

# 共享的MIME信息数据库

## X 桌面组

### 托马斯·伦纳德

<[tall97\\_at\\_users.sf.net](mailto:tall97_at_users.sf.net)>

---

## 目录

### 介绍

[版本](#)

[这个规格是什么？](#)

[本规范中使用的语言](#)

### 统一系统

[目录布局](#)

[源 XML 文件](#)

[媒体/子类型.xml文件](#)

[全球文件](#)

[魔术文件](#)

[XML命名空间文件](#)

[图标文件](#)

[树魔术文件](#)

[哑剧缓存文件](#)

[使用扩展属性存储 MIME 类型](#)

[子类化](#)

[建议的检查顺序](#)

[非常规文件](#)

[卷的内容类型](#)

[URI 方案处理程序](#)

[安全隐患](#)

[用户修改](#)

### 贡献

### 引用

## 介绍

### 版本

这是共享 MIME-info 数据库规范的 0.21 版，最后更新于 2018 年 10 月 2 日。

### 这个规格是什么？

许多程序和桌面使用 MIME 系统[MIME] 来表示文件类型。通常，有必要计算出 正确的文件MIME 类型。这通常是通过检查文件的 名称或内容，并在数据库中查找正确的 MIME 类型。

存储有关每种类型的信息也很有用，例如文本 它的说明，或可用于查看或编辑的应用程序列表 该类型的文件。

对于互操作性，不同的程序使用相同的程序很有用 数据库，以便不同的程序就文件类型达成一致，并且 信息不重复。对于应用程序作者来说，这也有助于 只需在一个地方安装新信息。

此规范试图统一当前 MIME 数据库系统 由 GNOME[GNOME]、KDE[KDE] 和 ROX[ROX]，并为将来的可扩展性提供空间。

MIME 数据库不存储用户首选项（例如用户的首选 用于处理特定类型文件的应用程序）。它可用于存储 静态信息，例如可以使用 查看某种类型的文件 特定应用程序。

## 本规范中使用的语言

关键词“必须”、“不得”、“要求”、“应”、“不应”、“应当”，本文档中的“不应”、“推荐”、“可以”和“可选”应为 按照 RFC 2119[RFC-2119] 中所述进行解释。

## 统一系统

在讨论GNOME，KDE和ROX使用的先前系统时（参见“历史及相关系统”文件），很明显差异 数据库之间只是因为它们是分开的结果，而不是到期 开发人员之间的任何根本分歧。每个人都渴望看到 他们合并了。

本规范提出：

- 应用程序安装新的 MIME 相关信息的标准方法。
- 获取文件的 MIME 类型的标准方法。
- 获取有关 MIME 类型的信息的标准方法。
- 所有文件的标准位置以及解决冲突的方法。

此外，现有数据库已合并为一个包 [共享哑剧]。

## 目录布局

MIME 数据库的存储方式有两个重要要求：

- 应用程序在安装时必须能够以任何方式扩展数据库，以添加用于确定类型的新规则以及有关特定类型的新信息。
- 必须能够在 /usr、/usr/local 和用户的主目录中安装应用程序（以正常的Unix方式）并使用 MIME信息。

此规范使用 XDG 基目录规范 [BaseDir] 来 定义用于存储数据库的前缀。在本文档的其余部分中，带有前缀的路径指示应从中加载文件 中每个目录的子目录：。 <MIME>mimeXDG\_DATA\_HOMEXDG\_DATA\_DIRS

例如，当使用默认路径时，“加载所有<MIME>/text/html.xml文件”意味着 load，和（如果它们存在，并按此顺序）。在目录中找到的信息将添加到上一个目录中找到的信息中 目录，除非使用 glob-deleteall 或magic-deleteall来覆盖 MIMETYPE 定义的部分内

容。 /usr/share/mime/text/html.xml/usr/local/share/mime/text/html.xml~/.local/share/mime/text/html.xml

每个希望为 MIME 数据库做出贡献的应用程序都将安装 单个 XML 文件，以应用程序命名，放入三个目录之一（取决于用户请求的位置 要安装的应用程序）。安装、卸载或修改后 文件，应用程序必须运行**更新-mime-database**命令，由 freedesktop.org 共享数据库[[SharedMIME](#)]提供。

<MIME>/packages/

**update-mime-database**被传递到包含子目录的目录，该子目录是 修改为其唯一参数。它扫描子目录中的所有 XML 文件，合并其中的信息，并创建许多输出文件。mimepackagespackages

如果这些文件中的信息冲突，则目录中的信息 列表中较低的优先。任何文件名都优先于所有其他文件 相同的目录。这可以由 允许用户编辑数据库以确保用户的 更改生效。Override.xmlpackages

