

C#文件编码检测：轻松解析文本文件编码及应用实例

作者：微信公众号：【架构师老卢】

12-18 12:19

22



概述：在C#中，通过`StreamReader`和`Encoding`类可检测文本文件编码。示例代码演示了读取文件并通过BOM检测文件编码，支持UTF-8、UTF-16等。此方法可用于处理不同编码的文本文件。

在C#中检测文本文件的编码可以使用`Encoding`类及`StreamReader`类。以下是一个示例，演示如何检测文本文件的编码：

```
1 using System;
2 using System.IO;
3 using System.Text;
4
5 class Program
6 {
7     static void Main()
8     {
9         string filePath = "路径\\到\\你的\\文件.txt";
10
11         // 读取文件内容
12         string content = ReadFile(filePath);
13
14         // 检测文件编码
15         Encoding encoding = DetectFileEncoding(filePath);
16
17         Console.WriteLine($"文件编码: {encoding.EncodingName}");
18     }
19
20     static string ReadFile(string filePath)
21     {
22         string content = "";
23
24         try
25         {
26             using (StreamReader reader = new StreamReader(filePath, true))
27             {
28                 content = reader.ReadToEnd();
29             }
30         }
31         catch (Exception ex)
32         {
33             Console.WriteLine($"读取文件时发生错误: {ex.Message}");
34         }
35
36         return content;
37     }
38
39     static Encoding DetectFileEncoding(string filePath)
40     {
41         byte[] buffer = new byte[4096];
42
43         try
44         {
45             using (FileStream fileStream = new FileStream(filePath, FileMode.Open, FileAccess.Read))
46             {
47                 fileStream.Read(buffer, 0, buffer.Length);
48             }
49         }
50         catch (Exception ex)
51         {
52             Console.WriteLine($"读取文件时发生错误: {ex.Message}");
53             return null;
54         }
55
56         return DetectEncoding(buffer);
57     }
58
59     static Encoding DetectEncoding(byte[] buffer)
60     {
61         // BOM (字节顺序标记) 检测
62         if (buffer.Length >= 2 && buffer[0] == 0xFF && buffer[1] == 0xFE)
63         {
64             return Encoding.Unicode; // UTF-16 little-endian
65         }
66         else if (buffer.Length >= 2 && buffer[0] == 0xFE && buffer[1] == 0xFF)
67         {
68             return Encoding.BigEndianUnicode; // UTF-16 big-endian
69         }
70         else if (buffer.Length >= 3 && buffer[0] == 0xEF && buffer[1] == 0xBB && buffer[2] == 0xBF)
71         {
72             return Encoding.UTF8; // UTF-8
73         }
74         else
75         {
76             // 默认为 ANSI 编码
77             return Encoding.Default;
78         }
79     }
80 }
```

上述代码中，`ReadFile`方法用于读取文件内容，而`DetectFileEncoding`方法通过读取文件头部来检测文件编码。`DetectEncoding`方法根据文件头的BOM（字节顺序标记）来判断文件编码。检测完成后，返回`Encoding`对象。

源代码获取：公众号回复消息【code: 71599】

相关代码下载地址



重要提示!：取消关注公众号后将无法再启用回复功能，不支持解封!

第一步：微信扫码关注公众号“架构师老卢”

第二步：在公众号聊天框发送 **code: 71599**，如：  **code: 71599**   获取下载地址

第三步：恭喜你，快去下载你想要的资源吧