

# Alterações e implementações

# calendario → Nova implementação

- Se o dia da semana for igual retorna score da categoria dates 100%
- Se o ano for igual retorna score 100%
- Se for “fim de semana” e no texto for “sabado/domingo” retorna 100%

**A alterar :** Se ano igual, **mas** dia diferente **ou** se ano diferente, **mas** dia igual retornar → 0 %

# Inside outside – nova implementação

- 1 → indoor do places365
- 2 → outdoor do places 365

- Se no texto for inside, retorna o score de todos os “1” somados - “indoor\_score”
- Se no texto for outside, retorna o score de todos os “ 2 “ somados - “outdoor\_score”

# atividade - alteração

- procura maior semelhança entre “verbo do texto” e a “atividade da imagem”
- procura maior semelhança entre “frase do texto” e a “atividade da imagem”
- A maior destas duas é a que conta para a confiança

**Exemplo:**

- using public transports vs transports → semelhança alta
- using vs transports → semelhança baixa

## Progressos

Foram analisadas 65.000 imagens para o tópico 1.

O algoritmo gera um ficheiro .txt com os resultados :

```
- , -----_-----_.jpg , -----  
1 , 20160905_185411_000.jpg , 0.6735127356202351  
1 , 20160905_185443_000.jpg , 0.6729643181912222  
1 , 20160905_185515_000.jpg , 0.6500754233531866  
1 , 20160905_185547_000.jpg , 0.6742467655881141  
1 , 20160905_185619_000.jpg , 0.7226317408482577  
1 , 20160905_185651_000.jpg , 0.6692918467013497  
1 , 20160905_185723_000.jpg , 0.7201390219417775  
1 , 20160905_185755_000.jpg , 0.6196854322567772  
1 , 20160905_185827_000.jpg , 0.5740021563901909  
1 , 20160905_185859_000.jpg , 0.6345257295393558  
1 , 20160905_185931_000.jpg , 0.544893147818281  
1 , 20160905_190003_000.jpg , 0.6463557140634857  
1 , 20160905_190035_000.jpg , 0.7005099861745135  
1 , 20160905_190107_000.jpg , 0.7076762110480448  
1 , 20160905_190139_000.jpg , 0.6745407174490332  
1 , 20160905_190211_000.jpg , 0.6472218055923369  
1 , 20160905_190243_000.jpg , 0.6245854274265485  
1 , 20160905_190315_000.jpg , 0.6173146174127252
```

**formato:** [tópico, nome de imagem, confiança]

## Métricas de avaliação :

F1 → Precision and cluster recall. Top 5, 10, 20, 30 , 40 e 50

Para calcular a precision :

		Actual Class y		
		Positive	Negative	
$h_{\theta}(x)$ Test outcome	Test outcome positive	True positive (TP)	False positive (FP, Type I error)	Precision = $\frac{\#TP}{\#TP + \#FP}$
	Test outcome negative	False negative (FN, Type II error)	True negative (TN)	Negative predictive value = $\frac{\#TN}{\#FN + \#TN}$
		Sensitivity = $\frac{\#TP}{\#TP + \#FN}$	Specificity = $\frac{\#TN}{\#FP + \#TN}$	Accuracy = $\frac{\#TP + \#TN}{\#TOTAL}$

Para calcular o cluster recall :

The image displays three screenshots of LibreOffice Writer documents used for image clustering evaluation:

- Left document (ground\_truth):** Lists image IDs and their assigned clusters. Example: `1, b00001213_2116bq_20150306_174403e.jpg, 1`.
- Middle document (ground\_truth):** Shows a list of clusters. Example: `1, 1`, `1, 2`, `1, 3`, `1, 4`, `2, 1`, `2, 2`, `3, 1`, `3, 2`, `4, 1`, `4, 2`, `4, 3`, `5, 1`, `5, 2`, `5, 3`, `5, 4`, `5, 5`, `5, 6`, `5, 7`, `5, 8`, `5, 9`, `6, 1`, `6, 2`, `7, 1`, `7, 2`, `7, 3`, `7, 4`, `7, 5`, `7, 6`, `7, 7`, `7, 8`, `7, 9`, `8, 1`, `8, 2`, `9, 1`, `10, 1`.
- Right document (resultados):** Shows the predicted clusters for each image. Example: `1, 20160905_184643_000.jpg, 0.6725467160473284`.

ficheiro da esquerda (ground\_truth) → [tópico, imagem, cluster a que pertence]

ficheiro do meio (ground\_truth) → [tópicos, numero de clusters existentes no tópico]

ficheiro da direita (resultados) → [tópico, imagem, confiança]

**cálculo :**  $[\text{n}^\circ \text{ diferente de clusters encontrados}] / [\text{n}^\circ \text{ de clusters existentes para o tópico}]$

Supondo que vou fazer para o top 5 imagens (ficheiro da direita) para o tópico 1.  
Vamos ao ficheiro da esquerda (ground\_truth) buscar os clusters dessas 5 imagens.  
Vamos ao ficheiro do meio (ground\_truth) buscar o número de clusters que o tópico 1 tem (4)

$$\begin{array}{c} 1, 2 \\ 1, 2 \\ 1, 3 \\ 1, 3 \\ 1, 3 \end{array}$$

## Dúvidas :

**“solução atual”** : Em vez de usar as top 50 imagens com confiança mais alta, estou a ir buscar as top 50 imagens com confiança mais alta, que também pertencem ao groundtruth. (o ficheiro groundtruth apenas tem cerca de 3500 imagens):

### Resultado :

[illegible][illegible]

Cluster das imagens :

```
['3', '1', '3', '4', '4', '4', '3', '4', '1', '4', '1', '4', '3', '3', '4', '3', '3', '4', '4', '3', '4', '3', '4', '3', '4', '4', '3', '3', '3', '3', '4', '3', '1', '4', '2', '3', '4', '3', '4', '4',  
'4', '3', '3']
```