

命名国际化电子邮箱的考虑事项

2020 年 11 月 2 日



目录

简介	3
执行摘要	4
需考量的政策主题	5
支持的文字	5
邮箱名称字符串的长度	5
文字混合	6
避免无效字符串和无法稳定呈现的字符串	8
从右到左 (RTL) 书写文字考虑事项	9
别名和显示名称考虑事项	11
标记和符号	12
统一码字符规范化	12
等效项考虑事项	13
其他考虑事项	15
术语表	16



简介

本文档阐述了关于命名国际化电子邮箱的考虑事项，目标读者为电子邮件系统管理员。电子邮件标准 [RFC6533 国际化电子邮件传递状态与处理通知](#)和 [RFC6430 电子邮件反馈报告类型](#) 值：非垃圾邮件现在允许邮箱名称（电子邮件地址中“@”符号前面的部分）和域名（“@”符号后面的部分）使用几乎所有语言或书写系统，而不再是仅局限于字母 A-Z 和数字 0-9。

对于使用且更喜欢使用非英语语言的用户，以及与这些用户进行邮件往来的用户而言，这是一项重大改进。但是，在允许使用的字符增多后，可能会导致混淆、错误和恶意滥用，为此，系统管理员需要制定政策来规范邮箱名称，以最大程度地发挥字符选择范围增大所带来的好处。

本文档的目标读者为希望为其用户配置国际化邮箱和/或希望配置并管理与国际化电子邮件地址兼容的系统的电子邮件系统管理员。本文档将帮助这些管理员在制定政策时提出适当的问题并做出正确的选择。

在本文档中，已默认假定读者熟悉电子邮件系统，且具有国际化电子邮件地址 (EAI) 第 1 阶段的经验，即，能够与国际化电子邮件地址相互收发电子邮件。本文档解决了 EAI 第 2 阶段的问题，例如，命名和托管国际化邮箱。

“术语表”部分提供了本文档中所用专业术语和缩略语的定义。



执行摘要

本文档的目标读者为电子邮件管理员、系统管理员和 IT 经理。本文档旨在汇总并阐述在创建国际化电子邮件地址邮箱名称、别名和显示名称时，在实际操作和政策方面需考量的主要事项。

邮箱名称（电子邮件地址中“@”符号前面的部分）对于电子邮件系统的实用性、用户友好性和安全性非常重要，因此，电子邮件管理员有必要采取行之有效的邮箱命名政策。电子邮件地址过去一直局限于拉丁字母和数字，但随着电子邮件技术的发展，现在已允许采用几乎所有语言和文字来书写邮箱名称和域名。这种灵活性使得选择命名政策变得更加复杂。为此，本文档旨在为电子邮件管理员提供指导，帮助他们调整邮箱命名政策，以涵盖使用非拉丁字母和数字的电子邮件地址。至于最终要选择什么样的政策，以及允许邮箱名称使用哪些文字（语言的书写系统）和使用多少个字符，将取决于电子邮件系统的业务目的，以及用户及其联系人的语言需求。

由于某些拼写、字符组合、标记或符号可能造成混淆或具有欺骗性，管理员采取的政策应禁止这些元素（相关细则则取决于邮箱名称所使用的语言）。有些名称在用户看来是一样的，但实际上却可能由不同的字符组成，此类技术问题需要特别注意。此外，从右到左 (RTL) 的文字可能会产生一些易混淆的特殊名称拼写，因此，如果系统允许使用此类文字，那么管理员还应解决额外的政策问题。

进行邮件往来时，收件人可能会在阅读电子邮件地址中使用的特定文字时感到困难。为了减少此类困难，建议实施相应政策，以允许用户选择显示名称（显示的个人姓名）以及使用别名（同一用户的备用地址）。

本文档还包含一些资源和参考资料，这些资源和参考资料提供了有关其中部分政策主题的更多详细信息。此外，本文档还包含一个术语表，在采用多种文字和语言命名邮箱时，管理员可能对有些术语并不熟悉。



