

# Trabalho Prático 1

SPLN 24-25, mestrado Informática

## Contents

1	1- Descarregar o arquivo (usando OAI-PMH)	1
2	2- Estudar a estrutura do documento	1
3	3- Calcular a árvore arquivística de fundos	1
4	4- Criar uma árvore de diretorias	2
5	4a- Preencher um wiki com esta informação	2
6	4b- Criar HTML estático	2
7	6- Script de procura	2
8	7- Entidades mencionadas	2
9	8- Focar em BiogHist e ScopeContent (outros?)	2
10	9- Explorar thesaurus/indices (ver campo “terms”)	2
11	10- Arranjar um tipo para cada nó	2

## 1 1- Descarregar o arquivo (usando OAI-PMH)

- Criar uma script que descarrega o arquivo inteiro (ou parte)

## 2 2- Estudar a estrutura do documento

- Dependências funcionais, chaves, campos constantes
- Arranjar uma função de prettyprint textual
  - ? (yaml?) ou (wiki) ou até html
  - ? basead num (yaml/estrutura python) de configuração
  - simplificar
  - agregar alguns campos numa notação mais legível data(inicio, fim, certeza)

## 3 3- Calcular a árvore arquivística de fundos

- Usando os atributos `Parent` e eventualmente `RootParent`
- cuidado pode haver né em falta (ex: chamar-lhe 777-fixme)
- F (fundo)
- SC (secção)
- SSC (subsecção)
- SR (Série)
- UI (unidade de instalação)
- D DC (documento, documento composto)

## 4 4- Criar uma árvore de diretorias

Exemplo:

```
31258-F-Arquivo Casa de Pindela
  31259-SC-Luís de Carvalho e Beatriz de Almeida
    31260-SSC-Luís de Carvalho e Beatriz Almeida
      54460-SR-Documentos
    31262-SC-Simão Pinheiro e Leonor Almeida e Helena Dias
      31263-SSC-Simão Pinheiro e Leonor Almeida
        54461-SR-Documentos
      31265-SSC-Simão Pinheiro e Helena Dias
        54462-SR-Documentos
    31287-SC-João Machado Fagundes da Guerra Pinheiro e Figueira e Mariana Josefa de Castro
      31288-SSC-João Machado Fagundes da Guerra Pinheiro e Figueira e Mariana Josefa de Castro
        54469-SR-Documentos
        98419-SR-Correspondência
          98896-UI-Jerónimo Vaz Vieira
          98903-UI-João Pereira de Sousa
          98922-UI-António Vaz Vieira
          99010-UI-António Francisco Valbom
```

## 5 4a- Preencher um wiki com esta informação

Ex zim-wiki (wiki desktop)

## 6 4b- Criar HTML estático

- html hiperligado (árvore de fundos + documentos + ligação ao arquivo original)

## 7 6- Script de procura

várias hipóteses. Script python usando

- 6a usando rg
- criar índice glimpse
- sqlite + freetext search

## 8 7- Entidades mencionadas

- Criar um índice de entidades mencionadas
  - pessoas (e sua profissão / título)
  - lugares, casas

## 9 8- Focar em BiogHist e ScopeContent (outros?)

- tabela de Pessoas e suas biografias

## 10 9- Explorar thesaurus/índices (ver campo “terms”)

- estrutura classificativa

## 11 10- Arranjar um tipo para cada nó

- formalizar (parcialmente) o tipo dos documentos, séries, DC, etc
- fotografia, carta, livro, seq(cartas)
- set(registo)