

# WAND SCORING EXAMPLE

(EFFICIENT QUERY EVALUATION  
USING A TWO-LEVEL  
RETRIEVAL PROCESS)

Query = "CATCHER IN THE RYE"

SCRITTO MOLTO BENE!

## Index Inverted

CATCHER → 140, 293, 356, 539, 463

RYE → 140, 140, 356, 400, 539, 700, 720, 763

IN → 90, 90, 356, 539, 600, 900

THE → 90, 100, 539, 600, 440, 300

## 0) INIZIALIZAZIONE

Inizializzo FINGERS, UPPER BOUNDS e THRESHOLD per ogni termine che compare nella query

↓  
CATCHER → 140, 293, 356, 539, 463

THRESHOLD = 0

$UB_{\text{CATCHER}} = 2,3$

↓  
RYE → 140, 140, 356, 400, 539, 700, 720, 763

$UB_{\text{RYE}} = 1,8$

↓  
IN → 90, 90, 356, 539, 600, 900

$UB_{\text{IN}} = 3,3$

↓  
THE → 90, 100, 539, 600, 440, 300

$UB_{\text{THE}} = 4,3$

Ad ogni iterazione dell'algoritmo posso accedere le seguenti due cose:

- Selezione in DOC-ID e mi calcolo il  $\text{max FULL RANK}^{\text{SCORE}}$ .  
In questo caso è possibile che la THRESHOLD e un UPPER BOUND si modificano.
- NON selezione nessun DOC-ID. In questo caso il THRESHOLD non cambia ma solo un UPPER BOUND potrebbe cambiare (diminuisce)

L'algoritmo si basa sulla seguente ITERAZIONE, che viene ripetuta fin a quando non ci sono più DOC-ID da processare

### 1) SORTING DEI TERMINI

I termini vengono ordinati rispetto al DOC-ID partendo dal finger

	THRESHOLD = 6
IN → 70, 90, 356, 539, 600, 900	$UB_{IN} = 3.3$
THE → 90, 100, 539, 600, 440, 300	$UB_{THE} = 4.3$
CATCHER → 140, 273, 356, 539, 763	$UB_{CATCHER} = 2.3$
RYE → 140, 140, 356, 400, 539, 700, 720, 763	$UB_{RYE} = 1.8$

## 2) CALCOLO DEL TERMINE PIVOT

Si calcola il primo termine che ci permette di superare il valore di THRESHOLD.

THRESHOLD = 6

IN → 90, 90, 356, 539, 600, 900

UB<sub>IN</sub> = 3.3

THE → 90, 100, 539, 600, 440, 300

UB<sub>THE</sub> = 4.3

TERMINI PIVOT

CATCHER → 140, 293, 356, 539, 463

UB<sub>CATCHER</sub> = 2.3

RYE → 140, 140, 356, 400, 539, 700, 720, 763

UB<sub>RYE</sub> = 1.8

Notiamo infatti che

$$3.3 + 4.3 = 7.6 > 6$$

## 3) VERIFICA ESISTENZA DOC-ID

Si verifica che il DOC-ID puntato dal FINGER del TERMINE PIVOT esiste nelle porting lists dei termini precedenti al TERMINE PIVOT

THRESHOLD = 6

IN → 90, 90, 356, 539, 600, 900

UB<sub>IN</sub> = 3.3

THE → 90, 100, 539, 600, 440, 300

UB<sub>THE</sub> = 4.3

TERMINI PIVOT

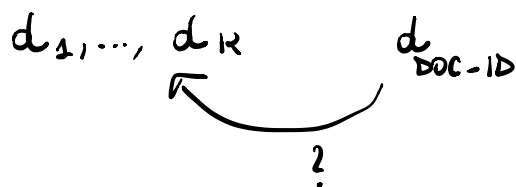
CATCHER → 140, 293, 356, 539, 463

UB<sub>CATCHER</sub> = 2.3

RYE → 140, 140, 356, 400, 539, 700, 720, 763

UB<sub>RYE</sub> = 1.8

Se il DOC-ID esiste allora lo si seleziona e ci si calcola lo SCORE COMPLETO. A questo punto confrontiamo lo score del doc. calcolato con lo score dei K doc scelti fino a questo punto, eventualmente rimpiazzando il nuovo doc con un vecchio.



Alla fine di questa operazione abbiamo K docs  $d'_1, d'_2, \dots, d'_K$ . Se nuovo THRESHOLD è quindi così def.

