

此页面由社区从英文翻译而来。了解更多并加入 MDN Web Docs 社区。

HTML 文本处理基础

HTML 的主要工作之一是赋予文本结构，使浏览器能够按照开发者的意图显示 HTML 文档。本文解释了 [HTML](#) 如何通过添加标题和段落、强调单词、创建列表等方式来构造文本。

前提:	阅读 开始学习 HTML ，了解基本的 HTML 知识。
目标:	学习如何用标记（段落、标题、列表、强调、引用等）来建立页面的基本文本结构和文本内容。

基础：标题和段落

大部分结构化文本由标题和段落组成。不管是小说、报刊、教科书还是杂志等。



内容结构化会使读者的阅读体验更轻松，更愉快。

在 HTML 中，每个段落是通过 `<p>` 元素标签进行定义的，比如下面这样：

```
HTML

<p>我是一个段落，千真万确。</p>
```

每个标题（Heading）都必须被包裹在一个标题元素中：

```
HTML

<h1>我是文章的标题</h1>
```

一共有六种标题元素标签——`h1`、`h2`、`h3`、`h4`、`h5` 和 `h6`。每个元素代表文档中不同级别的内容：`<h1>` 表示主标题，`<h2>` 表示二级子标题，`<h3>` 表示三级子标题，依此类推。

编辑结构层次

这里举一个例子。在一个故事中，`<h1>` 表示故事的名字，`<h2>` 表示每个章节的标题，`<h3>` 表示每个章节下的子标题，依此类推。

```
HTML
```

<h1>三国演义</h1>

<p>罗贯中</p>

<h2>第一回 宴桃园豪杰三结义 斩黄巾英雄首立功</h2>

<p>

话说天下大势，分久必合，合久必分。周末七国分争，并入于秦。及秦灭之后，楚、汉分争，又并入于汉.....

</p>

<h2>第二回 张翼德怒鞭督邮 何国舅谋诛宦竖</h2>

<p>

且说董卓字仲颖，陇西临洮人也，官拜河东太守，自来骄傲。当日怠慢了玄德，张飞性发，便欲杀之.....

</p>

<h3>却说张飞</h3>

<p>

却说张飞饮了数杯闷酒，乘马从馆驿前过，见五六十个老人，皆在门前痛哭。飞问其故，众老人答曰：“督邮逼勒县吏，欲害刘公；我等皆来苦告，不得放入，反遭把门人赶打！”.....

</p>

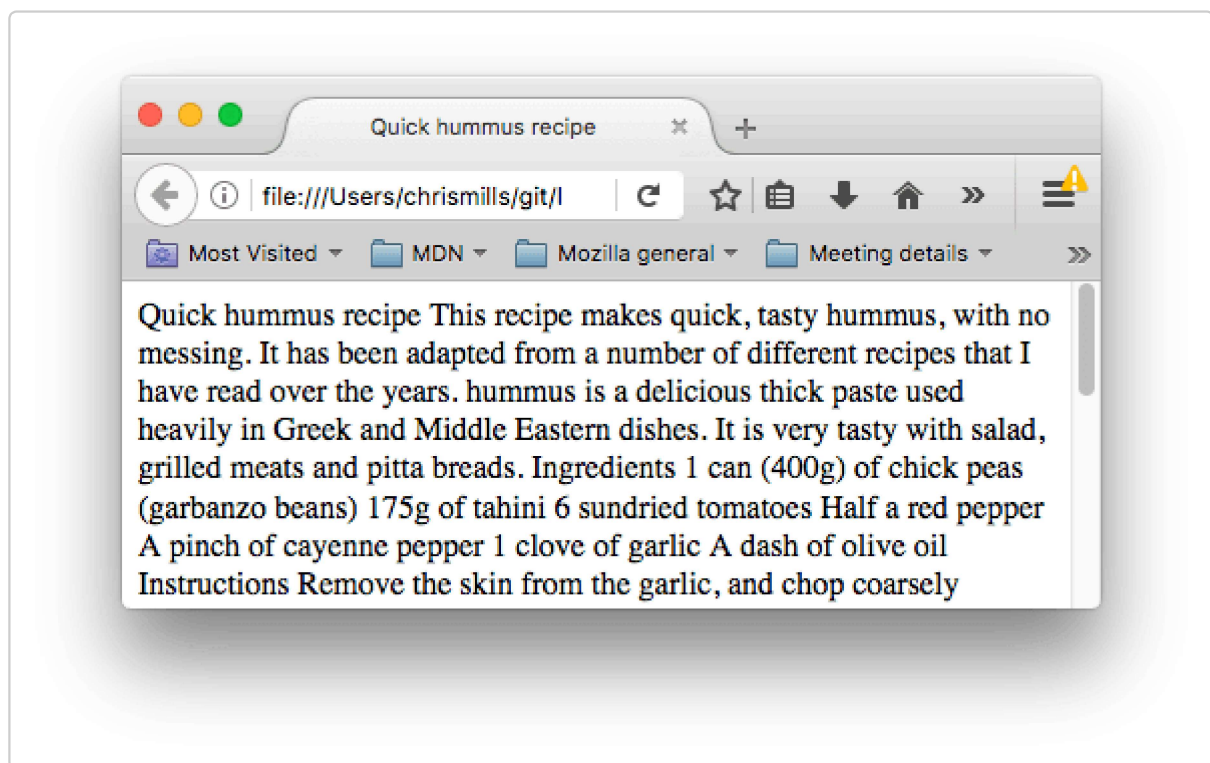
所涉及的元素具体代表什么，完全取决于作者编辑的内容，只要层次结构是合理的。在创建此类结构时，只需要记住一些最佳实践：

