



記事 D に対するクエリ q のスコア(下記式の t は形態素解析後の単語)

例: $q = \text{“申請書類”}$ $t = \{\text{“申請”, “書類”}\}$

$$\text{score}(q, D) = \sum_t \max\{k_{\text{title}} \cdot \text{score}_{\text{title}}(t, D), k_{\text{body}} \cdot \text{score}_{\text{body}}(t, D)\}$$

k_{title} : 記事タイトルに対する重み

k_{body} : 記事本文に対する重み

以下タイトルのスコア計算(本文も同様)

$$\text{score}_{\text{title}}(t, D) = \underbrace{\ln \left\{ 1 + \frac{N_{\text{title}} - n(t, \text{title}) + 0.5}{n(t, \text{title}) + 0.5} \right\}}_{\text{IDFに相当}} \cdot \underbrace{\frac{f(t, D) \cdot (k_1 + 1)}{f(t, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|_{\text{title}}}{\text{avgdl}_{\text{title}}} \right)}}_{\text{TFに相当}}$$

IDFに相当

TFに相当

TF (Term Frequency) :

→記事 D の中で単語 t が頻出するほど高いスコアとなる

IDF (Inverse Document Frequency) :

→単語 t について、記事 D で頻出し、
それ以外の記事で頻出しないほど高いスコアとなる

N_{title} : タイトルが存在する総記事数

$n(t, \text{title})$: 単語 t がタイトルに出現する記事数

$f(t, D, \text{title})$: 記事 D におけるタイトルでの単語 t の出現回数

k_1 : パラメータ (1.2に設定)

b : パラメータ (0.75に設定)

$|D|_{\text{title}}$: 記事 D のタイトルの総単語数

$\text{avgdl}_{\text{title}}$: 全記事のタイトルの平均単語数



記事 D に対するクエリ q のスコア (下記式の t は形態素解析後の単語)

例: $q = \text{“申請書類”}$ $t = \{\text{“申請”, “書類”}\}$

$$\text{score}(q, D) = \max\{k_{\text{title}} \cdot \text{score}_{\text{title}}(q, D), k_{\text{body}} \cdot \text{score}_{\text{body}}(q, D), \}$$

k_{title} : 記事タイトルに対する重み

k_{body} : 記事本文に対する重み

以下タイトルのスコア計算 (本文も同様)

$$\text{score}_{\text{title}}(q, D) = \underbrace{\left(\sum_t \ln \left\{ 1 + \frac{N_{\text{title}} - n(t, \text{title}) + 0.5}{n(t, \text{title}) + 0.5} \right\} \right)}_{\text{IDFに相当}} \cdot \underbrace{\frac{f(q, D) \cdot (k_1 + 1)}{f(q, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|_{\text{title}}}{\text{avgdl}_{\text{title}}} \right)}}_{\text{TFに相当}}$$

N_{title} : タイトルが存在する総記事数

$n(t, \text{title})$: 単語 t がタイトルに出現する記事数

$f(q, D, \text{title})$: 記事 D におけるタイトルでのフレーズ q の出現回数

k_1 : パラメータ (1.2に設定)

b : パラメータ (0.75に設定)

$|D|_{\text{title}}$: 記事 D のタイトルの総単語数

$\text{avgdl}_{\text{title}}$: 全記事のタイトルの平均単語数