

Cofinanciado por:



UNIÃO EUROPEIA

Fundo Social Europeu

API - Autocomplete

Introdução

A API de autocomplete é um serviço de *Backend* que devolve uma lista das sugestões mais prováveis de completar um *input* inicial.

A API assenta num princípio de acesso REST, mas como apenas se efetuam pesquisas, só está disponível o método GET.

Acesso

URL raiz da API

Todas as chamadas à API iniciam pelo url: <https://apife.ine.pt>

URL de posicionamento

Do ponto de vista do consumidor do serviço, segue-se o segmento “**dic**” (que indica que se pretende aceder a um dicionário). Por fim o segmento identificador do dicionário que se pretende usar para autocomplete.

Nesta fase, os dicionários de teste disponíveis são:

- Profissões: https://apife.ine.pt/dic/_CPP2010
- Atividade Económica: <https://apife.ine.pt/dic/CAEREV3>
- Cursos Superiores: <https://apife.ine.pt/dic/CURSOSUPC2021>

Utilização

Existem dois casos de uso disponíveis para consumo:

Prefetch

/prefetch (https://apife.ine.pt/dic/{identificador_de_dicionario}/prefetch)

Para o dicionário identificado, retorna uma lista de entradas mais frequentes. Pode ser invocado e guardado em cache no cliente do autocomplete.

Pesquisa

?q=XXXX (https://apife.ine.pt/dic/{identificador_de_dicionario}/?q={texto_a_pesquisar})

Cofinanciado por:



UNIÃO EUROPEIA

Fundo Social Europeu

Exemplo: https://apife.ine.pt/dic/_CPP2010/?q=baila

Retorna a lista das profissões que contêm a partícula “baila”. Como “Bailarino” não é uma profissão frequente, não constará da lista obtida no “prefetch”.

Estrutura

O *prefetch* e a pesquisa devolvem *arrays* em *JSON* com objetos que têm a estrutura:

```
[ { c : "AAA", d : "BBBB", t : "CCCC" }, ... ]
```

Em cada elemento:

- “c” contém o código;
- “d” a designação a apresentar como sugestão;
- “t” uma *string* de palavras separadas por espaços que denominaremos de *tokens*.

A ordem dos elementos no *array* reflete a sua ordenação por relevância (primeiro os mais relevantes).

Dicionários

A base de construção dos Dicionários, para além das listas oficiais de codificação (CAE Rev3, CPP 2010, CNAEF), reside em todo o histórico de codificação manual de mais de 30 operações estatísticas realizadas ao longo de cerca de 8 anos no âmbito dos Inquéritos às Famílias. À data, o total de entrevistas realizadas ultrapassava as 600000. Foram consideradas elegíveis para enriquecer os classificadores todas as expressões (1) com uma frequência igual ou superior a 10 e uma consistência de codificação de 90% e (2) com uma frequência igual ou superior a 5 e uma consistência de codificação de 100%. De seguida, foi calculada uma distância métrica entre as expressões já existentes no classificador e as restantes do histórico. Foi utilizada a Optical String Alingment - uma extensão da medida de Levenshtein - para o cálculo da distância a um intervalo de 1 a 3. Após validação, as expressões que se verificaram equivalentes no significado, mas distintas na grafia, foram integradas nos Dicionários.

Cofinanciado por:



UNIÃO EUROPEIA

Fundo Social Europeu

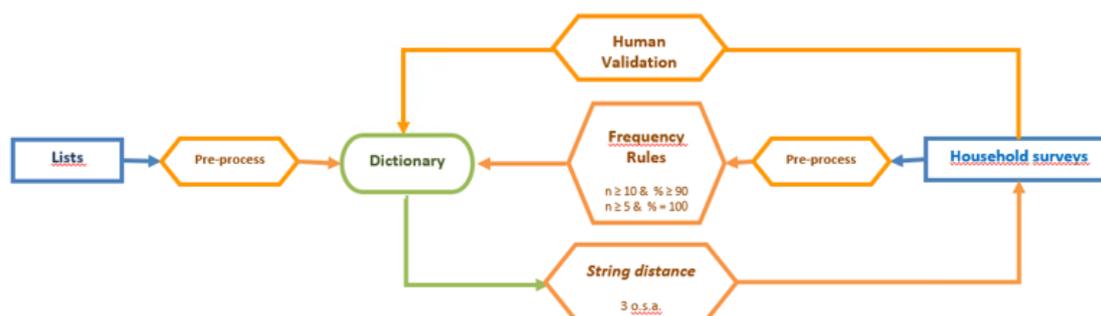


Figura 1- Esquema de Criação de Dicionários

Nomenclaturas

Como referido a API efetua a classificação de expressões com base em três nomenclaturas

- Profissões: https://apife.ine.pt/dic/_CPP2010
- Atividade Económica: <https://apife.ine.pt/dic/CAEREV3>
- Cursos Superiores: <https://apife.ine.pt/dic/CURSOSUPC2021>

Para a classificação das profissões é utilizada a Versão SMI : V02014- Classificação portuguesa das profissões, CPP 2010 que pode ser consultada em:

<https://smi.ine.pt/Versao/Detalhes/2014?modal=1>

Para a classificação da Atividade económica é utilizada a Versão SMI: V00554 - Classificação portuguesa das atividades económicas, revisão 3 que pode ser consultada em:

<https://smi.ine.pt/Versao/Detalhes/554?modal=1>

Para a classificação dos Cursos Superiores é utilizada a Versão SMI: V04477 - Qualificações do ensino superior, 2020 (Cursos - IINQE) que pode ser consultada em:

<https://smi.ine.pt/Versao/Detalhes/4477?modal=1>