

百度开放适配指南

目录

关于开放适配.....	2
开放适配的可选方案.....	2
一 自主适配.....	2
二 标注 Meta 声明	3
三 提交对应关系 sitemap	5
附录一 对应关系 Site map 文件格式及内容（ url 级别）	7
附录二 对应关系 Site map 文件格式及内容（ pattern 级别）	9

关于开放适配

背景：在移动终端上进行浏览，用户需要更符合移动终端特性的资源。相比传统的 PC 网页，专门为手机终端所设计的页面（xhtml/html5）以及 Webapp 在浏览及使用效果上更具优势。百度移动搜索会尽可能多地在搜索结果中给出手机页/Webapp 的结果，而对相对不适合移动终端的结果进行打压。

开放适配：在网站兼有 PC 站和手机站的前提下，同一内容的 PC 页与手机页之间往往存在简单和直接的对应关系。百度建议网站做好 PC 页与手机页之间的适配工作，以便百度将移动端对 PC 页的搜索请求引流至对应的手机页，一方面快速增加手机站流量，另一方面优化用户浏览效果，提升搜索排名。

开放适配的可选方案

百度共提供三种开放适配方案，网站可根据自身情况从中任意选择一种来操作，也可以同时选择几种方案来操作，并不影响适配效果。（注：适配只针对存在对应关系的页面，无对应关系的 PC 页面无需任何操作。）

开放适配可选方案分别为**自主适配**、**标注 Meta 声明**、**提交页面对应关系**。具体说明及操作方法如下：

一 自主适配

含义：

站点自己做好 PC 与手机之间的适配，以及手机各个版式之间的适配。当手机用户通过百度移动搜索或者其他方式访问站点的手机页（首页或者内容页）时，站点将其自动适配到该 PC 页面对应的手机页，并为之选择合适的页面版本。

举例：

- （1）当用户使用手机访问 <http://sports.sina.com.cn/nba/> 时，站点自行适配到 <http://nba.sina.cn>（频道页）
- （2）当用户使用手机访问 <http://bbs.gfan.com/android-4038549-1-1.html> 时，站点自行适配到 <http://bbs.gfan.com/mobile/android-4038549-1-1.html#>（内容页）

生效情况：

不仅限于百度移动搜索，当用户通过手动输入网址，或者其他移动搜索引擎、导航站等渠道访问站点时，适配同样有效。

二 标注 Meta 声明

含义：

站点如果自行适配有困难，可以在 PC 页面中做简单改造，百度协助实现适配效果。即：站长在站点 PC 页的源代码头部嵌入一行或多行 Meta 信息，由 Meta 信息来指明该 PC 页对应的手机页的 URL，以及该 URL 对应页面的格式，百度将根据用户终端类型选择最适合展示的页面。

站点标注完 Meta 声明后，还可选择提交一次性的对应关系 **sitemap**，以提高百度抓取效率。页面对应关系可以是 url 级别或者 pattern 级别（也可两种兼有），**且只需提交一次，无需后续维护。**

Meta 声明标注步骤

1. 梳理所有与手机页存在一一对应关系的 PC 页，用于添加 Meta 声明；无对应关系的 PC 页面无需添加 Meta 。

2. Meta 声明的格式：

```
<meta http-equiv="mobile-agent" content="format=[wml|xhtml|html5]; url=url">
```

注：红色字体部分是需要站点自定义的内容。

根据手机页的协议语言，选择[wml|xhtml|html5]中的一种。

url=url 中的后者，代表当前 PC 页所对应的手机页 url，两者必须是一一对应关系（而不是统一对应至手机站首页）

Meta 声明示例：

```
<meta http-equiv="mobile-agent" content="format=html5; url=http://3g.sina.com.cn/">
<meta http-equiv="mobile-agent" content="format=xhtml; url=http://sina.cn/">
```

3. 站长需要将 Meta 声明放在 PC 页源代码内部，如下：

```
<head>
<meta http-equiv="mobile-agent" content="format=html5;url=http://3g.sina.com.cn/">
.....
</head>
```

