

CÀLCUL DE LA PRECISIÓ

Objectiu:

Comprovar el funcionament del càlcul de la precisió seguint diversos criteris de ponderació de la freqüència estadística d'aparició dels termes.

Dades:

En el document A hi apareixen els següents termes:

- Economia: al títol (1), altres parts del document (3).
- Política: al títol (1), subtítol (1), altres parts del document (2).

En el document B hi apareixen els següents termes:

- Economia: al títol (1), metadades (1), primer paràgraf (1), altres parts del document (5).
- Política: a altres parts del document (2).

Preguntes:

1. Per a una consulta sobre “economia política”, quin és el document més rellevant seguint el principi de Luhn?

	Document A	Document B
Economia	4	8
Política	4	2
Resultat	8	10

Segons el principi de Luhn el Document B és el més rellevant per a una consulta sobre economia política.

2. Per a una consulta sobre “economia política”, quin és el document més rellevant seguint el principi de Salton? Cal tenir present que les ponderacions per als llocs d'aparició dels termes són les següents: títol (5), subtítol (4), metadades (3), primer paràgraf (2) i altres parts del document (1).

	Document A	Document B
Economia	Títol: 1 = 5 Altres parts del document : 1 = 1 Resultat = 6	Títol: 1 = 5 Metadada: 1 = 3 Primer paràgraf: 1 = 2 Altres parts del document: 5 = 5 Resultat = 16
Política	Títol: 1 = 5 Subtítol: 2 = 8 Altres parts del document: 1 = 1 Resultat = 14	Altres parts del document: 2 = 2 Resultat = 2
Resultat	20	18

Segons el principi de Salton el Document A és el més rellevant per a una consulta sobre economia política.

3. Per a una consulta sobre “economia política”, quin és el document més rellevant si es pondera tenint en compte la freqüència inversa del document (IDF)? Cal tenir present que hi ha 10.000 documents a la base de dades, que el terme “política” apareix en 5000 d'ells mentre que el terme “economia” només consta en 20 documents.

Num total Document = 10.000

Num Doc Economia = 20

Num Doc Política = 5.000

FID = Num total Document / Num Document amb el terme

Economia = FID = 10.000 / 20 = 500

Política = FID = 10.000 / 5.000 = 2

Rellevància = Freqüència (terme) * FID

	DOC A			DOC B		
	Freqüència	FID	Rellevància	Freqüència	FID	Rellevància
Economia	4	500	2.000	8	500	4.000
Política	4	2	8	2	2	4
Resultat			2.008			4.004

Segons la freqüència inversa del document (IDF) el Document B és el més rellevant per a una consulta sobre economia política.