需考量的政策主题

本文档阐述了可能需要包含在邮箱命名政策中的主题。每一部分都按如下方式介绍一个主题：

- 主题描述。
- 从以下几方面对考虑事项进行阐述：电子邮件系统环境、用户群和用户群需求。
- 政策建议（在某些情况下）。
- 应采取措施的检查清单。

支持的文字

设置电子邮件系统时，电子邮件管理员需要决定将哪些文字设为系统的默认文字。

此处的“文字”一词采用的是其在语言学中的含义，其定义为：用于书写一种或多种语言的规则和字符集。例如，拉丁文字是由字母 **A-Z** 和其他字符构成的集合，通常用于书写英语、法语，以及欧洲、北美洲、南美洲和非洲的许多其他语言。阿拉伯文字则用于书写中东、北非和南亚地区的多种语言。

考虑事项：

- 为了实现向后兼容性，大多数系统都将包含 **ASCII** 码点。**EAI** 允许管理员使用其他文字。

检查清单

[] 调查用户和目标市场的期望，以了解电子邮件地址中邮箱名称和域名部分所需使用的书写系统。

[] 查阅本文档，以了解提供其他文字可能引发的复杂问题（例如，可能引起的混淆、错误和恶意滥用），以及可能会对用户产生的任何影响。

[] 根据用户的期望，列出允许在邮箱名称中使用的文字。

邮箱名称字符串的长度

需考虑邮箱名称的长度（以字符为单位）。您希望允许的最小和最大长度分别是多少？

考虑事项：

- 邮箱名称字符串可能存在系统限制。虽然电子邮件标准允许一个邮箱名称最多包含 **64** 个八进制数，但有些系统会将字符串限制为少于标准字符限制。
- 对用户而言，短名称的好处在于更易记住和书写。而长名称也有好处，因为它们更具体，且不太可能与其他邮箱名称相冲突。您可能已经制定了一项政策，要求根据用户的个人姓名命名邮箱，例如，使用由“.”分隔的个人名字和姓氏作为邮箱名称。
- 可以考虑保留现有政策，但应扩大其涵盖范围，以包含更多字符和语言。此外，还应考虑用户及其联系人能够理解、记住并在电子邮件地址字段中输入的语言和字符。
- 当使用 **UTF8** 添加更多语言和文字时，一个字符可能会用多个八进制数来表示，这将增加邮箱名称的长度。例如，对于一个由 **ASCII** 字母构成且需要 **20** 个八进制数的邮箱名称而言，当使用另一种语言或文字来表示 **20** 个字母时，该邮箱名称的长度可能会增加到 **60** 个八进制数。



建议：无。

检查清单：

[] 了解系统限制。

[] 了解用户对国际化邮箱名称的预期。

[] 考虑在国际化邮箱名称长度方面应用与传统邮箱名称相同或相似的政策。

文字混合

文字混合是指不同于常规书写方式，在一个邮箱名称中使用来自多种文字的字符。尤其是一些来自其他文字的字符可能与主要文字中的字符非常相似，因而可能会导致混淆或恶意滥用。

考虑事项：

- 文字混合会引发安全问题以及用户混淆问题。例如，下图显示了几个看起来相似但实际不同的统一码 (Unicode) 码点，这些码点可能会导致用户混淆¹。

a a a a a a a a a a a a
b b b b b b b b b b b b
c c c c c c c c c c c c
d d d d d d d d d d d d
e e e e e e e e e e e e

- 您的联系人会将邮箱名称和域名写在一起，作为电子邮件地址。如果域名采用的是某种特定文字，那么邮箱名称自然会采用同一种文字。此时，应考虑是否可将任何其他文字与域名所用文字结合使用。
- 一些文字组合可能是有必要的，但有些组合可能会导致混淆或者恶意滥用。
- 在一些文化中，书写中同时采用多种文字是很常见的。例如，大多数日语文本都是由日语汉字、日语平假名、日语片假名和拉丁文字（“罗马”文字）混合而成的。然而，很少在日语文本中看到阿拉伯文字。
- 使用拉丁文字的邮箱名称通常不区分大小写。换言之，大写字母和小写字母是一样的，没有任何区别。对于您使用的文字，请考虑是否还有其他您的用户可能会使用的

