

informações e estuda a representação, o armazenamento, a organização e acesso a documentos, páginas *Web* e outros registros.

Baeza-Yates e Ribeiro-Neto (2013) afirmam também que a modelagem de RI pode se dividir em duas tarefas principais, que seriam a concepção de um arcabouço lógico e a definição de uma função de ordenamento para computar o grau de similaridade entre documentos e as consultas do usuário. Os modelos vetorial, booleano e probabilístico formam o conjunto de modelos clássicos de RI utilizados em arcabouços lógicos.

Um sistema de RI é uma quádrupla  $(D, Q, F, R(q_i, d_j))$ , onde  $D$  é o conjunto de visões lógicas da coleção;  $Q$  o conjunto das representações das consultas do usuário;  $F$  um arcabouço para modelagem do sistema e  $R(q_i, d_j)$  uma função de ordenação que associa um número real à representação de uma consulta  $q_i \in Q$  a cada documento  $d_j \in D$  (BAEZA-YATES; RIBEIRO-NETO, 2013).

*Tuxbot* é um exemplo de AC que faz uso de modelos de RI para busca de respostas. *Tuxbot* tem o objetivo de esclarecer dúvidas relacionadas ao sistema operacional *Linux* (TEIXEIRA, 2005).

#### 2.4.1 Modelo Probabilístico

O modelo probabilístico é um modelo de RI que propõe recuperar documentos relevantes ao usuário por intermédio da probabilidade. O modelo foi desenvolvido por *Robertson e Jones* (1976).

Dada uma consulta de usuário  $q$  e um documento  $d_j$  da coleção, o modelo probabilístico tenta estimar a probabilidade do usuário achar o documento  $d_j$  interessante (isto é, relevante). O modelo supõe que essa probabilidade de relevância depende apenas das representações da consulta e do documento, ou seja, das informações disponíveis ao sistema. (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 48).

A Equação 1 apresenta a equação de similaridade para obter a relevância de documentos quando não se sabe previamente o conjunto de documentos relevantes:

$$sim(d, BUSCA) = \sum_{i=1}^t \log_2 \left( \frac{N + 0,5}{n_i + 0,5} \right) \quad (1)$$