب ما قد يكون مفاجئاً للآخرين. ومن ناحية أخرى، عند الاستلام، اقبل أي شيء صالح من الناحية المعقولة. وهذا لا يعني تغيير قواعد التعامل مع الأخطاء الواضحة في حالات تنفيذ أخرى حيث يؤدي ذلك إلى فوضى غير موثقة و"غير قابلة للتصحيح".

الممارسات الجيدة لتطوير وتحديث برمجيات لتحقيق الجاهزية للقبول الشامل

القبول	
عرض الأسماء في صورة يونيكود كلما كان ذلك ممكناً.	✓
يجب السماح للمستخدمين، وليس شرطاً عليهم، إدخال أسماء النطاقات في صورة مسميات A وليس مسميات U. وعلى الرغم من ذلك، يجب عرض مسميات U بشكل افتراضي، وعرض مسميات A على المستخدم فقط عندما توفر فائدة.	
لا تقم باستخراج عناوين EAI بمسميات A ولكن كن قادرًا على التعامل معها في حالة عرضها عن طريق برامج لأشخاص آخرين.	!
ويجب على أي عنصر في واجهة المستخدم يطالب أي مستخدم بكتابة اسم نطاق أو عنوان بريد إلكتروني أن يقبل الأسماء الطويلة. حيث يمكن أن يكون لأسماء نطاقات نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII ما يصل إلى 63 حرفاً في كل تسمية ويمكن أن يكون في إجماله حتى 253 بايت. ويمكن أن تكون مسميات UTF-8 أطول من مسميات نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII وقد يصل الطول الإجمالي إلى 670 بايت. تذكر أن كود UTF-8 لغالبية نقاط كود يونيكود تأخذ أكثر من بايت واحد. ▪ راجع الحصول على الملاحظات والتعليقات رقم 1035.	✓

التصديق	
التوثيق فقط حسبما يتناسب.	✓
إجراء التوثيق فقط إذا كان مطلوباً من أجل تشغيل التطبيق أو الخدمة. وهذه الطريقة الأكثر اعتمادية في ضمان قبول جميع أسماء النطاقات الصحيحة في الأنظمة الخاصة بك.	
إدراك أن عمليات القيد الصحيحة من الناحية التركيبية قد تمثل أسماء نطاقات أو عناوين بريد إلكتروني قيد الاستخدام في الوقت الحالي على الإنترنت. قد يكون ذلك صالحاً أو قد لا يكون صالحاً اعتماداً على التطبيق.	✓



و عند قيامك بالتوثيق، يجب مراعاة ما يلي:

- التحقق من حصة TLD من أسماء النطاقات في مقابل قائمة موثوقة. تنشر هيئة الإنترنت للأرقام المخصصة IANA قائمة بنطاقات المستوى الأعلى على:
* <https://data.iana.org/TLD/tlds-alpha-by-domain.txt>
* راجع أيضًا <https://www.icann.org/en/system/files/files/sac-070-en.pdf>
- الاستعلام عن اسم النطاق في مقابل نظام أسماء النطاقات.
* يعتبر GETDNS API (<http://getdnsapi.net/>) طريقة متنقلة بشكل كبير في الاستعلام عن نظام أسماء النطاقات DNS.
* كما أن غالبية أنظمة التشغيل تحتوي على استعلام واجهات معاملة التطبيقات أصلي لنظام أسماء النطاقات.
- المطالبة بإدخال مكرر لعنوان بريد إلكتروني من أجل تجنب أخطاء الكتابة.
- توثيق الحروف في المسميات من خلال فحص كل مسمى يلي أسماء النطاقات المدوّلة في قواعد التطبيقات (IDNA 2008).
* انظر RFC 5892
- يجب تقييد توثيق المسميات نفسها على عدد صغير من قواعد المسميات الكلية المحددة في طلبات الحصول على التعليقات RFC.
* انظر RFC 5894
- تأكد أن المنتج أو الميزة تتعامل مع الأرقام بشكل صحيح.
* فعلى سبيل المثال: يجب التعامل مع الحروف الرقمية العربية-الهندية كأرقام في حقول الإدخال الرقمية، بالإضافة إلى أرقام نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII.
* لاحظ أن الأرقام العربية-الهندية صحيحة في مسميات U لكنها تعتبر مكافئة لأرقام نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII في ذلك السياق.

التخزين

يجب أن تدعم التطبيقات والخدمات معايير يونيكود الأحدث.	✓
ويجب تخزين المعلومات بنسق UTF-8 متى ما كان ذلك ممكناً. قد تتطلب بعض الأنظمة دعمًا قديمًا لنظام UTF-16 أيضًا، ولكن بشكل عام UTF-8 هو المفضل. علمًا بأن نسق UTF-7 عتيق الطراز، كما أن UTF-32 كبير الحجم بالنسبة لتخزين الملفات. ويجب تطبيع هذه السلاسل عندما يكون ذلك مناسبًا (قد يؤدي بعض التطبيع في سياق ما إلى فقد المعلومات).	✓
يجب مراعاة السيناريوهات المتكاملة قبل التحويل بين مسميات A ومسميات U عند التخزين. وفي التطبيقات الجديدة، من الأفضل الحفاظ فقط على مسميات U في ملف أو قاعدة بيانات، لأن هذا من شأنه تبسيط البحث والتصنيف والعرض. وعلى الرغم من ذلك، قد يكون للتحويل تأثيرات عند التشغيل البيئي مع التطبيقات والخدمات الأقدم التي ليس لها قدرة يونيكود.	!
قم بتمييز عناوين البريد الإلكتروني وأسماء النطاقات على ذلك النحو في الخزين من أجل سهولة الوصول إليها. أدت عناوين البريد الإلكتروني وأسماء النطاقات في حقل "المؤلف" الخاص بوثيقة أو "معلومات الاتصال" في ملف سجل إلى فقد العنوان الأصلي.	✓



وبصرف النظر عن الطريقة التي يتم بها تخزين العناوين وأسماء النطاقات، يجب أن تكون لك القدرة على مطابقة السلاسل بتنسيقات متعددة.



على سبيل المثال، البحث عن `example.みんな` يجب أن يعثر أيضًا على `example.xn--q9jyb4c`.

العملية

تأكد أن جميع ردود خادم الويب وبريد MIME محدد لها نسق UTF-8 في نوع المحتوى.	✓
حدد نسق UTF-8 في عنوان http بخادم الويب. ▪ من المهم التأكد من أن الترميز محدد في كل استجابة.	✓
يجب مراعاة المحتوى قبل تحويل مسميات A إلى مسميات U والعكس بالعكس خلال المعالجة. من المفضل الحفاظ على مسميات U فقط في ملف أو قاعدة بيانات، لأنه يعمل على تبسيط البحث والتصنيف. وعلى الرغم من ذلك، قد يكون للتحويل تأثيرات عند التشغيل البيئي مع التطبيقات والخدمات الأقدم التي ليس لها قدرة يونيكود.	!
تأكد أن المنتج أو الميزة تتناول ترتيب الفرز وعمليات البحث والمقارنة وفقًا لمواصفات المكان واللغة، وأنه يتناول البحث والفرز متعدد اللغات.	✓
لا تستخدم الترميز النسبي للمسميات في أسماء النطاقات: ▪ <code>example.みんな</code> صحيح ▪ <code>example.%E3%81%BF%E3%82%93%E3%81%AA</code> غير صحيح.	✗
وحيث إن معايير يونيكود يتوسع بشكل متواصل، فإن نقاط الأكواد غير المحددة عند إنشاء التطبيق أو الخدمة يجب التحقق منها من أجل ضمان عدم إنتاج أي مخرجات خاطئة أو تسبب اللبس. قد تؤدي الخطوط المفقودة في نظام التشغيل الأساسي إلى حروف غير قابلة للعرض (كثيرًا ما يتم استخدام مربع صغير لتمثيلها)، لكن يجب ألا يؤدي هذا الموقف إلى تعطل أو رسالة خطأ.	✓
استخدم واجهات برمجة التطبيقات المدعومة ذات قدرة يونيكود.	✓
استخدم أحدث أسماء النطاقات المدوّلة في مستندات بروتوكول وجداول التطبيقات (IDNA 2008) الخاصة بأسماء النطاقات المدوّلة: ▪ RFC 5891 ▪ RFC 5892	✓
قم بمعالجة النص بتنسيق UTF-8 متى ما كان ذلك ممكنًا.	✓
قم بتنسيق تحديثات التطبيقات والخدمات التي تعتمد عليها. إذا كان الخادم يعمل بنظام يونيكود وكان الوكيل غير يونيكود أو العكس بالعكس، فسوف يتوجب تحويل البيانات في كل معاملة، وهي عملية عرضة للخطأ وربما تكون بطيئة.	✓
وعند القيام بعملية تحويل للحروف، فقد تنمو سلاسل النص أو تنكمش بشكل كبير. ويمكن أن تكون نقطة رمز UTF-8 بداية من 1 إلى 4 بايت، وفي بعض الحالات قد يتوافق حرف واحد في ترميز آخر مع العديد من نقاط أكواد UTF-8 أو العكس بالعكس.	✓



العرض	
عرض جميع نقاط أكواد يونيكود المدعومة من خلال نظام التشغيل الرئيسي.	✓
تحتوي جميع أنظمة التشغيل الحديثة على دعم يونيكود، لكن محركات الإخراج الخاصة بها لا تكون صحيحة دائماً بالنسبة لجميع النصوص واللغات. توفير إخراج الحروف في التطبيقات فقط عندما لا يكون الإخراج الصحيح متوفراً من نظام (أنظمة) التشغيل المستهدفة.	✓
وعند تطوير تطبيق أو خدمة، يجب مراعاة اللغات المدعومة والتأكد من أن أنظمة التشغيل والتطبيقات تغطي تلك اللغات.	✓
قم بتحويل مسميات A إلى مسميات U قبل العرض.	✓
على سبيل المثال، يجب أن يرى المستخدم النهائي "example.みんな" وليس "example.xn--q9jyb4c". (هذا التحويل مثال على المعالجة الجاهزة لـ UA).	✓
اعرض أسماء النطاقات في نسق مسميات U بشكل افتراضي.	✓
لا تعرض مسميات A للمستخدم إلا عندما يوفر ميزة ما.	✓
لاحظ أنه من الممكن أن تكون هناك أسماء نطاقات ذات نصوص مختلطة. <ul style="list-style-type: none">قد تبدو بعض حروف يونيكود متشابهة للعين المجردة، لكنها مختلفة بالنسبة لأجهزة الكمبيوتر؛ على سبيل المثال، حرف O اللاتيني وحرف O باللغة السيريلية وحرف O الخامس عشر في الأبجدية الإغريقية.تشيع سلاسل النصوص المختلط في النصوص وثيقة المرتبطة ببعضها (على سبيل المثال، كانجي اليابانية، وكاتاكانا، هيراجانا، وروماجي). وإلا فيمكن أن تكون النصوص المختلطة موجهة لأغراض ضارة، مثل التصيد. استخدم معيار اليونيكود الفني رقم 39، "آليات أمن اليونيكود"⁷، للتحقق من أن النصوص في أي متتالية يونيكود تتبع التطبيق الصحيح.إذا استدعت واجهة المستخدم السلاسل لاهتمام المستخدم، فتأكد أنها تقوم بذلك بطريقة لا تضر بمستخدمي النصوص غير اللاتينية. اعرف المزيد حول اعتبارات أمن يونيكود على: http://unicode.org/reports/tr36 .	!
واحذر من الحروف أو أسماء النطاقات غير المخصصة وغير المسموح بها. <ul style="list-style-type: none">راجع RFC 5892	✓

