Alteraçoes e implementaçoes

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 \underline{ou} se ano diferente, **mas** dia igual retornar $\rightarrow 0 \%$

Inside outside – nova implementação

- 1 → indoor do places365
- $2 \rightarrow \text{outdoor do places } 365$
- → 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 - alteraçao

- → 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 :

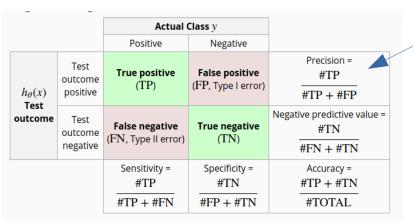
```
1 , 20160905_185411_000.jpg , 0.6735127356202351
1 , 20160905_185443_000.jpg ,
                                  0.6729643181912222
1 , 20160905_185515_000.jpg ,
                                  0.6500754233531866
1 , 20160905_185547_000.jpg , 0.6742467655881141
1 , 20160905_185619_000.jpg , 0.7226317408482577  
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 , 0.6463557<mark>1</mark>40634857
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]

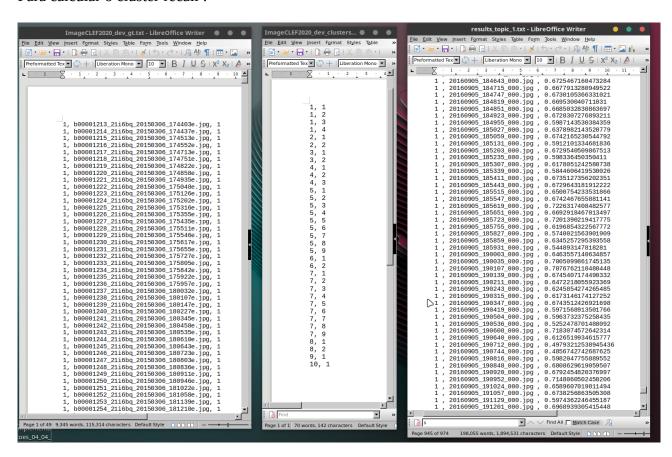
Métricas de avaliação:

 $F1 \rightarrow Precision and cluster recall. Top 5, 10, 20, 30, 40 e 50$

Para calcular a precision:



Para calcular o cluster recall:



ficheiro da esquerda (ground_truth) \rightarrow [tópico, imagem, cluster a que pertence] ficheiro do meio (ground_truth) \rightarrow [tópicos, numero de clusters existentes no tópico] ficheiro da direita (resultados) \rightarrow [tópico, imagem, confiança]

cálculo : [nº diferente de clusters encontrados] / [nº de clusters existentes para o tópico]

Exemplo:

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)

Ou seja, supondo que para o tópico 1 as minhas 5 imagens com mais confiança pertencem ao cluster 2 e 3. por exemplo :

- 1,2
- 1,2
- 1,3
- 1,3
- 1,3

o calculo sera $\rightarrow 2/4$

Dúvidas:

 \rightarrow Das 65.000 imagens que analisei, cerca de 270 deram confiança de 1.0. Como fazer top 5, 10 , 20 , 30 e 50 se todas tem a mesma confiança?

"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):

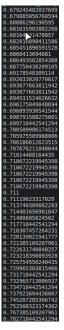
[problema → O groundtruth é apenas fornecido pelo imageclef nesta fase, pelo que isto apenas funcionara para esta fase**]**

Resultado:

top 50 imagens:



Score das imagens:



Cluster das imagens :