

## **Exercícios sobre Deduplicação e Blocagem**

### **Luiz Fernando Böhm**

**Resolva os exercícios abaixo utilizando a base de dados fornecida nas pág. 2 e 3. Registros com o campo Cod\_Dupl iguais representam registros duplicados.**

**1 - Calcule o número de comparações necessárias para definir quais são os registros duplicados sem a utilização de blocagem.**

**2 - Calcule Precisão, Recall e F-Measure utilizando como método de matching a comparação do campo sexo concatenado ao ano de nascimento.**

**3 – Qual o número de falsos positivos e falsos negativos?**

**4 – Calcule PC, RR e F-Score utilizando a chave de blocagem abaixo:**

**Bloco 1: Sexo = F**

**Bloco 2: Sexo = M**

**5 – Encontre uma chave de blocagem com PC maior do que o do exercício 4. Calcule o RR e o F-Score.**

# Base de Dados classificada por Cod\_Dupl

ID	Cod_Dupl	Nome	Nome Mãe	Data Nasc	Sexo
1	01	ADRIANA MARCONDES	CRISTIANE SOARES	30/08/1962	F
2	02	MARIA ADELAIDE ALVES DA SILVA	ZAIDA RODRIGUES ALVES	15/05/1967	F
3	02	MARIA ADELAIDE RODRIGUES ALVES	ZAIDA RODRIGUES ALVES	15/06/1967	F
4	03	ROBERTA FERNANDES	MARCIA FERNANDES	25/09/1988	F
5	04	JORGE SANTOS DA SILVA	JORGINA SANTOS DA SILVA	01/01/1990	M
7	05	VINICIUS SANTOS	ROSANE SANTOS	25/09/1988	M
6	05	VINICIOS SANTOS	ROSANNE SANTOS	15/09/1988	M
8	06	DERLY DA SILVA	MARIA DA SILVA	30/08/1962	F
10	07	MARIA APARECIDA RODRIGUES	JOAQUINA RODRIGUES	03/11/1945	F
9	07	MARIA A. RODRIGUES	JOAQUINA RODRIGUEZ	03/10/1945	M
11	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2000	M
12	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2001	M
13	09	BRUNO FREITAS DA SILVA	CARLA DA SILVA	01/01/1990	M
14	10	FERNANDA MARTINS PICCININI	ZAIRA PICCININI	05/07/1999	F
15	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
16	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
17	12	LETICIA MENDES DELLANO	N. INF	20/08/1988	F
18	13	THIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
19	13	TIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
20	14	NORBERTO G. FIGUEIREDO	HELENA G. FIGUEIREDO	10/10/1910	M
21	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE	30/03/1982	M
22	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE FERREIRA	30/03/1981	F
23	16	JOANA PEREIRA	JOANA PEREIR	12/12/2001	F
24	16	JOANA PEREIRA	JOANA PEREIRA	12/12/2001	F
25	17	DERLY DA SILVA	MARIA DA SILVA	22/12/1962	M
26	18	FABIO GUIMARAES	ZAIDA ALVES MENEZES	15/05/1967	M
27	18	PHABIO GUIMARAES	ZAIDA ALVES MENEZES	15/06/1967	M
28	19	MARIA ADELAIDE SOARES	CRISTIANE SOARES	10/10/1978	F
29	20	VINICIUS DOS SANTOS	N. INF	25/08/1988	M
30	21	JANA FERREIRA	JOANA FERREIRA	21/11/1949	F

