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" dones`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");
}
```