

RECUPERAÇÃO DE INFORMAÇÃO

Profa. Patrícia Proença

patricia.proenca@ifmg.edu.br

ATENÇÃO!!!

- ↓ O material a seguir é uma videoaula apresentada pela professora PATRÍCIA APARECIDA PROENÇA AVILA, como material pedagógico do IFMG, dentro de suas atividades curriculares ofertadas em ambiente virtual de aprendizagem. Seu uso, cópia e ou divulgação em parte ou no todo, por quaisquer meios existentes ou que vierem a ser desenvolvidos, somente poderá ser feito, mediante autorização expressa deste docente e do IFMG. Caso contrário, estarão sujeitos às penalidades legais vigentes”.
- ↓ Conforme Art. 2º§1º da Nota Técnica nº 1/2020/PROEN/Reitoria/IFMG (SEI 0605498, Processo nº 23208.002340/2020-04



Modelo booleano



Modelo Booleano

- O modelo Booleano é um modelo de recuperação de informação **simples baseado na teoria de conjuntos e na álgebra Booleana**;
- Como consequência, o modelo é bastante **intuitivo e possui uma semântica precisa**;
- Pela sua inerente simplicidade e formalismo, o modelo Booleano recebeu uma atenção considerável no passado e foi adotado por muitos dos primeiros sistemas bibliográficos comerciais.



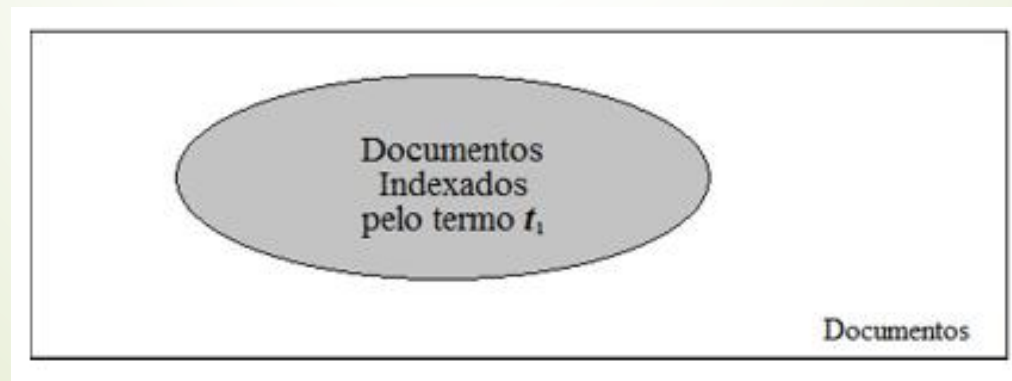
Modelo Booleano



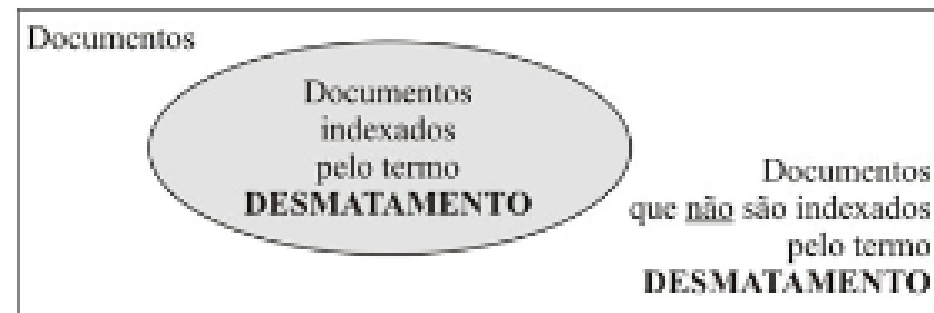
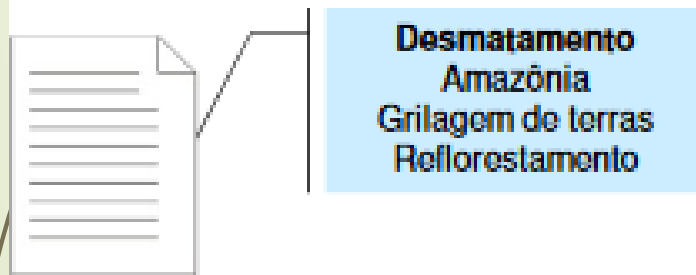
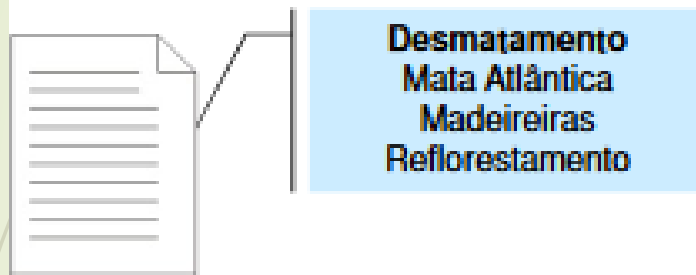
- O modelo Booleano **considera que os termos de indexação estão presentes ou ausentes nos documentos:**
 - ou seja, as frequências na matriz de termos por documentos são todas **binárias** (0 ou 1);
- Uma **consulta** q em um modelo booleano é composta por **termos de indexação ligados por três conectivos Booleanos: not, and e or.**
 - Uma consulta é essencialmente uma expressão Booleana convencional sobre termos de indexação.