由**更新-mime-database**创建的文件是：

- <MIME>/globs （包含从名称到 MIME 类型的映射）[已弃用 globs2]
- <MIME>/globs2 （包含从名称到 MIME 类型和球体权重的映射）
- <MIME>/magic （包含从文件内容到 MIME 类型的映射）
- <MIME>/subclasses （包含从 MIME 类型到它们继承的类型的映射）
- <MIME>/aliases （包含从别名到 MIME 类型的映射）
- <MIME>/icons （包含从 MIME 类型到图标映射的映射）
- <MIME>/generic-icons （包含从 MIME 类型到通用图标映射的映射）
- <MIME>/XMLnamespaces （包含来自 XML 的映射（命名空间URI，本地名称）与MIME类型配对）
- <MIME>/MEDIA/SUBTYPE.xml （每个 MIME 对应一个文件 类型，提供有关类型的详细信息，包括注释、图标和泛型图标）
- <MIME>/mime.cache （包含与,,,,和文件相同的信息，二进制、可映射格式）  
globs2magicssubclassesaliasesiconsgeneric-iconsXMLnamespaces

以下各节将介绍这些生成的文件和源文件的格式。此步骤有多种用途。首先，它允许 应用程序无需解析所有源 XML 文件即可快速获取所需数据（仅基本包就超过 700K）。其次，它允许数据库用于其他 目的（例如创建文件，如果 希望）。第三，它允许对输入数据执行验证，并且无需其他应用程序仔细检查输入 错误本身。packages/etc/mime.types

## 源 XML 文件

每个应用程序仅提供一个 XML 源文件，该文件如上所述安装在目录中。此文件是一个 XML 文件 其文档元素名为mime-info及其命名空间 URI 是<http://www.freedesktop.org/standards/shared-mime-info>。所有元素 本规范中描述的也必须具有此命名空间。packages

文档元素可能包含零个或多个MIME 类型的子节点，按任意顺序，每个描述单个 MIME 类型。每个元素都有一个类型 属性，用于提供它所描述的 MIME 类型。

每个mime 类型 节点可以包含以下元素的任意组合，并按任何顺序：

- glob元素具有模式 属性。任何文件 名称与此模式匹配的将被赋予此 MIME 类型（受 当然是其他文件）。还有一个可选的权重 属性，它在解决与其他 glob 匹配项的冲突时使用。默认权重值为 50，并且 最大值为 100。

KDE 的 glob 系统取代了 GNOME 和 ROX 的 ext/regex 字段，因为它检测 “\*.ext” 形式的模式并将其存储在 内部扩展哈希表。正则表达式的全部功能是不被任何桌面使用，全球模式更适合 无论如何文件名匹配。

第一个 glob 元素表示文件类型的 “主要” 扩展名。虽然这不会影响 mimetype 匹配算法，但此信息可能很有用 当 MIMETYPE 需要单个主扩展时，例如，使应用程序 可以在保存文件时选择适当的扩展名。

- 一个glob-deleteall元素，它指示模式来自 必须丢弃以前解析的目录。此文件中定义的模式（如果有）代替使用。
- 魔术元素包含匹配元素的列表（其中任何一个元素都可能匹配），以及所有包含的规则的可选优先级属性。低 数字应该用于更通用的类型（例如 “gzip 压缩数据”）以及特定子类型的较高值（例如，字处理器格式 碰巧使用 gzip 来压缩文件）。默认优先级值为 50，并且 最大值为 100。

每个匹配元素都有许多属性：

属性	必填?	价值
类型	是的	字符串，主机16，主机32，大16，大32，小16，小32或字节。
抵消	是的	字节偏移量 在要检查的文件中。这可以是单个数字或 形式 'start: end'，表示范围中的所有偏移量都应为 检查。范围是包含的。
价值	是的	要与文件内容进行比较的值，采用类型指示的格式 属性。字符串类型支持 C 字符转义（\0，\t，\n，\r，\xAB 表示十六进制，\777 表示八进制）。
面具	不	在将文件与文件进行比较之前，AND 中的值的数字 “价值”。数字类型的掩码可以是任意数字，而字符串的掩码 必须以 16 为基数，并以 0x 开头。

每个元素对应于file（1）的文件的一行。它们可以以相同的方式嵌套，以提供等效的延续线。也就是说，<a><b/><c/></a>表示 “a和（b或c）”。magic.mime