Unicode	
استخدم واجهات برمجة التطبيقات المدعومة ذات قدرة يونيكود.	✓

⁷ راجع https://www.unicode.org/reports/tr39/#Restriction_Level_Detection

✘	<p>استخدم واجهات معاملة التطبيقات قياسية وخالية تمامًا من العيوب من أجل:</p> <ul style="list-style-type: none"> ▪ عمليات تحويل أنساق السلاسل. ▪ تحديد أي السلاسل التي تحتوي على سلسلة. ▪ تحديد ما إن كانت سلسلة تحتوي على مزيج من النصوص. ▪ تطبيع / تفكيك يونيكود.
✘	<p>لا تستخدم UTF-7 وحدد الاستخدام على UTF-32.</p> <ul style="list-style-type: none"> ▪ فنسق UTF-7 عفا عليه الزمن. ▪ ونظام UTF-32 يستخدم أربعة أحرف لكل نقطة رمز. وحيث تستوعب كل نقطة رمز نفس مقدار المساحة ويمكن فهرستها مباشرة في مصفوفات، من الملائم استخدامها داخل كود برمجي، ولكنها قد تكون كبيرة أكثر من اللازم للتخزين في ملفات وقواعد بيانات.
✘	<p>لا تستخدم نظام UTF-16 إلا إذا كان مطلوبًا بشكل صريح (كما في بعض واجهات معاملة التطبيقات في نظام Windows وتطبيقات Javascript).</p> <p>وفي نظام UTF-16، يمكن لعدد 16 بت تمثيل حروف من 0x0 إلى 0xFFFF فقط. والقيم التي تكون أعلى من هذا النطاق (0x10000 إلى 0x10FFFF) تستخدم أزواج من الحروف الزائفة في صورة حروف بديلة. إذا لم يتم اختبار معالجة الأزواج البديلة بدقة، فقد يؤدي ذلك إلى أخطاء صعبة وثغرات أمنية محتملة.</p>
✔	<p>استخدم UTF-8 في ملفات تعريف الارتباط بحيث يمكن للتطبيقات قراءتها بشكل صحيح.</p>
✔	<p>استخدم بروتوكول IDNA 2008 ومستندات الجداول:</p> <ul style="list-style-type: none"> ▪ RFC 5891 ▪ RFC 5892
✘	<p>لا تستخدم بروتوكول IDNA لسنة 2003 والذي ألغاه وحل محله IDNA لسنة 2008.</p>
!	<p>حافظ على IDNA وجداول يونيكود المتسقة فيما يخص الإصدارات.</p> <p>على سبيل المثال، ما لم يتم التطبيق فعليًا بتنفيذ قواعد التصنيف في وثيقة الجداول لتفسير نقاط الأكواد وفقًا لما تم قيده (RFC 5892)، فيجب اشتقاق جداول IDNA من إصدار يونيكود المدعوم على النظام. لا يجب أن تعكس الجداول أحدث إصدار من يونيكود، ولكن يجب أن تكون متسقة.</p>
✔	<p>توثيق المسميات من خلال استخدام قواعد المسميات الكلية لبروتوكول IDNA 2008.</p> <ul style="list-style-type: none"> ▪ في بعض السياقات، قد يكون من المناسب إجراء مزيد التوثيق؛ على سبيل المثال، إذا كان التطبيق يعرف ماهية النصوص المسموح بها في أسماء النطاقات التي يستخدمها.

عام	
✔	<p>استخدم الموارد الموثوقة لتوثيق أسماء النطاقات.</p> <p>لا تستخدم افتراضات عتيقة مخصصة مثل "جميع نطاقات TLD مكونة من 6 أحرف أو أقصر".</p>
✔	<p>تأكد أن المنتج أو الميزة تتعامل مع الأرقام بشكل صحيح.</p> <p>على سبيل المثال، أرقام ASCII وتمثيلات الأرقام الأيديولوجية الصورية الآسيوية يجب أن تعامل جميعها معاملة الأرقام في سياقات الأرقام.</p>



<p>ابحث عن عناوين البريد التي قدم تكون عناوين EAI في الأماكن غير المتوقعة.</p> <ul style="list-style-type: none">البيانات الوصفية للفنان/المؤلف/المصور/حقوق النشر والتوزيع.البيانات الوصفية للخط.سجلات الاتصال لنظام أسماء النطاقات.معلومات الإصدار الثنائي.معلومات الدعم.معلومات اتصال OEM.التسجيل والتعليقات وغيرها من الصيغ.	!
<p>تقييد نقاط الرموز المسموح بها عن استخراج أسماء نطاقات وعناوين بريد إلكتروني جديدة:</p> <p>جميع المنتجات التي تستخدم عنوان البريد الإلكتروني يجب أن تقبل عنوان البريد الإلكتروني المدوّلة، بما يسمح لغالبية حروف طباعة UTF-8 في الجزء المحلي. وعلى الرغم من ذلك، لا يجب أن يتيح تطبيق أو خدمة جميع هذه الحروف عندما يقوم مستخدم بإنشاء اسم نطاق مدوّل أو عنوان EAI.</p> <p>قد يؤدي منع بعض أسماء النطاقات المدوّلة IDN أو عناوين البريد الإلكتروني من الإنشاء في المقام الأول إلى الحد من بعض المخاوف المحتملة المتعلقة بالأمن والقدرة على الوصول. (ملاحظة: تتطلب الممارسة الجيدة رغم ذلك أن يقبل البرنامج هذه السلاسل في حالة عرضها).</p>	!
<p>احذر من أن القبول الشامل قد لا يكون من الممكن دائماً قياسه من خلال حالات الاختبار التلقائية وحدها.</p> <p>على سبيل المثال، اختبار كيفية تعامل تطبيق أو بروتوكول مع موارد الشبكات قد لا يكون ممكناً دائماً ومن الأفضل في بعض الأحيان توثيق الامتثال من خلال مراجعة للمواصفة والتصميم الوظيفي.</p>	!
<p>لا نفترض أنه بسبب أن أحد المكونات لا يتصل مباشرة بواجهات معاملة التطبيقات ذات القدرة على حل الأسماء أو استعمال عنوان البريد الإلكتروني مباشرة، أنه لا يؤثر عليه.</p> <p>افهم كيفية الحصول على أسماء النطاقات من خلال المكون—فليس ذلك دائماً من خلال تفاعل المستخدم. فيما يلي بعض الأمثلة على كيفية حصول المكون على اسم نطاق:</p> <ul style="list-style-type: none">سياسة المجموعة.استعلام LDAP.ملفات التكوين.سجل Windows.التحويل من أو إلى مكون أو ميزة أخرى.	!
<p>قم بإجراء مراجعات للأكواد من أجل تجنب الهجوم بإغراق ذاكرة التخزين المؤقت.</p> <ul style="list-style-type: none">وفي يونيكود، يمكن أن تتمدد السلاسل أو تنكمش عند طي أو تطبيع الحالة.وعند القيام بعملية تحويل الحروف، قد يتعرض النص للنمو أو الانكماش بشكل كبير.	✓

تحديات أخرى

بعض تطبيقات البريد الإلكتروني القديمة استخدم تشفير حروف محلي ولم يكن لديها طريقة للتعرف على النص وتحويله من UTF-8 وإليه حسب الحاجة. وانطبق هذا الأمر خصيصاً على عناوين البريد الإلكتروني (إلى، نسخة إلى، نسخة مخفية إلى، الموضوع).

آلية للكشف عن مجموعات الأحرف وتحويلها



عندما يكون لدى مستخدم العديد من عناوين البريد الإلكتروني قد يكون من الصعب إدارة هذه العناوين كهوية مستخدم فردية.

إدارة عناوين البريد الإلكتروني المتعددة كهوية مستخدم فردية

يمكن لبرامج البريد الإلكتروني توجيه مرور البيانات الموجه إلى تلك الأسماء المستعارة إلى نفس صندوق البريد، ولكن قد تظل التطبيقات تعامل العناوين باعتبارها هويات مختلفة.

الموارد الموثوقة لأسماء النطاقات: منطقة جذر DNS وقوائم IANA

ثمة مصدران للحصول على قائمة معتمدة بنطاقات المستوى الأعلى. المصدر الأول وهو منطقة جذر نظام أسماء النطاقات نفسها. وهو منطقة موقعة بالامتدادات الأمنية لنظام أسماء النطاقات (DNSSEC)، ومن ثم يمكن توثيق محتوياتها من خلال خادم اسم على دراية بنظام الامتدادات الأمنية لنظام أسماء النطاقات، على الرغم من أن محتوياته من الصعب نسيبًا التعرف عليها في صورة ملف نصي. وثمة مصدر آخر وهو الملف النصي لنطاقات TLD الذي تقوم IANA على نشره (نطاقات TLD واحد لكل سطر بالترتيب الأبجدي). وهذه الملفات موجودة على خوادم وجيب https، لذلك فإن من الممارسات الجيدة أن تكون شهادة أمان طبقة النقل (TLS) صالحة عند تنزيلها للتأكد من الحصول على الملف الصحيح.