- 最好只对每个页面使用一次 <h1> ——这是顶级标题，所有其他标题位于层次结构中的下方。
- 请确保在层次结构中以正确的顺序使用标题。不要使用 <h3> 来表示副标题，后面再跟 <h2> 来表示二级副标题——这是没有意义的，会导致奇怪的结果。
- 在现有的六个标题层次中，除非觉得有必要，否则应该争取每页使用不超过三个。有很多层次的文件（例如，深层次的标题层次）会变得笨重，难以浏览。在这种情况下，如果可能的话，建议将内容分散到多个页面。

为什么我们需要结构化？

回答这个问题前，让我们先来看一段[文档示例](#) ——并从运行这段文档示例（美味的豆沙食谱）开始。你应该在本地机器上保存一份这个文件的副本，因为在以后的练习中会需要它。这个文件的主体目前包含了多条内容。它们没有任何标记，但都用换行符隔开（按回车键进入下一行）。

然而，当在浏览器中打开文档时，会看到文本显示为一整块！



这是因为没有元素给内容结构，所以浏览器不知道什么是标题，什么是段落。此外：

- 用户在阅读网页时，往往会快速浏览以查找相关内容，经常只是阅读开头的标题（我们通常在一个网页上会[花费很少的时间](#)）。如果用户不能在几秒内看到一些有用的内容，他们很可能会感到沮丧并离开。
- 对网页建立索引的搜索引擎将标题的内容视为影响网页搜索排名的重要关键字。没有标题，你的网页在[搜索引擎优化](#)方面效果不佳。
- 严重视力障碍者通常不会阅读网页；他们用听力来代替。完成这项工作的软件叫做[屏幕阅读器](#)。该软件提供了快速访问给定文本内容的方法。在使用的各种技术中，它们通过朗读标题来提供文档的概述，让用户能快速找到他们需要的信息。如果标题不可用，用户将不得不听到整个文档被大声朗读。
- 使用 [CSS](#) 样式化内容，或者使用 [JavaScript](#) 做一些有趣的事情，你需要包含相关内容的元素，使得 CSS / JavaScript 可以有效地定位它。

因此，我们需要给我们的内容进行结构性标记。

主动学习：编辑我们的内容结构

让我们直接进入一个运行实例。在下面的示例中，向[可编辑代码](#)字段中的原始文本添加元素，使其在[实时输出](#)字段中显示为标题和两个段落。

如果犯了错误，你可以使用[重置](#)按钮进行重置。如果你卡住了，请按[显示答案](#)按钮以查看答案。

Play

实时输出

静夜思 床前明月光 疑是地上霜 举头望明月 低头思故乡

可编辑代码

按 ESC 退出编辑区域，按 Tab 可插入制表符 '\t'

