

---

Question

[Jenna Poindexter](#) · Jun 1, 2016

## Best Way to Parse a HTML Page

I would like to write some code to parse a set of HTML pages from the internet in order to gather information from each web page.

All of the web pages are generated using a template, so the format of each of the web-pages is consistent with one-another and the information that I want to gather is always located in the same logical place within the page.

What is the best way to parse an html page in order to gather information at a specific place?

Can XML XPATH be used here? Does anyone have any examples of parsing HTML content?

[#XML](#) [#ObjectScript](#) [#Frontend](#) [#HTML](#) [#Caché](#)

---

Source URL: <https://community.intersystems.com/post/best-way-parse-html-page>