ويمكنك الحصول على قائمة بنطاقات TLD من أي من الروابط التالية:

- <https://www.internic.net/domain/root.zone> (ملف منطقة الجذر)
- <https://data.iana.org/TLD/tlds-alpha-by-domain.txt> (ملف TLD النصي)

البريد الإلكتروني مع نطاقات IDN ولماذا ليس هو نفس EAI

إن البريد بنظام تدويل عناوين البريد الإلكتروني (EAI) يفضل أسماء النطاقات بنظام UTF-8؛ أما استخدام مسميات A المشفرة فهو غير مفضل. قدمت بعض أنظمة البريد مخصصات جزئية لعناوين البريد الإلكتروني تشمل على أسماء النطاقات المدوّلة IDN بدلاً من توفير دعم EAI الكامل. ونظرًا لأنه يمكن تمثيل نطاقات IDN في صورة مسميات A بمنظومة نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII، تتيح بعض البرمجيات الحالية أسماء النطاقات المدوّلة IDN إمكانية تمثيل عنوان بريد إلكتروني في ASCII أو يونيكود. على سبيل المثال، سوف تتعامل بعض البرمجيات مع عنواني اسم نطاق مدوّل هذين بالتساوي لجميع الأغراض (الإرسال والاستقبال والبحث):

user@example.みんな = user@example.xn--q9jyb4c

وعلى الرغم من ذلك، سوف تتعامل بعض البرمجيات مع هذه العناوين بشكل متساوي حتى وإن كانا صحيحين، لأنها لا تحول مسمى A-Label الترميز بمنظومة الأسكي ("xn--q9jyb4c") إلى مسمى U المكافئ له ("みんな") قبل المقارنة. وقد يؤدي ذلك إلى تجربة مستخدم لا يمكن التنبؤ بها. وقد تصبح تجربة المستخدم مربكة على وجه الخصوص إذا قامت بعض البرامج بتحويل مسميات U إلى مسميات A من أجل "التوافق". ونظرًا لأن الرسائل يتم الرد عليها أو إعادة توجيهها، فقد تزداد العناوين المختلفة بشكل واضح للمستخدم، أو التي تفشل في البحث والفرز على النحو المتوقع.

وكما في المثال التالي، قد تحاول بعض البرمجيات تحويل الجزء المحلي في عنوان البريد الإلكتروني من خلال استخدام Punycode (بيونيكود)، وهو الخوارزمية المستخدمة في تحويل مسميات A إلى مسميات U (والعكس بالعكس). علمًا بأن هذا النوع من التحويل غير صالح وسوف يؤدي إلى عناوين غير صالحة ولا يمكن إرسالها.

فلا تحاول تحويل الجزء المحلي من عنوان بريد إلكتروني إلى صيغة مختلفة

用戶@example.みんな ✓

xn--youq53b@example.xn--q9jyb4c ✗

يجب أن تكون للبرامج والخدمات ذات جاهزية القبول الشامل القوية القدرة على معاملة ومعالجة جميع هذه الأنساق بشكل صحيح ويجب أن تكون لها القدرة على معالجة كل من أجزاء UTF-8 المحلية ومسميات U ذات نظام UTF-8 في العناوين، مع قبول مسميات A في الوقت ذلك في العناوين من أجل التوافق بالأثر الرجعي.



الربط وتحدياته

تتيح البرمجيات الحديثة في بعض الأحيان للمستخدم إنشاء رابط تشعبي تلقائيًا وذلك من خلال كتابة سلسلة تشبه عنوان ويب أو اسم بريد إلكتروني أو مسار شبكة. على سبيل المثال، فإن كتابة العنوان "www.icann.org" في رسالة بريد إلكتروني قد يؤدي تلقائيًا إلى إنشاء رابط قابل للنقر فوقه يوصل إلى <http://www.icann.org> إذا ما تعرف التطبيق على ".www" باعتبارها مسمى أول أو ".org" باعتباره نطاق TLD.

عملية الربط هي الإجراء التي يقبل التطبيق من خلاله أي سلسلة وتحدد بشكل ديناميكي ما إن كان من الواجب إنشاء رابط تشعبي لموقع إنترنت (<http://> أو <https://>) أو عنوان بريد إلكتروني (<mailto:>). ويجب أن تعمل عملية توفير الروابط -إذا ما وقعت- باتساق لجميع عناوين الويب وأسماء البريد الإلكتروني ومسارات الشبكات ذات التنسيق الجيد.

وتستخدم عملية توفير الروابط خوارزميات وقواعد يجري إنشاؤها بمعرفة مطوري البرمجيات من أجل تحديد ما إن كان من الواجب تفسير سلسلة باعتبارها رابطًا أم لا. ويرتبط بذلك الكيفية التي يمكن للناس من خلالها تحديد سلسلة بأنها اسم نطاق. وفي حين أن المتصفحات وكلاء البريد الإلكتروني وبرامج معالجة الكلمات أماكن واضحة، إلا أن هناك الكثير من التطبيقات الأخرى التي يمكنها اتخاذ تلك القرارات.

توصيات الممارسات الجيدة

1. حاول إجراء عملية توفير الروابط استنادًا إلى بادئات بروتوكول واضحة (على سبيل المثال <https://>، و <ftp://>، و <mailto:>) لكن لا تكمل الإجراء إلا إذا كانت بقية السلسلة ذات تكوين جيد.

مثال على سلسلة	السلوك / النتيجة المتوقعة
example.com	لا يحدث توفير للروابط لأن البروتوكول غائب وغير ظاهر.
http://example.com	قم بإنشاء رابط تشعبي لأن البروتوكول واضح.
http:example.com	لا تتوفر روابط بسبب التركيب السيئ (لا تتوفر //).
http://example.a	لا يتوفر رابط لأن "a" ليست نطاقًا من المستوى الأعلى.
http://example..ab	لا تتوفر روابط بسبب التركيب السيئ (نقاط متتالية).
http://普遍接受-测试.世界	قم بإنشاء رابط تشعبي لأن البروتوكول واضح.

2. حاول القيام بعملية الربط استنادًا إلى بادئات بروتوكول واضحة (على سبيل المثال "www" تشير إلى <http://www>).

مثال على سلسلة	السلوك / النتيجة المتوقعة
www.example.com	قم بإنشاء رابط تشعبي لأن البروتوكول مضمّن ⁸
label@example.com	قم بإنشاء mailto: label@example.com لأن البروتوكول مضمّن.

⁸ ملاحظة: قد يكون موقع الويب الفعلي نظام <https> فقط ويتطلب <https://> بدلاً من <http://>. وإذا كان هذا هو الحال، فقد لا يتم حل الرابط التشعبي إلى عنوانه المقصود.



3. وقد يشتمل نظام HTML الذي يحيط بروابط URL التي تحتوي على نص ثنائي الاتجاه على أكواد تؤثر على الاتجاه الذي يتم فيه عرض النص. ويجب أن يحافظ الإصدار المزود برابط على نفس اتجاه العرض.

4. وفي حالة استخدام نطاقات TLD باعتبارها "رمزًا خاصًا" لتحديد مدى توافر الروابط، فيجب عندئذ تضمين جميع نطاقات TLD. ويجب تحديث نطاقات TLD بصفة منتظمة.

الكود الموحد (اليونيكود) — الخلفية وسمات نقاط الأكواد

شهد معيار الكود الموحد (اليونيكود) تطورًا منذ أن تم نشره للمرة الأولى في صورة Unicode إصدار 1.0 في عام 1991. وقد أضاف كل إصدار منذ ذلك الإصدار المزيد من الحروف ونقاط الرموز للتعامل مع المزيد من اللغات والنصوص. الإصدار الحالي هو 12.1.

وفي نظام الكود الموحد (اليونيكود)، فإن لكل نقطة رمز مجموعة من السمات، مثل Uppercase_Letter أو Decimal_Number أو Nonspacing_Mark. وللعديد من الحروف خصائص نصوص مثل اللاتينية أو لغة الهان (الصينية) أو العربية، في حين ليس لغيرها تلك الخصائص، مثل علامات الترقيم.

ووفقًا للشرح التالي، تستخدم أسماء النطاقات المدوّلة في التطبيقات خصائص نقاط رموز من أجل تحديد الحروف المسموح بها في أسماء النطاقات المدوّلة. وتصف UAX#44، التي تعد قاعدة بيانات الكود الموحد (اليونيكود)، قاعدة بيانات خصائص نقاط الرموز.

UTF8 و UTF16 وغيرها من طرق التشفير

قد يكون لنقطة رمز يونيكود قيمة عددية تتراوح ما بين صفر إلى 0x10FFFF. وحيث إن أي بايت فردي بإمكانه أن يشغل قيمًا من صفر إلى 0xFF فقط، فإن بعض أنواع التشغيل متعددة البايتات بحاجة لتخزين نقاط رمز يونيكود.

يحتوي الإصدار الأصلي من يونيكود على أقل من 64 كيلو (0xFFFF) من نقاط الأكواد، لذلك يمكن أن تقبل كل نقطة كود عددًا صحيحًا قدره 16 بت. وقد أدى ذلك إلى ترميز ثنائي البايت يعرف باسم UCS أو UCS-2. عندما توسعت يونيكود لأبعد من نقاط أكواد سعة 64 كيلو، توسعت UCS إلى UTF-16⁹، والذي يستخدم أزواجًا من نقاط الرموز سعة 16 بت غير صالحة تُعرف باسم البدائل لتمثيل قيم أكبر من 64 كيلو. وفي حين أن هذا الأمر مفيد وناجح، إلا أنه قد أدى إلى مشكلات في حل العيوب حيث تضيف البدائل قدرًا من التعقيد لأي كود يقود بعد رقم نقاط الأكواد في أي سلسلة، أو التي تقوم بتصنيف السلاسل في ترتيب حسب نقاط الأكواد. هناك مشكلة إضافية تتمثل في أن بعض أجهزة الكمبيوتر كذلك التي تصنعها شركة IBM تخزن البايتات العالية بقيمة 16 بت أولاً ("big-endian")، وبعضها الآخر كذلك التي تصنعها شركة Intel تخزن البايتات المنخفضة أولاً ("little-endian"). ونتيجة لذلك، لدى UTF-16 متغيران للتخزين: UTF-16BE و UTF-16LE. وثمة أساليب للتعرف على البايتات العالية وتصحيحها لكنها قد تؤدي إلى أخطاء. وفي هذه النقطة، يستخدم UTF-16 بالأساس في التطبيقات الحالية مع واجهات معاملة التطبيقات الخاصة بنظام Microsoft Windows، ولغة Java و Javascript.

وثمة تشفير بديل ألا وهو UTF-8، والذي يقوم بتشفير كل نقطة رمز كسلسلة ذات طول متغير مكونة من بايت واحد إلى أربعة بايتات. علمًا بأن UTF-8 له العديد من الميزات على UTF-16، والتي تتمثل في أن مجموعة نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII الفرعية في الكود الموحد (اليونيكود) يتم تشفيرها كبايت واحد بحيث تكون كل سلسلة ASCII سلسلة UTF-8 تلقائيًا. وغالبًا ما يكون UTF-8 مدمجًا أكثر من UTF-16 المقابل له، ومن السهل تصنيفه نظرًا لأن سلاسل UTF-8 المخزنة بترتيب بايت تكون تلقائيًا بترتيب نقاط الأكواد. تتطلب أسماء النطاقات المدوّلة في التطبيقات IDNA و EAI تشفيرًا من نوع UTF-8.

