
Subject: problem when parse html with XMLParser
Posted by [akebee](#) on Fri, 22 Sep 2017 08:40:39 GMT
[View Forum Message](#) <> [Reply to Message](#)

```
xml =
"<html>"
"<head><meta content=\"always\" name=\"referrer\"></head>"
"<TITLE>A Midsummer Night's Dream</TITLE>"
"</html>";
```

```
XmlParser p(xml);
while(!p.IsTag())
  p.Skip();
p.PassTag("html");
while(!p.End())
  if(p.Tag("head"))
  {
    while(!p.End()) {
      if(p.TagE("meta")) {
        // crash here because "<meta" doesn't have "/>"
      }
      else
        p.Skip();
    }
  }
}
```

```
if(p.Tag("TITLE")) {
  String TITLE = p.ReadText();
  LOG(TITLE);
  p.PassEnd();
}
else
  p.Skip();
```

the "<meta" in "<head>" doesn't have "/>" so crashed
is there any way to skip "<head>"?

Subject: Re: problem when parse html with XMLParser
Posted by [mirek](#) on Sat, 23 Sep 2017 07:35:27 GMT
[View Forum Message](#) <> [Reply to Message](#)

First of all, it is not a crash, but exception:

CONSOLE_APP_MAIN

```

{
String xml =
"<html>"
"<head><meta content=\"always\" name=\"referrer\"></head>"
"<TITLE>A Midsummer Night's Dream</TITLE>"
"</html>";

XmlParser p(xml);
try {
while(!p.IsTag())
p.Skip();
p.PassTag("html");
while(!p.End())
if(p.Tag("head"))
{
while(!p.End()) {
if(p.TagE("meta")) {
// *exception* here because "<meta" doesn't have ">"
}
else
p.Skip();
}
}
}

if(p.Tag("TITLE")) {
String TITLE = p.ReadText();
LOG(TITLE);
p.PassEnd();
}
else
p.Skip();
}
catch(XmlError) {
LOG("ERROR!");
}
}

```

...that covers case when you need to detect invalid XML, but not need to parse it.

U++ can even parse invalid XML, using Relaxed or Raw flags:

```

CONSOLE_APP_MAIN
{
String xml =

```

```
"<html>"
"<head><meta content=\"always\" name=\"referrer\"></head>"
"<TITLE>A Midsummer Night's Dream</TITLE>"
"</html>";
```

```
XmlParser p(xml);
p.Relaxed();
try {
    while(!p.IsTag())
        p.Skip();
    p.PassTag("html");
    while(!p.IsEof())
        if(p.Tag("head"))
        {
            while(!p.End("head")) {
                if(p.Tag("meta")) {
                    DUMP(p["content"]);
                }
                else
                    p.Skip();
            }
        }
    else
        if(p.Tag("TITLE")) {
            String TITLE = p.ReadText();
            DUMP(TITLE);
            p.PassEnd();
        }
        else
            p.Skip();
}
catch(XmlError e) {
    LOG("ERROR: " << e);
}

LOG("---- Done");
}
```

Subject: Re: problem when parse html with XMLParser
Posted by [akebee](#) on Mon, 25 Sep 2017 01:19:14 GMT
[View Forum Message](#) <> [Reply to Message](#)

tk's very much!
