

Formal Formulation of TR

- **Vocabulary:** $V = \{w_1, w_2, \dots, w_N\}$ of language
- **Query:** $q = q_1, \dots, q_m$, where $q_i \in V$
- **Document:** $d_i = d_{i1}, \dots, d_{im_i}$, where $d_{ij} \in V$
- **Collection:** $C = \{d_1, \dots, d_M\}$
- **Set of relevant documents:** $R(q) \subseteq C$
 - Generally unknown and user-dependent
 - Query is a “hint” on which doc is in $R(q)$
- **Task** = compute $R'(q)$, an approximation of $R(q)$