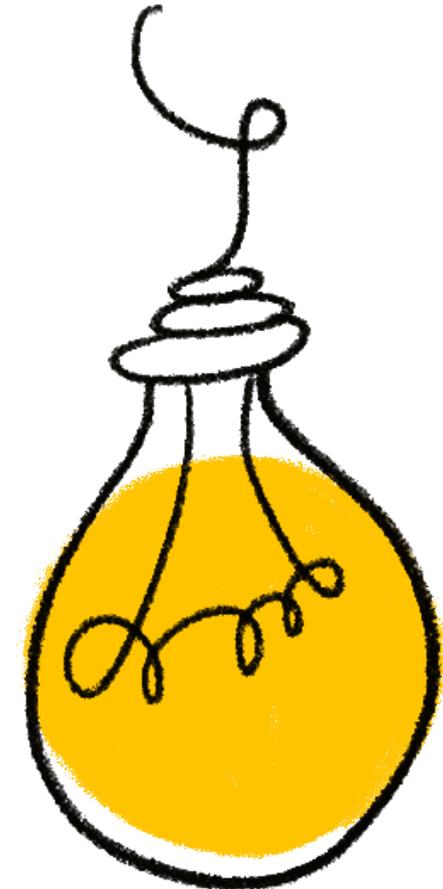


**Motor multifunções:
pesquisa
terminológica
bilíngue
e assistente de escrita
académica com base
em dados científicos
abertos**



Micaela
Aguiar
José Monteiro
Sílvia Araújo
Universidade
do Minho
(Portugal)



01 PortLinguE

é um projeto financiado que usa dados abertos para criar recursos de apoio à literacia académica e às línguas de especialidade



os recursos serão disponibilizados na plataforma online Lang2Science

02 Motor Multifunções



O objetivo é criar um motor multifunções que faça o aproveitamento da quantidade inumerável de conteúdo científico em regime de acesso aberto disponível em repositórios científicos.



01 PESQUISA BILÍNGUE DE
TERMINOLOGIA

02 ASSISTENTE DE
ESCRITA ACADÊMICA



Arquitetura do Motor

O sistema assenta na *framework* Django.

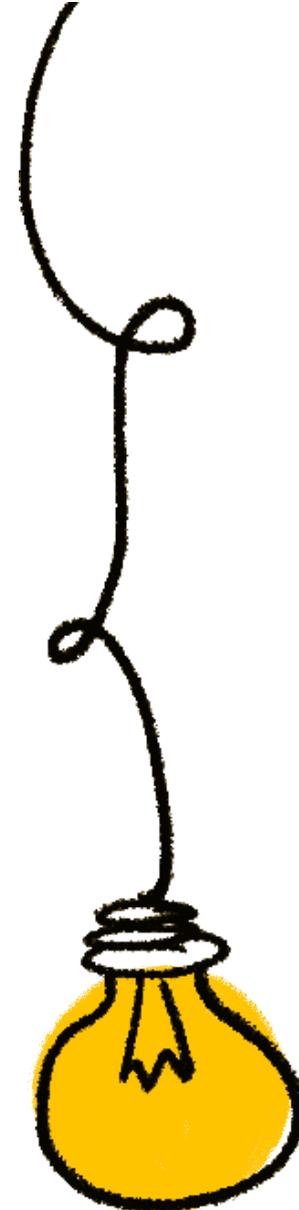
Com esta *framework* conseguimos estabelecer o tipo de pedidos que o utilizador poderá realizar mas também os resultados que são devolvidos ao utilizador.

03 Arquitetura do Motor



BERT é um modelo de Processamento de Linguagem Natural que analisa corpus de texto em termos das semelhanças ao nível do significado das palavras, das colocações e das frases e distribui os dados processados com base na similaridade semântica, gerando assim vetores semânticos.

o motor funciona com modelos de machine learning BERT



03 Arquitetura do Motor

Para a pesquisa terminológica bilíngue

são usados dois modelos, um pré-treinado com um corpus português e outro com um corpus inglês, o que permitirá transformar as pesquisas dos utilizadores e os artigos extraídos dos repositórios científicos em vetores semânticos

Para o assistente de escrita

apenas o modelo português será necessário para transformar em vetores semânticos as pesquisas do utilizador e as frases-modelo

Os artigos, as frases-modelo e os vetores semânticos são armazenados numa base de dados Elasticsearch.

É difícil encontrar
textos online
alinhados com as suas
traduções

Para identificar
textos
comparáveis, é
usado o BERT.

corpus de artigos
científicos
originalmente não
paralelos

pesquisa bilíngue

um motor de
pesquisa
"semântico"

É capaz de
encontrar
resultados por
contexto e não
apenas por
palavras-chave
específicas

o motor está restrito
a repositórios
científicos

assistente de escrita



Parte da criação
de um banco de
frases académicas
em Português
Europeu.



O assistente
de escrita usa
a mesma
tecnologia.



O motor será capaz de
analisar semanticamente
o input do utilizador e
sugerir outras frases
semelhantes ou com
funções similares.



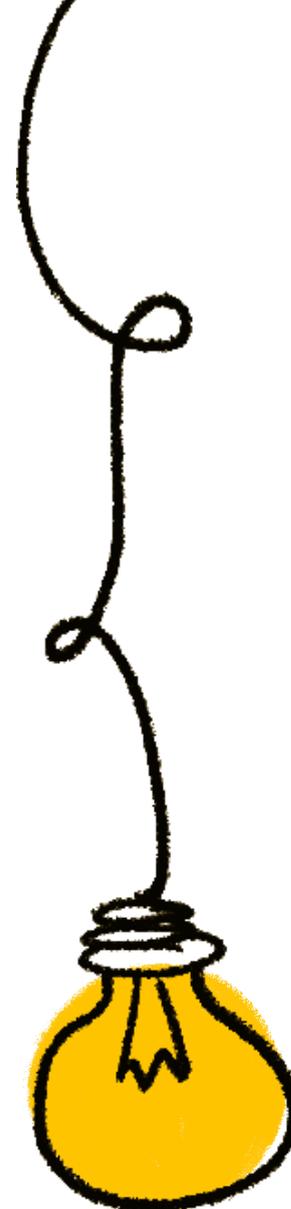
DESIGN

O design da interface facilitará uma utilização intuitiva para o utilizador, à semelhança de motores de pesquisa, como o Google.

04

CONCLUSÕES

O motor multifunções visa oferecer um recurso gratuito de apoio à literacia académica e científica e será disponibilizado na plataforma online Lang2Science.



Agradecimentos



Este trabalho foi realizado no âmbito do projeto PortLinguE (PTDC/LLT-LIG/31113/2017), financiado pelo FEDER no âmbito do Portugal 2020 e por fundos nacionais através da Fundação para a Ciência e a Tecnologia, I.P. (FCT, I.P.).