

Análise de texto de fontes desestruturadas e Web

Maciel Calebe Vidal

macielcv@insper.edu.br

Na aula de hoje

- Expressões regulares (Regex)

Motivação

- Transformações no texto
 - Limpeza
- Busca por padrões
 - Extrair informações relevantes



Sobre o autor

Heverton Peixoto é CEO-Presidente da Wiz. Graduado em Engenharia Civil, com MBA em Corporate Finance no Insead, foi consultor da McKinsey & Company de 2008 a 2013 em projetos estratégicos no mercado bancário e de seguros da América Latina.

Mais colunas

Com tantas mudanças e incertezas, como não sentir cansaço ou ansiedade? Nunca se discutiu tanto sobre o cuidado com as equipes, recomendações para descanso e racionalização do número de reuniões, entre tantas medidas para conter a sensação generalizada de esgotamento, mais conhecida como burnout. A pandemia nos trouxe esse desafio adicional e vem mostrando o quanto uma cultura institucional consolidada e orientada à geração de resultados com propósito é essencial para lidarmos com tudo isso.

O Brasil é o quinto país com maior agravamento da saúde mental, segundo pesquisa da Ipsos. Entre os brasileiros, 53% disseram que sua saúde mental piorou nos últimos 12 meses. Estudo feito pela Harris para a Microsoft mostra que a maior sensação de esgotamento entre pessoas em trabalho remoto ocorre no Brasil, onde 44% dizem que a pandemia aumentou a exaustão.

Isso não apenas tem afetado a qualidade de vida no trabalho como também dificulta a criação de senso de propósito nas empresas. Nas últimas colunas, falamos sobre a importância de uma cultura de engajamento e propósito para os colaboradores e do quanto isso traz energia e sentido para o trabalho ao passo em que gera melhores resultados. Agora, falamos da proteção da qualidade de vida dos nossos colaboradores, da saúde mental de cada um dos que estão em suas casas, distantes e muitas vezes isolados. Cultura corporativa tornou-se elemento crítico, algo que temos todos que discutir e no que devemos atuar, independentemente da área ou departamento.

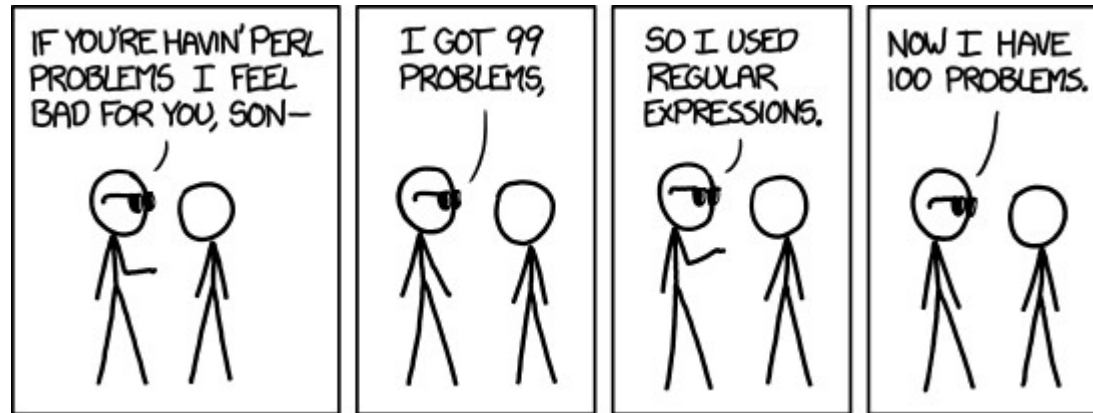
Como uma empresa pode combater este efeito e proteger seus colaboradores? Voltamos ao principal ativo que uma companhia pode criar, sua cultura. Uma cultura que gere propósito e seja verdadeiramente focada em resultado pode ser vital na proteção dos nossos times, atuando em quatro pilares principais:

Expressões regulares

- Poderosa
- Flexível
- Eficiente

Expressões regulares

- Notação
- Representar padrões



Relembrando...

- Na aula 02

```
In [33]: # Esta não é a melhor forma de fazer isso, veremos outras formas mais simples no futuro!
def cep_ok(cep):
    if len(cep) != 9 or cep[5] != '-':
        return False
    elif any(not char.isnumeric() for char in cep.replace('-', '', 1)):
        return False
    else:
        return True

cep_ok('78663-001')
```

Out[33]: True



Dúvidas?

Obrigado pela participação!