

Relatório SPLN

Rúben Cerqueira	Júlio Alves
PG47626	PG47390
Universidade do Minho	Universidade do Minho

April 7, 2022

1 Introdução

Neste relatório iremos apresentar diferentes módulos para Python, todos eles com áreas de atuação distintas, e que nos acreditamos que sejam módulos com bastante potencial de utilização e que são capazes de simplificar bastante o trabalho daqueles que trabalhem sobre as áreas respetivas de cada módulo. Para a limpeza de texto iremos apresentar o Clean-Text, para a criação, manipulação, formatação e conversão de datas, horas e timestamps o Arrow e para a monitorização de eventos no sistema de ficheiros o Watchdog.

2 Clean-Text

O clean-text é um módulo bastante simples, cujo objetivo é eliminar texto que tenha sido corrompido ou para fazer uma limpeza da informação contida num texto.

A sua instalação é bastante simples, bastando correr o comando *pip3 install -U clean-text*. Recomenda-se a utilização da flag -U de forma a atualizar o pacote para a versão mais recente.

Não oferece suporte para todas as linguagens, sendo que apenas inglês e alemão são totalmente suportados, no entanto, deverá funcionar em grande parte das linguagens ocidentais.

Oferece funções que nos permitem, entre outras, remover emails, números de telemóvel, dígitos, pontuação, do input fornecido, e possui uma função principal *clean*, que possui uma grande variedade de flags que podemos escolher.

3 Arrow

Arrow é um módulo extremamente competente na sua missão de criar, manipular, formatar e converter datas, horas e timestamps.

Tal como o clean-text, para o obter, basta correr o comando *pip3 install -U*

arrow.

Aparece como substituto para o módulo *datetime* e oferece suporte para todas as versões superiores a Python3.6, inclusive.

Tem, por padrão, noção de fuso horário e de UTC.

4 Watchdog

O módulo Watchdog permite implementar um observer capaz de gerar eventos que reflitam alterações do path observado. Estes eventos podem ser criação, edição, remoção e movimento de ficheiros ou diretorias.

Além disso também fornece uma classe abstracta que consome esses eventos para o programador poder realizar operações mediante o evento em questão.

Um possível cenário consiste na leitura de um ficheiro que foi adicionado nesse mesmo instante. Esta lógica pode ser útil no processamento de dados vindos de uma outra aplicação dependente da atual.

A simplicidade deste módulo é uma das razões que é muitas vezes escolhido para funções de monitorização de diretorias.

5 Conclusão

Em suma, estes três módulos são uma mais valia no contexto de processamento de linguagem natural uma vez que trabalham dados de uma forma muito simplista.

O facto de estes módulos serem todos bastante simples, tendo uma curva de aprendizagem perto de nula, ou até mesmo nula, e o facto de serem módulos que permitem simplificar as tarefas de um utilizador, levando a que possam ser utilizados em qualquer escala, permitem que sejam facilmente adotados por qualquer tipo de programador, seja ele um iniciante a desenvolver os seus primeiros projetos ou um programador experiente a produzir software bastante complexo. Por estes motivos, e pelos motivos abordados anteriormente no relatório, estes módulos apresentam-se como módulos com um potencial de utilização bastante elevado, dando bastantes possibilidades ao seu utilizador final, e que devem ser, pelo menos, conhecidos da comunidade, uma vez que poderão ser bastante úteis a algum membro da mesma em algum ponto da sua jornada pelo mundo da programação.