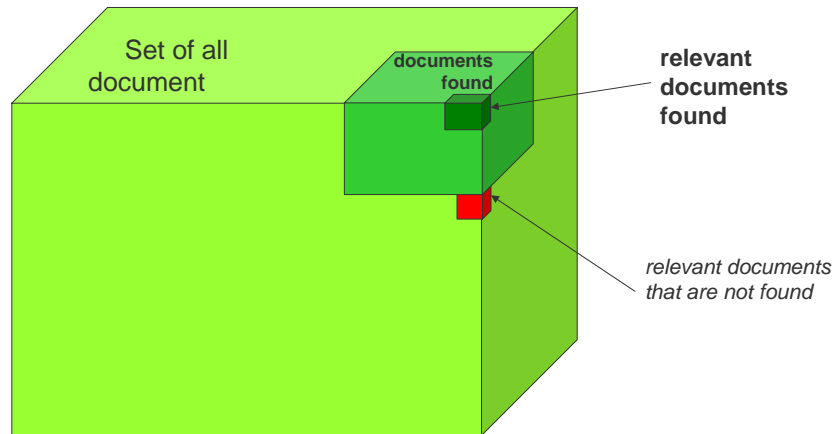
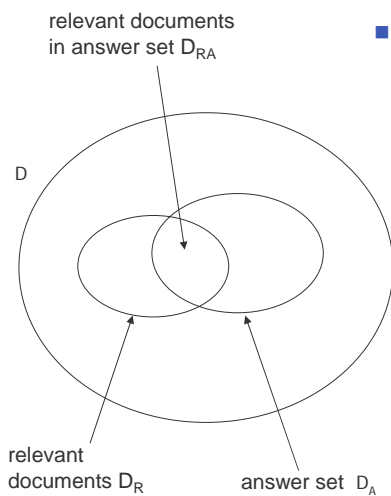


## 2.4 Evaluating Search Methods



## Performance Measure of Information Retrieval: Recall und Precision



- Several different measures for evaluating the performance of information retrieval systems have been proposed; two important ones are:

**Recall:** fraction of the relevant documents that are successfully retrieved.

$$\text{recall} = \frac{| \{ \text{relevant documents} \} \cap \{ \text{retrieved documents} \} |}{| \{ \text{relevant documents} \} |}$$

$$R = \frac{|D_{RA}|}{|D_R|}$$

**Precision:** fraction of the documents retrieved that are relevant to the user's information need

$$\text{precision} = \frac{| \{ \text{relevant documents} \} \cap \{ \text{retrieved documents} \} |}{| \{ \text{retrieved documents} \} |}$$

$$P = \frac{|D_{RA}|}{|D_A|}$$

## F-Measure

- The F-measure is a mean of precision and recall

$$F = 2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall}).$$

- In this version, precision and recall are equally weighted.
- The more general version allows to give preference to recall or precision

$$F_\alpha = (1 + \alpha) \cdot (\text{precision} \cdot \text{recall}) / (\alpha \cdot \text{precision} + \text{recall}).$$

- ◆  $F_2$  weights recall twice as much as precision
- ◆  $F_{0.5}$  weights precision twice as much as recall



## Computing Recall and Precision

- Evaluation: Perform a predefined set of queries
  - ◆ The search engines delivers a ranked set of documents
  - ◆ Use the first X documents of the result list as answer set
  - ◆ Compute recall and precision for the first X documents of the ranked result list.
- How do you know, which documents are relevant?
  1. A general reference set of documents can be used. For example, TREC (Text REtrieval Conference) is an annual event where large test collections in different domains are used to measure and compare performance of information retrieval systems
  2. For companies it is more important to evaluate information retrieval systems using their own documents
    1. Collect a representative set of documents
    2. Specify queries and associated relevant documents
    3. evaluate search engines by computing recall and precision for the query results