- 一个magic-deleteall元素，它指示 magic 匹配来自 必须丢弃以前解析的目录。此文件中定义的魔术（如果有）代替使用。
- 别名元素指示类型有时也是 由另一个名称已知，由type属性给出。为 例如，音频/MIDI具有 Audio/X-MIDI 的别名。请注意，不应有一个mime 类型的元素定义每个别名;单曲 元素定义类型的规范名称并列出其所有别名。
- 元素的子类表示此元素的任何数据 type 也是其他类型，由type属性给出。请参阅[名为 “子类化” 的部分](#)。
- 注释元素提供了人类可读的MIME文本描述 类型，通常由文件扩展名的首字母缩略词和简短描述组成，例如 “ODS 电子表格”。其中可能有许多元素具有不同的xml: lang属性 以多种语言提供文本。
- 首字母缩略词元素使有经验的用户对文档内容有一个简洁的了解。例如 “ODS” 、 “GEDCOM” 、 “JPEG” 和 “XML” 。其中可能有许多元素具有不同的xml: lang属性 以多种语言提供文本，尽管只有在绝对必要时才应使用这些语言。

- 扩展首字母缩略词元素是首字母缩略词元素的扩展版本，例如“OpenDocument Spreadsheet”、“GEnealogical Data COMmunication”和“eXtensible Markup Language”。这些元素的目的是为用户提供一种查找有关各种 MIME 类型或 第三方资源中的文件格式。其中可能有许多元素具有不同的xml: lang属性 以多种语言提供文本，尽管只有在必要时才应使用这些语言。
- 图标元素指定要用于此特定 MIME 类型的图标，给定 按名称 属性。通常，创建用于 mimetype 的图标 通过将 “/” 字符映射到 “-” 来基于 MIME 类型，但用户可以通过使用 用于自定义特定 MIMETYPE 的图标的图标元素。此元素不用于系统数据库，而仅用于用户覆盖的数据库。只允许一个图标元素。
- 泛型图标元素指定要用作此图标的通用图标 特定的 MIME 类型，由name属性给出。如果存在，则使用此选项 没有特定的图标（有关如何找到这些图标，请参阅图标）。这些是 用于可以使用通用图标的类似类型的类别（如电子表格或存档）。图标命名规范列出了一组此类图标名称。如果未指定此元素 然后，MIMETYPE 用于通过使用顶级媒体类型（例如 “video/ogg” ）中的 “video” ）并附加 “-x-generic” （即上一个示例中的 “video-x-generic” ）。只允许一个泛型图标元素。
- 根 XML元素具有命名空间URI和localName属性。如果文件被标识为 XML 文件，这些规则允许根据命名空间和本地名称选择更具体的 MIME 类型 的文档元素。

如果localName存在但为空，则文档元素可能具有 任何名称，但命名空间仍必须匹配。

- 树魔术元素包含树匹配元素列表，其中任何一个都可能匹配，并且所有 包含规则。默认优先级值为 50，最大值为 100。

每个树匹配元素都有许多属性：

属性	必填?	价值
路径	是的	挂载的卷/文件系统上必须存在的路径。该路径被解释为从测试卷/文件系统的根目录开始的相对路径
类型	不	路径的类型。可能的值：文件，目录，链接
匹配案例	不	路径是否应区分大小写匹配。可能的值：真，假
可执行	不	文件是否必须可执行。可能的值：真，假
非空	不	目录是否必须为非空。可能的值：真，假
哑剧型	不	路径处文件的 mimetype

Treematch 元素可以嵌套，这意味着必须同时满足外部和内部 Treematch 才能实现 “匹配” 。

应用程序还可以定义自己的元素，前提是它们具有命名空间以防止冲突。未知元素将像注释元素一样直接复制到输出 XML 文件中。其典型用途是指示默认处理程序 特定桌面的应用程序 （“Galeon 是 GNOME 默认的文本/html 浏览器”）。请注意，这不会 指示用户的首选应用程序，仅指示（固定）默认值。

下面是一个示例源文件，名为：diff.xml

```
<?xml version="1.0"?>
<mime-info xmlns='http://www.freedesktop.org/standards/shared-mime-info'>
  <mime-type type="text/x-diff">
```

```

<comment>Differences between files</comment>
<comment xml:lang="af">verskilte tussen lêers</comment>
...
<magic priority="50">
  <match type="string" offset="0" value="diff\t"/>
  <match type="string" offset="0" value="***\t"/>
  <match type="string" offset="0" value="Common subdirectories: "/>
</magic>
<glob pattern="*.diff"/>
<glob pattern="*.patch"/>
</mime-type>
</mime-info>

```

在实践中，常见的类型（如text/x-diff）由共享 freedesktop.org 提供 数据库。此外，只需要提供新信息，因为此信息将被合并 以及有关同一类型的其他信息。

