

HTML → BibTeX(ML)

Představení a motivace

- BibTeX

```
@MISC { someID,  
  author = {Name Surname and  
  Another One},  
  title = {Title Abc},  
  year = {2013},  
  note = {to appear, CHI} }
```

- BibTeXML

(strict X ext.)

- Python 3

```
<bibtex:file  
xmlns:bibtex="http://bibtexml  
.sf.net/">  
<bibtex:entry id="someID">  
<bibtex:misc>  
<bibtex:author>Name Surname;  
Another One</bibtex:author>  
<bibtex:title>Title  
Abc</bibtex:title>
```

Seznámení...

Weby diametrálně odlišné

=> profily

Sekce neodpovídají BibTeX

=> odhad

Řádky si neodpovídají

=> rezignace

Hackování

přímo BibTeX – téměř nikde

>95% citací – >1 zásadních chyb

(zad. BP)

parser pro každý web/řádek – realita

python – regexpy nerad

XPath – záchrana?

Řešení

Profil + trocha fuzzy + XPath prefix

**unix-like batch chování (okamžitě
chyba)**

lxml, BeautifulSoup

Řešení – pokr.

locals() + bez tříd

pojmenované argumenty + static

look-behind

regex precompile + simulace

lxml close tags

BibTeX parser

Zajímavosti

BibTeXML

TODO try to distinguish between 'Name X.,' and 'Name, X.'

impossible: Chambers, N., J. Allen, et al.

possible?: Chambers, Nol., J. Allen,

possible?: Chambers, Nolan, J. Allen,

impossible: Allen, J. F. and C. R. Perrault

impossible: Hinkelman, E. and J.F. Allen

impossible: Traum, D. and Allen, J.F.

The image features a large, pixelated green question mark centered on a dark gray background. The background is composed of several overlapping, semi-transparent shapes in various shades of gray and brown, creating a layered, abstract effect. A bright green rectangular border frames the entire scene. The question mark itself is rendered in a bold, pixelated style, with a thick green outline and a solid green fill. It is positioned in the center of the frame, slightly offset from the top and bottom edges.

?