علمًا بأن UTF-32 عبارة عن نسق بسيط يقوم بتخزين كل نقطة رمز في عدد صحيح من 32 بت. وهو ملائم للمعالجة الداخلية في البرامج لأن نقاط الأكواد في أي مصفوفة من نوع UTF-32 يمكن فهرستها بشكل مباشرة، لكن من النادر استخدامها في التخزين بسبب حجمها.

⁹ انظر القسم 3.10 في معيار الكود الموحد (اليونيكود) للتعرف على التفاصيل الفنية لنظام UTF-8 و UTF-16 و UTF-32، على <https://www.unicode.org/versions/Unicode12.0.0/ch03.pdf>.



أسماء النطاقات المدوّلة في التطبيقات IDNA - سرد تاريخي موجز وعرض للحالة الراهنة

تم تعريف أسماء النطاقات المدوّلة في التطبيقات (IDNA) في بادئ الأمر من خلال فريق عمل هندسة الإنترنت في عام 2003 فيما يعرف الآن باسم IDNA2003¹⁰. وقد اشتملت على خوارزمية لتخطيط نقاط أكواد يونيكود في نموذج معياري في مسميات أسماء النطاقات المعروفة باسم Nameprep، بالإضافة إلى خوارزمية لتشير مسميات نقاط رموز الكود الموحد (اليونيكود) في نظام ASCII المعروف باسم Punycode (بيونيكود). وتشتمل Nameprep على عمليات تحويل مثل تخطيط الحروف الكبيرة إلى الصغيرة.

وبعد قدر من الخبرات والتعرف على أسماء النطاقات المدوّلة في التطبيقات، قام فريق عمل هندسة الإنترنت بتطوير ونشر مواصفة منقحة معروفة باسم المعيار IDNA2008 في عام 2010¹¹. وقد أدى المعيار IDNA2008 إلى إنشاء مصطلحات مثل مسميات U ومسميات A وأدى إلى إزالة خطوة Nameprep، مشيرًا إلى أن التطبيقات يجب أن تقوم بعملية تخطيط مناسبة من أجل موقع وبيئة التطبيق. تم تحديث المعيار IDNA2008 ليناسب الكود الموحد (اليونيكود) الإصدار 6.0 من خلال طلب الحصول على الملاحظات والتعليقات رقم 6452 في 2011 واستمر فريق عمل هندسة الإنترنت في مراجعته.

ومن ناحية الممارسة العملية، لا تزال العديد من عمليات التنفيذ تستخدم المعيار IDNA2003. والقليل من المكتبات لا تستخدم قوائم (مثل القوائم المشمولة في المعيار IDNA2003) التي تم إنشاؤها من أجل المعيار IDNA2008. ولا توجد مخططات حسب الموقع من أجل المعيار IDNA2008 باستثناء القواعد المعيارية الخاصة بالتعاضدي عن حالة الأحرف والتطبيع المشمولة في معيار اليونيكود.

وثمة استثناء واحد وهو أن هناك القليل من التخطيط من UTS#46، معالجة توافق أسماء النطاقات المدوّلة في التطبيقات ليونيكود. فهذا يحدد ما إن كانت بضعة حروف مشتركة يتم تخطيطها في IDNA2003 لكنها مسموحة كحروف في المعيار IDNA2008 من الواجب قبولها أو وضع مخطط لها. ومن المهم بالنسبة للتطبيقات أن تتعامل مع هذه الحروف طبقًا للمعيار IDNA2008 وليس المعيار IDNA2003، وأنه في حال كان UTS#46 قيد الاستخدام أن يتم استخدامه بطريقة تكون متوافقة مع المعيار IDNA2008.

استخدام الحالات للاختبار

يجب أن تخضع البرمجيات المخصصة للتعامل مع أسماء النطاقات المدوّلة IDN وعناوين البريد EAI للاختبار مع مجموعة واسعة من أسماء النطاقات والعناوين. انظر UASG 004، استخدام الحالات من أجل تقييم جاهزية القبول الشامل، للتعرف على مجموعة من حالات الاختبار.

ترقية البرامج من أجل EAI

تتطلب مطابقة EAI ترقية لخواص البريد، وبرامج الإرسال والتسليم، ووكلاء مستخدمي البريد، وبريد الويب، وأي تطبيق يتعامل مع عناوين البريد الإلكتروني ويرسل البريد.

للحصول على نظرة عامة تفصيلية حول EAI ومشكلاتها وكيفية تنفيذها، راجع UASG 012، تدويل عناوين البريد الإلكتروني (EAI): نظرة عامة فنية.

¹⁰ وهذا التعريف وارد في طلبات الحصول على الملاحظات والتعليقات رقم 3490، ورقم 3491، ورقم 3492.
¹¹ التعريف وارد في طلبات الحصول على الملاحظات والتعليقات رقم 5890، ورقم 5891، ورقم 5892، ورقم 5893، ورقم 5894، ورقم 5895.



موضوعات متقدمة

النصوص المعقدة

قد تكون تفاصيل البرامج النصية المعقدة ذات أهمية محدودة لمن ليسوا مطورين يقومون بإنشاء تحليل السلاسل أو مكتبات العرض الخاصة بهم. ومع ذلك، تم هنا تضمين ملخص لضمان من أن جميع القراء لديهم وعي كافٍ للتعرف على الأخطاء البرمجية المتعلقة بهذه البرامج النصية عند مواجهتها في تجارب المستخدمين.

أما بالنسبة لنص HTML المنسق في صفحات الويب وفي البريد الإلكتروني، فإن معايير HTML تحظى بميزات تفصيلية لمعالجة وعرض النصوص المعقدة وثنائية الاتجاه، والتي يجب على المطورين فهمها واستخدامها لإنتاج نص. راجع القسم الخاص بمعايير WHATWG HTML حول الإنتاج¹²، والقسم المقابل له الخاص بمعيار W3C HTML¹³.

اللغات المكتوبة من اليمين إلى اليسار وتوافق الكود الموحد (اليونيكود)

بعض نصوص الكتابة مثل اللاتينية والديفغارية تعرض حروفاً من اليسار إلى اليمين عندما يتم عرض النص في خطوط أفقية. وهناك نصوص أخرى، مثل العربية والعبرية تعرض الحروف من اليمين إلى اليسار. كما يمكن أن يكون النص ثنائي الاتجاه عندما يستخدم نصاً من اليمين إلى اليسار حروفاً يتم كتابتها من اليسار إلى اليمين أو عندما تستخدم كلمات مضمنة من الإنجليزية أو غيرها من اللغات التي يتم كتابتها باستخدام نصوص من اليسار إلى اليمين.

يمكن أن تحدث تحديات وحالات غموض عندما يكون الاتجاه الأفقي للنص غير موحد. ولحل هذه المشكلة، ثمة خوارزمية لتحديد الاتجاه بالنسبة لنص الكود الموحد (اليونيكود) ثنائي الاتجاه.

وثمة مجموعة من القواعد التي يجب تطبيقها من خلال التطبيق لتقديم ترتيب صحيح في وقت العرض الموصوف من خلال خوارزمية الكود الموحد (اليونيكود) ثنائي الاتجاه. ونشير عموماً إلى هذا الأمر باسم "الخوارزمية ثنائية الاتجاه".

الخوارزمية ثنائية الاتجاه

تصف الخوارزمية ثنائية الاتجاه الطريقة التي يجب أن تتعامل بها البرمجيات مع النص الذي يحتوي على كل من تسلسل الحروف من اليسار إلى اليمين (LTR) ومن اليمين إلى اليسار (RTL). وسوف يحدد الاتجاه القاعدي¹⁴ المحدد للصياغة الترتيب الذي يتم به عرض النص. ويمكن أن يكون ذلك إما من اليسار إلى اليمين أو من اليمين إلى اليسار ويحدد الترتيب الذي يتم به عرض تسلسل الحروف. وفي هذه الوثيقة، فإن الاتجاه القاعدي من اليسار إلى اليمين ولذلك فإن جميع تسلسلات الحروف معروضة بالتسلسل الأول إلى اليسار من التسلسل الثاني.

ولمعرفة ما إن كان أي تسلسل من اليسار إلى اليمين أو من اليمين إلى اليسار، فإن كل حرف في الكود الموحد (اليونيكود) يحتوي على خاصية اتجاه مرتبطة به. علمًا بأن غالبية الأحرف يتم كتابتها (الحروف القوية) بصيغة LTR (من اليسار إلى اليمين) أو RTL (من اليمين إلى اليسار) استناداً إلى النص الذي تكون جزءاً منه. وأي تسلسل لحروف مكتوبة بنظام RTL بقوة سيتم عرضه من اليمين إلى اليسار. ويأتي ذلك مستقلاً عن الاتجاه القاعدي المحيط. فعلى سبيل المثال:

(من اليسار إلى اليمين) example - مثال (من اليمين إلى اليسار).

أما النصوص ذات الاتجاه المختلف فيمكن أن تكون مختلطة في السطر. وفي تلك الحالات، تقدم الخوارزمية ثنائية الاتجاه مساراً اتجاهياً منفصلاً عن كل سلسلة من الحروف المتاخمة ذات نفس الاتجاه.

¹² متاح على <https://html.spec.whatwg.org/multipage/rendering.html>

¹³ متاح على <https://www.w3.org/TR/2018/WD-html53-20181018/rendering.html>

¹⁴ أما في HTML، فإن الاتجاه القاعدي إما أن يكون متأصل من الاتجاه الافتراضي للوثيقة، والذي يكون من اليسار إلى اليمين، أو محدد صراحة من خلال أقرب عنصر أعلى يستخدم خاصية الاتجاه "dir".



أما المسافات وغالبية علامات الترقيم فلا يتم كتابتها بقوة بنسق LTR أو RTL في الكود الموحد (اليونيكود) لأنه يمكن استخدامها في أي من نوعي النصوص. ومن ثم يتم تصنيفها في فئة الحروف المحايدة أو الضعيفة. والحروف الضعيفة هي الحروف التي يتم استخدامها بشكل عام في اتجاه واحد، ولكن يمكن استخدامها في بعض السياقات في الاتجاه الآخر. وتشمل الأمثلة على هذا النوع من الحروف:

- الأرقام الأوروبية.
- الأرقام العربية-الهندية الشرقية.
- الرموز الحسابية ورموز العملات.
- رموز علامات الترقيم الشائعة في العديد من النصوص، مثل النقطتان المتعامدتان والفاصلة والنقطة والمسافة غير المقطوعة.

اتجاه الحروف المحايدة يعد غامضًا بدون السياق. وتشمل بعض الأمثلة ما يلي:

- علامات التبويب.
- فواصل الفقرات.
- غالبية حروف المسافات البيضاء الأخرى.

