Challenges in Creating a Test Collection

Relevance Judgments

Queries: representative & many

Q1 Q2 Q3
… Q50 …
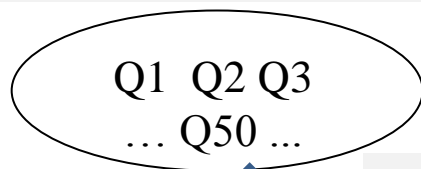
Existence of relevant docs
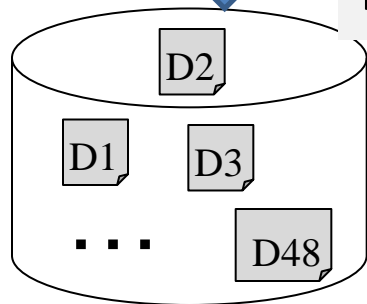
Judgments:
completeness vs.
minimum human work

D2

D1   D3

· · ·   D48

Measures: capture the perceived utility by users

Docs: representative & many

Q1 D8
…
Q2 D1 –
Q2 D2 +
Q2 D3 +
Q2 D4 –
…
Q50 D1 –
Q50 D2 –
Q50 D3 +

3