A2q4:

Judgements on the relevance of results returned by the search engine are as follows:
1. Relevant
2. Relevant
3. Relevant
4. Not relevant
5. Relevant
6. Relevant
7. Not relevant
8. Not relevant
9. Relevant
10. Relevant
11. Not relevant
12. Not relevant
13. Relevant
14. Not relevant
15. Not relevant
16. Not relevant

The total number of relevant documents in the collection = 25

A. Calculate precision, recall, and F-measure for the returned results.
   The total number of relevant documents in the set = 8
   Therefore, **precision (P)** = 8/16 = **0.5**
   Given that the total number of relevant documents in the collection = 25
   Therefore, **recall (R)** = 8/25 = **0.32**
   F-measure (F) = 2PR / (P+R) = (2 * 0.5 * 0.32) /(0.5 + 0.32) = 0.32/0.82 ≈ **0.3902**

B. **Precision Recall Curve**
   Set 1: {R}
   P = 1/1 = 1, R = 1/25 = 0.04

   Set 2: {R,R}
   P = 2/2 = 1, R = 2/25 = 0.08

   Set 3: {R,R,R}
   P = 3/3 =1, R = 3/25 = 0.12

Set 4: {R,R,R,NR}
P = 3/4 = 0.75, R = 3/25 = 0.12

Set 5: {R,R,R,NR,R}
P = 4/5 = 0.8, R = 4/25 = 0.16

Set 6: {R,R,R,NR,R,R}
P = 5/6 ≈ 0.83, R = 5/25 = 0.2

Set 7: {R,R,R,NR,R,R,NR}
P = 5/7 ≈ 0.71, R = 5/25 = 0.2

Set 8: {R,R,R,NR,R,R,NR,NR}
P = 5/8 = 0.625, R = 5/25 = 0.2

Set 9: {R,R,R,NR,R,R,NR,NR,R}
P = 6/9 ≈ 0.66, R = 6/25 = 0.24

Set 10: {R,R,R,NR,R,R,NR,NR,R,R}
P = 7/10 = 0.7, R = 7/25 = 0.28

Set 11: {R,R,R,NR,R,R,NR,NR,R,R,NR}
P = 7/11 ≈ 0.63, R = 7/25 = 0.28

Set 12: {R,R,R,NR,R,R,NR,NR,R,R,NR,NR}
P = 7/12 ≈ 0.58, R = 7/25 = 0.28

Set 13: {R,R,R,NR,R,R,NR,NR,R,R,NR,NR,R}
P = 8/13 ≈ 0.61, R = 8/25 = 0.32

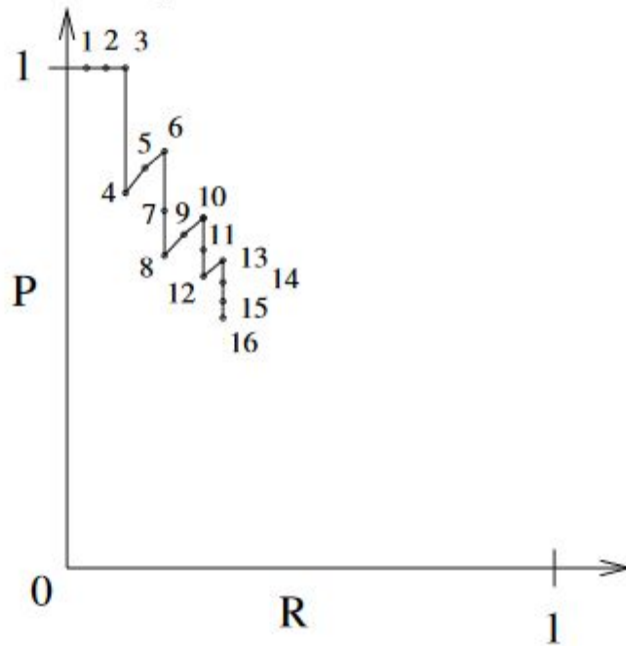Set 14: {R,R,R,NR,R,R,NR,NR,R,R,NR,NR,R,NR}
P = 8/14 ≈ 0.57, R = 8/25 = 0.32

Set 15: {R,R,R,NR,R,R,NR,NR,R,R,NR,NR,R,NR,NR}
P = 8/15 ≈ 0.53, R = 8/25 = 0.32

Set 16: {R,R,R,NR,R,R,NR,NR,R,R,NR,NR,R,NR,NR,NR}
P = 8/16 = 0.5, R = 8/25 = 0.32

Using these 16 coordinates, the precision-recall curve would be as follows:

To calculate the interpolated precision-recall curve, we use the formula:

$$\text{IntPrec}(r) = \max_{x, R(k) \geq r} P(k)$$

On putting $r = 0$, $\text{IntPrec}(0) = \max\limits_{x, R(k) \geq 0} P(k) = 1$ as $R(k) \geq 0$ for all k, and $\max P(k)$ is 1.

If we increase r starting from 0, we see that the maximal precision will remain 1 for all points R1, R2 and R3.
This is how we get the following values for the Interpolated Precision:
$0 \leq r \leq R3 = 0.12 \Rightarrow \text{IntPrec}(r) = 1$

For the values $r > R3$, the next maximum ≈ 0.83, and that is how we obtain the next interval: $R3 < r \leq R6 = 0.2 \Rightarrow \text{IntPrec}(r) \approx 0.83$

For the values $r > R6$, the next maximum $= 0.7$, and that is how we obtain the next interval:
$R6 < r \leq R10 = 0.28 \Rightarrow \text{IntPrec}(r) = 0.7$

For the values r > R10, the next maximum ≈ 0.6153846153846, and that is how we obtain the next interval:

R10 < r ≤ R13 = 0.32 ⟹ IntPrec(r) ≈ 0.6153846153846

The interpolated precision-recall curve would be as follows: