

Modelo Booleano

Wendel Melo

Faculdade de Computação
Universidade Federal de Uberlândia

Recuperação da Informação

Adaptado do Material da Prof^a Vanessa Braganholo - IC/UFF

Modelo Booleano

- Modelo simples;
- Baseado em teoria dos conjuntos e álgebra booleana;
- Documentos e consultas são representados como vetores binários;
- Consultas são especificadas como expressões booleanas

Modelo Booleano

- Cada documento é representado logicamente por um vetor de pesos binários ;
- Seja w_{ij} o peso do termo k_i no documento d_j . Assim:

$$w_{ij} = \begin{cases} 1, & \text{se } k_i \text{ aparece em } d_j, \\ 0, & \text{caso contrário} \end{cases}$$

- Desse modo, o número de componentes do vetor de pesos é dado pelo número de termos do vocabulário da base.

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

▶ Índice terá os seguintes termos (supondo eliminação de termos irrelevantes):

- ▶ casa
- ▶ campo
- ▶ é
- ▶ linda
- ▶ azul
- ▶ amarela
- ▶ carro
- ▶ Marcelo

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

Índice terá os seguintes termos (supondo eliminação de *stopwords*):

- ▶ casa
- ▶ campo
- ▶ é
- ▶ linda
- ▶ azul
- ▶ amarela
- ▶ carro
- ▶ Marcelo

Cada documento será modelado como um vetor de pesos:

D1 = (1,1,1,1,1,1,0,0)

D2 = (0,0,1,0,1,0,1,1)

Modelo Booleano

Considerando uma consulta q :

$$q = ka \wedge (kb \vee \neg kc)$$

Modelo Booleano

Considerando uma consulta q :

$$q = ka \wedge (kb \vee \neg kc)$$

- ▶ Como um documento pode satisfazer a esta consulta?
 - ▶ Se ka , kb e kc estiverem presentes, **OU**
 - ▶ Se ka e kb estiverem presentes, **OU**
 - ▶ Se ka estiver presente e kc não estiver presente

Modelo Booleano

Considerando uma consulta q :

$$q = ka \wedge (kb \vee \neg kc)$$

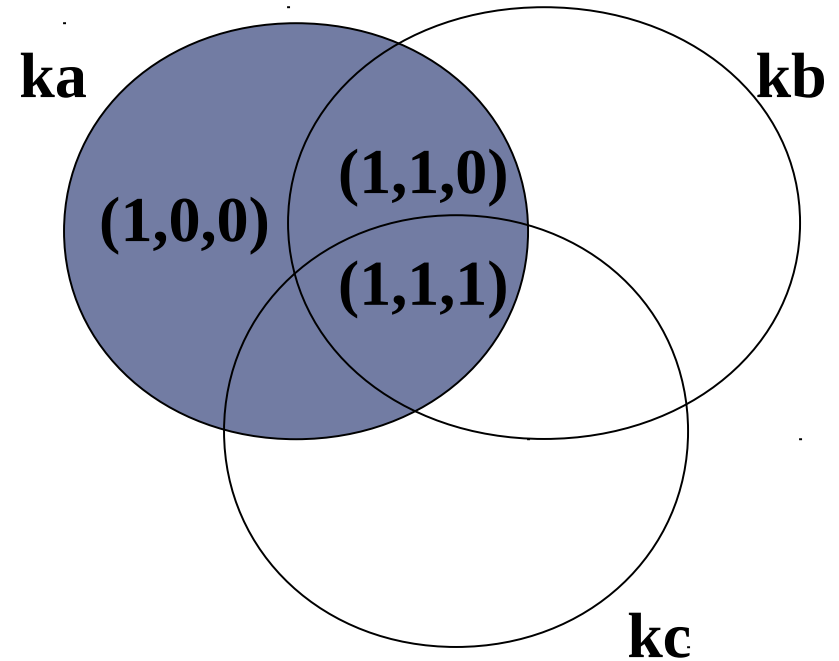
- ▶ Como um documento pode satisfazer a esta consulta?
 - ▶ Se ka , kb e kc estiverem presentes **(1,1,1)**, OU
 - ▶ Se ka e kb estiverem presentes **(1,1,0)**, OU
 - ▶ Se ka estiver presente e kc não estiver presente **(1,0,0)**