4. 站长在添加过程中应减少因标注错误而产生的负面影响，需要注意的事项有

- a. 标注正确的手机站 url，避免对应错误和拼写错误
- b. 确保适配到的手机页自身的可用性

页面对应关系 sitemap 格式及提交

url 级别 sitemap: 通过 XML 文件描述 url 级别的 PC 页与手机页的对应关系，网站当前有对应关系的所有 PC 页面 url 均需要给出。（格式及内容请参见[附录一](#)）

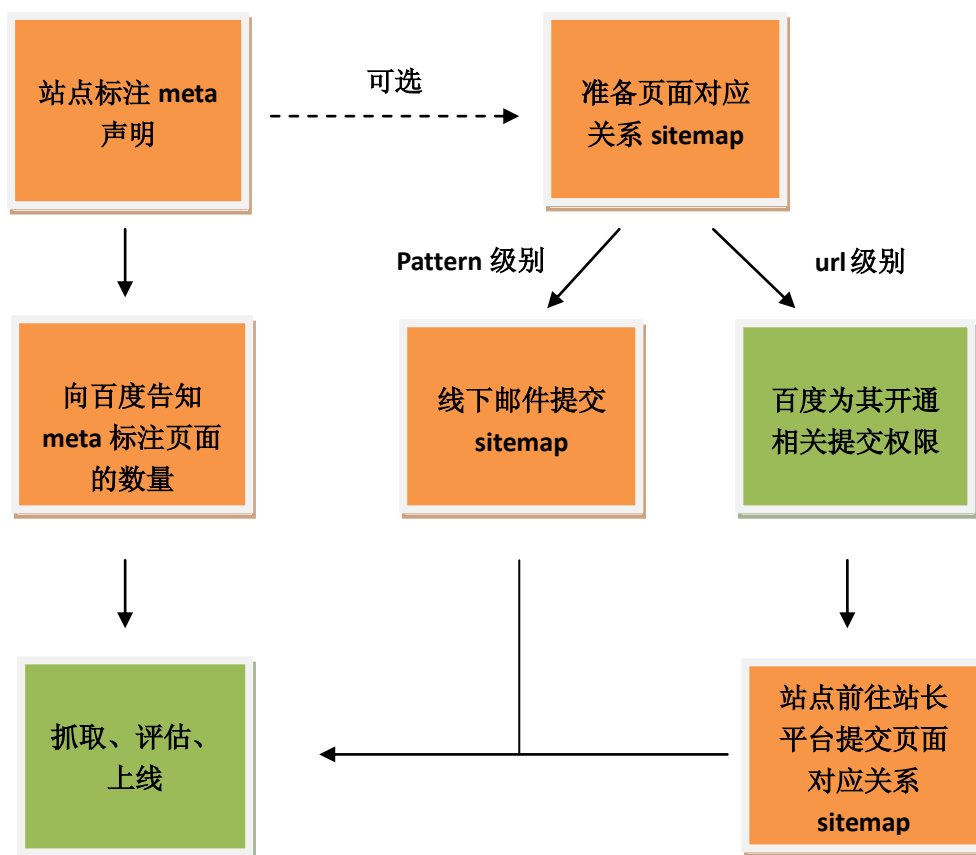
目前的提交方式为——与百度商务相关负责人联系，从而开通权限，在站长平台<http://zhanzhang.baidu.com/>提交 sitemap。

Pattern 级别 sitemap: 一个 pattern 级别的对应关系，往往可以表示成千上万的 url 对应关系。通过 pattern 的形式批量的提交对应关系。能够大大减少站长需要提交的数据量，提高 meta 对应关系的生效速度。（格式及内容请参见[附录二](#)）

目前的提交方式为——与百度商务相关负责人联系，通过线下邮件提交。

注：站点也可以根据自身情况提交兼有 url 和 pattern 的 sitemap，将具备简单对应关系的 url 聚合成 pattern，非简单对应关系的则仍以 url 对应形式存在。这种混合级别的 sitemap 提交方式为，url 级别部分提交至站长平台，pattern 级别部分线下邮件提交。