(IL THRESHOLD HA UN COMPORTAMENTO NON-DECRESCENTE)

$$\text{THRESHOLD} = \min_{i=1, \dots, K} \text{SCORE}(d'_i, q)$$

(QUESTA SCELTA GARANTISCE IL SAFE RANKING)

Nel nostro particolare caso aumentiamo la seguente modifica

(PRIMA DI AVER SELEZIONATO IL DOC 90)

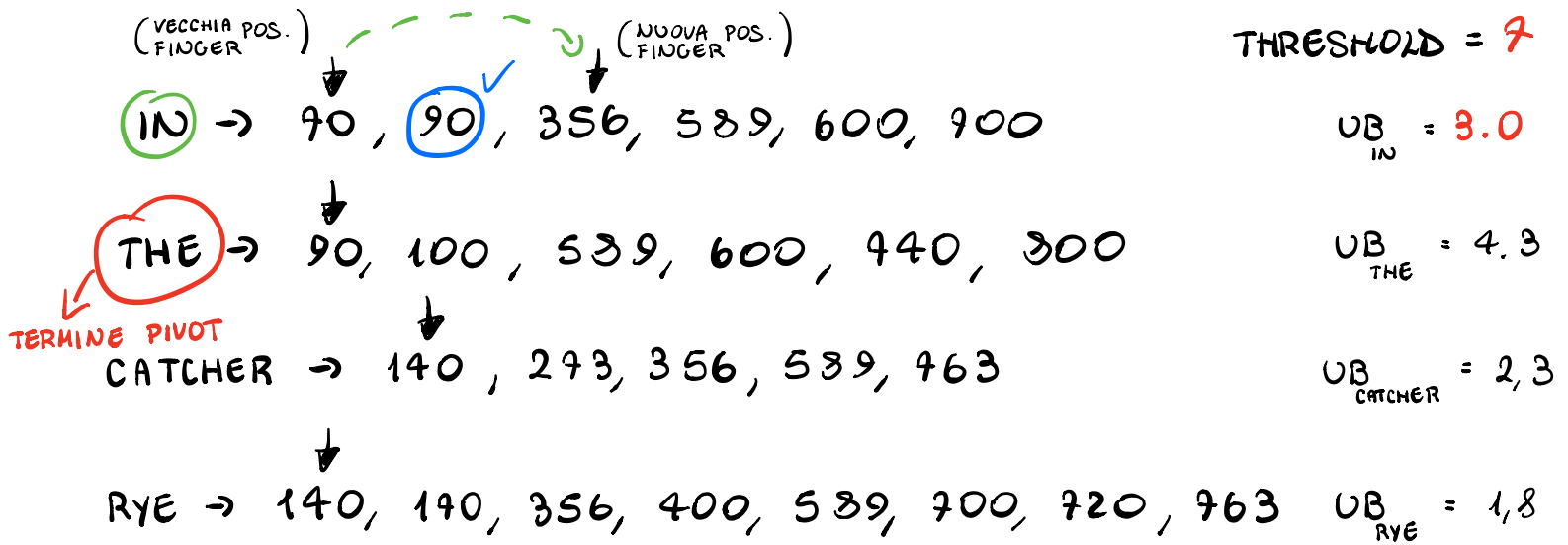
$$\text{THRESHOLD} = 6 \rightarrow \text{THRESHOLD} = 7$$

(DOPO AVER SELEZIONATO IL DOC 90)

#### 4) SPOSTAMENTO FINGER

Come ultimo step, per prepararci all'iterazione successiva, si sceglie UN TERMINE tra quelli che precedono il TERMINE PIVOT e si muove il FINGER del termine scelto al primo DOC-ID  $x$  t.c.  $x > \text{DOC-ID}$

Questo è il DOC-ID puntato dal FINGER del PIVOT



Dopo aver mosso il FINGER del termine scelto dobbiamo aggiornare l'UPPER BOUND

$$\left( \begin{array}{l} \text{PRIMA DI AVER} \\ \text{SPOSTATO IL FINGER} \end{array} \right) UB_{IN} = 3.3 \xrightarrow{\text{SPOSTO IL FINGER}} UB_{IN} = 3.0 \left( \begin{array}{l} \text{DOPO AVER} \\ \text{SPOSTATO IL FINGER} \end{array} \right)$$

L'UPPER BOUND dei termini ha un comportamento NON-CRESCENTE.

La scelta di quale termine scegliere è arbitraria ai fini della CORRETTEZZA del metodo. Alcune EURISTICHE potrebbero essere più efficienti di altre.