## 媒体/子类型.xml文件

这些文件具有mime 类型 元素作为根节点。格式为 如上所述。它们是通过合并源文件中的所有 MIME 类型元素并为每个MIME 类型 创建一个输出文件来创建的。每个文件可能包含 来自多个源文件的信息。魔术， 全球和根XML元素将 已被删除。

上面给出的示例源文件将（单独）创建一个输出文件，其中包含以下内容：<MIME>/text/x-diff.xml

```

<?xml version="1.0" encoding="utf-8"?>
<mime-type xmlns="http://www.freedesktop.org/standards/shared-mime-info" type="text/x-diff">
<!--Created automatically by update-mime-database. DO NOT EDIT!-->
  <comment>Differences between files</comment>
  <comment xml:lang="af">verskilte tussen lêers</comment>
  ...
</mime-type>

```

## 全球文件

globs2 文件是包含粗细、MIME 类型和模式的行的简单列表，用冒号分隔。这些线按球体重量排序。例如：

```

# This file was automatically generated by the
# update-mime-database command. DO NOT EDIT!
...
55:text/x-diff:*.patch
50:text/x-diff:*.diff
50:text/x-c++src:*.C:cs
...

```

glob 文件是包含 MIME 类型和模式的行的简单列表，用冒号分隔。是的 已弃用，以支持 globs2 文件，该文件还列出了 glob 规则的权重。这些线按球体重量排序。例如：

```

# This file was automatically generated by the
# update-mime-database command. DO NOT EDIT!
...
text/x-diff:*.patch
text/x-diff:*.diff
...

```

应用程序必须不区分大小写地匹配 globs，除非区分大小写的属性 设置为 true。这样，例如将被视为C++文件， 但仍会使用 \*.gif 模式。main.CIMAGE.GIF

如果相同重量的多个图案匹配，则应使用最长的图案。特别是，具有多个扩展名（例如）的文件必须与最长的扩展名序列匹配（例如， "\*.tar.gz" 优先于 "\*.gz" ）。文字模式（例如， "生成文

件” ) 必须 在所有其他匹配之前匹配。建议以 “\*” 开头的模式。并且不包含其他特殊字符 (‘\*?’ [‘] ) 应放在哈希中 用于高效查找的表, 因为这涵盖了大多数模式。因此 此形式的模式应先于其他通配符模式匹配。Data.tar.gz

如果匹配模式由两个或多个 MIME 类型提供, 则应用程序 不应该依赖其中之一。相反, 他们应该使用魔术数据 (见下文) 以检测实际的 MIME 类型。例如, 这是必需的 处理容器格式, 如Ogg或AVI, 映射各种视频和/或 音频编码数据到一个扩展。

可能有多个规则映射到同一类型。它们都应该合并。如果相同的模式被定义两次, 那么它们必须由如上所述, 规则来自目录。

glob-deleteall元素, 这意味着实现 应该丢弃以前目录中的信息, 使用 \_\_NOGLOBS\_\_作为模式。例如:

```
0:text/x-diff:__NOGLOBS__
50:text/x-diff:*.diff
...
```

在上面的例子中, mimetype text/x-diff被重新定义 (例如在用户的 ~/.local/share/mime) 仅与模式 \*.diff 相关联, 因此其他模式 像 \*.patch 被删除了。\_\_NOGLOBS\_\_线前面的权重将被忽略。在给定的 globs2 文件中, 始终写入给定 mimetype 的\_\_NOGLOBS\_\_行 在这种哑剧类型的任何其他球体之前。

以 “#” 开头的行是注释, 应忽略。一切从 换行符的 “: ” 字符是模式的一部分;空格不应 剥夺。该文件采用 UTF-8 编码。球形图案的格式 至于 fnmatch (3) 。该格式不允许模式包含文本 换行符, 但这预计不会成为问题。

常见类型 (如MS Word文档) 将在X桌面中提供 组的包, 所有使用此包的应用程序都必须需要该包规范。由于每个应用程序将仅提供信息 关于它自己的类型, 冲突应该是罕见的。

第四个字段 (第一个 globs2 示例中的 “cs” ) 包含逗号分隔标志的列表。当前定义的标志为: cs (区分大小写) 。实现应忽略 未知标志。

实现也应该忽略其他字段, 以便 globs2 文件的语法 将来可以扩展。示例: “50: text/x-c++src: \*.C: cs, newflag: newfeature: somethingelse” 当前应解析为 “50: text/x-c++src: \*.C: cs” 。