静夜思 床前明月光 疑是地上霜 举头望明月 低头思故乡

重置

显示答案

为什么我们需要语义？

在我们身边的任何地方都要依赖语义——我们依靠以前的经验来告诉我们一个日常物品的功能是什么；当我们看到某个东西时，我们知道它的功能是什么。举个例子，我们知道红色交通灯表示“停止”，绿色交通灯表示“通行”。如果运用了错误的语义，事情会迅速地变得非常棘手（难道有某个国家使用红色代表通行？我不希望如此）

同样的道理，我们需要确保使用了正确的元素来给予内容正确的含义、作用以及外形。在这里，[h1](#) 元素也是一个语义元素，它所包裹的文本具有“页面上的顶级标题”的作用（或意义）。

HTML

```
<h1>这是一个顶级标题</h1>
```

一般来说，浏览器会给它一个更大的字形来让它看上去像个标题（虽然也可以使用 CSS 让它变成任何你想要的样式）。更重要的是，它的语义值将以多种方式被使用，比如通过上文提到过的搜索引擎和屏幕阅读器。

在另一方面，你可以让任一元素看起来像一个顶级标题，考虑如下：

HTML

```
<span style="font-size: 32px; margin: 21px 0; display: block;"
>这是顶级标题吗? </span>
>
```

这是一个 `` 元素，它没有语义。当想要对它用 CSS（或者 JS）时，可以用它包裹内容，且不附加任何额外的意义（在未来的课程中你会发现更多这类元素）。我们已经对它使用了 CSS 来让它看起来像一个顶级标题。然而，由于它没有语义值，所以它不会有任何上文提到的帮助。最好的方法是使用相关的 HTML 元素来标记这个项目。

列表

现在让我们把注意力转向列表。在生活中，列表无处不在——从你的购物清单到你每天下意识地遵循的方向列表，再到你在这些教程中遵循的指令列表！在网络上，列表也无处不在，我们有三种不同的类型需要注意。

无序列表

无序列表用于标记列表项目顺序无关紧要的列表——让我们以购物清单为例。

豆浆
油条
豆汁
焦圈

每份无序的清单从 `` 元素开始，需要包裹清单上所有被列出的项目：

HTML

```
<ul>
  豆浆
  油条
  豆汁
  焦圈
</ul>
```

然后就是用 `` 元素把每个列出的项目单独包裹起来：

```
<ul>
  <li>豆浆</li>
  <li>油条</li>
  <li>豆汁</li>
  <li>焦圈</li>
</ul>
```

主动学习：标记无序列表

尝试编辑下面的示例来创建一个 HTML 无序列表。

Play

实时输出

豆浆 油条 豆汁 焦圈

可编辑代码

按 ESC 退出编辑区域，按 Tab 可插入制表符 '\t'

```
豆浆
油条
豆汁
焦圈
```

重置

显示答案

有序列表

有序列表需要按照项目的顺序列出来——让我们以一组方向为例：

沿这条路走到头

右转

直行穿过第一个十字路口

在第三个十字路口处左转

继续走 300 米，学校就在你的右手边

这个标记的结构和无序列表一样，除了需要用 `` 元素将所有项目包裹，而不是 ``：

```
HTML
<ol>
  <li>沿这条路走到头</li>
  <li>右转</li>
  <li>直行穿过第一个十字路口</li>
  <li>在第三个十字路口处左转</li>
  <li>继续走 300 米，学校就在你的右手边</li>
</ol>
```

主动学习：标记有序列表

尝试编辑下面的示例来创建一个 HTML 有序列表：

Play

实时输出

沿这条路走到头 右转 直行穿过第一个十字路口 在第三个十字路口处左转 继续走 300 米，学校就在你的右手边

可编辑代码

按 ESC 退出编辑区域，按 Tab 可插入制表符 't'

沿这条路走到头
右转
直行穿过第一个十字路口
在第三个十字路口处左转
继续走 300 米，学校就在你的右手边

重置 显示答案

主动学习：标记我们的食谱

到了这里，你拥有了所有你需要的信息来标记我们的食谱样例。你可以选择从 [text-start.html](#) 复制一份文件并保存在本地，打开它进行编辑，或者在下面的例子中进行编辑。在本地做可能会更好，因为这样你就可以保存你正在做的工作，而如果你把它填到可编辑的例子中，在你下次打开页面时，它就会丢失。各有利弊吧。

