



UNIVERSIDADE FEDERAL DO TOCANTINS  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

**RAFAEL BRUSTULIN**

**INCLUSÃO DE ETAPA DE PÓS-PROCESSAMENTO DETERMINÍSTICO  
PARA AUMENTO DE PERFORMANCE DO RELACIONAMENTO (*LINKAGE*)  
PROBABILÍSTICO.**

PALMAS-TO  
2018

**RAFAEL BRUSTULIN**

**INCLUSÃO DE ETAPA DE PÓS-PROCESSAMENTO DETERMINÍSTICO  
PARA AUMENTO DE PERFORMANCE DO RELACIONAMENTO (*LINKAGE*)  
PROBABILÍSTICO.**

Dissertação apresentada ao Mestrado Profissional em Ciências da Saúde da Universidade Federal do Tocantins como requisito parcial à obtenção do grau de Mestre.

Orientadora: Dr<sup>a</sup> Poliana Guerino Marson.

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**Sistema de Bibliotecas da Universidade Federal do Tocantins**

---

B912i Brustulin, Rafael.

INCLUSÃO DE ETAPA DE PÓS-PROCESSAMENTO DETERMINÍSTICO PARA AUMENTO DE PERFORMANCE DO RELACIONAMENTO (LINKAGE) PROBABILÍSTICO. / Rafael Brustulin. – Palmas, TO, 2018.

75 f.

Dissertação (Mestrado Acadêmico) - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Pós-Graduação (Mestrado) em Ciências da Saúde, 2018.

Orientadora : Poliana Guerino Marson

1. Base de Dados. 2. Software. 3. Sistemas de Informação. 4. Saúde Coletiva. I. Título

**CDD 610**

---

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

**Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).**

UNIVERSIDADE FEDERAL DO TOCANTINS  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE  
CÂMPUS DE PALMAS

Avenida NS 15, Quadra 109 Norte | Plano Diretor Norte  
Complexo de Estudos Geoambientais e de Saúde, Câmpus de Palmas | 77001-090 | Palmas/TO  
(63) 3229-4687 | www.uft.edu.br/ppgcs/ | mprofisaude@uft.edu.br



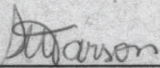
**ATA N.º 04/2018 – ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO**

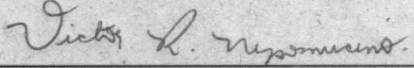
No décimo sexto dia do mês de março de 2018, realizou-se na Universidade Federal do Tocantins, no Campus de Palmas, às 14h00, a defesa de Dissertação de Mestrado do aluno **Rafael Brustulin**, regularmente matriculado no Curso de Mestrado Profissional em Ciências da Saúde, com o título “**INCLUSÃO DE ETAPA DE PÓS-PROCESSAMENTO DETERMINÍSTICO PARA AUMENTO DE PERFORMANCE DO RELACIONAMENTO (LINKAGE) PROBABILÍSTICO**”. O discente expôs seu trabalho por cerca de 30 minutos e foi arguido pelos componentes da Comissão Julgadora, como segue: Dra. Poliana Guerino Marson, Dr. Victor Rodrigues Nepomuceno e Dr. Fernando Almeida Machado. Recebeu sugestões e em seguida foi emitido o resultado final:

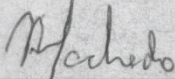
APROVADO  
 REPROVADO

A Comissão Julgadora retornou à sala e assim o resultado foi proclamado. Desta forma, os trabalhos foram encerrados por volta das 16h00.

Banca:

