Jaccard Coefficient for Ranked Retrieval

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

Jaccard (A,B)

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

$$\text{Jaccard}(A,B) = |A \cap B|$$

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

$$\text{Jaccard}(A,B) = |A \cap B| \,/\, |A \cup B|$$

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

$$\text{Jaccard}(A,B) = |A \cap B| \,/\, |A \cup B|$$

$$A = \{0,1,2,3,4,5\}$$

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

A = {0,1,2,3,4,5}
B = {0,2,4,6}

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

A = {0,1,2,3,4,5}
B = {0,2,4,6}

Jaccard (A,B) =

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

Jaccard (A,B) = |A∩B| / |A∪B|

A = {0,1,2,3,4,5}
B = {0,2,4,6}

Jaccard (A,B) = 3

Jaccard Coefficient

- Commonly used to measure the overlap of two sets

Jaccard (A,B) = |A∩B| / |A∪B|

A = {0,1,2,3,4,5}
B = {0,2,4,6}

Jaccard (A,B) = 3 / 7

Jaccard Coefficient for Ranked Retrieval

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1)

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2 / 5

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2 / 5 = 0.4

Analytics Vidhya

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2 / 5 = 0.4

Jaccard (q,d2)

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2 / 5 = 0.4

Jaccard (q,d2) = 1

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2 / 5 = 0.4

Jaccard (q,d2) = 1 / 12

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2 / 5 = 0.4

Jaccard (q,d2) = 1 / 12 = 0.08

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = $|A \cap B| / |A \cup B|$

Jaccard (q,d1) = 2 / 5 = 0.4

Jaccard (q,d2) = 1 / 12 = 0.08

Jaccard Coefficient for Ranked Retrieval

Query (q): Analytics book

Doc 1 (d1): This book is on Analytics

Doc 2 (d2): Big-Data Analytics is the process of examining large volume of data

Jaccard (A,B) = |A∩B| / |A ∪ B|

Jaccard (q,d1) = 2 / 5 = 0.4

Jaccard (q,d2) = 1 / 12 = 0.08

Jaccard Coefficient

- Commonly used to measure the overlap of two sets
- Range: [0,1]

Jaccard Coefficient

- Commonly used to measure the overlap of two sets
- Range: [0,1]
- Higher the value of Jaccard Coefficient, more the relevance

Analytics Vidhya

Challenges: Jaccard Coefficient

- Does not consider the frequency of a term

Challenges: Jaccard Coefficient

- Does not consider the frequency of a term
- Generally prefer to retrieve documents having less words

Challenges: Jaccard Coefficient

- Does not consider the frequency of a term
- Generally prefer to retrieve documents having less words

$$\text{Jaccard}(A,B) = |A \cap B| \,/\, |A \cup B|$$

Thank You

Analytics
Vidhya