# Base de dados classificada por Sexo / Nome

ID	Cod_Dupl	Nome	Nome Mãe	Data Nasc	Sexo
1	01	ADRIANA MARCONDES	CRISTIANE SOARES	30/08/1962	F
8	06	DERLY DA SILVA	MARIA DA SILVA	30/08/1962	F
22	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE FERREIRA	30/03/1981	F
14	10	FERNANDA MARTINS PICCININI	ZAIRA PICCININI	05/07/1999	F
30	21	JANA FERREIRA	JOANA FERREIRA	21/11/1949	F
23	16	JOANA PEREIRA	JOANA PEREIR	12/12/2001	F
24	16	JOANA PEREIRA	JOANA PEREIRA	12/12/2001	F
17	12	LETICIA MENDES DELLANO	N. INF	20/08/1988	F
2	02	MARIA ADELAIDE ALVES DA SILVA	ZAIDA RODRIGUES ALVES	15/05/1967	F
3	02	MARIA ADELAIDE RODRIGUES ALVES	ZAIDA RODRIGUES ALVES	15/06/1967	F
28	19	MARIA ADELAIDE SOARES	CRISTIANE SOARES	10/10/1978	F
10	07	MARIA APARECIDA RODRIGUES	JOAQUINA RODRIGUES	03/11/1945	F
4	03	ROBERTA FERNANDES	MARCIA FERNANDES	25/09/1988	F
11	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2000	M
12	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2001	M
13	09	BRUNO FREITAS DA SILVA	CARLA DA SILVA	01/01/1990	M
15	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
16	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
21	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE	30/03/1982	M
25	17	DERLY DA SILVA	MARIA DA SILVA	22/12/1962	M
26	18	FABIO GUIMARAES	ZAIDA ALVES MENEZES	15/05/1967	M
5	04	JORGE SANTOS DA SILVA	JORGINA SANTOS DA SILVA	01/01/1990	M
9	07	MARIA A. RODRIGUES	JOAQUINA RODRIGUEZ	03/10/1945	M
20	14	NORBERTO G. FIGUEIREDO	HELENA G. FIGUEIREDO	10/10/1910	M
27	18	PHABIO GUIMARAES	ZAIDA ALVES MENEZES	15/06/1967	M
18	13	THIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
19	13	TIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
6	05	VINICIOS SANTOS	ROSANNE SANTOS	15/09/1988	M
29	20	VINICIUS DOS SANTOS	N. INF	25/08/1988	M
7	05	VINICIUS SANTOS	ROSANE SANTOS	25/09/1988	M

# Base de dados classificada por Nome

ID	Cod_Dupl	Nome	Nome Mãe	Data Nasc	Sexo
1	01	ADRIANA MARCONDES	CRISTIANE SOARES	30/08/1962	F
11	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2000	M
12	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2001	M
13	09	BRUNO FREITAS DA SILVA	CARLA DA SILVA	01/01/1990	M
15	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
16	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
21	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE	30/03/1982	M
8	06	DERLY DA SILVA	MARIA DA SILVA	30/08/1962	F
25	17	DERLY DA SILVA	MARIA DA SILVA	22/12/1962	M
22	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE FERREIRA	30/03/1981	F
26	18	FABIO GUIMARAES	ZAIDA ALVES MENEZES	15/05/1967	M
14	10	FERNANDA MARTINS PICCININI	ZAIRA PICCININI	05/07/1999	F
30	21	JANA FERREIRA	JOANA FERREIRA	21/11/1949	F
23	16	JOANA PEREIRA	JOANA PEREIR	12/12/2001	F
24	16	JOANA PEREIRA	JOANA PEREIRA	12/12/2001	F
5	04	JORGE SANTOS DA SILVA	JORGINA SANTOS DA SILVA	01/01/1990	M
17	12	LETICIA MENDES DELLANO	N. INF	20/08/1988	F
9	07	MARIA A. RODRIGUES	JOAQUINA RODRIGUEZ	03/10/1945	M
2	02	MARIA ADELAIDE ALVES DA SILVA	ZAIDA RODRIGUES ALVES	15/05/1967	F
3	02	MARIA ADELAIDE RODRIGUES ALVES	ZAIDA RODRIGUES ALVES	15/06/1967	F
28	19	MARIA ADELAIDE SOARES	CRISTIANE SOARES	10/10/1978	F
10	07	MARIA APARECIDA RODRIGUES	JOAQUINA RODRIGUES	03/11/1945	F
20	14	NORBERTO G. FIGUEIREDO	HELENA G. FIGUEIREDO	10/10/1910	M
27	18	PHABIO GUIMARAES	ZAIDA ALVES MENEZES	15/06/1967	M
4	03	ROBERTA FERNANDES	MARCIA FERNANDES	25/09/1988	F
18	13	THIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
19	13	TIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
6	05	VINICIOS SANTOS	ROSANNE SANTOS	15/09/1988	M
29	20	VINICIUS DOS SANTOS	N. INF	25/08/1988	M
7	05	VINICIUS SANTOS	ROSANE SANTOS	25/09/1988	M