¹ 资料来源：<https://thenextweb.com/wp-content/blogs.dir/1/files/2014/08/confusables-3.png>



等效文字。澄清一点，我们通常不将大写字母和小写字母的混合视为文字混合；有关更多详细信息，请参阅[等效项考虑事项](#)部分。

- 在域名系统 (DNS) 中，国际化域名 (IDN) 可以用两种格式表示，即：U-标签和 A-标签。在本文档中，不认为使用 A-标签属于文字混合；有关更多详细信息，请参阅[别名考虑事项](#)部分。

建议：

- 为了最大程度地提高可用性，建议电子邮件地址中的邮箱名称与域名部分使用相同的文字，除非您的组织为实现业务目标而另有要求。
- 避免在邮箱名称中混合使用采用同一种文字书写的多种语言，进而避免在邮箱名称中使用或产生看起来相似但实际采用的是不同码点的字符。
- 对于采用某种特定语言的邮箱名称，只应允许在邮箱名称中使用可在该语言的常规书写中一起使用的字符。在邮箱名称中，应禁止混合使用不同文字。例如，阿拉伯语的邮箱名称应只允许包含阿拉伯文字符，而日语邮箱名称可以允许包含日语汉字、日语平假名、日语片假名和拉丁文字符。

示例 1：使用阿拉伯文字的电子邮件地址：مستخدم@رسيل.السعودية

示例 2：混合了不同文字的日文电子邮件地址：誰か@日本語ドメイン名試験.jp

检查清单：

[] 考虑由于在邮箱名称内，以及邮箱名称与域名之间混合使用文字而可能导致的混淆、错误和恶意滥用情况。

[] 根据当地实践和市场需求，允许在当用户有明确需求时进行文字混合。例如，零售电子邮件服务的用户可能需要使用按某些特定方式进行文字混合的邮箱名称，而组织成员则可能更喜欢使用未进行文字混合的有限邮箱名称。

[] 如果允许用户请求其邮箱名称，请为用户提供示例。



避免无效字符串和无法稳定呈现的字符串

在一些书写系统中，字符能以多种方式进行组合，有些组合方式是有效的，能够保证有效通信，但有些组合方式是无效的，可能导致混淆和引发安全风险。以无效方式组合字符而构成的字符串在不同平台和应用程序之间无法以统一方式呈现，换言之，这些字符串在不同的系统或显示器上看起来会有所不同。本主题将介绍避免无效字符串的方法。

考虑事项：

- 对于各语言和文字中可以使用哪些字符以及应避免使用哪些字符，DNS 已经制定了明确的规则。在一些文字中，尤其是东南亚和南亚文字，字符和符号可以多种方式进行组合。
- 例如，请比较泰文字符串：“**ดี**”与“**ดี**”。第一个字符串“**ดี**” U+0E14(ด) + U+0E35 (ี)+ U+0E35 (ี) 是一个无效字符串，因为它看起来与第二个字符串 **ดี** U+0E14(ด) + U+0E35 (ี) 很像。泰文用户社群已经明确了 U+0E35 后不能跟 U+0E35 的规则。
- 域名系统中使用的这些规则称为标签生成规则 (LGR)。您可能希望将这些规则应用于系统中的邮箱。
- 但是，LGR 的限制程度可能过高，不适合您的系统。例如，根区 LGR 禁止使用数字，但实际上在邮箱名称中是可以使用数字的。有些文字目前还没有制定 LGR，因此不能将它们用作参考。
- 您可以参考现有的二级参考 LGR 以验证邮箱名称，可在网上找到 LGR 验证工具，网址为 <https://lgrtool.icann.org/>。该工具的源代码也可以在 GitHub 上找到，您可以将该工具下载并集成到您的系统中。