Modelo Booleano

- Uma expressão de busca que utiliza apenas um **termo t_1** terá como resultado o conjunto de documentos indexados por t_1 ;



Modelo Booleano



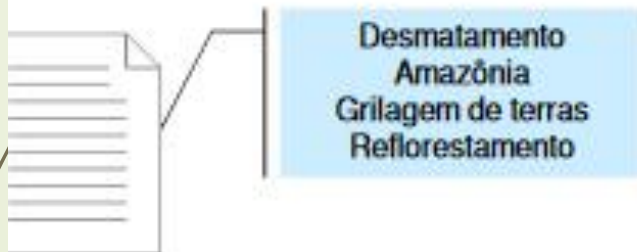
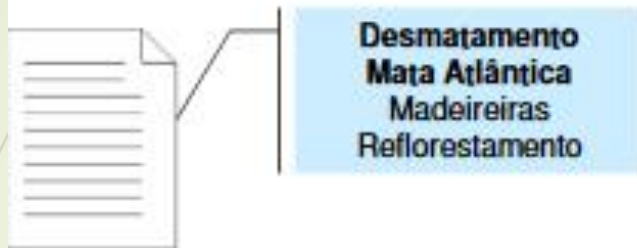
Desmatamento



Modelo Booleano

- Uma expressão conjuntiva de enunciado t_1 AND t_2 recuperará documentos indexados por ambos os termos (t_1 e t_2).
- Esta operação equivale à interseção do conjunto dos documentos indexados pelo termo t_1 com o conjunto dos documentos indexados pelo termo t_2 .