## 魔术文件

魔术数据以二进制格式存储, 以便于解析。旧的魔法数据库 有复杂的逃生规则;这些现在由**update-mime-database** 处理。

该文件以魔术字符串 “MIME-Magic\0\n” 开头。文件中没有版本号。不兼容的更改将由 以新格式创建当前的 “magic” 文件和较新的 “magic2” 。在可能的情况下, 将仅进行兼容的更改。所有数字都是大端序的, 因此需要在小端数机器上进行字节交换。

文件的其余部分由一系列小部分组成。每个部分都通过在括号中给出优先级和类型来介绍, 然后是换行符。优先级较高的条目排在第一位。例:

```
[50:text/x-diff]\n
```

该部分中的每一行采用以下形式:

```
[ indent ] ">" start-offset "=" value
[ "&" mask ] [ "~" word-size ] [ "+" range-length ] "\n"
```



部分	例	意义
订货单	1	筑巢 规则的深度，对应于传统文件格式中的 “>” 字符数。
“>” 开始偏移	>4	偏移量进入 文件以查找匹配项。
“=” 值	=\0x0\0x2\0x55\0x40	两个字节给出值的（大端序）长度，后跟值本身。
“&” 掩码	&\0xff\0xf0	掩码，其长度（如果存在）与值完全相同。
“~” 字大小	~阿拉伯数字	在小端机器上， 要进行字节交换的每个组的大小。
“+” 范围长度	+8	区域的长度 在要检查的文件中。

请注意，值、值长度和掩码都是二进制的，而所有内容 否则是文本。每个元素都以单个字符开头 标识它，缩进级别除外。

字大小用于字节交换。小端系统应该逆转 值和掩码中字节组的顺序（如果大于 1）。这只会影响 “主机” 匹配（“big32” 条目的字大小仍为 1，例如，因为不需要交换，而 “host32” 具有字大小的 4）。

缩进、范围长度、字大小和掩码组件是可选的。如果 缺少，缩进默认为 0，范围长度为 1，字大小为 1，以及 屏蔽到所有 “一个” 位。

缩进对应于规则的嵌套深度。顶级规则具有 缩进为零。条目的父项是前面带有缩进的条目 比条目少一个。

如果在需要换行符的地方找到未知字符，则整个 行应忽略（新的之后将没有二进制数据 字符，因此下一行在下一个 “\n” 字符之后开始）。这是为了 未来的扩展。

上面的 text/x-diff 示例将（单独）创建此魔术文件：

```
00000000  4d 49 4d 45 2d 4d 61 67  69 63 00 0a 5b 35 30 3a  |MIME-Magic..[50:|
00000010  74 65 78 74 2f 78 2d 64  69 66 66 5d 0a 3e 30 3d  |text/x-diff].>0=|
00000020  00 05 64 69 66 66 09 0a  3e 30 3d 00 04 2a 2a 2a  |..diff..>0=..***|
00000030  09 0a 3e 30 3d 00 17 43  6f 6d 6d 6f 6e 20 73 75  |..>0=..Common su|
00000040  62 64 69 72 65 63 74 6f  72 69 65 73 3a 20 0a     |bdirectories:..|
```

magic-deleteall属性，这意味着实现 应该丢弃以前目录中的信息，使用 \_\_NOMAGIC\_\_为值：

```
>0=__NOMAGIC__\n
```

这之后可以遵循 mimetype 的其他魔术规则。

XML命名空间文件

每个文件是以下形式的行列表：XMLnamespaces

```
namespaceURI " " localName " " MIME-Type "\n"
```

例如：

```
http://www.w3.org/1999/xhtml html application/xhtml+xml
```

这些行已排序（在 C 语言环境中使用 strcmp），并且没有具有相同命名空间 URI 和 本地名称在一个文件中。如果 localName 为空，则后面将有两个空格 命名空间 URI。



图标文件

文件是以下形式的行列表：iconsgeneric-icons

```
MIME-Type ":" icon-name "\n"
```

例如：

```
application/msword:x-office-document
```

树魔术文件

树魔术数据存储在格式与魔术文件格式非常相似的文件中。

该文件以魔术字符串“MIME-TreeMagic\0\n”开头。文件中没有版本号。不兼容的更改将通过创建当前的“树魔法”和更新的“树魔法2”来处理 在新格式中。在可能的情况下，将以兼容的方式进行更改。

文件的其余部分由一系列小部分组成。每个部分都通过给予 优先级和键入括号，后跟换行符。更高优先级的条目来了 第一。例：

```
[50:x-content/image-dcf]\n
```

该部分中的每一行采用以下形式：

```
[ indent ] ">" "\" path "\" "=" type [ ",", option ]* "\n"
```