建议：

- 避免因使用无效字符串及其等效字符串创建邮箱名称而引发的潜在安全问题。
- 如果您打算使用某种语言提供邮箱名称，请查阅基于该语言的 LGR。例如，如果您希望提供德语邮箱名称，那么请采用德语 LGR。
- 如果您打算用一种可能涵盖多种语言的文字提供邮箱名称，请参考基于该文字的 LGR。例如，如果您希望提供德语、法语和瑞典语的邮箱名称，请使用拉丁文字 LGR 的适当子集。
- 对于不在二级参考 LGR 涵盖范围内的文字，您可以参考根区 LGR 或遵循 [RFC6912](#) 中的总体原则。有关使用从右到左文字的名称中的数字相关规则，请参阅 [RFC5893](#) “国际化域名应用 (IDNA) 的从右到左书写文字”。

检查清单：

[] 确定“二级参考 LGR”是否能够满足您希望使用的邮箱名称字符串要求。如果能，请采用该参考 LGR。如果不能完全满足您的要求，请根据需要进行更新。

[] 将标签生成规则 (LGR) 作为一种机制进行共享，让最终用户了解哪些字符及字符组合将被视为有效。

[] 如果适用，请使用 LGR 验证工具手动验证邮箱名称字符串。

[] 思考是希望就地使用 LGR 验证工具，还是希望将其部署到内部服务器，以简化工作流程并将其处理的邮箱名称保留在网络中。



从右到左 (RTL) 书写文字考虑事项

阿拉伯文、希伯来文、波斯文和乌尔都文是当前广泛使用的从右到左书写系统。随着时间的推移，阿拉伯文字的使用日益广泛，为适应许多其他语言（如普什图语、波斯语、乌尔都语等）的发音特征，用于书写阿拉伯语的 28 个字符得到了补充。此外，除了用于书写希伯来语之外，希伯来文字母也用于书写其他语言，如意第绪语。

在从右到左 (RTL)、从上到下书写的文字中，书写从页面右侧开始，一直写到页面左侧。这与从左到右 (LTR) 的书写系统形成对比，在 LTR 系统中，书写从页面左侧开始，一直写到页面右侧。

但是，请注意，在从右到左的文字系统中，有些文本可以是左向右书写的。例如，在阿拉伯文中，拉丁文字中的数字和单词是从左向右书写的。因此，这些文字有时被称为“双向文字”。

考虑事项：

- 与从左到右书写的文字相比，从右到左书写的文字可能会造成混淆。因此，针对此类文字的标签生成规则 (LGR) 需仔细考量并反映这些可能造成的混淆情况。出于一些合理但不明显的原因，有必要认真考量针对此类文字的 LGR 需要做出哪些规定或者需要禁止哪些情况。
- 当使用从右到左书写的地址与看不懂从右到左书写文字的联系人进行电子邮件往来时，更加难以确保联系人可以看懂地址。在这种情况下，别名可能特别有用。有关更多信息，请参阅[别名和显示名称考虑事项](#)部分。
- 在一些从右到左书写的文字（例如，阿拉伯文和希伯来文）中，“相同”标签可能具有多种不同的表示形式。这些替代形式被视为电子邮件地址的等效地址，而不是独立的地址。有关更多详细信息，请参阅[等效项考虑事项](#)部分。
- 从右到左文字系统中的文本通常是双向的。例如，拉丁文字（例如，“Cola”）和阿拉伯数字（例如 123）将从左向右书写。此外，一些标点符号会采用周围文本的方向。这意味着包含从左到右书写文字的邮箱名称更有可能是出现了输入错误。
- 阿拉伯文字除了使用阿拉伯数字之外，还可以使用 ASCII 数字。如果邮箱名称中包含数字，需考虑要使用的数字类型，以及使用另一种类型的数字的名称是否应被视为“等效地址”。文字混合考虑事项适用于同时包含这两种数字的邮箱名称。
- 在双向文字中，字符的顺序可能看起来不是很清晰明了。通常，会按口头读出的顺序输入和存储文本，但这个顺序可能与显示的地址中从右到左的文本排列顺序并不是简单对应的。对于能够阅读地址中的文字并理解文字方向变化的用户而言，这通常不是问题。但是，对于无法阅读地址中的文字的用户而言，他们可能很难从显示的地址中判断出输入顺序。
- 某个地区使用的键盘是根据当地文字调整过的，因此可能并不包括电子邮件地址中使用的所有字符，而这会造成输入困难。有些键盘中不包含句点（“.”）键或 at 符号（“@”）键。这会使输入电子邮件地址变得更加困难。为此，需考虑用于键入邮箱名称的键盘类型，以及是否允许使用较难键入的字符。
- 许多从右到左的文字中包含表示元音和发音的发音符号。这些符号在本质上是可选的。虽然允许在邮箱名称中包含这些发音符号，但用户可能会忘记包含它们。对用户而言，允许没有这些发音符号的名称更为简单。
- 阿拉伯文中具有用于强制或阻止将字符连接起来的可选字符，但在书写中，阿拉伯文通常会显示连在一起的字符。这些可选字符被称为“零宽连接符”和“零宽非连接符”。这两种字符是一种鲜为人知的机制，是计算机编码文本的产物，许多用户可能



