

# Why Evaluation?

- Reason 1: Assess the actual utility of a TR system
  - Measures should reflect the utility to users in a real application
  - Usually done through user studies (interactive IR evaluation)
- Reason 2: Compare different systems and methods
  - Measures only need to be correlated with the utility to actual users, thus don't have to accurately reflect the exact utility to users
  - Usually done through test collections (test set IR evaluation)