



# Terminologia e Tecnologias da Informação

## Módulo 3: Ferramentas para Tradução e Terminologia

### Exercício prático:

- 1) Cria um ficheiro **JSON** chamado “tm.json”, que simule uma **Memória de Tradução** e represente a seguinte hierarquia de dados:

**document**

**metadata**

**autor (=nome do criador)**

**date (=data da criação)**

**translationUnits**

**translationUnit**

**sourceLang**

**text (=texto original)**

**language (=língua)**

**translatedtLang**

**text (=texto traduzido)**

**language (=língua)**

Popula o dataset com 10 translationUnit fictícios.

Critério de avaliação: adequada estruturação dos dados e domínio do conceito de Memória de Tradução.

- 2) Desenvolve um programa em Python com o nome “nlp.py”, capaz de executar as seguintes tarefas de Tokenization nesta ordem:
- Abrir o ficheiro “tm.json” e carregar os dados numa variável para leitura.
  - Contar quantas translationUnit existem no dataset, armazenar numa variável e efetuar print no ecrã do valor com o texto “Number of translationUnit: xx”.
  - Realizar um loop sobre cada translationUnit e em cada iteração efetuar o seguinte: armazenar o “text” da translatedtLang numa variável, contar quantas palavras individuais não repetidas existem nesse texto utilizando uma função, e exibir na tela essa contabilização com os seguintes dados: “Translated Text: xxx” quebra de linha “Individual Tokens: xx”.

Enviar os 2 ficheiros para paulo.jorge.pm@elach.uminho.pt com nome e nº de aluno.