都不知道如何使用它们。对用户而言，要正确键入使用连接符的邮箱名称可能更加困难。

建议：

- 不建议使用连字符，尤其是在阿拉伯文字中。例如，应使用 محمد，而不是 مُحَمَّد，应使用 شؤون، 而不是 شُؤُونْ。
- 使用从右到左书写文字的邮箱名称不应以数字开头。放在文本开头的数字可能无法按照文本方向规则正确放置。
- 避免由无效字符串及其“等效字符串”引起的潜在安全问题。无效字符串及其等效字符串在阿拉伯文中尤为常见。相关信息，请参阅上面的[避免无效字符串和无法稳定呈现的字符串部分](#)，以及下面的[等效项考虑事项部分](#)。
- 如果您打算使用某种语言提供邮箱名称，请查阅基于该语言的 LGR。例如，如果您希望提供阿拉伯语的邮箱名称，那么请采用阿拉伯语 LGR。有关详细信息，请参阅[等效项考虑事项部分](#)。
- 如果您打算用一种可能涵盖多种语言的文字提供邮箱名称，请参考基于该文字的 LGR。例如，如果您希望提供阿拉伯语、普什图语和乌尔都语的邮箱名称，请使用[阿拉伯文字 LGR](#)。
- 对于不在[二级参考 LGR](#) 涵盖范围内的文字，您可以参考根区 LGR 或 [RFC6912](#) 中的总体原则。
- 有关使用从右到左文字的邮箱名称中的数字相关规则，请参阅 [RFC5893](#)，其中提供了有关如何在国际化域名应用 (IDNA) 中使用从右到左书写文字的指南。

检查清单：

- [] 避免混合使用包含从右到左书写的文字，因为这会导致混淆和安全问题。
- [] 使用二级参考 LGR 或根区 LGR。
- [] 避免允许使用两个名称互为等效项的邮箱名称。



别名和显示名称考虑事项

在使用某些电子邮件系统时，用户可以轻松地拥有多个电子邮件邮箱名称或多个显示名称，这些名称可以将邮件定向到同一个邮箱，从而通过同一个邮箱收发邮件。这些别名和显示名称有多种用途，可以用于若干 EAI 使用案例。

使用案例包括：

- 用户需要不同语言的电子邮件地址，例如，需要本地语言和拉丁语的双语名片。
- 接收或发送电子邮件时需使用不兼容的系统。
- 需要让发件人能够更轻松地区分其地址簿中的电子邮件地址和别名。

考虑事项：

- 评估从您的用户到其联系人，以及从联系人到您的用户的电子邮件传递路径。请注意，按这两个方向传递邮件时可能会遇到不同的系统，其中某个方向可能可以正常传递邮件，但另一个方向可能会遇到问题。
- 您的系统中往来邮件的联系人越多、传递路径越多，您就越有可能遇到无法与国际化电子邮件地址进行邮件往来的邮件传递系统。
- 允许别名或显示名称使用与邮箱名称不同的语言可能会导致安全问题；但是，将 EAI 地址用于不同的语言社群时，这也是有益的。