Play

实时输出

宫保鸡丁的做法 宫保鸡丁，川菜系中的传统名菜，由鸡丁、干辣椒、花生米等炒制而成。由于其入口鲜辣，鸡肉的鲜嫩配合花生的香脆，广受大众欢迎。相传宫保鸡丁是清朝光绪年间的署理四川总督丁宝桢所发明，是他招待客人时叫家厨煮的菜肴。由于丁宝桢后来被封为东宫少保（太子少保），所以被称为“丁宫保”，而这道菜亦被称为“宫保鸡丁”

原料 去骨鸡胸肉：一斤八两 干红辣椒：八钱 炸花生米：一两五钱 花椒粒：两大匙 葱：两根（切段） 蛋白：一个 淀粉：三大匙 酱油：两大匙 蒜末：半茶匙 糖：半茶匙 白醋：一茶匙 色拉油：适量 盐：两茶匙

做法 先用蛋白一个、盐半茶匙及淀粉两大匙搅拌均匀，调成“腌料”，鸡胸肉切成约一厘米见方的碎丁并用“腌料”搅拌均匀，腌渍半小时。用酱油一大匙、淀粉水一大匙、糖半茶匙、盐四分之一茶匙、白醋一茶匙、蒜末半茶匙调拌均匀，调成“综合调味料”。鸡丁腌好以后，色拉油下锅烧热，先将鸡丁倒入锅内，用大火快炸半分钟，炸到变色之后，捞出来沥干油汁备用。在锅里留下约两大匙油，烧热后将切好的干辣椒下锅，用小火炒香后，再放入花椒粒和葱段一起爆香。随后鸡丁重新下锅，用大火快炒片刻后，再倒入“综合调味料”继续快炒。如果你采用正宗川菜做法，最后只需加入花生米，炒拌几下就可以起锅了。如果你在北方，可加入黄瓜丁、胡萝卜丁和花生米，翻炒后起锅。

大千鸡 张大千居加拿大期间，曾按自己喜好改变宫保鸡丁的做法，并传授当地厨师，厨师将之命名为“大千鸡”，以兹纪念。大千鸡与宫保鸡丁不同之处，是使用经细工去皮、出骨、剔膜的鸡腿肉，以干辣椒、豆瓣酱为味，而且不用花生。

可编辑代码

按 ESC 退出编辑区域，按 Tab 可插入制表符'\t'

宫保鸡丁的做法
宫保鸡丁，川菜系中的传统名菜，由鸡丁、干辣椒、花生米等炒制而成。由于其入口鲜辣，鸡肉的鲜嫩配合花生的香脆，广受大众欢迎。
相传宫保鸡丁是清朝光绪年间的署理四川总督丁宝桢所发明，是他招待客人时叫家厨煮的菜肴。由于丁宝桢后来被封为东宫少保（太子少保），所以被称为“丁宫保”，而这道菜亦被称为“宫保鸡丁”

原料
去骨鸡胸肉：一斤八两
干红辣椒：八钱
炸花生米：一两五钱
花椒粒：两大匙
葱：两根（切段）
蛋白：一个

重置 显示答案

如果你感到棘手，你可以随时按下[显示答案按钮](#)，或者在我们的 GitHub 仓库上检查 [text-complete.html](#) 样例。

嵌套列表

将一个列表嵌入到另一个列表是完全可以的。你可能想让一些子项目列在一级项目之下。让我们从食谱示例中获取第二个列表：

HTML

```
<ol>
  <li>
    先用蛋白一个、盐半茶匙及淀粉两大匙搅拌均匀，调成“腌料”，鸡胸肉切成约一厘米见方的碎丁并用“腌料”搅拌均匀，腌渍半小时。
  </li>
  <li>
    用酱油一大匙、淀粉水一大匙、糖半茶匙、盐四分之一茶匙、白醋一茶匙、蒜末半茶匙调拌均匀，调成“综合调味料”。
  </li>
  <li>
    鸡丁腌好以后，色拉油下锅烧热，先将鸡丁倒入锅内，用大火快炸半分钟，炸到变色之后，捞出来沥干油汁备用。
  </li>
  <li>
    在锅里留下约两大匙油，烧热后将切好的干辣椒下锅，用小火炒香后，再放入花椒粒和葱段一起爆香。随后鸡丁重新下锅，用大火快炒片刻后，再倒入“综合调味料”继续快炒。
  </li>
  <li>如果你采用正宗川菜做法，最后只需加入花生米，炒拌几下就可以起锅了。</li>
  <li>如果你在北方，可加入黄瓜丁、胡萝卜丁和花生米，翻炒后起锅。</li>
</ol>
```

