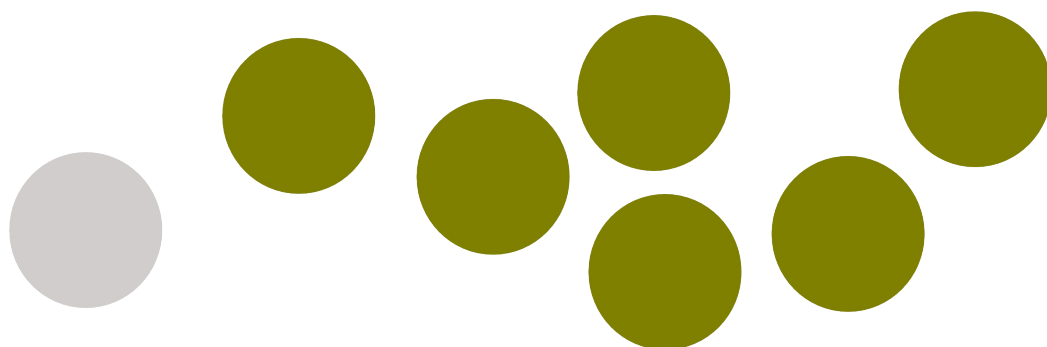




mais parlamento



logo não oficial



Rua Conde de Vizela, 95
Porto, Portugal
[#hv](http://irc.freenode.org)



logo não oficial

Transformar a informação pública do parlamento.pt em algo que os cidadãos possam consultar, relacionar, escrutinar.

Aproximar o cidadão da actividade parlamentar.

Criar uma API que permita a outros interessados e a aplicações variadas reutilizar os dados recolhidos.



Fase 1 - Source

Fonte: parlamento.pt

Escrito em Html

Cada entidade tem o seu id único – deputados, legislaturas, etc.



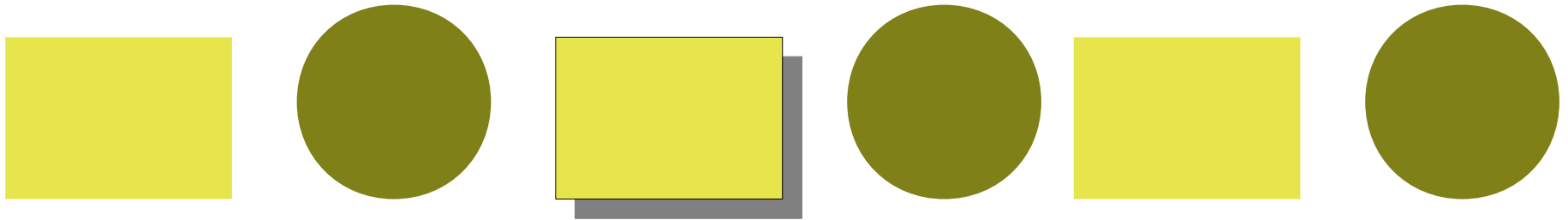
Fase 2 a - Crawler

Criado em php

Informação recolhida da página de Deputados - biografia

Fase 2 b - Feeds

Ensaio em rss e atom

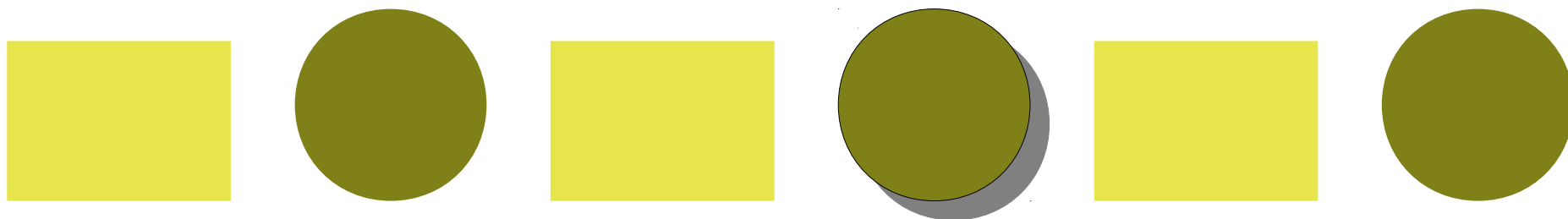


Fase 3 – Tabelas BD

Criadas 3 tabelas gerais

- MP (identificação de deputados)
- Caucus (Legislaturas dos deputados)
- Facts (Cargos dos deputados)

Possibilidade de aplicar filtros



Fase 4 – BD centralizada

Limpeza e uniformização de dados (ex. profissões, nomes)

Introdução de dados obtidos em fontes que não o parlamento.pt (ex. rede social dos deputados)

Actualização e estado último dos dados

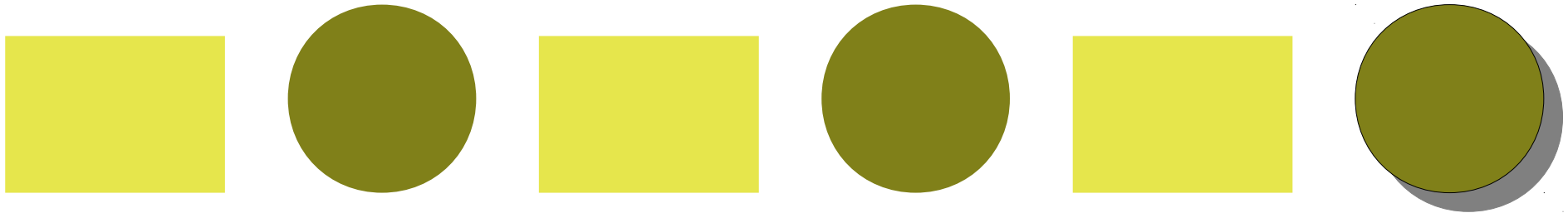
Consultas à BD por todos os participantes, mediante introdução de chave



Fase 5 – BD(s) centralizada(s)

Integração e ou criação de nova BD com informação recolhida dos pdf das sessões legislativas desde 1998 à data (~100/ ano)

- Download massivo dos pdf com addon do Mozilla “DownThemAll”
- Conversão de pdf em csv com “pdf2txt” do ubuntu
- Parsing com script em python
- Análise temática e linguística



Fase 6 – Tratamento, consultas, visualização, apresentação

- Interface tipo Google?
- Interface com múltiplos widgets?
- Website Drupal?
 - Consultas simples/ listagens
 - Tags, taxonomias
 - Visualização de dados (VIDI)
 - Mapas com informação associada / Openlayers