وعندما يكون حرف محايد بين حرفين مكتوبين بقوة ولهما نفس نوع الاتجاه، فسوف يأخذ ذلك الاتجاه هو الآخر. على سبيل المثال، أي حرف محاي بين حروف RTL سوف تتم معاملته معاملة حرف RTL نفسه، وسوف يكون له أثر تمديد طول الاتجاه:

- مثال. نطاق

حتى وإن كانت هناك الكثير من الحروف المحايدة بين حرفين مكتوبين بقوة، فسوف تتم معاملتهم جميعًا بنفس الطريقة.

وعندما تقع مسافة أو علامة ترقيم بين حرفين مكتوبين بقوة ولهما اتجاه مختلف، فإن الحرف (الحروف) المحايدة سوف تتم معالجتها كما لو كان لها نفس اتجاه الاتجاه القاعدي السائد. فعلى سبيل المثال:

- مثال. example.

لا تنس أن هذه الوثيقة لها اتجاه من اليسار إلى اليمين مثل تسلسلها القاعدي لذلك فإن كلمة *example* هي نطاق المستوى الثاني وكلمة مثال هي نطاق المستوى الأعلى.

وما لم يكن هناك إلغاء للاتجاه حاضرًا، فدائمًا ما يتم تشفير الأرقام وإدخالها في صورة رقم عالي الترتيب أولاً، ويتم تحويل الأرقام إلى LTR. وينطبق الاتجاه الضعيف فقط على وضع الأرقام في مجملها.

يشار إلى أن التفاصيل الكاملة للخوارزمية ثنائية الاتجاه موصوفة في التقرير الفني للكود الموحد (اليونيكود) رقم 9.

قاعدة الاتجاه الثنائي لأسماء النطاقات

اسم النطاق ثنائي الاتجاه هو ذلك الذي يحتوي على مسمى واحد على أقل تقدير بنظام RTL من اليمين إلى اليسار. وقاعدة الاتجاه الثنائي لأسماء النطاقات، المحددة في طلب الحصول على الملاحظات والتعقيبات رقم 5893¹⁵، تقيد نقاط الأكواد في الأسماء بحيث لا تكون هناك أسماء في صورة تسلسلات مختلفة من نقاط الأكواد لكن تعرض نفس الشيء بسبب قواعد العرض ثنائي الاتجاه.

وحدات الوصل

بعض اللغات تستخدم نصوصًا يتم فيها كتابة وحدة الكلام الفردية من خلال استخدام حرفين يطلق عليهما اسم صوت واحد من حرفين. وبمعنى آخر، فإن الصوت الواحد المكون من حرفين عبارة عن مجموعة من حرفين متتابعين يمثلان صوتًا واحدًا (أو مقطعًا صوتيًا).

¹⁵ نصوص الاتجاه من اليمين إلى اليسار المخصصة لأسماء النطاقات المدوّلة للتطبيقات (IDNA)، طلب الحصول على الملاحظات والتعقيبات رقم 5893، <https://www.rfc-editor.org/info/rfc5893>.



أمثلة على المقاطع الصوتية المكونة من حرفين في اللغة الإنجليزية

sh (في كلمة shoe) وأيضًا gh (في كلمة rough)	th (في كلمة then) وأيضًا th (في كلمة think)	ch (كما في كلمة church) وأيضًا ph (كما في كلمة phony)
---	---	---

وبعض المقاطع الصوتية ثنائية الأحرف تكون مرتبطة بالكامل في صورة حروف مزدوجة. وفي الكتابة والطباعة، تحدث الحروف المزدوجة عندما تمتاز وحدتان كتابيتان أو حرفان أو أكثر في شكل فردي منقوش. والمثال على ذلك حرف العطف (&)، والذي تطور من الحروف اللاتينية الملتصقة e والحرف t (أي "et" بمعنى "و"). وفي تنضيد الحروف الإنجليزية، يتم عرض كل من ffi و ffi في صورة حروف مزدوجة.

وإذا كان للحروف المزدوجة والمقاطع الصوتية ثنائية الأحرف نفس التفسير في جميع اللغات التي تستخدم نصًا محددًا، يقوم تطبيع الكود الموحد (اليونيكود) بحل الاختلافات ويجعلها متوافقة. وعندما يكون لها تفسيرات مختلفة، فيجب أن تستخدم المطابقة طرقًا بديلة (يتم اختيارها على الأرجح في مستوى السجل) أو يجب توجيه المستخدمين إلى فهم أن تلك المطابقة لن تحدث. وأحد الأمثلة على التفسير المختلف يمكن العثور عليه في القسم 4.3 من طلب الحصول على الملاحظات والتعليقات رقم 5894¹⁶. يدرج اتحاد يونيكود إستراتيجيتان رئيسيتان من أجل تحديد سلوك الانضمام لحرف ما محدد بعد تطبيق الخوارزمية ثنائية الاتجاه للتعامل مع حروف وحدة انضمام بدون أي عرض معروفة باسم ZWJ و ZWNJ. (لمعرفة المزيد حول هذه الروابط، راجع <http://www.unicode.org/L2/L2005/05307-zwj-zwnj.pdf>).

- عند الصياغة، يمكن أن يشير التنفيذ مرة أخرى إلى مخزن الدعم الأصلي للتعرف على ما إن كانت هناك أي حروف ZWNJ أو ZWJ متاخمة.
- وبدلاً من ذلك، يمكن للتطبيق أن يستبدل ZWJ و ZWNJ بخاصية حروف خارج الإطار تكون مرتبطة بتلك الحروف المتاخمة، بحيث لا تتداخل المعلومات مع الخوارزمية ثنائية الاتجاه ويتم حفظ المعلومات عبر إعادة ترتيب تلك الحروف. وبمجرد تطبيق الخوارزمية ثنائية الاتجاه، يمكن للمعلومات خارج الإطار تلك أن تستخدم بعد ذلك من أجل الصياغة الصحيحة.

سجلات أسماء النطاقات وأي كيان آخر يتيح إنشاء أسماء النطاقات (على سبيل المثال التطبيقات التي تقوم بإنشاء مسميات من المستوى الثالث والمنخفض) يجب أن تتبع قاعدة الاتجاه الثنائي لأسماء النطاقات لضمان أن الأسماء سوف تعرض بشكل متسق ولمنع الأسماء المسببة للبس والتي يمكن أن تستخدم في عمليات هجوم الألفاظ المجانسة.

ولمعرفة المزيد حول وحدات الوصل، راجع اقسام 4.3 في RFC طلب الحصول على الملاحظات والتعليقات رقم 5894.

الرموز الصوتية والحروف المماثلة

الحروف الصوتية عبارة عن حروف تبدو بسبب التشابهات في الحجم والشكل- متطابقة أو متشابهة لحد اللبس. وتحدث كثيرًا عند خلط النصوص اللاتينية والسيريلية والإغريقية. على سبيل المثال، الحرف اللاتيني "o" (الرمز U+006f)، والحرف السيريللي الصغير "o" (الرمز U+043e)، والحرف الخامس عشر الصغير في الأبجدية الإغريقية "o" (الكود U+03bf). وفي بعض الحالات، ثمة رموز صوتية في نص واحد، مثل الحرف الكرواتي الصغير "ja" (الرمز U+01c9) والحرفان "ja" (الرمز U+006c U+006a). راجع القائمة في <http://homoglyphs.net/> للحصول على مزيد من الأمثلة.

ولمنع أسماء النطاقات ذات الرموز الصوتية، يجب على السجلات استخدام قواعد استخراج المسميات (LGR) التي تقيد نقاط الرموز في أحد المسميات على مجموعة من النصوص الفردية أو النصوص المتوافقة. ويجب أن يكون لكل سجل قواعد استخراج المسميات LGR لكل نص يقبل من خلال التسجيلات¹⁷.

¹⁶ أسماء النطاقات المدوّلة في التطبيقات (IDNA): الخلفية والشرح والحيثيات، طلب الحصول على الملاحظات والتعليقات رقم 5894، <https://www.rfc-editor.org/rfc/rfc5894.html#section-4.2>

¹⁷ ولدى هيئة الإنترنت للأرقام المخصصة IANA مجموعة من قواعد استخراج المسميات LGR الخاصة بالسجلات في مستودع ممارسات أسماء النطاقات الخاص بها على <https://www.iana.org/domains/idn-tables>.



ولمعرفة المزيد حول آليات أمن الكود الموحد (اليونيكود) للاكتشاف المسبب للبس، راجع:

▪ http://www.unicode.org/reports/tr39/#Confusable_Detection

ولمعرفة المزيد حول الحروف المتشابهة لحد اللبس والممارسات الجيدة، راجع:

- نظرة عامة ودرس تعليمي حول إساءة استخدام الكود الموحد (اليونيكود) لنظام M3AAWG على <https://www.m3aawg.org/sites/default/files/m3aawg-unicode-tutorial-2016-02.pdf>
- أفضل ممارسات M3AAWG لمنع إساءة استخدام الكود الموحد (اليونيكود) <https://www.m3aawg.org/sites/default/files/m3aawg-unicode-best-practices-2016-02.pdf>

التطبيع وقواعد التغاضي عن حالة الأحرف وإعداد السلسلة

يساعد تطبيق الكود الموحد (اليونيكود) في تحديد ما إن كانت أي سلسلتان من الكود الموحد (اليونيكود) متكافئتان مع بعضهما الآخر أم لا كما يوفر معايير قياسية يمكن استخدامها لمعالجة السلاسل وتخزينها. يمكن تمثيل بعض الحروف في الكود الموحد (اليونيكود) من خلال العديد من تسلسلات الأكواد. ويطلق على هذا الأمر اسم تكافؤ الكود الموحد (اليونيكود). ويوفر الكود الموحد (اليونيكود) نوعان من التكافؤ:

- قانوني
- توافقي

تسمى المتواليات التي تمثل نفس الحرف المرئي مكافئاً قانونياً ولهذه التسلسلات نفس المظهر والمعنى عند طباعتها أو عرضها. فعلى سبيل المثال:

$\tilde{n} = U + 006E$ (الحرف اللاتيني الصغير "n") متبوعاً بـ 0303 + U (مسافة الدمج " ")

$\tilde{n} = U+00F1$ (حالة الحرف الصغير "ñ" في الأبجدية الإسبانية)

يحدد الكود الموحد (اليونيكود) NFC (نموذج التطبيع C) في شكل تفكيك قانوني، متبوعاً بتركيب قانوني. وهذا من شأنه تقليل النص إلى أقل عدد من نقاط الأكواد مع عدم تغيير مظهره. تجدر الإشارة إلى أنه في هذا المثال، فإن ثلاثة أحرف أعلاه من الصحيح استخدامها طبقاً للمعيار IDNA2008.