由于最后两项与它们的前一项非常密切相关（它们看起来更像该项的子项或选项），将它们编辑成无序列表，并嵌套在该项的子项中可能更合理。就像下面这样：

HTML

```
<ol>
  <li>
    先用蛋白一个、盐半茶匙及淀粉两大匙搅拌均匀，调成“腌料”，鸡胸肉切成约一厘米见方的碎丁并用“腌料”搅拌均匀，腌渍半小时。
  </li>
  <li>
    用酱油一大匙、淀粉水一大匙、糖半茶匙、盐四分之一茶匙、白醋一茶匙、蒜末半茶匙调拌均匀，调成
```

“综合调味料”。

鸡丁腌好以后，色拉油下锅烧热，先将鸡丁倒入锅内，用大火快炸半分钟，炸到变色之后，捞出来沥干油汁备用。

在锅里留下约两大匙油，烧热后将切好的干辣椒下锅，用小火炒香后，再放入花椒粒和葱段一起爆香。随后鸡丁重新下锅，用大火快炒片刻后，再倒入“综合调味料”继续快炒。

如果你采用正宗川菜做法，最后只需加入花生米，炒拌几下就可以起锅了。

如果你在北方，可加入黄瓜丁、胡萝卜丁和花生米，翻炒后起锅。

尝试回到上一个实践操作的例子中，并更新第二个列表。

重点强调

在日常用语中，我们常常会加重某个字的读音，或者用加粗等方式突出某句话的重点。与此类似，HTML 也提供了相应的标签，用于标记某些文本，使其具有加粗、倾斜、下划线等效果。下面，我们将学习一些最常见的标签。

强调

在口语表达中，我们有时会强调某些字，用来改变这句话的意思。同样地，在书面用语中，我们可以使用斜体字来达到同样的效果。例如，下面两个句子便有不同的意思：

我很庆幸你没有迟到。

我很庆幸你没有*迟到*。

第一句话听起来真的像松了一口气因为没有迟到。相反，第二句话听起来具有讽刺性而且有隐含的攻击性，表达对一个人迟到的恼怒。

在 HTML 中我们用 `` (emphasis) 元素来标记这样的情况。这样做既可以让文档读起来更有趣，也可以被屏幕阅读器识别，并以不同的语调发出。浏览器默认样式为斜体，但你不应该纯粹为了获得斜体风格而使用这个标签。如果仅仅为了获得斜体样式而不增加语义辅助，你应该使用 `` 元素和一些 CSS，或者是 `<i>` 元素（见下文）。

HTML

```
<p>我很<em>庆幸</em>你没有<em>迟到</em>。</p>
```

强烈的重要性

为了强调重要的词，在口语方面我们往往用重音强调，在书面用语中则是用粗体字来达到强调的效果。例如下面这段：

这杯液体**毒性很大**。
就指望你了，千万**不要**迟到！

在 HTML 中我们用 `` (strong importance) 元素来标记这样的情况。除了使文档更有用之外，也可以被屏幕阅读器识别，并以不同的语调发出。浏览器默认样式为粗体，但你不应该纯粹为了获得粗体风格而使用这个标签。如果仅仅为了获得粗体样式而不增加语义辅助，你应该使用 `` 元素和一些 CSS，或者是 `` 元素（见下文）。

HTML

```
<p>这杯液体<strong>毒性很大</strong>。</p>
```

```
<p>就指望你了，千万<strong>不要</strong>迟到！</p>
```

如有需要你可以将 strong 元素和 em 元素嵌套在其他的标签中：

HTML

```
<p>  
  这杯液体<strong>毒性很大</strong>——如果饮用了它，你<strong>可能<em>会死</em></strong>。  
</p>
```

主动学习：我们是重要的！

在这个主动学习部分，我们提供了一个可编辑的示例。在这个示例中，我们希望你能尝试在你认为需要斜体强调和着重强调的词上添加斜体强调和着重强调的内容，只是为了进行一些练习。

Play

实时输出

AlphaGo 李世乭五番棋

2016 年 3 月 8 日到 3 月 15 日，韩国职业棋士李世乭（이세돌）九段与由 Google DeepMind 开发的计算机围棋软件 AlphaGo 对弈的五局三胜制围棋比赛在韩国首尔举行。结果为 AlphaGo 以四胜一负的战绩击败李世乭。赛后韩国棋院授予 AlphaGo 荣誉九段的称号。

可编辑代码

按 ESC 退出编辑区域，按 Tab 可插入制表符 '\t'

AlphaGo 李世乭五番棋

<p>2016 年 3 月 8 日到 3 月 15 日，韩国职业棋士李世乭（이세돌）九段与由 Google DeepMind 开发的计算机围棋软件 AlphaGo 对弈的五局三胜制围棋比赛在韩国首尔举行。结果为 AlphaGo 以四胜一负的战绩击败李世乭。赛后韩国棋院授予 AlphaGo 荣誉九段的称号。</p>

重置

显示答案

斜体字、粗体字、下划线

到目前为止，我们所讨论的元素都有明确的相关语义。[](#)、[<i>](#) 和 [<u>](#) 的情况却有点复杂。它们出现于人们要在文本中使用粗体、斜体、下划线但 CSS 仍然不被完全支持的时期。像这样仅仅影响表象而且没有语义的元素，被称为**表象元素**（presentational elements）并且不应该再被使用。因为正如我们在之前看到的，语义对无障碍、SEO（搜索引擎优化）等非常重要。

HTML5 重新定义了 ``、`<i>` 和 `<u>`，赋予了它们新的但有点令人困惑的语义角色。

最好的经验法则是：只有在没有更合适的元素时，才适合使用 ``、`<i>` 或 `<u>` 来表达传统上用粗体、斜体或下划线表达的意思；而通常情况下是有更合适的元素可供使用的。``、``、`<mark>` 或 `` 可能是更加合适的选择。

始终保持无障碍的开发理念。斜体的概念对使用屏幕阅读器的人或使用拉丁字母以外的书写系统的人没有什么帮助。

- `<i>` 被用来传达传统上用斜体表达的意义：外国文字，分类名称，技术术语，一种思想.....
- `` 被用来传达传统上用粗体表达的意义：关键字，产品名称，引导句.....
- `<u>` 被用来传达传统上用下划线表达的意义：专有名词，拼写错误.....

备注：人们强烈地将下划线与超链接联系起来。因此，在网页中，最好只给链接加下划线。在语义上合适的时候使用 `<u>` 元素，但要考虑使用 CSS 将默认的下划线改为在网页中更合适的东西。下面的例子说明了如何做到这一点：

HTML

```
<!-- 学名 -->
```

```
<p>
```

红喉北蜂鸟（学名：`<i>Archilocus colubris</i>`）是北美东部最常见的蜂鸟品种。

```
</p>
```

```
<!-- 舶来词 -->
```

```
<p>
```

菜单上有好多舶来词汇，比如 `<i lang="uk-latn">vatrushka</i>`（东欧乳酪面包）、`<i`

`lang="id"`

`>nasi goreng</i`

`>`（印尼炒饭）以及 `<i lang="fr">soupe à l'oignon</i>`（法式洋葱汤）。

```
</p>
```

```
<!-- 已知的错误书写 -->
```

```
<p>总有一天我会改掉写<u class="spelling-error">措字</u>的毛病。</p>
```

```
<!-- 在定义中，被定义的术语 -->
```

```
<dl>
```

`<dt>`语义化 HTML`</dt>`

`<dd>`根据元素的``语义``意义而不是外观来使用它们。`</dd>`

```
</dl>
```

技能测试！

你已经到了本文的结尾，但你能记住最重要的信息吗？你可以找到一些进一步的测试，以验证你在继续前进之前已经保留了这些信息——参见[技能测试：HTML 文本基础知识 \(en-US\)](#)。

总结

现在就到此为止了！这篇文章应该已经让你对如何开始在 HTML 中标记文本有了一个很好的概念，并且向你介绍了这个领域中的一些最重要的元素。这个领域还有很多语义元素要介绍，我们将在本课程稍后的[文本格式化进阶](#)一文中看得更清楚。在下一篇文章中，作为 Web 上最重要的元素，我们将详细介绍如何[创建超链接](#)。

Help improve MDN

Was this page helpful to you?

[Learn how to contribute.](#)



This page was last modified on 2024年3月16日 by [MDN contributors](#).