操作流程（绿色框为百度方需要操作的内容，黄色框为站点方需要操作的内容）



生效情况：

标注 Meta 声明这一适配方案仅限适用于百度移动搜索，即只有当用户通过百度移动搜索访问站点时，适配才会生效；通过其他渠道则不生效。

三 提交对应关系 sitemap

含义：

当站点推动 PC 页标注有困难时，可以选择直接提交页面对应关系 sitemap。

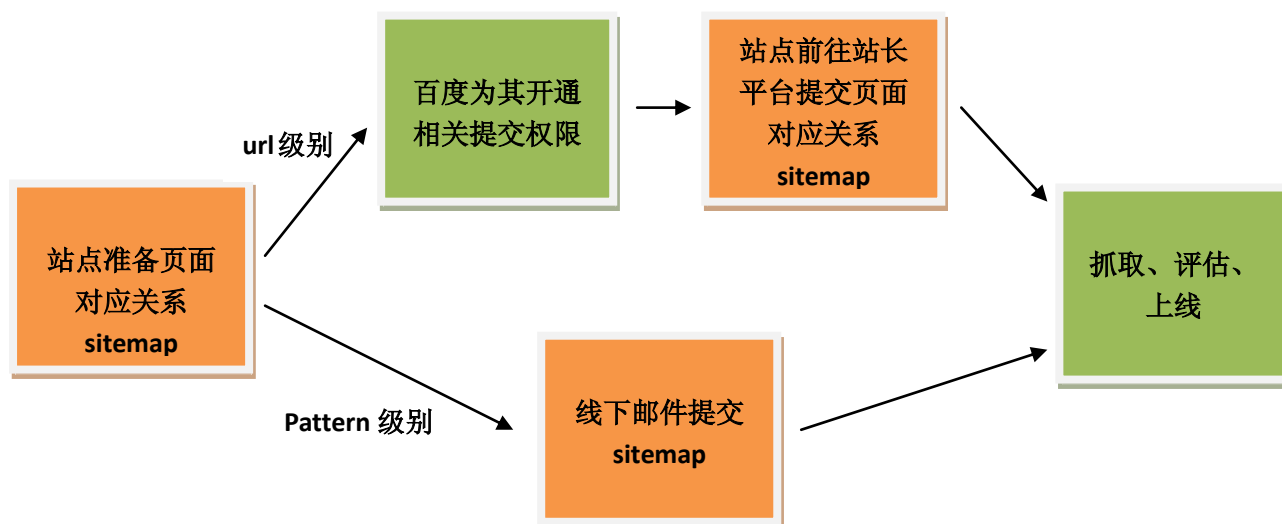
此方案的对应关系 sitemap 格式和内容与上一方案提到的相同，既可以是 url 级别或 pattern 级别，也可以两者兼有。（具体情况请见[附录](#)）