Questo TERMINA la singola iterazione



-----

Schematicamente quindi ogni ITERAZIONE effettua i seguenti  
i seguenti passi:

- 1) SORTING DEI TERMINI (RISPETTO AI DOC-IDS PUNTATI DAI FINGERS)
- 2) CALCOLO TERMINE PIVOT
- 3) VERIFICA ESISTENZA DOC-ID
- 4) SPOSTAMENTO FINGER

-----

Per finire, vediamo anche una seconda iterazione.  
In questa iterazione però, a differenza della prima,  
il metodo non sceglie nessun doc.

Iniziamo dalla seguente situazione

		THRESHOLD = 7
IN	→ 90, 90, 356, 589, 600, 900	$UB_{IN} = 3.0$
THE	→ 90, 100, 589, 600, 440, 300	$UB_{THE} = 4.3$
CATCHER	→ 140, 273, 356, 589, 463	$UB_{CATCHER} = 2.3$
RYE	→ 140, 140, 356, 400, 589, 700, 720, 763	$UB_{RYE} = 1.8$

Procediamo quindi con i vari steps.

## 1) SORTING DEI TERMINI

↓	THRESHOLD = 7
THE → 90, 100, 539, 600, 440, 300	$UB_{THE} = 4.3$
↓	
CATCHER → 140, 273, 356, 539, 463	$UB_{CATCHER} = 2.3$
↓	
RYE → 140, 140, 356, 400, 539, 700, 720, 763	$UB_{RYE} = 1.8$
↓	
IN → 90, 90, 356, 539, 600, 900	$UB_{IN} = 3.0$

## 2) CALCOLO TERMINE PIVOT

↓	THRESHOLD = 7
THE → 90, 100, 539, 600, 440, 300	$UB_{THE} = 4.3$
↓	
CATCHER → 140, 273, 356, 539, 463	$UB_{CATCHER} = 2.3$
↓	
<b>RYE</b> → 140, 140, 356, 400, 539, 700, 720, 763	$UB_{RYE} = 1.8$
↓	
IN → 90, 90, 356, 539, 600, 900	$UB_{IN} = 3.0$

TERMINI  
PIVOT

Sinfatti,

$$4.3 + 2.3 + 1.8 = 8.4 > 7$$

### 3) VERIFICA ESISTENZA DOC-ID

↓	THRESHOLD = 7
THE → 90, 100, <del>539</del> , 600, 440, 300	$UB_{THE} = 4.3$
↓	
CATCHER → <u>140</u> , 273, 356, 539, 463	$UB_{CATCHER} = 2.3$
↓	
<u>RYE</u> → 140, 140, 356, 400, 539, 700, 720, 763	$UB_{RYE} = 1.8$
↓	
IN → 70, 90, 356, 539, 600, 900	$UB_{IN} = 3.0$

Im questo caso il DOC-ID 140 non è presente in tutte le ranking lists dei termini che precedono il termine pivot.

Perche non viene selezionato, il THRESHOLD non cambia e si passa al passo successivo

### 4) SPOSTAMENTO FINGER

Im questo caso moviamo l'indice da THE che CATCHER.

Se scegliamo THE otteniamo

↓	THRESHOLD = 7
<u>THE</u> → 90, 100, <del>539</del> , 600, 440, 300	$UB_{THE} = 4.1$
↓	
CATCHER → <u>140</u> , 273, 356, 539, 463	$UB_{CATCHER} = 2.3$
↓	
<u>RYE</u> → 140, 140, 356, 400, 539, 700, 720, 763	$UB_{RYE} = 1.8$
↓	
IN → 70, 90, 356, 539, 600, 900	$UB_{IN} = 3.0$



Se invece scegliamo CATCHER otteniamo

THRESHOLD = 7

↓  
THE → 90, 100, 539, 600, 440, 300

$UB_{THE} = 4.3$

↓  
CATCHER → 140, 273, 356, 539, 463

$UB_{CATCHER} = 2.3$

↓  
RYE → 140, 140, 356, 400, 539, 700, 720, 763

$UB_{RYE} = 6.8$

TERMINE  
PIVOT

↓  
IN → 70, 90, 356, 539, 600, 900

$UB_{IN} = 3.0$

L'importante è che mostriamo il FINGER del termine  
sotto al minimo valore  $x$  che sia STRETTAMENTE MAGGIORE  
del DOC-ID puntato dal FINGER del TERMINE PIVOT.

---

E questo è tutto.

MADE WITH ❤ BY LEONARDO TAHIANO.