**# 2.3**

1. **df of all terms**

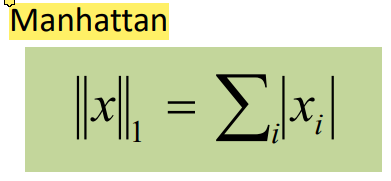
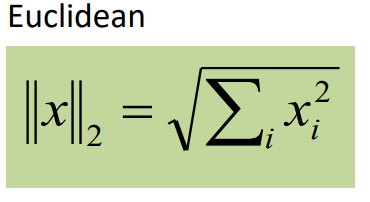
|  |  |
| --- | --- |
| **redirect** | **2** |
| **coop** | **1** |
| **movein** | **1** |
| **hall** | **1** |
| **recycling** | **1** |
| **davis** | **1** |
| **residence** | **1** |
| **drive** | **1** |
| **food** | **1** |

1. **idf of all terms ---**

|  |  |
| --- | --- |
| **redirect** | **0** |
| **coop** | **0.6931** |
| **movein** | **0.6931** |
| **hall** | **0.6931** |
| **recycling** | **0.6931** |
| **davis** | **0.6931** |
| **residence** | **0.6931** |
| **drive** | **0.6931** |
| **food** | **0.6931** |

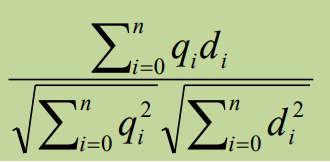
1. **What is the term frequency of the words food and residence?**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **food**  **Davis\_Food...f** | **residence Davis\_Food...f** | **food Resource...f** | **residence Resource...f** |
| **tf** | **1** | **0** | **0** | **1** |

1. **What are the lengths of these documents (round to 4 decimal places)?**
2.  

|  |  |  |  |
| --- | --- | --- | --- |
|  | **food residence** | **Davis\_Food…f** | **Resource…f** |
| **Euclidean, tf** | **1.4142** | **2** | **2.4494** |
| **Manhattan, tf** | **2** | **4** | **6** |
| **Euclidean, tf x idf** | **0.9802** | **1.2005** | **1.5498** |
| **Manhattan, tf x idf** | **1.3862** | **2.0793** | **3.4655** |

1. **What is the cosine similarity between the query and the two documents (in the specified spaces using the specified normalization, rounded to 4 decimal places)?**



|  |  |  |
| --- | --- | --- |
|  | **Davis\_Food…f** | **Resource…f** |
| **Euclidean, tf** | **0.3536** | **0.2887** |
| **Manhattan, tf** | **0.1250** | **0.0833** |
| **Euclidean, tf x idf** | **0.4082** | **0.3162** |
| **Manhattan, tf x idf** | **0.1667** | **0.1000** |

1. **What is the cosine similarity (rounded to 4 decimal places) if the query coordinates are considered to be (1,1)?**

|  |  |  |
| --- | --- | --- |
|  | **Davis\_Food…f** | **Resource…f** |
| **Euclidean, tf** | **0.3536** | **0.2887** |
| **Manhattan, tf** | **0.1250** | **0.0833** |
| **Euclidean, tf x idf** | **0.4082** | **0.3162** |
| **Manhattan, tf x idf** | **0.1667** | **0.1000** |

**# 2.4**

**Precision-recall graph**

**图表, 折线图

描述已自动生成**

* **result of 1.5**

**Precision: good/(good + errors) = 9/22 = 0.41**

**Recall: good/(good+misses) = 9/100 = 0.09**

* **result of 2.4**

precision at 10, 20, 30, 40, and 50

|  |  |
| --- | --- |
| 10 | 0.50 |
| 20 | 0.25 |
| 30 | 0.23 |
| 40 | 0.275 |
| 50 | 0.32 |

recall at 10, 20, 30, 40, and 50

|  |  |
| --- | --- |
| 10 | 0.05 |
| 20 | 0.05 |
| 30 | 0.07 |
| 40 | 0.11 |
| 50 | 0.16 |

-Which precision is the highest?

Precision at 10

-Are there any trends?

Decrease

-Which recall is the highest?

Recall at 50

-Is there any correlation between precision at 10, 20, 30, 40, 50, recall at 10, 20, 30, 40, 50?

Decrease and increase.

**#2.7**

图表, 折线图

描述已自动生成