ومكافئات التوافق عبارة عن سلاسل يمكن أن تظهر بشكل مختلف، ولكن بنفس المعنى في بعض السياقات. وهو نوع أضعف من التكافؤ بين الحروف أو تسلسل الحروف. فعلى سبيل المثال:

$ff = U+FB00$ (الحروف المزدوجة الطباعية "ff")

$ff = U+0066 U+0066$ (حرفان "f" باللاتينية)

في المثال السابق، تم تحديد نقطة الرمز القيمة U+FB00 بأنها متوافقة، لكنها غير مكافئة قانونياً مع متتالية U+0066 U+0066. والمتتاليات المتكافئة قانونياً متوافقة أيضاً، لكن العكس غير صحيح دائماً.

وتجدر الإشارة إلى نقطة الكود U+FB00 غير صحيحة طبقاً للمعيار IDNA2008. ويحدد الكود الموحد (اليونيكود) NFKC (نموذج التطبيع KC) تفكيكاً للتوافق، متبوعاً بتجميع قانوني. وهذا من شأنه تقليل النص إلى نقاط أكواد حسب المعيار وربما يتغير في مظهره. على سبيل المثال، يحول NKFC الحرف المزدوج "ff" إلى الحرفين "f f" ورمز ما قبل الظهر am. أي (U+33C2) إلى أربعة حروف "a.m." وهي (U+002E U+006D U+002E U+0061).



ولتجنب مشكلات عدم القدرة على التشغيل المتبادل التي تنشأ عن استخدام متواليات حروف مكافئة قانونياً ولكنها مختلفة، توصي الجمعية العالمية لشبكة الويب باستخدام NFC لجميع النصوص.

للاطلاع على قائمة بجميع الحروف التي قد تتغير في أي نماذج تطبيع، راجع:
<http://www.unicode.org/charts/normalization>

بعض النقاط الأخرى التي يجب ملاحظتها:

- الحروف المستخدمة في مسميات أسماء النطاقات المدوّلة يجب أن تكون في صيغة NFC.
- وعندما يتشارك تطبيقان في بيانات كود موحد (اليونيكود)، لكن يقومان بتطبيعها بشكل مختلف، فقد تحدث أخطاء وضياع للبيانات.
- ويؤكد اتحاد الكود الموحد أن نماذج التطبيع يجب أن تظل مستقرة بمرور الوقت. وبمعنى آخر، يجب أن تظل السلسلة في وضعها الطبيعي في جميع الإصدارات المستقبلية من الكود الموحد (اليونيكود) من أجل التوافق مع الإصدارات السابقة.
- كما أشرنا سابقاً، كن متحفظاً عند النظر في نقاط الرموز التي يجب السماح بها في اسم نطاق.

ملاحظات لمطوري البرمجيات

لا تحاول التطبيع عن طريق التحويل إلى أحرف كبيرة أو تجاهل الأحرف غير التباعدية، لأن هذا يجعل الفرز واستيراد وتصدير نسخ البيانات، واسترداد البيانات بواسطة تطبيقات العميل أمراً صعباً وقد يؤدي إلى فقد البيانات أو تلفها.	✘
لا تسمح مطلقاً بنقاط رموز في أسماء النطاقات غير المسموح بها وفقاً لمعيار IDNA2008.	✘

لمعرفة المزيد حول تطبيع الكود الموحد (اليونيكود)، راجع:

- <http://www.w3.org/TR/charmod-norm>
- <http://unicode.org/reports/tr15>

قواعد التغاضي عن حالة الأحرف والتخطيط

التغاضي عن حالة الأحرف والتخطيط هي عملية تحويل كل الأحرف في سلسلة إلى نفس الحالة، وعادةً ما تكون حالة الأحرف الصغيرة. فتخطيط حالة الأحرف الكبيرة [A-Z] إلى الحالة الصغيرة [a-z] يفيد بالنسبة للمستندات النصية ذات نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII فقط، لكنه أكثر تعقيداً بكثير في اللغات التي تستخدم حروفاً تقليدية. ويمكن أن يكون تخطيط حالة الأحرف معتمداً على السياق، مع اعتماد الحرف المخطط على السياق الذي يحدث فيه، مثل، الأشكال المختلفة من الحرف الثاني عشر في الأبجدية اليونانية. ويمكن أيضاً أن يعتمد على الموقع، مع اعتماد الحرف المخطط على الموقع والمكان الذي يتم فيه تفسير النص، مثل، حرف | التركي المنقط وغير المنقط في الحالة الكبيرة والصغيرة. أما التغاضي عن حالة الأحرف فيعتمد على الموقع والمكان، بالنسبة للسلاسل التي سيتم تفسير عن طريق البرمجيات، في حين أن تعيين حالة الأحرف يعتمد على الموقع والمكان ويهدف إلى قراءة النص من قبل البشر. وأخيراً، فإن تعيين الحالة على الأحرف الكبيرة وتعيين الحالة على الأحرف الصغيرة ليستا وظيفيتين معكوستين.

أما بالنسبة لأسماء النطاقات المدوّلة، يتيح المعيار IDNA2008 للتطبيقات استخدام أي تعيين مناسب لحالة الأحرف لأن تعيين الحالة يتم قبل توثيق نقاط الأكواد. ومن الناحية العملية، فإن عمليات تعيين حالة المعرفات الخاصة بمكان ومنطقة محددة غير موجودة والجميع يستخدموا عمليات تعيين الحالة من UTS#46 في الكود الموحد (اليونيكود)¹⁸.

¹⁸UTS#46، معالجة التوافق لأسماء النطاقات المدوّلة في التطبيقات (IDNA) للكود الموحد (اليونيكود)،
<https://www.unicode.org/reports/tr46/#Mapping>



ملاحظات لمطوري البرمجيات

يجب مراعاة الهدف المرغوب قبل محاولة إجراء تعيين حالة الأحرف: هل هو مخطط عام للمسميات، أو سلسلة في لغة غير معروفة أو شيء آخر؟	✓
يجب إجراء تطبيع الكود الموحد (اليونيكود) قبل التغاضي عن حالة الأحرف.	✓

المسرد والموارد الأخرى

قاموس المصطلحات

الترميز بمنظومة الأسكي	التمثيل المشفر المتوافق (ACE) المتوافق مع نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII لأي مسمى في اسم نطاق مدول، يستخدم على المستوى الداخلي داخل بروتوكول نظام أسماء النطاقات. مسميات A دائماً ما تبدأ ببادئة الترميز بمنظومة الأسكي "xn--". يمكن تحويل تسمية A إلى تسمية U وإليها دون فقد المعلومات.
بادئة ACE	بادئة متوافقة مع نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII للتشفير "xn--".
نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII	نظام الترميز المعياري الأمريكي لتبادل المعلومات. يشتمل نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII على حروف لاتينية بدون تأكيد في نطاقها بالإضافة إلى أرقام أوروبية-عربية. نظام الترميز المعياري الأمريكي لتبادل المعلومات ASCII عبارة عن مجموعة فرعية من الكود الموحد (اليونيكود): كل حرف ASCII يعتبر أيضاً حرفاً من الكود الموحد (اليونيكود).
API	واجهات برمجة التطبيقات (API) عبارة عن مجموعة من البرامج الفرعية والبروتوكولات والأدوات لبناء البرمجيات والتطبيقات. وقد تكون واجهات برمجة التطبيقات للأنظمة المستندة إلى الويب أو أنظمة التشغيل أو أنظمة قواعد البيانات، كما أنها توفر وسائل لتطوير التطبيقات لذلك النظام من خلال استخدام لغة برمجة محددة.
مساحة الكود	نطاق يحدد قيود الحالة العليا والصغرى لأي ترميز.
نقطة الرمز	نقطة الرمز عبارة عن قيمة عددية في مساحة كود. يتم استخدام نقاط الكود لتمييز القيمة العددية عن الترميز الخاص بها في صورة سلسلة من البتات، ولتمييز حرف مجرد عن تمثيل رسومي معين لها (شكل منقوش).
منطقة جذر نظام أسماء النطاقات	منطقة الجذر هي الدليل المركزي لنظام أسماء النطاقات، وهي مكون رئيسي في البحث عن الأشياء في نظام أسماء النطاقات، على سبيل المثال، ترجمة أسماء المضيف إلى عناوين IP.
تدويل عناوين البريد الإلكتروني EAI	يتيح تدويل عناوين البريد الإلكتروني حروف UTF-8 في أي عنوان بريد إلكتروني— اسم النطاق أو الجزء المحلي أو كليهما.



<p>IANA</p> <p>Internet Assigned Numbers Authority (هيئة أرقام الإنترنت المخصصة). تشتمل وظائفها ما يلي:</p> <ul style="list-style-type: none">إدارة جذر نظام أسماء النطاقات، نطاق .int. ونطاق .arpa، ومصدر موارد .IDN.تنسيق المجموعة العامة من أرقام IP وAS، وتزويدهما بالأساس بسجلات الإنترنت الإقليمية (RIR).تدار نظم ترقيم بروتوكولات الإنترنت بالاشتراك مع هيئات المعايير.	
<p>ICANN</p> <p>تتمثل مهمة ICANN في ضمان شبكة إنترنت عالمية مستقرة وأمنة وموحدة. فلأجل الوصول الى شخص آخر عبر الإنترنت ينبغي عليك أن تقوم بكتابة عنوان - اسم أو رقم - من خلال حاسوبك أو أي جهاز آخر. يجب أن يكون هذا العنوان فريداً من نوعه لكي يتسنى للكمبيوترات أن تجد بعضها البعض. تقوم ICANN بتنسيق تلك المعرفات الفريدة من نوعها عبر كافة أنحاء العالم. تقوم ICANN تم تأسيس مؤسسة ICANN في عام 1998 كشركة غير ربحية ذات منفعة عامة ومجتمع من المشاركين من كافة أنحاء العالم.</p>	
<p>IDN</p> <p>اسم النطاق المدوّل. أسماء النطاقات المدوّلة IDN عبارة عن أسماء نطاقات تشمل حروف UTF-8 تتخطي الحروف الستة والعشرين في الأبجدية اللاتينية الأساسية "a-z"، والأرقام 0-9 والشرطة "-".</p>	
<p>IDNA</p> <p>أسماء النطاقات المدوّلة في التطبيقات.</p>	
<p>IDN ccTLD</p> <p>نطاق من المستوى الأعلى برمز دولة ويشتمل على حروف تتخطي الحروف الستة والعشرين في الأبجدية اللاتينية الأساسية "a-z".</p> <p>أمثلة:</p> <ul style="list-style-type: none">.pڤ (روسيا).مصر (مصر).السعودية (المملكة العربية السعودية)	
<p>IETF</p> <p>يمثل فريق عمل هندسة الإنترنت (IETF) مجتمعاً عالمياً كبيراً مفتوحاً، يتكون من مصممي الشبكات ومشغليها وموزعيها والباحثين ذوي الاهتمامات بتطوير هندسة الإنترنت والتشغيل السلس للإنترنت. فهو مجتمع مفتوح لكل مهتم بذلك. كما يقوم فريق عمل هندسة الإنترنت IETF على تطوير معايير الإنترنت، وبصورة خاصة، المعايير ذات الصلة بحزمة بروتوكول الإنترنت (TCP/IP) والبروتوكولات المستخدمة للويب مثل HTTP و TLS.</p>	
<p>اللغة</p> <p>طريقة تواصل البشر، سواء المنطوقة أو المكتوبة، وتتألف من استخدام الكلمات بطريقة مركبة وتقليدية.</p>	
<p>Punycode (بيونيكود)</p> <p>عبارة عن خوارزمية تمثل UTF-8 في مجموعة فرعية ذات أحرف محدودة من ASCII المدعومة من نظام أسماء النطاقات (DNS). يستخدم Punycode (بيونيكود) في تسميات A في إطار أسماء النطاقات الدولية في إطار عمل (IDNA) الخاص بالتطبيقات.</p>	
<p>أمين السجل</p> <p>منظمة يتم فيها تسجيل أسماء النطاقات من قبل المستخدمين. يحتفظ أمين السجل بسجلات تضم معلومات الاتصال ويقوم بتقديم المعلومات التقنية إلى سجل مركزي يعرف باسم "السجل".</p>	
<p>السجل</p> <p>قاعدة البيانات الرئيسية الرسمية لجميع أسماء النطاقات المسجّلة في كل نطاق من المستوى الأعلى (TLD).</p>	