Modelo Booleano

$$q = ka \wedge (kb \vee \neg kc)$$

Esta consulta pode ser representada na forma normal disjuntiva, da seguinte forma:

$$q = (1,1,1) \vee (1,1,0) \vee (1,0,0)$$



Modelo Booleano

- Para que um documento seja classificado como relevante, basta que seu vetor de pesos case com alguma das disjunções da FND da consulta.
- Não há ranqueamento entre os documentos. Cada documento é apenas classificado como relevante ou não relevante.

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

$D1 = (1,1,1,1,1,1,0,0)$

$D2 = (0,0,1,0,1,0,1,1)$

$Q = \text{azul}$

Índice:

casa

campo

é

linda

azul

amarela

carro

Marcelo

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

$D1 = (1,1,1,1,1,1,0,0)$

$D2 = (0,0,1,0,1,0,1,1)$

$Q = \text{azul} (1)$

$D1 = (1)$

$D2 = (1)$

Índice:

casa

campo

é

linda

azul

amarela

carro

Marcelo

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

$D1 = (1,1,1,1,1,1,0,0)$

$D2 = (0,0,1,0,1,0,1,1)$

$Q = \text{azul} (1)$

$D1 = (1)$

$D2 = (1)$

Índice:

casa

campo

é

linda

azul

amarela

carro

Marcelo

Resposta: D1 e D2

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

$D1 = (1,1,1,1,1,1,0,0)$

$D2 = (0,0,1,0,1,0,1,1)$

$Q = \text{azul} \wedge \text{amarela}$

Índice:

casa

campo

é

linda

azul

amarela

carro

Marcelo

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

$D1 = (1,1,1,1,1,1,0,0)$

$D2 = (0,0,1,0,1,0,1,1)$

$Q = \text{azul} \wedge \text{amarela} (1,1)$

$D1 = (1,1)$

$D2 = (1,0)$

Índice:

casa

campo

é

linda

azul

amarela

carro

Marcelo

Exemplo

D1

A casa de campo é linda, azul e amarela.

D2

O Carro azul é de Marcelo.

$D1 = (1,1,1,1,1,1,0,0)$

$D2 = (0,0,1,0,1,0,1,1)$

$Q = \text{azul} \wedge \text{amarela} (1,1)$

$D1 = (1,1)$

$D2 = (1,0)$

Índice:

casa

campo

é

linda

azul

amarela

carro

Marcelo

Resposta: D1

Vantagens do Modelo Booleano

- Fácil compreensão e implementação;
- Semântica precisa: resultados previsíveis
- Suporte nativo ao operador *NOT*:
 - Muitas vezes, sistemas de RI baseados em outros modelos implementam a operação de MENOS em vez de *NOT* devido a dificuldade de fornecer suporte a esse operador.

Desvantagens do Modelo Booleano

- A semântica precisa faz com que o casamento entre documento e consulta precise ser exato para o documento ser considerado relevante;
 - Isto torna este modelo mais próximo a recuperação de dados que recuperação de informação.
- O critério binário de decisão pode implicar em qualidade ruim para a recuperação: Se uma consulta exigir a presença de 10 termos, e um documento d contiver apenas 9, a simples falta de um fará com que ele não entre no resultado;
 - Talvez d fosse considerado relevante pelo usuário, especialmente se não houver nenhum documento que case exatamente com a consulta.

Desvantagens do Modelo Booleano

- Usuários podem se confundir com uso de operadores lógicos;
- Consultas booleanas formuladas pelos usuários frequentemente são simplistas;
- Em consequência, a quantidade de resultados retornados pode ser muito pequena ou muito grande;
- A falta de ranqueamento torna o modelo não prático;