部分	意义
订货单	规则的嵌套深度。
路径	要匹配的路径。
类型	所需的文件类型，“文件”、“目录”、“链接”或“任意”之一
选择	对于树匹配 元素的可选属性是可选的。可能的值为“可执行文件”、“匹配大小写”、“非空”或 MIME 类型

哑剧缓存文件

文件包含与二进制文件中的,,和文件相同的信息， 可映射格式：

```
mime.cacheglobs2magicsubclassesaliasesXMLnamespaces
```

```
Header:
2          CARD16          MAJOR_VERSION    1
2          CARD16          MINOR_VERSION    2
4          CARD32          ALIAS_LIST_OFFSET
4          CARD32          PARENT_LIST_OFFSET
4          CARD32          LITERAL_LIST_OFFSET
4          CARD32          REVERSE_SUFFIX_TREE_OFFSET
4          CARD32          GLOB_LIST_OFFSET
4          CARD32          MAGIC_LIST_OFFSET
4          CARD32          NAMESPACE_LIST_OFFSET
4          CARD32          ICONS_LIST_OFFSET
4          CARD32          GENERIC_ICONS_LIST_OFFSET

AliasList:
4          CARD32          N_ALIASES
8*N_ALIASES      AliasListEntry
```

AliasListEntry:		
4	CARD32	ALIAS_OFFSET
4	CARD32	MIME_TYPE_OFFSET
ParentList:		
4	CARD32	N_ENTRIES
8*N_ENTRIES	ParentListEntry	
ParentListEntry:		
4	CARD32	MIME_TYPE_OFFSET
4	CARD32	PARENTS_OFFSET
Parents:		
4	CARD32	N_PARENTS
4*N_PARENTS	CARD32	MIME_TYPE_OFFSET
LiteralList:		
4	CARD32	N_LITERALS
12*N_LITERALS	LiteralEntry	
LiteralEntry:		
4	CARD32	LITERAL_OFFSET
4	CARD32	MIME_TYPE_OFFSET
4	CARD32	WEIGHT in lower 8 bits
		FLAGS in rest:
		0x100 = case-sensitive
GlobList:		
4	CARD32	N_GLOBS
12*N_GLOBS	GlobEntry	
GlobEntry:		
4	CARD32	GLOB_OFFSET
4	CARD32	MIME_TYPE_OFFSET
4	CARD32	WEIGHT in lower 8 bits
		FLAGS in rest:
		0x100 = case-sensitive
ReverseSuffixTree:		
4	CARD32	N_ROOTS
4	CARD32	FIRST_ROOT_OFFSET
ReverseSuffixTreeNode:		
4	CARD32	CHARACTER
4	CARD32	N_CHILDREN
4	CARD32	FIRST_CHILD_OFFSET
ReverseSuffixTreeLeafNode:		
4	CARD32	0
4	CARD32	MIME_TYPE_OFFSET
4	CARD32	WEIGHT in lower 8 bits
		FLAGS in rest:
		0x100 = case-sensitive
MagicList:		
4	CARD32	N_MATCHES
4	CARD32	MAX_EXTENT
4	CARD32	FIRST_MATCH_OFFSET
Match:		
4	CARD32	PRIORITY
4	CARD32	MIME_TYPE_OFFSET
4	CARD32	N_MATCHLETS
4	CARD32	FIRST_MATCHLET_OFFSET
Matchlet:		
4	CARD32	RANGE_START

2022/11/11 15:19

共享的MIME信息数据库

4	CARD32	RANGE_LENGTH
4	CARD32	WORD_SIZE
4	CARD32	VALUE_LENGTH
4	CARD32	VALUE_OFFSET
4	CARD32	MASK_OFFSET (0 if no mask)
4	CARD32	N_CHILDREN
4	CARD32	FIRST_CHILD_OFFSET
NamespaceList:		
4	CARD32	N_NAMESPACES
12*N_NAMESPACES	NamespaceEntry	
NamespaceEntry:		
4	CARD32	NAMESPACE_URI_OFFSET
4	CARD32	LOCAL_NAME_OFFSET
4	CARD32	MIME_TYPE_OFFSET
GenericIconsList:		
IconsList:		
4	CARD32	N_ICONS
8*N_ICONS	IconListEntry	
IconListEntry:		
4	CARD32	MIME_TYPE_OFFSET
4	CARD32	ICON_NAME_OFFSET

对文件中的列表进行排序，以启用二叉搜索。的清单 别名按别名排序，文字列表按 字面。后缀树节点同级按字符排序。 命名空间列表按命名空间 uri 排序。图标列表 按哑剧类型排序。

