

Given a query q , where the relevant documents are $d_3, d_{12}, d_{15}, d_{21}, d_{22}$, and d_{30} , an IR system retrieves the following ranking: $d_5, d_3, d_{21}, d_{36}, d_{30}, d_{45}, d_{80}, d_{28}, d_{23}$, and d_{12} .

1. What are the precision and recall values at each retrieved document for this ranking?

	Recall	Precision
d_5	$0/6 = 0$	$0/1 = 0$
d_3	$1/6 = 0.166$	$1/2 = 0.50$
d_{21}	$2/6 = 0.33$	$2/3 = 0.66$
d_{36}	$2/6 = 0.33$	$2/4 = 0.50$
d_{30}	$3/6 = 0.50$	$3/5 = 0.60$
d_{45}	$3/6 = 0.50$	$3/6 = 0.50$
d_{80}	$3/6 = 0.50$	$3/7 = 0.42$
d_{28}	$3/6 = 0.50$	$3/8 = 0.37$
d_{23}	$3/6 = 0.50$	$3/9 = 0.33$
d_{12}	$4/6 = 0.66$	$4/10 = 0.40$

2. Interpolate the precision scores at 11 recall levels. Note: The interpolated precision at the j -th standard recall level is the maximum known precision at any recall level between the j -th and $(j + 1)$ -th level:

Recall	Interpolated Precision
0%	0.66
10%	0.66
20%	0.66
30%	0.66
40%	0.6
50%	0.6
60%	0.4
70%	0
80%	0
90%	0
100%	0

Explanation (uP is the uninterpolated precision):

$$\begin{aligned}
 P(100\%) &= \max_{100\% \leq r \leq 100\%} \{uP(100\%)=0\} = 0 \\
 P(90\%) &= \max_{90\% \leq r \leq 100\%} \{P(100\%)=0\} = 0 \\
 P(80\%) &= \max_{80\% \leq r \leq 90\%} \{P(90\%)=0\} = 0 \\
 P(70\%) &= \max_{70\% \leq r \leq 80\%} \{P(80\%)=0\} = 0 \\
 P(60\%) &= \max_{60\% \leq r \leq 70\%} \{uP(66\%)=0.40, P(70\%)=0\} = 0.40 \\
 P(50\%) &= \max_{50\% \leq r \leq 60\%} \{uP(50\%)=0.60, P(60\%)=0.40\} = 0.60 \\
 P(40\%) &= \max_{40\% \leq r \leq 50\%} \{P(50\%)=0.60\} = 0.60 \\
 P(30\%) &= \max_{30\% \leq r \leq 40\%} \{uP(33\%)=0.66, P(40\%)=0.60\} = 0.66
 \end{aligned}$$

$$P(20\%) = \max_{20\% \leq r \leq 30\%} \{P(30\%) = 0.66\} = 0.66$$

$$P(10\%) = \max_{10\% \leq r \leq 20\%} \{uP(16.6\%) = 0.50, P(20\%) = 0.66\} = 0.66$$

$$P(0\%) = \max_{0\% \leq r \leq 10\%} \{P(10\%) = 0.66\} = 0.66$$

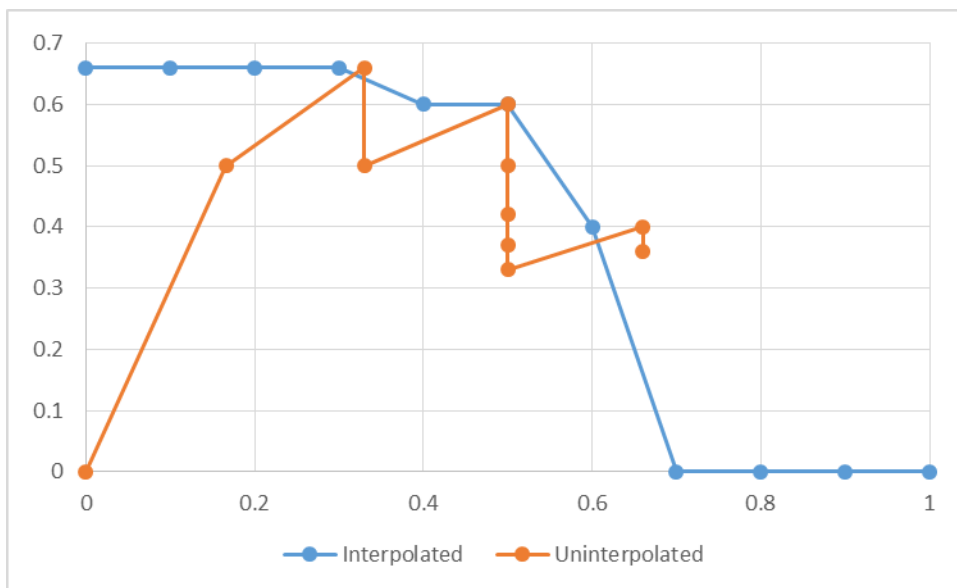
3. Why is interpolation of precision scores necessary when evaluation an IR system?

To be able to evaluate the precision for many queries at the same levels of recall (queries tend to have different recall points).

4. What is the R-precision? (precision at first R retrieved documents where R is the total number of relevant documents)

R-Precision	$3/6 = 0.5$

Could also draw a P-R graph with the uninterpolated points (4 points or all 10 points) and the interpolated points (11 points) for this question:



$$AvPre(\text{uninterpolated}) = (0.50 + 0.66 + 0.60 + 0.40)/6 = 0.36$$

$$AvPre(\text{interpolated}) = (0.66 + 0.66 + 0.66 + 0.66 + 0.60 + 0.60 + 0.40)/11 = 0.38$$

5. If we have two users, the first user decided that d12, d15, d21 are relevant to the query, and the second user decided that d3, d12, d15, d21, d22 are relevant to the query, what is the coverage ratio and the novelty ratio for this user? (Remember that the coverage ratio is the proportion of relevant items retrieved out of the total relevant documents known to a user prior to the search. The novelty ratio is the proportion of retrieved items, judged relevant by the user, of which they were previously unaware.)

	Coverage ratio	Novelty ratio
User 1	$2/3 = 0.66$	$2/4 = 0.50$
User 2	$3/5 = 0.60$	$1/4 = 0.25$

Explanation:

User1 knows 3 documents (d12, d15, d21). The system retrieves 4 relevant documents, 2 of them are the ones expected by user1 (d12, d21). Therefore the coverage ratio for user1 is 2 retrieved out of 3 expected. Among the 4 relevant documents retrieved by the system 2 are known to user1 and 2 are new. So the novelty ratio for user1 is 2 out of 4.

For user2, please compute yourself.