Sitemap 的提交方式也与 Meta 方案中提到的相同。

但与 Meta 方案不同的是，**直接提交对应关系 sitemap**，需要后续的更新和维护，一方

面站长要定期对提交到站长平台的 **sitemap** 进行更新，另一方面也要在站长平台上选择适当的更新周期，方便百度按时抓取。

操作流程（绿色框为百度方需要操作的内容，黄色框为站点方需要操作的内容）



生效情况：

与 meta 方案相同，仅限于百度移动搜索中生效。

更多常见问题，请点击官网了解

<http://open.shouji.baidu.com/?page=guide&action=mbsearch>

附录一 对应关系 Site map 文件格式及内容（url 级别）

通过 XML 文件描述 url 级别的 PC 页与手机页的对应关系，网站当前有对应关系的所有 PC 页面 url 均需要给出。格式如下：

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset>
<url>
<!--必填标签,表示 PC 页的 url 地址 -->
<loc>http://www.yoursite.html/yoursite.html</loc>
<data>
<display>
<!--可选标签,表示该 PC 页对应的 html5 版式的手机页的 url 地址 -->
<html5_url>http://html5.yoursite.html/yoursite.html</html5_url>
<!--可选标签,表示该 PC 页对应的 wml 版式的手机页的 url 地址 -->
<wml_url>http://wml.yoursite.html/yoursite.html</wml_url>
<!--可选标签,表示该 PC 页对应的 xhtml 版式的手机页的 url 地址 -->
<xhtml_url>http://xhtml.yoursite.html/yoursite.html</xhtml_url>
</display>
</data>
</url>
</urlset>
```

例如：

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset>
<url>
<loc>http://www.sina.com.cn/</loc>
<data>
<display>
<html5_url>http://html5.sina.com.cn/</html5_url>
<wml_url>http://wml.sina.com.cn/</wml_url>
<xhtml_url>http://3g.sina.com.cn/</xhtml_url>
</display>
</data>
</url>
<url>
<loc>http://www.sohu.com/</loc>
<data>
<display>
<html5_url>http://html5.sohu.com/</html5_url>
<wml_url>http://wml.sohu.com/</wml_url>
<xhtml_url>http://3g.sohu.com/</xhtml_url>
</display>
</data>
</url>
</urlset>
```


附录二 对应关系 Site map 文件格式及内容(pattern 级别)

格式

Pattern 级别对应关系提交支持 XML sitemap 的形式，每个 xml sitemap 文件可包含 1 组或者多组 pattern 对应关系。格式说明如下：

标签名称	标签说明	标签限制	标签路径	可选/必选
urlset	唯一的根节点，标记整个文档的开始和结尾	/	/	必选
url	url 标记每组 pattern 的开始和结束	1 个或多个	/urlset	必选
loc	表示 PC 站点的网址首页	最大长度 256 个字符，以 "http://" 开头	/urlset/url	必选
pc_url_pattern	表示该 PC 页的 url pattern	最大长度 256 个字符，以 "http://" 开头	/urlset/url/data/display	必选
html5_url_pattern	表示 PC 页对应的 html5 版式的手机页的 url pattern	最大长度 256 个字符，以 "http://" 开头	/urlset/url/data/display	可选
wml_url_pattern	表示 PC 页对应的 wml 版式的手机页的 url pattern	最大长度 256 个字符，以 "http://" 开头	/urlset/url/data/display	可选
xhtml_url_pattern	表示 PC 页对应的 xhtml 版式的手机页的 url pattern	最大长度 256 个字符，以 "http://" 开头	/urlset/url/data/display	可选

示例：

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset>
  <!-- 表示 news.163.com 下的一组 pattern 对应关系 -->
  <url>
    <loc>http://news.163.com/</loc>
    <data>
    <display>
    <pc_url_pattern>http://news.163.com/\(\d+\)/\(\d+\)/\(\d+\)/\(\w+\).html</ pc_url_pattern>
    <xhtml_url_pattern >http://3g.163.com/news/\${1}/\${2}/\${3}/\${4}.html</xhtml_url_pattern>
    <wml_url_pattern >http://wap.163.com/wml/page/ntes/\${1}/\${2}/\${3}/\${4}.wml</wml_url_p
```