  
\_\_\_\_\_  
Dra. Poliana Guerino Marson  
Orientadora

  
\_\_\_\_\_  
Dr. Victor Rodrigues Nepomuceno  
Examinador

  
\_\_\_\_\_  
Dr. Fernando Almeida Machado  
Examinador

## **EPÍGRAFE**

*In the middle of difficulty lies opportunity (Albert Einstein)*

## RESUMO

O objetivo do presente estudo foi demonstrar a aplicação de uma etapa de pós-processamento determinístico, baseada em medidas de similaridade, para aumentar a performance do relacionamento probabilístico com e sem a etapa de revisão manual. As bases de dados utilizadas no estudo foram o Sistema de Informações de Agravos de Notificação e o Sistema de Informação Sobre Mortalidade no período de 2007 a 2015 do município de Palmas, Tocantins. O software probabilístico utilizado foi o *OpenRecLink*; foi desenvolvida e aplicada uma etapa de pós-processamento determinístico aos dados obtidos por três diferentes estratégias de pareamento probabilístico. As três estratégias foram comparadas entre si e acrescidas da etapa de pós-processamento determinístico. A sensibilidade das estratégias probabilísticas sem revisão manual variou entre 69,1% e 77,8%, enquanto que as mesmas estratégias, acrescidas da etapa de pós-processamento determinístico, tiveram uma variação entre 92,9% e 96,3%. A sensibilidade de duas estratégias probabilísticas com revisão manual foi semelhante aos obtidos pela etapa de pós-processamento determinístico, no entanto o número de pares destinados a revisão manual pelas duas estratégias probabilísticas variou entre 1.177 e 1.132 registros, contra 149 e 145 após a etapa de pós-processamento determinístico. Nossos resultados sugerem que a etapa de pós-processamento determinístico é uma opção promissora, tanto para aumentar a sensibilidade quanto para reduzir o número de pares que precisam ser revisados manualmente, ou mesmo para eliminar sua necessidade.

**Palavras-chave:** Base de Dados; Software; Sistemas de Informação.

## ABSTRACT

The objective of the present study was to demonstrate the application of a deterministic post-processing step, based on similarity measures, to increase the performance of the probabilistic relationship with and without the clerical review. The databases used in the study were the Information System of Notifiable Diseases and the Mortality Information System in the period from 2007 to 2015 of the municipality of Palmas, Tocantins, Brazil. The probabilistic software used was OpenRecLink; a deterministic post-processing step was developed and applied to the data obtained by three different probabilistic matching strategies. The three strategies were compared to each other and added to the deterministic post-processing step. The sensitivity of the probabilistic strategies without manual revision varied between 69.1% and 77.8%, while the same strategies, added to the deterministic post-processing step, ranged from 92.9% to 96.3%. The sensitivity of two probabilistic strategies with manual revision was similar to those obtained by the deterministic post-processing step. However, the number of pairs destined for manual revision by the two probabilistic strategies varied between 1,177 and 1,132 registers, against 149 and 145 after the post-processing step. Our results suggest that the deterministic postprocessing step is a promising option both to increase sensitivity and to reduce the number of pairs that need to be revised manually or even to eliminate their need.

**Keywords:** Data Base; Software; Information Systems.

## **LISTA DE ABREVIATURAS E SIGLAS**

EPPD	Etapa de Pós-Processamento Determinístico
IC <sub>95%</sub>	Intervalo de Confiança de 95%
SIM	Sistema de Informação de Mortalidade
SINAN	Sistema de Informação de Agravos de Notificação
SINASC	Sistema de Informações sobre Nascidos Vivos
SIS	Sistemas de Informação em Saúde
SUS	Sistema Único de Saúde
VPP	Valor Preditivo Positivo



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>10</b>
1.1	SISTEMAS DE INFORMAÇÃO .....	10
1.2	RELACIONAMENTO DE DADOS.....	12
1.3	JUSTIFICATIVA .....	15
<b>2</b>	<b>OBJETIVO</b> .....	<b>16</b>
2.1	OBJETIVOS GERAIS.....	16
2.2	OBJETIVOS ESPECÍFICOS.....	16
<b>3</b>	<b>METODOLOGIA</b> .....	<b>17</b>
3.1	BASES DE DADOS E PRÉ-PROCESSAMENTO .....	17
3.2	RELACIONAMENTO PROBABILÍSTICO.....	17
3.3	ROTINA PARA REVISÃO MANUAL .....	18
3.4	PONTOS DE CORTE PARA OS ESCORES DO RELACIONAMENTO PROBABILÍSTICO .....	19
3.5	ETAPA DE PÓS-PROCESSAMENTO DETERMINÍSTICO (EPPD).....	19
3.6	AVALIAÇÃO DO RELACIONAMENTO DE DADOS .....	21
<b>4</b>	<b>RESULTADOS</b> .....	<b>22</b>
4.1	ARTIGO.....	23
4.2	INFORMAÇÕES SUPLEMENTARES .....	40
<b>5</b>	<b>CONCLUSÃO</b> .....	<b>46</b>
<b>6</b>	<b>REFERÊNCIAS</b> .....	<b>47</b>
	<b>APÊNDICE A - MANUAL</b> .....	<b>51</b>
	<b>ANEXO A – PARECER DO CEP</b> .....	<b>72</b>