Modelo Booleano



Documentos



**Desmatamento
AND
Mata Atlântica**

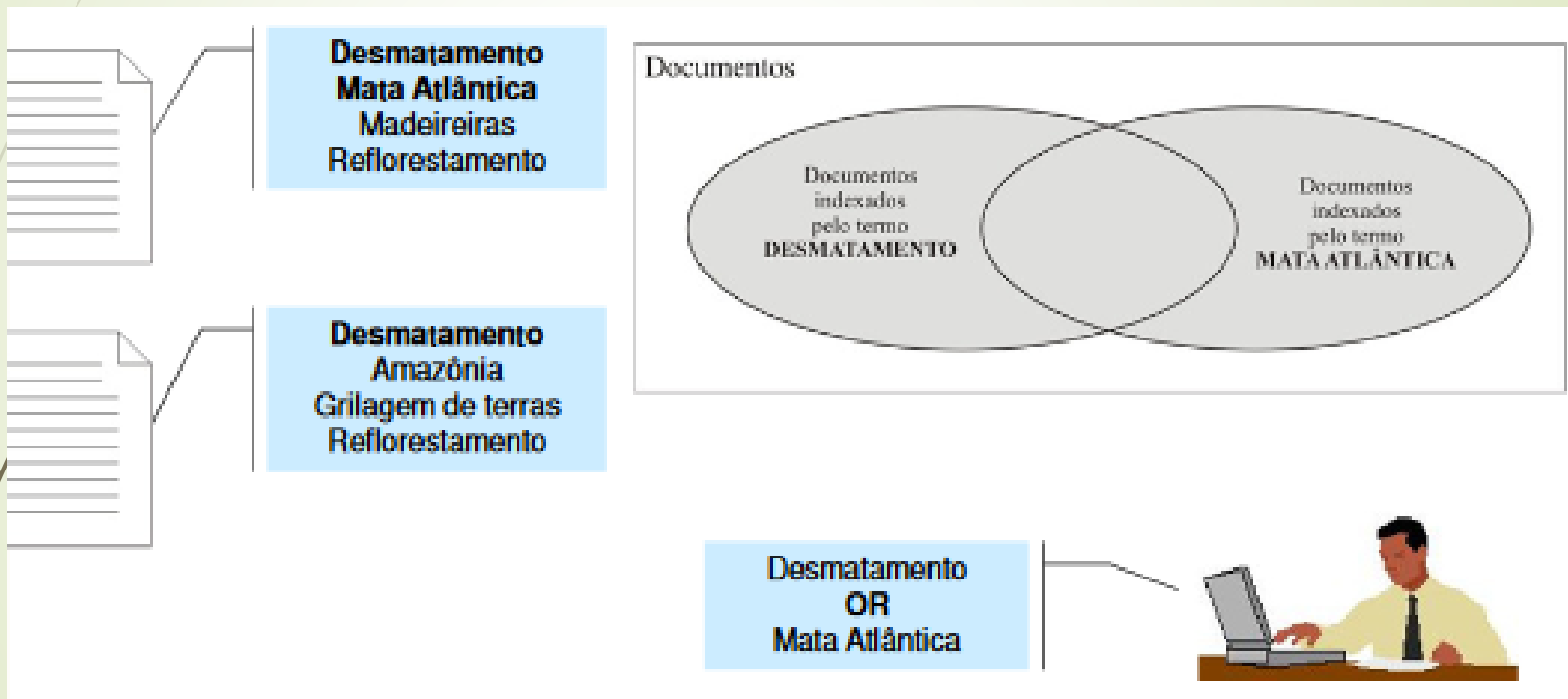




Modelo Booleano

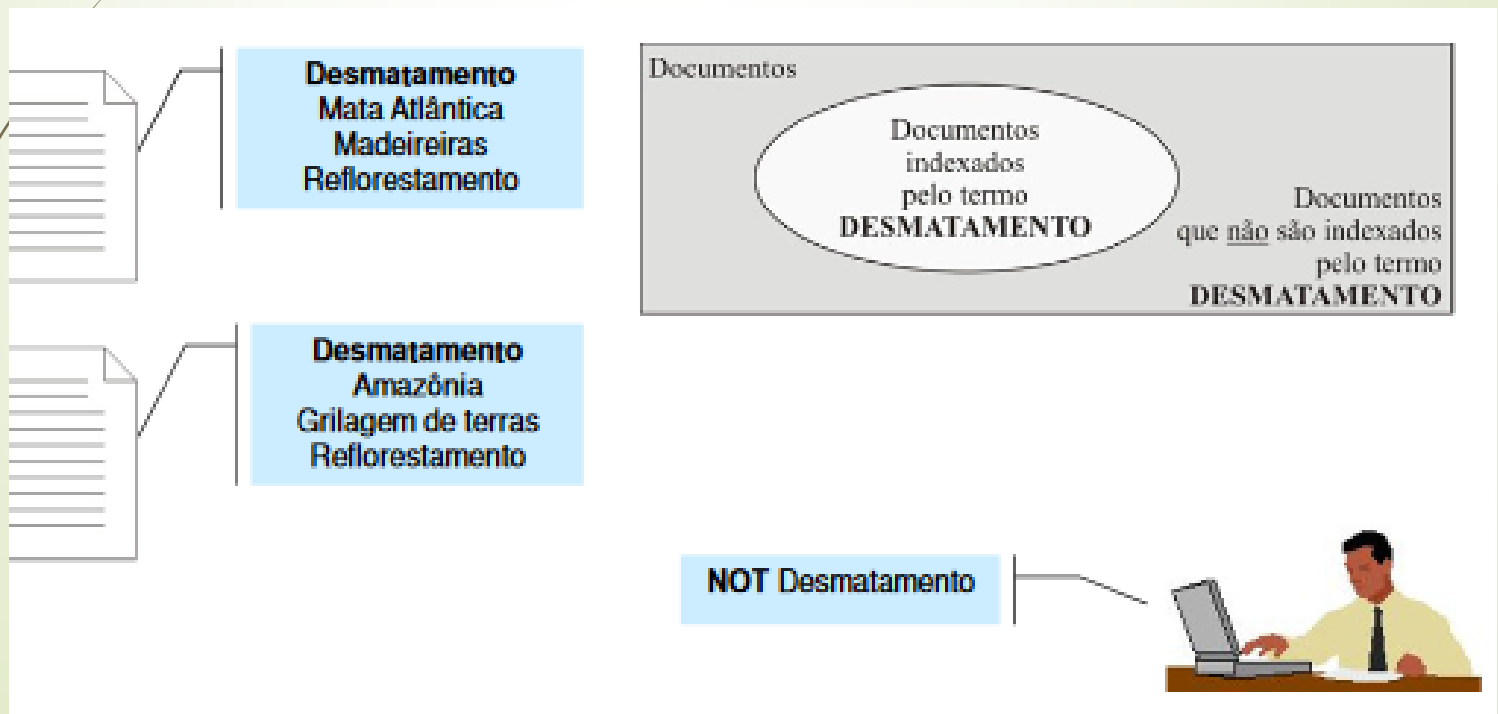
- Uma expressão disjuntiva t_1 OR t_2 recuperará documentos indexados pelo termo t_1 ou pelo termo t_2 .
- Esta operação equivale à união entre o conjunto dos documentos indexados pelo termo t_1 com o conjunto dos documentos indexados pelo termo t_2 .