以下示例是某个英文电子邮件地址的阿拉伯文显示名称：

"مشرف@اختبار - سجل.مصر" <adminmail@datamail.asia>
显示名称 电子邮件地址

建议：

- 考虑提供一个 ASCII 电子邮件地址作为国际化电子邮件地址的别名。别名应将电子邮件发送到与国际化地址相同的位置。用户应该能够选择是通过国际化地址，还是通过 ASCII 别名地址收发邮件。

检查清单：

[] 了解您的系统是否支持别名。

[] 如果您的系统支持别名，请在选择邮箱名称的过程中向用户界面添加一个别名创建选项。届时，系统将告知用户他们应该有一个 ASCII 别名，并允许其创建一个与 EAI 邮箱名称相关的别名。

[] 如果您的系统支持，可以选择添加一个用户界面，以允许用户稍后添加其他别名。

[] 如果将邮箱名称指定给用户，而不是允许他们自行选择邮箱名称，则需在指定 EAI 邮箱名称时指定 ASCII 别名。



标记和符号

ASCII 邮箱名称中允许使用某些标记和符号，例如点 (.)、下划线 (_)、连字符 (-) 和加号 (+)。在确定或选择 EAI 邮箱名称时，管理员还需要考虑在支持的文字中应允许使用哪些统一码标记或符号。

考虑事项：

- 应考虑包含原则（请参阅[标签生成规则程序](#)）。包含原则建议从一个空列表开始，只包含用户需要的字符集，例如，标记和符号。此外，应避免使用可能引发特定于实施的安全问题的符号。
- 应仅允许使用您的用户及其联系人可以阅读和输入的标记和符号。邮箱名称应该可以使用常用的输入法编辑器 (IME) 输入。
- 以一致且易理解的方式，为拉丁文邮箱名称和国际化邮箱名称制定相关且具有相似限制的策略。

建议：

- 避免可能产生安全问题的数学符号，例如，#、\$、%、=、“、*、< 和 >。
- 应禁止使用表情符号（请参阅 ICANN SSAC 公告 [SAC95](#)）。
- 应禁止使用不符合可键入原则的连字符，例如，阿拉伯文上喉塞音符 (U+0654)。

检查清单：

[] 默认情况下，避免使用数学符号。

[] 默认情况下，避免使用用户键盘/输入设备中不存在的标记和符号。某些标记和符号在特定市场的输入设备中不存在。

[] 了解用户和当地市场期望。如果您的市场需要标记和符号，可将点 (.)、下划线 (_)、连字符 (-) 和加号 (+) 等常用符号添加到字汇中。

[] 检查额外符号的使用情况（如果需要），并确保它们不会给最终用户带来安全问题，包括导致客户混淆、文字混合和其他安全问题。

统一码字符规范化

统一码标准定义了一个用于替换等效字符序列的文本规范化流程，目的是将任意两个等效文本简化为相同的码点序列。

由于存在不同类型的标准化形式，例如，规范化形式 D (NFD) 和规范化形式 C (NFC)，统一码定义了可以用多种方式进行编码的字符的规范化形式。

示例：Å 可以编码为 [212B](#) 或 [00C5](#)。这两种编码应该指的是同一个邮箱，而不是不同的邮箱，并且应规范化为同一种 NFD 和 NFC。

请注意，虽然这些示例使用的是拉丁字符，但在其他文字中也可以找到类似的示例。

有关其他规范化类型的更多详细信息及示例，请访问[统一码联盟网站](#)。



建议

- 邮箱名应该使用 **NFC**，一是因为要与域名保持一致，二是因为 **NFC** 字符串在大多数情况下比 **NFD** 字符串要短。
- 通常，标准软件已经在其库中进行了统一码规范化编码。电子邮件管理员应该知道其操作的系统所采用的规范化类型，以便能够指定和管理其邮箱。
- 域名使用规范化形式 **C (NFC)**。构建系统时，请确保库支持 **NFC** 以确保一致性。

