

# Tanja Crijns s4204999 - Exercises week 3: Evaluation

The following list of Rs and Ns represent relevant (R) and nonrelevant (N) returned documents in a ranked list of 20 documents retrieved in response to a query from a collection of 10,000 documents. The top of the ranked list is on the left of the list. This list shows 6 relevant documents. Assume there are 8 relevant documents in total in the collection.

RRNNNNNNNRNRNNNNNRNNNNR

## 1. Calculate Precision, Recall and F1 for this result list

Precision:

# retrieved relevant documents / # retrieved documents

$$6/20 = 0,3$$

Recall:

# retrieved relevant documents / # relevant documents

$$6/8 = 0,75$$

f1:

$2 * (\text{precision} * \text{recall} / (\text{precision} + \text{recall}))$

$$2 * (0,3 * 0,75 / (0,3 + 0,75)) = 0,4285714$$

## 2. Calculate Average Precision

Average precision:

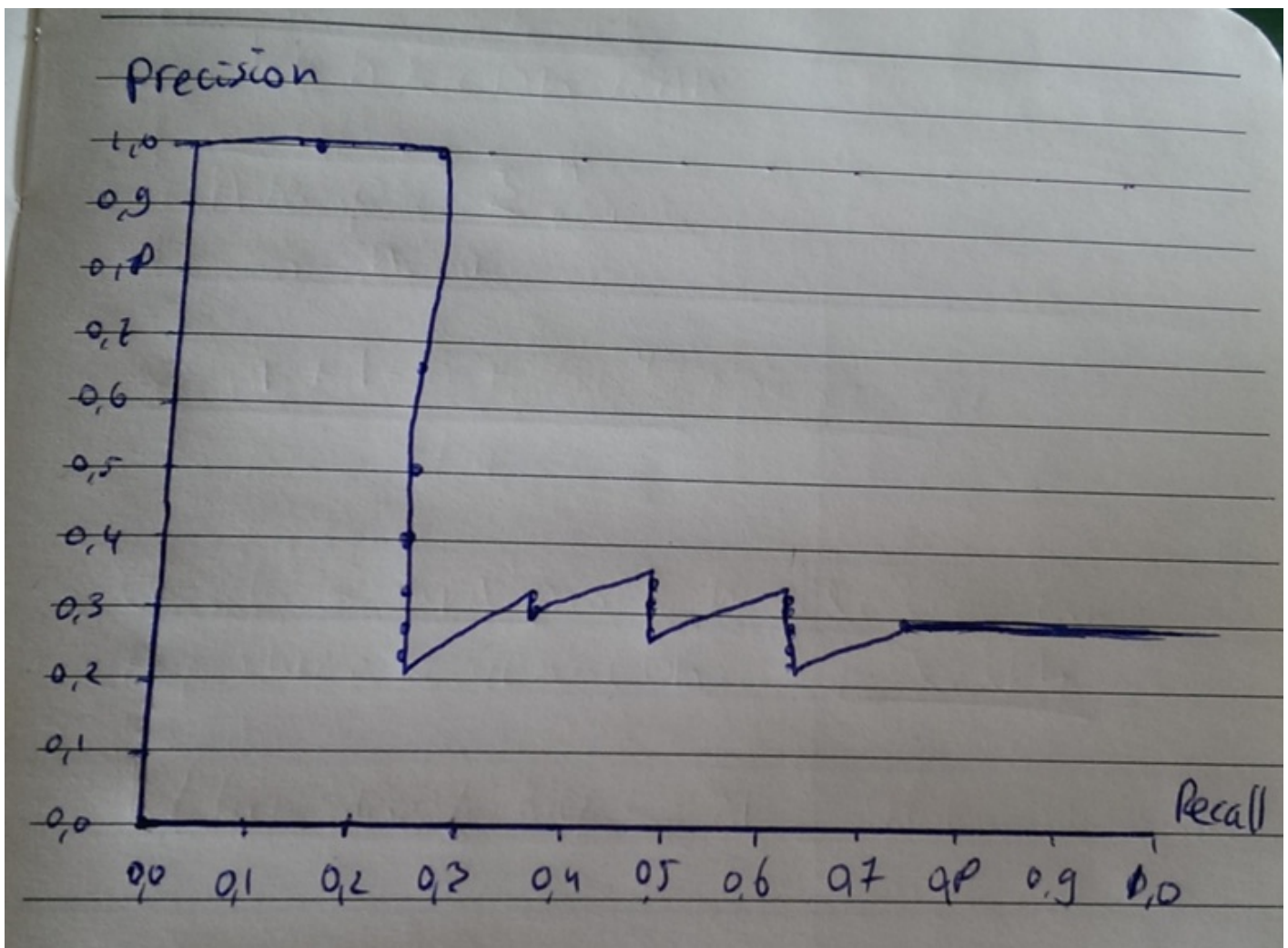
$$(1/1 + 2/2 + 3/9 + 4/11 + 5/15 + 6/20) / 8 = 0,4163$$

Which is based on the following table:

	k	R	P
x	1	1/8	1/1
x	2	2/8	2/2
	3	2/8	2/3
	4	2/8	2/4
	5	2/8	2/5
	6	2/8	2/6
	7	2/8	2/7
	8	2/8	2/8
x	9	3/8	3/9

	10	3/8	3/10
x	11	4/8	4/11
	12	4/8	4/12
	13	4/8	4/13
	14	4/8	4/14
x	15	5/8	5/15
	16	5/8	5/16
	17	5/8	5/17
	18	5/8	5/18
	19	5/8	5/19
x	20	6/8	6/20

### 3. Draw the Precision-Recall graph



#### 4. Assume that R=1 and N=0

- Calculate CG@20

$$CG(L) = \sum_{i=1}^n r_i$$

For this problem: CG(L):  $1 + 1 + 1 + 1 + 1 + 1 = 6$

- Calculate DCG@20

$$DCG(L) = r_1 + \sum_{i=2}^n \frac{r_i}{\log_2 i}$$

For this problem: DCG(L):  $1 + 1 + 1/2\log(9) + 1/2\log(11) + 1/2\log(15) + 1/2\log(20) = 3,092$

- Calculate nDCG@20

$$nDCG(L) = \frac{DCG(L)}{iDCG}$$

Where iDCG = the DCG for the ideally ranked list.

We have 8 relevant documents; in an ideal list, those documents will be returned first. There is no further ranking for those 8 documents, they are equally relevant. The iDCG is  $16,3$ , which makes the nDCG(L):  $3,09/16,3 = 0,19$ .