Mimetype 通过附加后缀存储在后缀树中 以 “\0” 作为字符的树叶节点。这些节点显示在 子项列表的开头。

所有偏移量都以文件开头的字节为单位。

字符串以零结尾。

所有数字均按网络（大端序）顺序排列。这是必要的，因为数据将存储在 独立于 arch 的目录，甚至在用户的主目录中。 /usr/share/mime

缓存文件必须以原子方式写入 - 写入临时名称，然后移动旧文件 - 所以 打开旧缓存文件并 mmap'ed 的客户端不会获得损坏的数据。

## 使用扩展属性存储 MIME 类型

实现还可以从user.mime\_type扩展属性获取文件的 MIME 类型。此处给出的类型通常应在 优先选择任何猜测的类型，因为用户可以显式设置它。 应用程序可以选择在保存文件时设置类型。由于许多应用程序和文件系统不支持扩展属性， 实现不得依赖于此方法可用。

## 子类化

类型是另一种类型的子类，如果第一个类型的任何实例是 也是第二个实例。例如，所有 image/svg+xml 文件也是 application/xml、text/plain 和 application/octet-stream 文件。子类是关于 格式，而不是数据的类别（例如，没有 所有电子表格都继承自的“通用电子表格”类）。

一些子类规则是隐式的：

- 所有文本/\* 类型都是文本/纯文本的子类。

- 所有可流式传输类型（即，除 inode/\* 类型之外的所有类型）是应用程序/八位字节流的子类。

除了这些规则之外，还可以使用 元素的子类。

请注意，某些文件格式也是压缩文件（应用程序/x-jar 文件也是应用程序/压缩文件）。但是，这与诸如 压缩的后脚本文件，它本身不是有效的后脚本文件（所以 application/x-gzpostscript 不继承自 application/postscript，因为可以处理后者的应用程序可能无法处理前者）。

某些类型可能是也可能不是其他类型的实例。例如，一个 电子表格文件可能被压缩或不压缩。这是一个有效的电子表格文件 无论哪种方式，但在一种情况下仅从应用程序/gzip 继承。这 信息不能静态表示；而是应用程序 对这些信息感兴趣应该运行所有魔术规则，并且 使用返回的类型列表作为子类。

## 建议的检查顺序

由于不同的应用程序有不同的要求，他们可能会选择 按任意顺序使用本规范提供的各种方法。然而，执行检查的建议顺序为：

- 如果显式提供了MIME类型（例如，通过ContentType HTTP标头，则MIME 电子邮件附件，扩展属性或其他一些方式）那么应该 被使用而不是猜测。
- 否则，请首先对文件名进行全局匹配。只保留重量最大的球体。如果模式不同，请仅保留具有最长模式的球体，如前所述。如果在此之后，有一个或多个匹配的球体，以及所有匹配的球体结果是相同的 MIMETYPE，请使用该 MIMETYPE 作为结果。
- 如果 glob 匹配失败或导致多个冲突的 mimetype，请阅读 文件的内容，并对其进行魔术嗅探。如果没有魔术规则与数据匹配（或者如果 内容不可用），使用默认类型的应用程序/八位字节流 二进制数据，或文本数据的文本/纯文本。如果没有全局匹配，请使用魔术匹配 结果。

注意：检查文件的前 128 个字节是否有 ASCII 控制字符是 猜测文件是二进制文件还是文本文件的好方法，但请注意，具有高位集的文件 字符仍应被视为文本，因为这些字符可以出现在 UTF-8 文本中，与控制字符不同。

- 如果由 glob 匹配产生的任何 mimetype 等于 或 魔术嗅探的结果，用这个作为结果。例如，这使我们能够 区分名为 “foo.doc” 的文本文件与同名的 MS-Word 文件，如 MS-Word文件的魔术匹配将是MS-Word类型的应用程序/ x-ole-storage。继承。
- 否则，请使用权重最高的 glob 匹配的结果。

在魔法之前检查球形图案有几个原因。首先，做魔术嗅探是非常昂贵的，因为读取文件的内容 导致大量搜索，这是非常昂贵的。其次，某些应用程序不检查 魔力完全（有时内容不可用或阅读速度太慢），以及这个 使两者更有可能获得相同的类型。

此外，用户可以轻松理解为什么调用他们的 文本文件使系统认为它是MP3，而他们很难理解为什么他们的计算机认为是一个PostScript文件。如果系统猜错了，用户通常可以重命名文件来解决问题。README.mp3README.txt

## 非常规文件

有时将 MIME 类型分配给文件系统中的其他对象很有用，例如目录、套接字和设备文件。这在查找时可能很有用 类型的图标，或用于提供这些对象之一的文本说明的图标。为此提供了媒体类型

“inode” ， 以下类型对应于 到 Unix 文件系统中的标准对象类型：

索引节点/块设备  
索引节点/字符设备  
索引节点/目录  
inode/fifo  
索引节点/挂载点  
索引节点/套接字  
索引节点/符号链接

inode/mount-point 是 inode/directory 的子类。添加额外内容时可能很有用 这些目录的操作，例如“装载”或“弹出”。挂载的目录可以是 通过将目录的“st\_dev” 与其父目录的“” 进行比较来检测。如果 它们不同，它们来自不同的设备，目录是挂载 点。

## 卷的内容类型

传统的 MIME 类型适用于单个文件或字节流。它通常很有用 在对可装入卷的内容进行分类时应用相同的方法，或 文件系统。为此引入了 x 内容类型。典型示例 是 X-Content/Audio-DVD、X-Content/Blank-CD 或 X-Content/Image-DCF。

内容类型的匹配适用于树 魔术 元素，这些元素 类似于用于 MIME 类型匹配的魔术 元素。而不是在文件中寻找字节序列，树 魔术 元素 允许在目录中查找具有某些名称，权限或MIME类型的文件 等级制度。

## URI 方案处理程序

URI 方案处理（例如处理 mms:// URI 的电影播放器或播客程序 处理 feed:// URI）通过处理 x-scheme-handler/foo 的应用程序进行处理 mime 类型，其中 foo 是有问题的 URI 方案。

此方案允许 URI 方案处理享受与 mime 类型处理程序相同的优势， 例如更改默认处理程序、跨桌面支持的能力，以及 更轻松地启动应用程序。

请注意，此虚拟 mime 类型不用于列出应用程序的 URI 方案 可以从中加载文件。例如，电影播放器不会列出 x-scheme-handler/http 在其支持的 MIME 类型中，但如果它支持 从 RTSP 位置播放。

## 安全隐患

本文档中描述的系统旨在允许不同的程序 以查看与具有相同类型的相同文件。这是为了帮助互操作性。以这种方式确定的类型只是一个猜测，应用程序不得 仅基于其 MIME 类型信任文件。例如，下载器应该 未经确认，不要将文件直接传递给启动器应用程序 因为类型看起来“无害”（例如，文本/纯文本）。

不要依赖两个应用程序为同一文件获取相同的类型，即使 如果他们都使用这个系统。该规范在实现中允许一些回旋余地， 在任何情况下，程序都可能遵循不同版本的规范。

## 用户修改

MIME 数据库不用于存储用户首选项。用户不应 编辑数据库。如果他们希望更正或提供 MIME 条目本身不提供这些的软件，他们应该通过以下方式提供 覆盖.xml在名为 [“目录布局”](#) 的部分中提到。诸如此类的信息 “text/html 文件需要用 Mozilla 打开” 不应该进入数据库。

## 贡献

托马斯·伦纳德<[tall97@users.sf.net](mailto:tall97@users.sf.net)>  
大卫·富尔<[faure@kde.org](mailto:faure@kde.org)>  
亚历克斯·拉尔森<[alex1@redhat.com](mailto:alex1@redhat.com)>  
塞思·尼克爾<[snickell@stanford.edu](mailto:snickell@stanford.edu)>  
基思·帕卡德<[keithp@keithp.com](mailto:keithp@keithp.com)>  
菲利普·范·雷姆东克<[mechanix@debian.org](mailto:mechanix@debian.org)>  
克里斯托斯·祖拉斯<[christos@zoulas.com](mailto:christos@zoulas.com)>  
马蒂亚斯·克拉森<[mclasen@redhat.com](mailto:mclasen@redhat.com)>  
巴斯蒂安·诺塞拉<[hadess@hadess.net](mailto:hadess@hadess.net)>

## 引用

[侏儒]GNOME 桌面, <http://www.gnome.org>  
[KDE]KDE 桌面, <http://www.kde.org>  
[罗克斯]ROX桌面, <http://rox.sourceforge.net>  
[桌面条目]桌面入门规范, <http://www.freedesktop.org/standards/desktop-entry-spec.html>  
[共享MIME]共享MIME信息数据库, <http://www.freedesktop.org/standards/shared-mime-info.html>  
[RFC-2119]在 RFC 中用于指示需求级别的关键字 <http://www.ietf.org/rfc/rfc2119.txt?number=2119>  
[BaseDir]XDG 基本目录规范<http://www.freedesktop.org/standards/basedir/draft/basedir-spec/basedir-spec.html>  
[ACAP]ACAP 媒体类型数据集类<ftp://ftp.ietf.org/internet-drafts/draft-ietf-acap-mediatype-01.txt>