Three Basic Problems in Content-Based Filtering

- Making **filtering decision** (Binary classifier)
  - Doc text, profile text $\rightarrow$ yes/no
- **Initialization**
  - Initialize the filter based on only the profile text or very few examples
- **Learning** from
  - Limited relevance judgments (only on "yes" docs)
  - Accumulated documents
- All trying to maximize the utility