

# TD03 Corrected - Information Retrieval

## Third Year ISIL - University of M'Sila

### Exercise 01

#### Document Collection

- **Doc1:** The quick brown fox jumped over the lazy dog.
- **Doc2:** The brown dog chased the fox, but the fox was too quick.

#### 1. Positional Index

**Normalization:** Lowercase, remove punctuation.

- Doc1: [the, quick, brown, fox, jumped, over, the, lazy, dog]
- Doc2: [the, brown, dog, chased, the, fox, but, the, fox, was, too, quick]

**Positional Index:**

- the  $\rightarrow$  Doc1:[0,6], Doc2:[0,4,7]
- quick  $\rightarrow$  Doc1:[1], Doc2:[11]
- brown  $\rightarrow$  Doc1:[2], Doc2:[1]
- fox  $\rightarrow$  Doc1:[3], Doc2:[5,8]
- jumped  $\rightarrow$  Doc1:[4]
- over  $\rightarrow$  Doc1:[5]
- lazy  $\rightarrow$  Doc1:[7]
- dog  $\rightarrow$  Doc1:[8], Doc2:[2]
- chased  $\rightarrow$  Doc2:[3]
- but  $\rightarrow$  Doc2:[6]
- was  $\rightarrow$  Doc2:[9]
- too  $\rightarrow$  Doc2:[10]

## 2. Phrase Query: "the fox"

Match:

- Doc1: "the" at 0 and 6; "fox" at 3 → Not adjacent
- Doc2: "the" at 0, 4, 7; "fox" at 5, 8 → Match at (4,5) and (7,8)

**Result: Doc2[4,7]**

## 3. Phrase Query: "quick brown"

Match:

- Doc1: quick at 1, brown at 2 → Match:
- Doc2: quick at 11, brown at 1 → Not adjacent

**Result: Doc1[1]**

## 4. Proximity Query: "fox" within 3 words of "dog"

- Doc1: dog at 8, fox at 3 → Distance = 5 → Not match
- Doc2: dog at 2, fox at 5 → Distance = 3 →

**Result: Doc2[2]**

## 5. Proximity Query: "quick" within 2 words of "brown"

- Doc1: quick at 1, brown at 2 → Distance = 1 →
- Doc2: quick at 11, brown at 1 → Distance = 10 → Not match

**Result: Doc1[1]**

## Exercise 02

### Positional Index

| Term   | Document IDs and Positions  |
|--------|-----------------------------|
| apple  | Doc1: 1, 3, 5    Doc2: 2, 4 |
| banana | Doc1: 2, 4    Doc2: 1, 3, 5 |
| cherry | Doc1: 3    Doc2: 2, 4       |

## 1. Phrase Query: "apple banana"

We look for positions where 'banana' appears immediately after 'apple' (+1 offset).

- **Doc1:**

- apple: 1, 3, 5
- banana: 2, 4
- Matches at (1,2) and (3,4)  $\Rightarrow$

- **Doc2:**

- apple: 2, 4
- banana: 1, 3, 5
- No 'banana' at position 'apple+1'  $\Rightarrow$

**Result: Doc1**

## 2. Proximity Query: "banana \* cherry" (within distance $\leq 2$ )

We look for positions where the distance between 'banana' and 'cherry' is  $\leq 2$ .

- **Doc1:**

- banana: 2, 4
- cherry: 3
- Distance between 2 and 3 = 1, 4 and 3 = 1  $\Rightarrow$

- **Doc2:**

- banana: 1, 3, 5
- cherry: 2, 4
- Distance: 1-2 = 1, 3-2 = 1, 3-4 = 1, 5-4 = 1  $\Rightarrow$

**Result: Doc1, Doc2**

## Exercise 03

1. "fools rush in"  $\rightarrow$  Doc2[1], Doc4[8], Doc7[3,13]
2. "fools rush in" AND "angels fear to tread"  $\rightarrow$  Doc4