检查清单

[] 了解电子邮件系统所采用的规范化类型。

[] 确保电子邮件程序可以进行独立于规范化形式的名称比较。

[] 如果可以选择规范化类型，请选择 **NFC**，除非需要使用其他规范化类型。

等效项考虑事项

在社群中被视为“相同”或“可互换使用”的域名标签称作“变体”域名标签。对于在社群中被视为“相同”的电子邮件邮箱字符串，本文档称之为“等效”邮箱名称。

在有些情况下，等效名称很有用，可以用作主邮箱名称的别名。但在其他情况下，如果未识别出等效名称或将等效名称分配给了不同的用户，则会导致混淆，并可能引发安全问题。

考虑事项：

- 在 **ASCII** 电子邮件地址中，一些邮件服务可能会将大写字母和小写字母视为等效项。
- 为制定清晰明确的标签生成规则 (**LGR**)，包括明确各种文字中的变体标签，**DNS** 社群已经做了大量的工作。请查阅 **LGR**，确认是否希望将 **LGR** 用作定义邮箱名称等效项的基础。

例如：

1. 繁体中文和简体中文等效项：

名称 表示 “name” (U+540D U+79F0)

名稱 表示 “name” (U+540D U+7A31)

名称 (**Míngchēng**，英文为 “Name”) 在中国大陆使用，而名稱 (**Míngchēng**，英文为 “Name”) 在中华台北和香港地区使用，但是，中文社群中的所有成员都认为这两个词是一样的。

2. 阿拉伯文等效词 (该词在阿拉伯语中表示 “网络”)：

شبكة (统一码 U+06C3 U+06A9 U+0628 U+0634)

شبكة (统一码 U+0629 U+06A9 U+0628 U+0634)

شبكة (统一码 U+0629 U+0643 U+0628 U+0634)



3. 日文和中文中的全角字符与半角字符。“JIM”与“JIM”、“1 2 3”与“123”、“アイオ”与“アイオ”。

4. 大小写折叠等效项：

John.Callen
john.callen
John.callen
john.Callen

5. 名称与分隔符构成的变体（也可视为等效项）：

john.callen
johncallen

- 对于一个电子邮件系统，您可以考虑将发送到等效邮箱名称的所有电子邮件均发送到同一个邮箱。这样可以避免混淆、文字混合和其他安全问题。
- 可以将更多等效邮箱名称作为别名指定给同一个邮箱，但是这些等效邮箱名称可能会非常多（数百或数千个），从而造成管理困难。因此，需谨慎权衡等效项的有用性与处理等效项的开销。
- 请注意，电子邮件地址中域名部分的变体可能会对用户产生类似的影响。域名及其变体不在本文档的讨论范围内。
- 需注意当用户在系统中键入电子邮件地址时可能出现的各种错误或混淆。本文档讨论了与地址中邮箱名称部分有关的混淆。
- 关于地址中域名部分的正确拼写或规范化，也可能存在混淆。域名变体不在本文档的讨论范围内。

建议：

- 在分配了邮箱名称之后，应该识别该邮箱名称的等效名称，并将其标记为已注册，以阻止其他用户注册。
- 可考虑将某个邮箱名称的最为有效的等效名称用作别名（请参阅本文档中的[别名考虑事项](#)部分），并将要发送到等效邮箱的电子邮件发送到原始邮箱名称所指向的地址。然而，等效邮箱名称的管理可能是一项难题，所以要谨慎权衡等效名称的有用性与处理等效名称的开销，并根据需要进行取舍。
- 考虑对等效邮箱名称采用基于语言的可行等效规则。例如，对于 ASCII 邮箱名称而言，将所有大写和小写变体视为等效项是可行的。对于简体中文邮箱名称（例如表示“China”的中国），可考虑将具有相同含义的繁体中文名称（如中國）用作等效项。
- 所有[大小写折叠](#)排列都应被视为同一个邮箱。
- 当同一字汇（用于创建邮箱名称的一系列可用码点）中使用了多个数字集时，这些数字应映射为等效码点。例如，印地语邮箱管理员可能会将 abc123 和 abc१२३ 视为等效的邮箱名称：