RFC	طلب الحصول على الملاحظات والتعقيبات (RFC) عبارة عن وثيقة رسمية من فريق عمل هندسة الإنترنت (IETF) تأتي نتيجة صياغة لجنة ومراجعة تالية من الأطراف المعنية. بعض طلبات الحصول على الملاحظات والتعقيبات (وليس جميعها) تقوم بتوثيق معايير الإنترنت المعتمدة.
النص	مجموعة الحروف أو الأحرف المستخدمة في الكتابة، والتي تمثل أصوات لغة ما.
اسم نطاق في المستوى الثاني	الترتيب الهرمي لنظام أسماء النطاقات (DNS)، نطاق من المستوى الثاني (SLD) أو (2LD) وهو نطاق أقل بشكل مباشر من نطاق المستوى الأعلى (TLD). على سبيل المثال،، في example.com، فإن example هو نطاق المستوى الثاني لنطاق المستوى الأعلى .com.
U-label	تسمية U هي سلسلة IDNA صالحة مكونة من رموز يونيكود، تتضمن رمزًا واحدًا على الأقل خلاف رموز ASCII. ويمكن تحويلها إلى تسمية A وبالعكس دون فقد المعلومات.
البرامج ذات جاهزية القبول الشامل أو جاهزية القبول الشامل	البرامج ذات القدرة على قبول وتخزين ومعالجة والتحقق من وعرض جميع نطاقات المستوى الأعلى، وأسماء النطاقات المدوّلة وعاوين البريد الإلكتروني على قدم المساواة.
Unicode	معياري ترميز عالمي للحروف. إنه يحدد الطريقة التي يتم بها تمثيل الأحرف الفردية في الملفات النصية وصفحات الويب وأنواع المستندات الأخرى. تم تصميم الكود الموحد (اليونيكود) لدعم الأحرف من جميع اللغات في جميع أنحاء العالم. يمكن أن تدعم ما يقرب من 1,000,000 حرف. انظر: http://unicode.org
UTF	تنسيق تحويل الكود الموحد (اليونيكود). عبارة عن طريقة لتمثيل نقاط رمز الكود الموحد (اليونيكود) كمسار من البايتات. UTF-8 هي نسق UTF المفضل للتعامل مع EAI و IDN. حيث تقوم UTF-8 بتحويل الكود الموحد (اليونيكود) إلى بايتات مكونة من 8 بت.
M3AAGW	مجموعة العمل المختصة بمكافحة الرسائل والبرمجيات الخبيثة وسوء استخدام الهواتف النقالة (M ³ AAWG) وهي المكان الذي تلتقي فيه الصناعة من أجل العمل ضد شبكات بوت نت والبرامج الضارة والرسائل العشوائية والفيروسات وهجمات رفض الخدمة وغيرها من أشكال الاستغلال على الإنترنت. انظر: https://www.m3aawg.org/
W3C	الاتحاد العالمي لشبكة الويب (W3C) عبارة عن جمعية دولية تقوم فيها المنظمات الأعضاء، بالإضافة إلى فريق عمل بالدوام الكامل، والجمهور بالعمل معًا من أجل وضع معايير الويب مثل HTML. انظر: https://www.w3.org/
WHATWG	مجموعة عمل تكنولوجيا تطبيقات النص التشعبي للويب (WHATWG) عبارة عن مجموعة من الأشخاص المهتمين بتطوير الويب من خلال معايير واختبارات. وقد تم تأسيس WHATWG من قبل أفراد من شركة Apple وشركة Mozilla Foundation وشركة Opera Software في عام 2004، بعد ورشة عمل الجمعية العالمية لشبكة الويب. انظر https://whatwg.org/
ZWJ	الواصله بدون أي عرض عبارة عن حرف غير طباعي يستخدم في التنضيد المحوسب لبعض النصوص، ويشمل ذلك العربية وجميع النصوص ذات المنشأ الهندي. وعند وضعها بين حرفين لا يمكن ربطهما بخلاف ذلك ، تؤدي واصلة ZWJ إلى طباعتها في نسق متصل.



ZWNJ

وحدة غير واصلة بدون أي عرض عبارة عن حرف غير طباعي يستخدم في حوسبة أنظمة الكتابة التي تستغل الحروف المزدوجة. بالنسبة لبعض اللغات والنصوص، ترتبط العديد من حروف الأبجدية بشكل طبيعي بالحرف التالي عند كتابتها بكلمة، بما يكون حرفاً مزدوجاً. من أجل عرض بعض البادئات واللاحقات والكلمات المركبة بشكل صحيح، يتم استخدام وحدة ZWNJ لتجاوز هذا السلوك الطبيعي المتمثل في ضم الحروف ومنعها من الانضمام إلى الحرف التالي (ولكن دون إضافة مسافة بين الاثنین).

للحصول على مسرد ICANN الكامل، انتقل إلى: <https://www.icann.org/icann-acronyms-and-terms/>.

طلبات الحصول على الملاحظات والتعليقات والمعايير الأساسية

طلبات الحصول على الملاحظات والتعليقات لأسماء النطاقات المدوّلة

<p>Punycode (يونيكود): ترميز Bootstring للكود الموحد (اليونيكود) لأسماء النطاقات المدوّلة في التطبيقات (IDNA)</p> <p>يصف طلب الحصول على الملاحظات والتعليقات رقم 3492 كود Punycode (بيونيكود) بما يلي:</p> <p>"عبارة عن قواعد بناء ترميز نقل بسيط وفعال مصمم للاستخدام مع أسماء النطاقات المدوّلة في التطبيقات (IDNA)"</p> <p>يحول Punycode (بيونيكود) بشكل فريد وبما يمكن عكسه سلسلة الكود الموحد (اليونيكود) إلى سلسلة ASCII. ويحدد طلب الحصول على الملاحظات والتعليقات هذا خوارزمية عامة يطلق عليها اسم Bootstring. تتيح هذه الخوارزمية سلسلة من نقاط الرموز الأساسية من أجل تمثيل أي سلسلة من نقاط الرموز مستمدة من مجموعة أكبر تمثيلاً فريداً.</p> <p>https://tools.ietf.org/html/rfc3492</p>	<p>RFC 3492</p>
<p>أسماء النطاقات المدوّلة في التطبيقات (IDNA): التعريفات وإطار عمل الوثائق</p> <p>يصف طلب الحصول على الملاحظات والتعليقات سياق وبروتوكول الاستخدام لمراجعة أسماء النطاقات المدوّلة في التطبيقات (IDNA).</p> <p>https://tools.ietf.org/html/rfc5890</p>	<p>RFC 5890</p>
<p>بروتوكول أسماء النطاقات المدوّلة في التطبيقات (IDNA)</p> <p>يحدد طلب الحصول على الملاحظات والتعليقات هذا آلية البروتوكول، التي تسمى أسماء النطاقات المدوّلة في التطبيقات (IDNA)، لتسجيل أسماء النطاقات المدوّلة IDN والبحث عنها بطريقة لا تتطلب تغييرات على نظام اسم النطاق DNS نفسه.</p> <p>https://tools.ietf.org/html/rfc5891</p>	<p>RFC 5891</p>



<p>نقاط الكود الموحد (اليونيكود) وأسماء النطاقات المدوّلة في التطبيقات (IDNA)</p> <p>يحدد طلب الحصول على الملاحظات والتعليقات رقم 5892 قواعد لتقرير ما إذا كانت القيمة الرقمية للترميز، التي تُعتبر بمعزل عن غيرها أو في سياقها، مرشحة لإدراجها في اسم النطاق المدوّل (IDN).</p> <p>https://tools.ietf.org/html/rfc5892</p>	<p>RFC 5892</p>
<p>النصوص المكتوبة من اليمين إلى اليسار لأسماء النطاقات المدوّلة في التطبيقات (IDNA)</p> <p>يوفر طلب الحصول على الملاحظات والتعليقات قاعدة جديدة للنظام ثنائي الاتجاه لمسميات أسماء النطاقات المدوّلة للتطبيقات (IDNA)، من أجل استخدام النصوص المكتوبة من اليمين إلى اليسار في أسماء النطاقات المدوّلة.</p> <p>https://tools.ietf.org/html/rfc5893</p>	<p>RFC 5893</p>
<p>أسماء النطاقات المدوّلة في التطبيقات (IDNA): الخلفية والشرح والحيثيات</p> <p>توفر هذه الوثيقة التعريفية نظرة عامة على نظام منقح من أجل التعامل مع إصدارات أحدث من الكود الموحد (اليونيكود) وتوفر مواد شرح لمكوناته.</p> <p>https://tools.ietf.org/html/rfc5894</p>	<p>RFC 5894</p>
<p>تخطيط الرموز لأسماء النطاقات المدوّلة في التطبيقات (IDNA) 2008</p> <p>يصف طلب الحصول على الملاحظات والتعليقات هذا الإجراءات التي يمكن اتخاذها من خلال التنفيذ ما بين تلقي إدخال المستخدم وتمرير القيم الرقمية للترميز المسموح بها إلى بروتوكول أسماء النطاقات المدوّلة في التطبيقات IDNA الجديد (2008). ويصف العملية التي سيتم تطبيقها على إدخال المستخدم من أجل إعداد إدخال المستخدم للاستخدام في بروتوكول "على الشبكة". كما يشتمل على إجراءات تنفيذ عامة لتخطيط الأحرف.</p> <p>https://tools.ietf.org/html/rfc5895</p>	<p>RFC 5895</p>
<p>طلبات الحصول على الملاحظات والتعليقات لنظام EAI</p>	
<p>نظرة عامة على البريد الإلكتروني المدوّل وإطار عمله</p> <p>يقدم هذا المعيار سلسلة مواصفات تحدد الآليات وامتدادات البروتوكول اللازمة لدعم عناوين البريد الإلكتروني المدوّلة بشكل كامل. يصف هذا المستند كيفية توافق العناصر المختلفة لتدويل البريد الإلكتروني معاً والعلاقات بين المواصفات الأساسية المرتبطة بنقل الرسائل وتنسيقات العنوان والمعالجة.</p> <p>https://tools.ietf.org/html/rfc6530</p>	<p>RFC 6530</p>
<p>امتداد بروتوكول نقل البريد البسيط (SMTP) للبريد الإلكتروني المدوّل</p> <p>تحدد الوثيقة امتداد بروتوكول نقل البريد البسيط بحيث يمكن للخوادم الإعلان عن القدرة على قبول ومعالجة عناوين البريد الإلكتروني المدوّلة وعناوين البريد الإلكتروني المدوّلة.</p> <p>https://tools.ietf.org/html/rfc6531</p>	<p>RFC 6531</p>



