Information Retrieval

Precision and Recall

Dr. Seán Russell

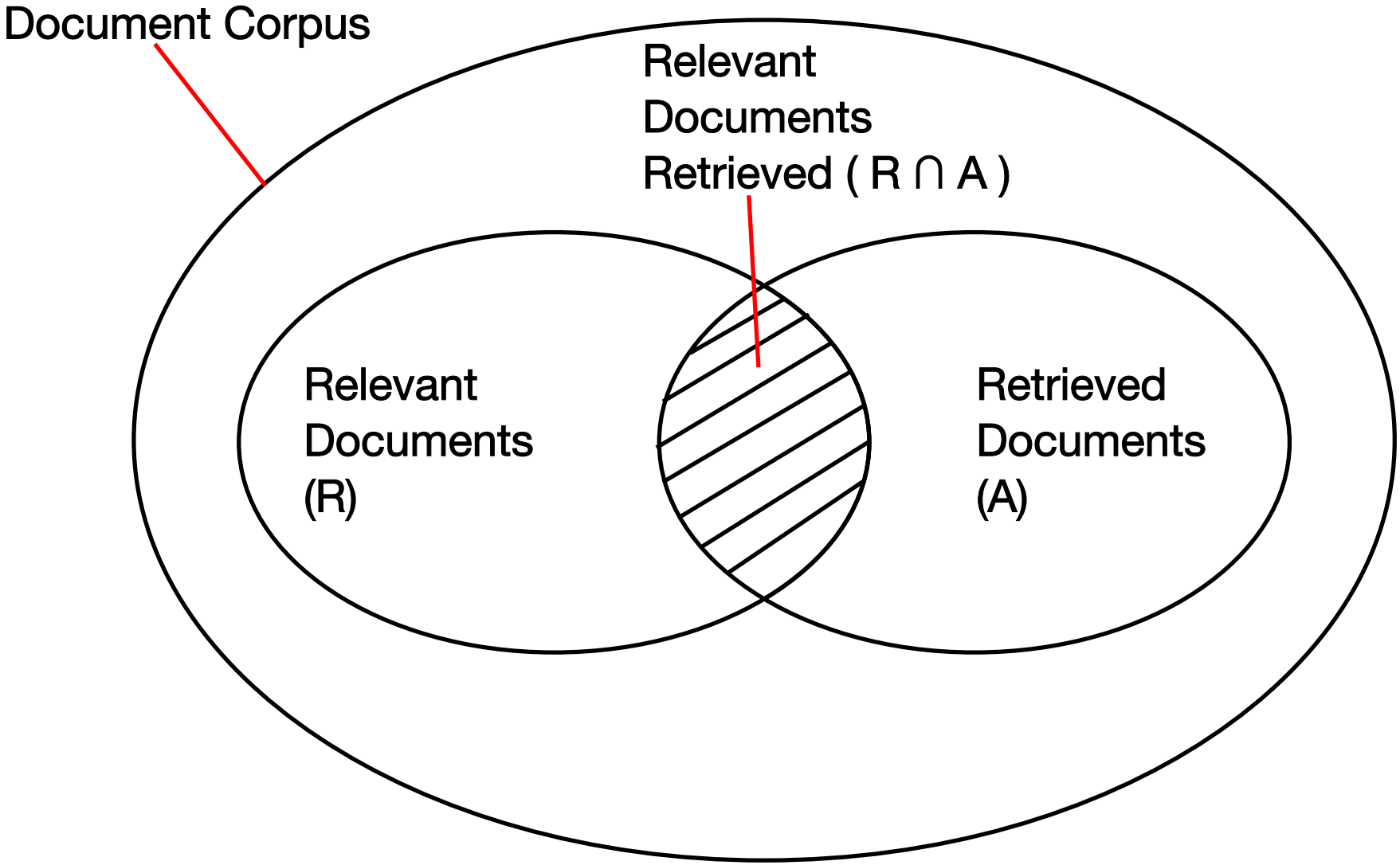
School of Computer Science,
University College Dublin

Week 6

Table of Contents

Section Contents

- Precision and Recall are the two most basic (and most widely used) evaluation metrics in IR, upon which many others are based
- Precision is the fraction of the set of retrieved documents (Ret) that are relevant (i.e. that are also in Rel)
- Recall is the fraction of the relevant documents (Rel) that have been retrieved (i.e. they are also in Ret).



$$\text{Precision} = \frac{|R \cap A|}{|A|}$$

$$\text{Recall} = \frac{|R \cap A|}{|R|}$$

- Rel $= \{d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{56}, d_{71}, d_{89}, d_{123}\}$
- Ret:

1. d_{123}
2. d_{84}
3. d_{56}
4. d_6
5. d_8
6. d_9
7. d_{511}
8. d_{129}
9. d_{187}
10. d_{25}
11. d_{38}
12. d_{48}
13. d_{250}
14. d_{113}
15. d_3

- Number of relevant documents: $|Rel| = 10$
- Number of retrieved documents: $|Ret| = 15$
- Number of relevant documents retrieved: $|Rel \cap Ret| = 5$
- Precision $= \frac{|Rel \cap Ret|}{|Ret|} = \frac{5}{15} = 0.33$
- Recall $= \frac{|Rel \cap Ret|}{|Rel|} = \frac{5}{10} = 0.5$

- The two metrics of precision and recall are often **inversely** related: as one increases the other decreases
- Precision will be high whenever a system is good at avoiding non-relevant documents
  - A system can achieve very high precision by retrieving very few documents
- Recall will be high whenever a system finds many relevant documents
  - A system can achieve 100% recall simply by retrieving all the documents in the collection

- Which is most important?
  - That is task-dependent!
- Ideally every IR system would have both recall and precision of 100% (the answer set is equal to the relevant set)

- Users searching the web want **high precision** (i.e. they want the returned documents to be relevant)
  - Because of the size of the web, there are very many documents that will help satisfy the information need, so the user does not need to examine all of them
  - Instead, the user wishes to avoid wasting time looking at non-relevant documents

- A patent lawyer researching a patent must ensure that they get all of the relevant documents and hence they want **high recall**
  - If any relevant documents are missed, this may have serious consequences, so it is essential that all relevant documents are returned
  - They will tolerate lower precision to facilitate this (i.e. they are more likely to be willing to read through some non-relevant documents)

- For older, smaller document collections, the calculation of precision and recall was easy
- Nowadays, because document collections are so huge, it is very difficult to identify all the relevant documents for every query
- This makes it often impossible to calculate recall accurately
- Thus, precision tends to be preferred, in addition to other metrics based on precision

- Basic Precision is a set-based, unranked retrieval metric (as is recall)
- It is a single-value metric based on the entire list of results that was returned by an IR system
- It does not assess the way the results are ranked
- Most IR systems return ranked lists, not sets of documents, so users are most likely to look at the top of the list first

Rank	IR 1	IR 2
1	N	R
2	N	R
3	N	N
4	R	N
5	R	N

- N is non-relevant, R is relevant
- Precision for both systems is 40%
- But clearly IR system 2 is better!