

EXTRAÇÃO DE DADOS RELACIONADOS A SAÚDE MENTAL

Utilização da biblioteca *Scrapy*
para extração de dados
relacionados a saúde mental dos
repositórios da UFPE

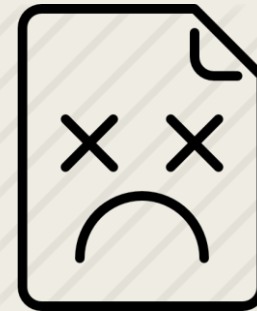


Problema



Muitos repositórios não
relacionados

+

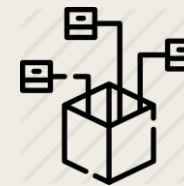


Dados inconsistentes

Solução



WEB CRAWLER

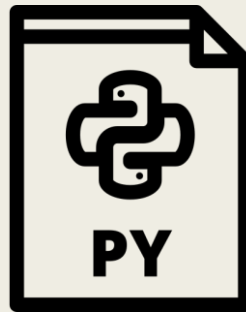


Buscar todos os repositórios
relacionados a saúde mental



Padronizar os dados

Ferramentas



Linguagem Python

+



Scrapy

Biblioteca *Scrapy*

Desenvolvimento



Definição
da URL
inicial

+



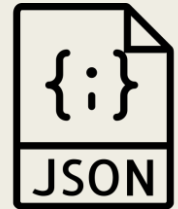
Seleção dos
Artigos

+



*Regular
Expressions*
para extrair
dados

+



Armazenamento
dos resultados
em arquivo JSON

Regular Expressions

Categorías de Metacaracteres:

- Quantificador
- Grupo

Modelo:

- `regex_title = re.compile('Title: </td><td
class=\"metadataFieldValue\">([^\<]+)')`
- `regex_author = re.compile('Authors: </td><td
class=\"metadataFieldValue\">(?:<\s*a[^\>]*>(.*),(.*)<\s*/\s*a
>))')`

Resultados

```
{
  "title": "Demanda de Saúde Mental na Estratégia Saúde
da Família em Recife-PE: Aspectos Epidemiológicos",
  "author": "Aluska Mirtes Araujo Souza",
  "keywords": [
    "Saúde Mental",
    "Programa Saúde da Família",
    "Educação em Saúde",
    "Enfermagem Psiquiátrica"
  ],
  "date": "Dec-2013",
  "abstract": "A Saúde Mental no Brasil é constituída
por serviços assistenciais de base comunitária,
implantados ..."

  "uri":
  "https://repositorio.ufpe.br/handle/123456789/11061",
  "type": "Dissertações de Mestrado"
}
```



OBRIGADO

Heitor Gabriel
Mateus Rodrigues
Muriel Souza da Cruz