邮箱名称：abc123

等效邮箱名称：abc१२३

检查清单：

[] 根据书写系统、用户期望和实施人员的技术能力，制定一项政策以用于确定将应用于系统的“相同”或等效邮箱名称。制定的政策将需要考量上述考虑事项中提及的 LGR、大小写折叠、分隔符、数字和符号。



[] 确定二级参考 LGR 是否能够满足您对等效有效邮箱名称字符串的需求。如果不能，请根据需要进行调整。

[] 分享制定的政策，让最终用户了解哪些字符和字符组合被视为是有效的，以及哪些可以具有等效项。

[] 当使用等效字符串时，考虑是希望就地使用 **LGR** 验证工具，还是希望将其部署到内部服务器，以简化工作流程并将其处理的邮箱名称保留在网络中。

其他考虑事项

电子邮件地址中的国际化域名部分可以使用统一码字符拼写，也可以用以“xn-”开头的等效纯 ASCII 名称拼写。

考虑事项：

- 域名的“xn-”拼写是一个隐含的技术细节。理想情况下，用户永远不需要看到此拼写。然而，如果在过渡期内不是每个电子邮件系统都支持国际化的电子邮件地址，采用“xn-”拼写可能是一个有用的解决方法。

建议：

- 使用国际化的非 ASCII 名称拼写域名。避免显示“xn-”等效名称。
- 有些电子邮件客户端可能不会自动关联电子邮件地址邮箱名称的 U-标签和 A-标签形式，因此，请确保这两种标签已相互映射。



术语表

别名：可将邮件发送到与另一个电子邮件地址指向相同位置的电子邮件地址。通常，别名是一个 **ASCII** 电子邮件地址，可在当联系人因邮件传递系统不支持 **EAI** 而无法向 **EAI** 地址发送邮件时提供 **EAI** 地址的回退地址。

ASCII：指由不含重音、数字和一些标点符号的拉丁字母组成的有限字符集。**ASCII** 这一术语是计算机行业的一个缩略语。

显示名称：电子邮件地址旁边显示的名称。

域名：用于标识网站或电子邮件服务器的名称序列，例如“**uasg.tech**”或“**icann.org**”。

EAI：国际化电子邮件地址。

国际化电子邮件地址：使用所有实际存在的语言和文字，而不仅仅是使用 **ASCII** 或拉丁文字书写的电子邮件地址。

等效项：两个或更多统一码码点具有同一个含义或被地方当局视为等效语言，由此产生的“字符变体”。例如，简体中文和繁体中文之间存在等效项，而且这些等效项可能会让使用这两种文字的人感到困惑。

IDN：国际化域名。

IME（输入法编辑器）：**IME** 是一种应用程序，允许使用标准键盘（例如 **US-101** 键盘）键入在键盘上没有直接表示形式的字符和符号。

国际化域名：含非 **ASCII** 字符的域名。

标签生成规则：有关域名中所用字符的规则。

LGR：标签生成规则。

邮箱名称：电子邮件地址中用于标识域名中邮箱位置的部分。可与其他术语互换使用，例如“邮箱名称”、“帐户名称”、“用户名”、“**UTF-8** 部分”、“本地部分”。

国际化域名编码：一种算法，可在用于互联网主机名的有限 **ASCII** 字符子集中对统一码进行编码。

EAI 第 1 阶段：一个电子邮件系统，可以成功地与国际化电子邮件地址之间收发电子邮件，但不能托管使用国际化名称的邮箱。

EAI 第 2 阶段：一个电子邮件系统，可以成功托管国际化邮箱名称，并与国际化电子邮件地址之间成功地收发电子邮件。



呈现：计算机显示系统将字符序列转换为屏幕上显示的可视内容的过程。

文字：用于书写一种或多种语言的规则和字符集。

Script.txt：与统一码字符数据库相关联的数据文件，用于标识所有码点的文字属性值 (<http://www.unicode.org/Public/12.1.0/ucd/Scripts.txt>)。

统一码：一种字符编码标准，涵盖了世界上大多数的文字和书面语言。请参阅 <http://unicode.org>。

变体名称：被社群视为“相同”的名称。