Base de dados classificada por Nome da Mãe

ID	Cod Dupl	Nome	Nome Mãe	Data Nasc	Sexo
11	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2000	M
12	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2001	M
13	09	BRUNO FREITAS DA SILVA	CARLA DA SILVA	01/01/1990	M
1	01	ADRIANA MARCONDES	CRISTIANE SOARES	30/08/1962	F
28	19	MARIA ADELAIDE SOARES	CRISTIANE SOARES	10/10/1978	F
18	13	THIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
19	13	TIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
20	14	NORBERTO G. FIGUEIREDO	HELENA G. FIGUEIREDO	10/10/1910	M
30	21	JANA FERREIRA	JOANA FERREIRA	21/11/1949	F
23	16	JOANA PEREIRA	JOANA PEREIR	12/12/2001	F
24	16	JOANA PEREIRA	JOANA PEREIRA	12/12/2001	F
10	07	MARIA APARECIDA RODRIGUES	JOAQUINA RODRIGUES	03/11/1945	F
9	07	MARIA A. RODRIGUES	JOAQUINA RODRIGUEZ	03/10/1945	M
5	04	JORGE SANTOS DA SILVA	JORGINA SANTOS DA SILVA	01/01/1990	M
21	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE	30/03/1982	M
22	15	DIAINE MARISTELA DUARTE	MARCIA DUARTE FERREIRA	30/03/1981	F
4	03	ROBERTA FERNANDES	MARCIA FERNANDES	25/09/1988	F
8	06	DERLY DA SILVA	MARIA DA SILVA	30/08/1962	F
25	17	DERLY DA SILVA	MARIA DA SILVA	22/12/1962	M
17	12	LETÍCIA MENDES DELLANO	N. INF	20/08/1988	F
29	20	VINICIUS DOS SANTOS	N. INF	25/08/1988	M
7	05	VINICIUS SANTOS	ROSANE SANTOS	25/09/1988	M
15	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
16	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
6	05	VINICIOS SANTOS	ROSANNE SANTOS	15/09/1988	M
26	18	FABIO GUIMARAES	ZAIDA ALVES MENEZES	15/05/1967	M
27	18	PHABIO GUIMARAES	ZAIDA ALVES MENEZES	15/06/1967	M
2	02	MARIA ADELAIDE ALVES DA SILVA	ZAIDA RODRIGUES ALVES	15/05/1967	F
3	02	MARIA ADELAIDE RODRIGUES ALVES	ZAIDA RODRIGUES ALVES	15/06/1967	F
14	10	FERNANDA MARTINS PICCININI	ZAIRA PICCININI	05/07/1999	F

Base de dados classificada por Data Nasc

ID	Cod Dupl	Nome	Nome Mãe	Data Nasc	Sexo
20	14	NORBERTO G. FIGUEIREDO	HELENA G. FIGUEIREDO	10/10/1910	M
9	07	MARIA A. RODRIGUES	JOAQUINA RODRIGUEZ	03/10/1945	M
10	07	MARIA APARECIDA RODRIGUES	JOAQUINA RODRIGUES	03/11/1945	F
30	21	JANA FERREIRA	JOANA FERREIRA	21/11/1949	F
1	01	ADRIANA MARCONDES	CRISTIANE SOARES	30/08/1962	F
8	06	DERLY DA SILVA	MARIA DA SILVA	30/08/1962	F
25	17	DERLY DA SILVA	MARIA DA SILVA	22/12/1962	M
2	02	MARIA ADELAIDE ALVES DA SILVA	ZAIDA RODRIGUES ALVES	15/05/1967	F
26	18	FABIO GUIMARAES	ZAIDA ALVES MENEZES	15/05/1967	M
3	02	MARIA ADELAIDE RODRIGUES ALVES	ZAIDA RODRIGUES ALVES	15/06/1967	F
27	18	PHABIO GUIMARAES	ZAIDA ALVES MENEZES	15/06/1967	M
28	19	MARIA ADELAIDE SOARES	CRISTIANE SOARES	10/10/1978	F
22	15	DIAINE MARISTELA DUARTE	MARCIA DUARTE FERREIRA	30/03/1981	F
21	15	DAIANE MARISTELA DUARTE	MARCIA DUARTE	30/03/1982	M
17	12	LETÍCIA MENDES DELLANO	N. INF	20/08/1988	F
29	20	VINICIUS DOS SANTOS	N. INF	25/08/1988	M
6	05	VINICIOS SANTOS	ROSANNE SANTOS	15/09/1988	M
4	03	ROBERTA FERNANDES	MARCIA FERNANDES	25/09/1988	F
7	05	VINICIUS SANTOS	ROSANE SANTOS	25/09/1988	M
15	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
16	11	CARLOS ROBERTO DOS SANTOS	ROSANE SANTOS	25/09/1988	M
5	04	JORGE SANTOS DA SILVA	JORGINA SANTOS DA SILVA	01/01/1990	M
13	09	BRUNO FREITAS DA SILVA	CARLA DA SILVA	01/01/1990	M
14	10	FERNANDA MARTINS PICCININI	ZAIRA PICCININI	05/07/1999	F
18	13	THIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
19	13	TIAGO DIAS SOARES	EUNICE DIAS SOARES	05/07/1999	M
11	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2000	M
12	08	BRUNO FREITAS DA SILVA	CARLA DA SILVA	12/12/2001	M
23	16	JOANA PEREIRA	JOANA PEREIR	12/12/2001	F
24	16	JOANA PEREIRA	JOANA PEREIRA	12/12/2001	F