Modelo Booleano



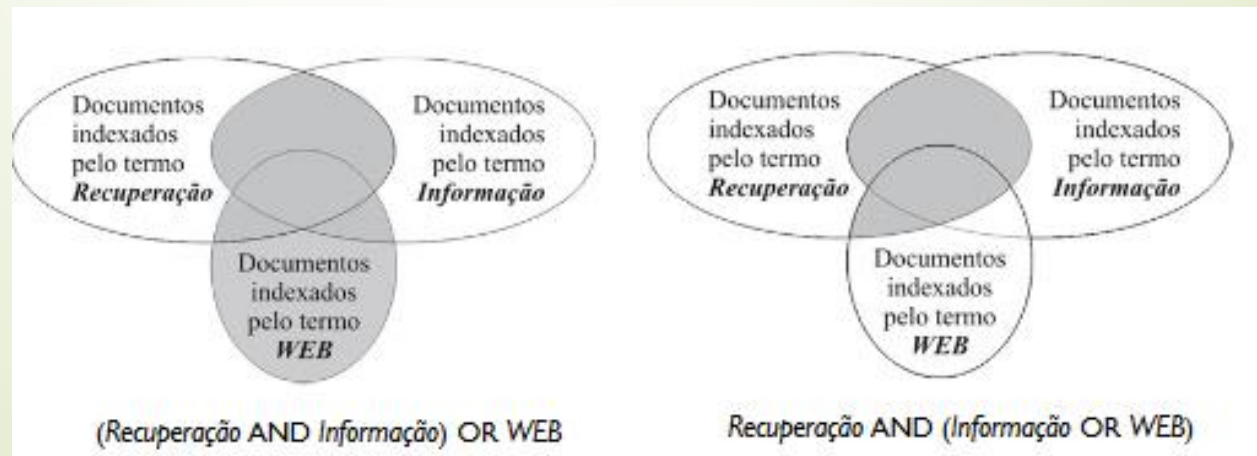
Modelo Booleano

- A expressão NOT t_1 recuperará documentos que não são indexados pelo termo t_1 .



Modelo Booleano

- **Termos e operadores booleanos podem ser combinados** para especificar buscas mais detalhadas ou restritivas.
- **Sendo que a ordem de execução das operações lógicas de uma expressão influencia no resultado da busca;**





Modelo Boleano

- O modelo Booleano prevê que cada documento seja **relevante** ou **não relevante**;
- Não existe **satisfação parcial** das condições da consulta;
- Esse critério binário de decisão, sem nenhuma noção de grau, impede uma boa qualidade na recuperação de informação.

