

Recuperação de Informação Câmeras Digitais



Bruno Cavalcanti
Guilherme Henrique
Pedro Henrique

Índice Invertido

...

Download das páginas

Mil páginas positivamente classificadas foram baixadas ao todo

No total 21250 páginas foram baixadas

Harvest ratio -> 0,0483

Pré-processamento automático

Atributos mais frequentes escolhidos:

Nome

Preço

Modo de armazenamento

Sensibilidade

Velocidade do obturador

Discretização dos atributos numéricos

Detalhes do índice invertido

Atributos com vários nomes são divididos, gerando múltiplas chaves no índice invertido

Compressão

Serialização sem compressão -> 198 kilobytes

Serialização com compressão -> 170 kilobytes

```
name.Sony%1=37,1;  
name.5DS%3=499,1;507,1;509,1;  
Shutter Speed.Via%4=836,1;891,1;896,1;909,1;  
Storage Mode.Eye-Fi%61=2,1;10,1;18,1;24,1;25,1;27,1;30,1;31,1;
```

```
name.Sony%1=37,1;  
name.5DS%3=499,1;8,1;2,1;  
Shutter Speed.Via%4=836,1;55,1;5,1;13,1;  
Storage Mode.Eye-Fi%61=2,1;8,1;8,1;6,1;1,1;2,1;3,1;2,1;2,1;1,1;
```

Processamento de Consulta

...

Calcular os pesos dos documentos para cada termo

$$\vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$$

Com TfIdf

Sem TfIdf

Tamanho dos vetores = 827

Calcular os pesos dos termos da consulta

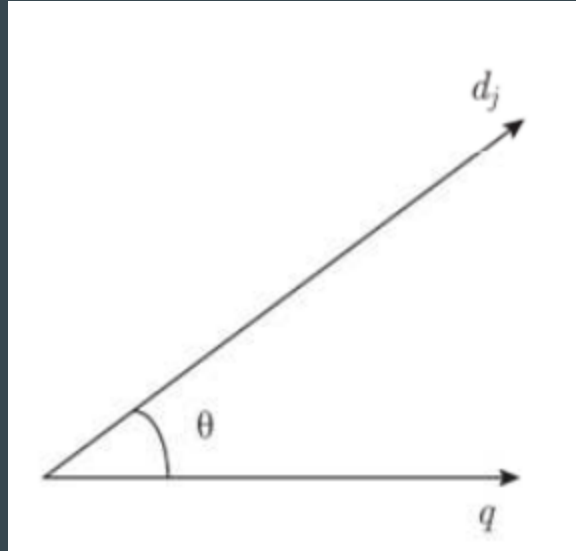
$$\vec{q} = (w_{1,q}, w_{2,q}, \dots, w_{t,q})$$

Com TfIdf

Sem TfIdf

Tamanho do vetor = 827

Modelo vetorial



Modelo vetorial

$$\begin{aligned} \text{sim}(d_j, q) &= \frac{\vec{d}_j \bullet \vec{q}}{|\vec{d}_j| \times |\vec{q}|} \\ &= \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \end{aligned}$$

$$\text{sim}(d_j, q) = \text{rank}$$

Spearman Correlation

$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^K (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}$$

name.Sony x name.Compact

Spearman Correlation = 0.76

Spearman Correlation

$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^K (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}$$

shutter speed.Manual x shutter speed.Scene

Spearman Correlation = 0.81

Spearman Correlation

$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^K (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}$$

shutter speed.Manual x shutter speed.Scene

name.Sony x name.Compact

Spearman Correlation = 0.61

Spearman Correlation

$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^K (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}$$

shutter speed.Manual x shutter speed.Scene

name.Sony x name.Compact

sensitivity.Using x sensitivity.1EV

storage mode.Recording x storage mode.SD

Spearman Correlation = 0.47

Spearman Correlation

$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^K (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}$$

shutter speed.Manual x shutter speed.Scene

name.Sony x name.Compact

sensitivity.Using x sensitivity.1EV

Spearman Correlation = 0.62

Composição da Resposta

...

Project Structure and Framework



Recomendação

```
RECOMMENDATIONS = INVERTED_INDEX != null ? INVERTED_INDEX.attributes.stream()
    .collect(Collectors.toMap(
        Function.identity(),
        attribute -> INVERTED_INDEX.termDocuments.entrySet().stream()
            .filter(entry -> entry.getKey().split("\\.")[0].equals(attribute))
            .sorted((e1, e2) -> e2.getValue().size() - e1.getValue().size())
            .limit(3)
            .map(entry -> entry.getKey().split("\\.")[1])
            .collect(Collectors.toList())
    )
    : null;
```