Summary of Categorization Evaluation

- Evaluation is always very important, so get it right!
- Measures must reflect the **intended use** of the results for a particular application (e.g., spam filtering vs. news categorization)
  - Consider: How will the results be further processed (by a user)?
  - Ideally associate a different cost with each different decision error
- Commonly used measures for **relative** comparison of different methods:
  - Accuracy, precision, recall, F score
  - Variations: per-document, per-category, micro vs. macro averaging
- Sometimes **ranking** may be more appropriate