

ScalaでWebスクレイパー作ってみた

Scala超☆初心者セッション
教えて!偉い人!?



@qtamaki

自己紹介

当方、rubyプログラマ

35歳男

Java初心者

Scalaは超☆初心者です(はあと)

Scalaの実力はおそらくピクミン級



@qtamaki

自己紹介



そんな平民出の俺に
やっと巡ってきた幸運か
それとも破滅の罠か

そんな僕が、今日は、セッションを持
たせていただくことができとても光
栄です。

orz



@qtamaki

Webスクレイパーの流れ

スクレイパーっても大したものじゃなくて、単にHTMLからリストを抜き出して、RSSを取得したいだけの簡単なものです。



@qtamaki

動機



このサイト

RSS対応していねえ

早くなんとかしないと・・・

とあるニュースサイトの新着をRSSで
取ろうとしたら、RSS対応していなかつ
たので、作ってみようかと。



@qtamaki

実装

というか、悩み

- HTML取得、どーするの？
- XML解析、どーするの？
- XML構築、どーするの？
- 正規表現、どーするの？
- Webフレームワークは？
- ホスティングできる？

続きはコードで ➡➡



@qtamaki

ちなみにRubyの場合

当方、rubyプログラマ

Rubyだったら、 こんな目じゃないぜ？

```
agent = HTTPClient.new
res = agent.get("http://www.ryutsuu.biz/")
doc = Nokogiri::HTML(res.body.content)
arr = []
doc.xpath('//td[@id="main"]/ul/li').each do |x|
  arr << { :href => li.xpath('a').first['href'], :text =>
x.xpath('a').first.text, :date => li.xpath('span').first.text}
end
```



@qtamaki