```

attemn>
</display>
</data>
</url>

```

<!-- 下一组 pattern 放在一个新的 url 节点内 -->

```

</urlset>

```

字段详细说明

pc_url_pattern: 表示 PC 页 pattern，在 PC 页 url 的基础上，首先确定 url 中哪些路径或参数是可替换的。然后根据其类型，使用正则匹配符号(\d+)或者(\w+)表示该路径或参数。(\d+)表示纯数字字符串，(\w+)表示数字或字母组成的字符串。

xhtml_url_pattern / html5_url_pattern/ wml_url_pattern: 表示 xhtml/html5/wml 版式的手机页 pattern，在手机页 url 的基础上，根据可替换参数在对应的 PC 页 pattern 中出现的顺序，依次用\${1}，\${2}，.....表示该参数。

举例

以网易的新闻页“漫游接电话移动也不要钱了”为例，其 PC 页 url 地址为 <http://news.163.com/09/1001/07/5KH8DE1F000120GR.html>，其对应的 xhtml 格式的手机页 url 地址为 <http://3g.163.com/news/09/1001/07/5KH8DE1F000120GR.html>。

步骤一: 确定 PC 页中的可替换参数或者路径，得到其位置序号和类型。

PC 页url: <http://news.163.com/09/1001/07/5KH8DE1F000120GR.html>



序号	可替换路径/参数	类型
1	09	纯数字
2	1001	纯数字
3	07	纯数字
4	5KH8DE1F000120GR	字母或者数字

根据网站自身 url 的层次结构，其中 09,1001,07 和 5KH8DE1F000120GR 为动态可替换的路径。除 5KH8DE1F000120GR 为字母和数字混合外，其余均为纯数字。

步骤二: 根据可替换参数或路径的类型，得到 PC 页 pattern (即 sitemap 中的 pc_url_pattern)。使用正则匹配符号(\d+)或者(\w+)表示该路径或参数。(\d+)表示纯数字字符串，(\w+)表示数字或字母组成的字符串。

PC页url: <http://news.163.com/09/1001/07/5KH8DE1F000120GR.html>

PC页pattern: [http://news.163.com/\(\d+\)/\(\d+\)/\(\d+\)/\(\w+\).html](http://news.163.com/(\d+)/(\d+)/(\d+)/(\w+).html)

步骤三：根据手机页 url，以及可替换参数在步骤一中的位置序号，依次用\${1}，\${2}，.....表示替换掉手机页 url 中的可替换参数或路径，得到手机页 pattern。

手机页url: <http://3g.163.com/news/09/1001/07/5KH8DE1F000120GR.html>

手机页pattern: [http://3g.163.com/news/\\${1}/\\${2}/\\${3}/\\${4}.html](http://3g.163.com/news/${1}/${2}/${3}/${4}.html)

例子中，依次用\${1},\${2},\${3}和\${4}替换掉 09,1001,07 和 5KH8DE1F000120GR，得到手机页 pattern 为 [http://3g.163.com/news/\\${1}/\\${2}/\\${3}/\\${4}.html](http://3g.163.com/news/${1}/${2}/${3}/${4}.html)。

步骤四：根据手机页的页面类型，将手机页 pattern 写入对应的字段。如示例中该 pattern 的手机页面为 xhtml 版式，将 pattern [http://3g.163.com/news/\\${1}/\\${2}/\\${3}/\\${4}.html](http://3g.163.com/news/${1}/${2}/${3}/${4}.html) 写入到 xhtml_url_pattern 字段中。

最终得到 sitemap 文件如下：

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset>
  <!--表示 news.163.com 下的一组 pattern 对应关系-->
  <url>
    <loc>http://news.163.com/</loc>
    <data>
      <display>
        <pc_url_pattern>http://news.163.com/\(\d+\)/\(\d+\)/\(\d+\)/\(\w+\).html</ pc_url_pattern>
        <xhtml_url_pattern >http://3g.163.com/news/\${1}/\${2}/\${3}/\${4}.html</xhtml_url_pattern>
        <!-- 一组 PC pattern 可能含有多个版式的手机页 pattern -->
      </display>
    </data>
  </url>
  <!-- 下一组 pattern 放在一个新的 url 节点内 -->
</urlset>
```

备注

- 通过 `sitemap` 只能提交具有简单对应关系的 `pattern`，也即，在 PC 页 url 与手机页路径或参数之间存在直接的对应关系，使得机器能根据 PC 页 url 按照一定的正则 `pattern` 自动计算出与其对应的手机页的 url。如
<http://news.sina.com.cn/c/2012-05-27/235824487992.shtml> 和
<http://news.sina.cn/?sa=t124v71d6155414&pos=102&vt=4> 虽然是对应的，但是从 url 本身很难找到联系，就不属于简单对应关系。
- 正则匹配符号不可嵌套使用，如 `(\d+(\w+))` 这种形式不合法。
- `Pattern` 中，站长无需考虑特殊符号的正则转义等问题，百度会帮您转义成机器可以识别的正则表达式。