<p style="text-align: center;">عناوين البريد الإلكتروني المدول</p> <p>تحدد هذه الوثيقة تحسيناً لتنسيق رسالة الإنترنت ولامتدادات البريد الإلكتروني متعدد الأغراض MIME الذي يسمح باستخدام الكود الموحد (اليونيكود) في عناوين البريد ومعظم محتوى حقول الترويسة. يحدد هذا المستند تعزيزاً لتنسيق رسائل الإنترنت (RFC 5322) ولنظام MIME الذي يسمح بتوجيه استخدام UTF-8، بدلاً من ASCII فقط في قيم حقول العنوان، بما في ذلك عناوين البريد. يتم تعريف نوع وسائط جديد، رسالة/عمومي، للرسائل التي تستخدم هذا التنسيق الموسع. كما ترفع هذه المواصفة حظر MIME على الحصول على تشفيرات نقل محتوى غير هوية على أي نوع فري من نوع المستوى الأعلى للرسالة بحيث يمكن نقل تلك الأجزاء في الرسالة/العمومية بأمان عبر البنية التحتية الحالية للبريد</p> <p style="text-align: center;">https://tools.ietf.org/html/rfc6532</p>	<p style="text-align: center;">RFC 6532</p>
<p style="text-align: center;">إشعارات حالة التسليم المدولة والترتيب</p> <p>وتضيف هذه المواصفة نوعاً جديداً من العناوين لعناوين البريد الإلكتروني المدولة بحيث يمكن الاحتفاظ بعنوان المستلم الأصلي برموز غير ASCII بشكل صحيح حتى بعد الخفض. ويوفر هذا أيضاً أنواع وسائط لإرجاع المحتوى محدثة لإشعارات حالة التسليم وإشعارات ترتيب الرسائل لدعم استخدام نوع العنوان الجديد.</p> <p style="text-align: center;">https://tools.ietf.org/html/rfc6533</p>	<p style="text-align: center;">RFC 6533</p>
<p style="text-align: center;">عناوين البريد الإلكتروني المدولة في شهادات X.509</p> <p>تُعرّف هذه الوثيقة نموذج اسم جديد لإدراجه في حقل الاسم الآخر لامتداد الاسم البديل للموضوع X.509 واسم الاسم البديل للمصدر الذي يسمح لموضوع شهادة بأن يرتبط بعنوان بريد إلكتروني مدول.</p> <p style="text-align: center;">https://tools.ietf.org/html/rfc8398</p>	<p style="text-align: center;">RFC 8398</p>
<p style="text-align: center;">تحديثات التدويل على طلب الحصول على الملاحظات والتعليقات رقم 5290</p> <p>توفر التحديثات على RFC 5280 الموضحة في هذه الوثيقة التوافق مع مواصفات 2008 لأسماء النطاقات المدولة (IDN) وإضافة دعم لعناوين البريد الإلكتروني المدولة في شهادات X.509.</p> <p style="text-align: center;">https://tools.ietf.org/html/rfc8399</p>	<p style="text-align: center;">RFC 8399</p>

المعايير الأساسية



<p>لتوفير أساس تقني مشترك لمعالجة المعلومات الإلكترونية بلغات مختلفة، وضعت المنظمة الدولية للتوحيد القياسي (ISO) معيارًا دوليًا للترميز يسمى ISO 10646. ويوفر ISO 10646 معيارًا موحدًا لترميز الأحرف في جميع اللغات الرئيسية في العالم، بما في ذلك حروف اللغة الصينية التقليدية والمبسطة. كما تسمى مجموعة الأحرف الكبيرة هذه مجموعة الأحرف العالمية (UCS). ويتم تعريف نفس مجموعة الأحرف بواسطة معيار يونيكود، الذي يحدد خصائص الأحرف الإضافية وتفصيل التطبيق الأخرى ذات الأهمية الكبرى للمنفذين.</p> <p>واليونيكود عبارة عن نظام لترميز الأحرف تم تصميمه بواسطة اتحاد الكود الموحد لدعم تبادل النصوص المكتوبة لجميع اللغات الرئيسية في العالم ومعالجتها وعرضها. ويحدد كل من المنظمة الدولية للمعايير 10646 والكود الموحد (اليونيكود) العديد من نماذج الترميز في الأدوار المشتركة الخاصة بكل منها: UTF-8 و UCS-2 و UTF-16 و UCS-4 و UTF-32.</p> <p>http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=63182</p>	<p>ISO 10646 (يونيكود)</p>
<p>GB 18030-2000 عبارة عن خاص بالحوكمة الصينية يحدد صفحة كود ممتدة للاستخدام في السوق الصينية بالإضافة إلى UTF-8. رمز المعالجة الداخلية لمجموعة الأحرف التي يمكن ويجب أن تكون في الكود الموحد (اليونيكود)؛ ومع ذلك، ينص المعيار على أن مقدمي البرامج يجب أن يضمنوا رحلة ذهاب وإياب ناجحة بين GB18030 ورمز المعالجة الداخلية. يجب على جميع المنتجات التي يتم بيعها حاليًا أو المراد بيعها في الصين أن تخطط لنقل صفحة الكود لدعم GB18030 دون استثناء. إن GB18030 عبارة عن "معيار إلزامي" وتقوم الحكومة الصينية بتنظيم عملية التصديق لتعزيز نشر GB18030.</p> <p>http://icu-project.org/docs/papers/unicode-gb18030-faq.html</p>	<p>GB18030 (الصين)</p>

الموارد على الإنترنت

<p>واجهات برمجة التطبيقات (API)</p> <p>https://www.msdn.microsoft.com/enus/library/windows/desktop/ff818516%28v=vs.85%29.aspx</p> <p>واجهات برمجة تطبيقات SharePoint</p> <p>https://msdn.microsoft.com/en-us/library/office/jj860569.aspx</p> <p>قائمة اللاهقات العامة</p> <p>https://publicsuffix.org/list/public_suffix_list.dat</p> <p>قائمة نطاقات المستوى الأعلى المعتمدة من ICANN</p> <p>http://data.iana.org/TLD/tlds-alpha-by-domain.txt</p> <p>واجهات برمجة تطبيقات Android</p> <p>http://developer.android.com/guide/index.html</p> <p>واجهات برمجة تطبيقات MAC IOS</p> <p>https://developer.apple.com/library/mac/navigation</p> <p>Net Framework.</p>	<p>APIs</p>
--	--------------------



<p>https://msdn.microsoft.com/en-us/library/system.text.encoding(v=vs.110).aspx</p>	
<p>اعتبارات أمن اليونيكود Unicode http://www.unicode.org/reports/tr36</p> <p>آليات أمن الكود الموحد (اليونيكود) http://www.unicode.org/reports/tr39</p>	أمن الكود الموحد (اليونيكود)
<p>حدود رموز الكود الموحد (اليونيكود) https://www.unicode.org/versions/Unicode12.0.0/ch02.pdf؛ الصفحات 44-54</p> <p>نظرة عامة على GB18030 http://icu-project.org/docs/papers/gb18030.html</p> <p>جدول التعيين الموثق بين BG18030-2000 والكود الموحد (اليونيكود) http://source.icu-project.org/repos/icu/data/trunk/charset/data/xml/gb-18030-2000.xml</p> <p>تطبيع الكود الموحد (اليونيكود) https://unicode.org/reports/tr15/</p>	مجموعات أحرف يونيكود
<p>القسم 3.1، "UTF-8 حالات استغلال" في التقرير الفني للكود الموحد (اليونيكود) رقم 36 http://unicode.org/reports/tr36/#UTF-8_Exploit</p> <p>أفضل ممارسات M3AAGW لمنع انتهاك الكود الموحد (اليونيكود) https://www.m3aawg.org/sites/default/files/m3aawg-unicode-best-practices-2016-02.pdf</p> <p>نظرة عامة ودرس تعليمي حول إساءة استخدام الكود الموحد (اليونيكود) من M3AAGW https://www.m3aawg.org/sites/default/files/m3aawg-unicode-tutorial-2016-02.pdf</p> <p>راجع أيضاً: http://www.unicode.org</p>	حالات استغلال الكود الموحد (اليونيكود)
<p>روابط URI http://tools.ietf.org/html/rfc3986</p> <p>نظام أسماء النطاقات: شرح غير تقني—ما أهمية الحل الشامل http://www.internic.net/faqs/authoritative-dns.html</p> <p>مسرد ICANN https://www.icann.org/icann-acronyms-and-terms/</p>	متفرقات



هل تحتاج إلى المزيد من المعلومات؟

إن مجموعة توجيه القبول الشامل (UASG) والمجتمع متفرغان لتقديم النصائح إلى مطوري البرمجيات والمنفذين.

👍 اتصل بنا لإطلاعنا على ما لديك من أفكار واقتراحات حول الموضوع على [.info@uasg.tech](mailto:info@uasg.tech)

👍 انضم إلى قائمة مناقشات القبول الشامل على [.http://tinyurl.com/ua-discuss](http://tinyurl.com/ua-discuss)

👍 من أجل معرفة المزيد حول الجهود، تفضل بزيارة [.http://www.icann.org/universalacceptance](http://www.icann.org/universalacceptance)