**NARASIMMAN SAIRAM**
**N13296703 / ns3184**
**Web Search Engines**
**Problem Set – 3**

**1. Kappa measure**

| | | Judge 2 Relevance | | |
|---|---|---|---|---|
| | | YES | NO | TOTAL |
| YES | | 2 | 2 | 6 |
| NO | | 4 | 2 | 6 |
| TOTAL | | 6 | 6 | 12 |

P(A) = (2 + 2) /12 = 1/3 = 0.333
P(Non relevance) = 6 + 6 / 12 +12 = 12/24 = 0.5
P(Relevance) = 6 + 6 / 12 + 12 = 12 /24 = 0.5

P( E ) = ¼ + ¼ = 0.5

Kappa measure; k = p(A) – P ( E ) / (1 – P ( E )) = 0.33 – 0.5/(1 – 0.5) = - 0.18/0.5 = -0.36

**b. relevance = when both the judges agree on the relevance**

Total number of relevant documents retrieved = 1
Total number of documents retrieved = 5
Total number of relevant documents = 4

Precision = # of relevant documents retrieved / # of documents retrieved
**Precision = 1 / 5 = 0.2**

Recall = # of relevant documents retrieved / # relevant documents
**Recall = 1 /4 = 0.25**

F1 = 2PR/(P+R)
= 2 * (0.2) * (0.25) / (0.2 + 0.25) = 2/9
**= 0.222**

**C. Relevance = When either of the judges agree**

P = 5/5 = 1
R = 5/10 = 0.5
F1 = 2 * 1 * 0.5 / (1 + 0.5) = 2/3 = 0.6667

## 2. Evaluating an image search engine

a. Each word in the query should have a representation in the image. There should be a one-one mapping between each query term and the entities present in the image.
   Eg. Query = red book on a table

   Ideally, the retrieved images should contain a book that is red in color that's placed on a table

   (Or) images that contain red book and table

   Search results are of high quality if the top results contain all the query term representations

b. Similar to text search, Image search results will be ranked based on the relevancy of the query terms with the features represented in the image. An image with most or all features that represent the query terms will be ranked higher.

c. Highly ranked results will have close relevance to the attributes of the query terms (if any)

   Eg: dark big room

   Here the context is a room that is big and dark.
   So, the results should have a room that represents those features of the room.

   Eg: red book (a book that is red)

   So, as opposed to text search, it is not enough to just have the features representing the query terms somewhere in the document. Image search engine should tag these attributes to the corresponding objects in the image and return the results that represents best!

## 3. Clustering Scheme

| | A | B | C |
|---|---|---|---|
| X | 10 | 40 | 20 |
| Y | 50 | 10 | 20 |

TP = (10 * 9/2) + (40 * 39/2) + (20 * 19/2)
+ (50 * 49/2) + (10 * 9/2) + (20 * 19/2)
= 45 + 780 + 190 + 1225 + 45 + 190
**= 2475**

FP = (10 * 40) + (10 * 20) + (40 * 20)
+ (50 * 10) + (50 * 20) + (10 * 20)
= 400 + 200 + 800 + 500 + 1000 + 200
**= 3100**

FN = (10 * 50) + (40 * 10) + (20 * 20)
= 500 + 400 + 400
**= 1300**

TN = (10 * 10) + (10 * 20) + (40 * 20)
+ (40 * 50) + (10 * 20) + (50 * 20)
= 100 + 200 + 800 + 2000 + 200 + 1000
**= 4300**

a. Purity = (1/N) * (sum of the number of majority classes of all clusters)
= (1/150) * (40 + 50)
= 90/150
**= 0.6**

b. RI = (TP + TN)/(TP + FP + FN + TN)
= (2475 + 4300) / (11175)
= 6775/11175
**= 0.6062**

P = TP/(TP + FP)
= 2475/(2475 + 3100)
= 2475/5575
**= 0.4439**

R = TP / (TP + FN)
= 2475 / (2475 + 1300)
= 2475/3775
= **0.6556**

F1 = 2*P*R/(P + R)
= (2*0.4439 * 0.6556) / (0.4439 + 0.6556)
**= 0.5293**