Exemplo

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	1	1	0	0	0	1
Brutus	1	1	0	1	0	0
Caesar	1	1	0	1	1	1
Calpurnia	0	1	0	0	0	0
Cleopatra	1	0	0	0	0	0
mercy	1	0	1	1	1	1
worser	1	0	1	1	1	0

Brutus AND Caesar but NOT Calpurnia

1 if play contains
word, 0 otherwise



Exemplo

- Solução:
- Brutus AND Caesar AND NOT Calpurnia
- 110100 AND
- 110111 AND
- 101111 (**inverso do conjunto de Calpurnia**)
- **100100 ::: RESULTADO**



Modelo booleano x Sistema de RI

- Consulta do usuário
 - Uso de expressões booleanas;
- Recuperação de documentos
 - Somente os documentos que satisfazem a consulta são recuperados;
- Ranqueamento dos documentos
 - Impossível!



Modelo booleano

Vantagens e desvantagens

- Vantagens

- Formalismo claro;
- Simplicidade;
- Fácil de implementar;

- Desvantagens

- Impossibilidade de realizar ranqueamento dos documentos;
- Formulação de consultas booleanas pode ser inconveniente para os usuários.

ATIVIDADE PARA ENTREGAR – 25/05

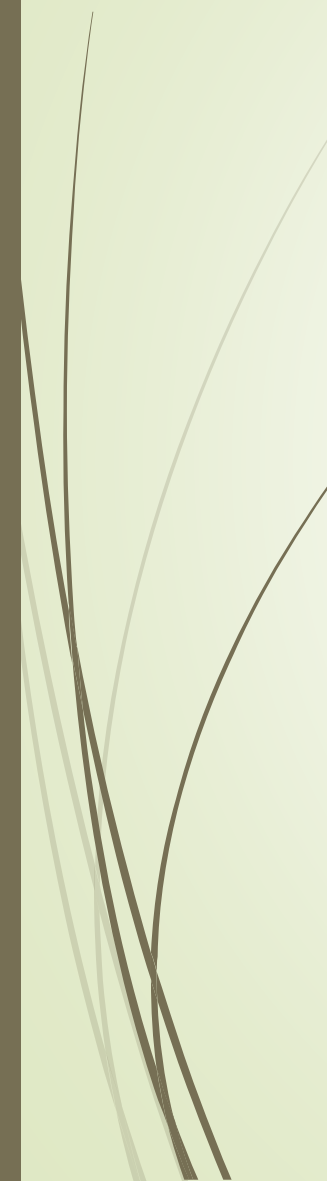

- Suponha que o vocabulário da coleção seja dado por $V = \{k_a, k_b, k_c\}$. Seja d_j um documento que contém os termos k_a e k_c . Ou seja, $d_j = \{1, 0, 1\}$. Considere a consulta $q = k_a \text{ AND } (k_b \text{ OR NOT } k_c)$.
- Pergunta 1: o documento d_j satisfaz a consulta q ?
- Pergunta 2: que documento satisfaz a consulta q ?

ATIVIDADE PARA ENTREGAR – 25/05

- Considere três documentos D1, D2 e D3. D1 é a primeira estrofe do hino à bandeira, D2 é a primeira estrofe do hino da independência e D3 é a primeira estrofe do hino nacional. Considere somente os substantivos em cada documento.
- a) Encontre o vocabulário dessa coleção.
- b) Monte a matriz de termos e documentos.
- c) Encontre a similaridade entre os documentos e cada uma das consultas a seguir: $q1 = \{\text{liberdade brasil}\}$, $q2 = \{\text{patria}\}$ e $q3 = \{\text{nobre heroico}\}$.

ATIVIDADE PARA ENTREGAR – 25/05

- D1 = {Salve, lindo pendão da esperança, Salve, símbolo augusto da paz! Tua nobre presença à lembrança A grandeza da Pátria nos traz. }
- D2 = {Já podeis, da Pátria filhos, Ver contente a mãe gentil; Já raiou a liberdade No horizonte do Brasil.}
- D3 = {Ouviram do Ipiranga as margens plácidas De um povo heroico o brado retumbante, E o sol da Liberdade, em raios fúlgidos Brilhou no céu da Pátria nesse instante.}



*Se você acredita que é
capaz, ignore a opinião dos
outros e siga em frente.
Nem sempre é bom saber o
que outros pensam.*

frasesdobem