

SCD DW: Proposta de um *Data Warehouse* para a Doença Anemia Falciforme

Arthur Emanuel de Oliveira Carosia

Instituto Federal de São Paulo (IFSP), São João da Boa Vista, SP, Brasil

Abstract. The Sickle Cell Disease is a genetic and hereditary disease that causes malformation of red blood cells. This disease has no cure currently and therefore there is a need to find ways that can mitigate it. Thus, it is important to gather and exploit knowledge reported in medical articles and treat this disease, in order to assist the development of new research and treatments. In this context, the development of a Data Warehouse can help significantly in the decision-making process on Sickle Cell Disease. This paper aims to propose a set of multidimensional schemes to facilitate trend analysis of aspects related to Sickle Cell Disease, thereby contributing to the development of new research in the medical field. In addition, this paper discusses the development of an application, called SCD Data Warehouse, which allows the specialist to access and analysis data stored in the Data Warehouse

Resumo. A Anemia Falciforme é uma doença genética e hereditária que causa a má formação das hemácias. Esta doença não possui cura atualmente e, portanto, há a necessidade de encontrar meios que possam atenuá-la. Para isso, deve-se reunir e explorar o conhecimento relatado em artigos científicos da área médica e que tratam dessa doença, com o objetivo de auxiliar o desenvolvimento de novas pesquisas e tratamentos. Nesse contexto, o desenvolvimento de um *Data Warehouse* pode auxiliar de maneira significativa no processo de tomada de decisão sobre a Anemia Falciforme. Este trabalho tem como objetivo propor um conjunto de esquemas multidimensionais que visam facilitar a análise de tendências de aspectos relacionados à doença Anemia Falciforme, contribuindo, dessa forma, para o desenvolvimento de novas pesquisas na área medica. Além disso, este trabalho aborda o desenvolvimento de um aplicativo, denominado SCD *Data Warehouse*, que permite ao especialista médico acessar análises dos dados armazenados no *Data Warehouse*.

Palavras-chave - *Data Warehouse*, Sistemas de informação médicos, Banco de dados

I. INTRODUÇÃO

A anemia falciforme (em inglês, *sickle cell disease* - SCD) é uma doença genética e hereditária que causa a má

formação das hemácias, que assumem formato de foices (ilustrado na Figura 1). No Brasil, existem poucas pesquisas no que visam encontrar uma solução atenuadora da doença. Esta doença não possui cura atualmente e, portanto, há a necessidade de encontrar meios que possam moderá-la.

Diversos estudos sobre a doença podem ser encontrados em artigos científicos internacionais publicados em periódicos da medicina, os quais relatam experiências sobre a doença em países europeus considerando: diferentes grupos de pacientes, áreas geográficas, tratamentos, efeitos causados pelos tratamentos, e assim por diante [1, 2, 3, 4]. Conseqüentemente, o conhecimento armazenado nesses artigos científicos consiste em uma importante fonte de informação, que pode ser explorada visando auxiliar o desenvolvimento de novas pesquisas e a adequação das experiências internacionais ao cenário brasileiro.

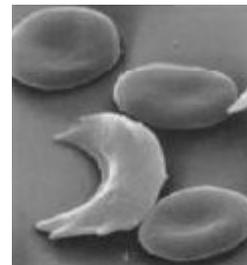


Figura 1. Exemplo de hemácias falciformes. Adaptada de [5].

Surge, então, a necessidade de desenvolver técnicas que auxiliem a extração de informações relevantes destes artigos, para que pesquisas diferenciadas possam ser realizadas. Neste sentido, está em desenvolvimento um ambiente para a análise de dados da doença Anemia Falciforme, composto por dois sistemas principais. O primeiro deles, chamado de DORS-SCD (*Data Organizing and Recovering System for Sickle Cell Disease*), tem como objetivo extrair dados de interesse de artigos médicos sobre a Anemia Falciforme e armazená-los em um banco de dados. O trabalho descrito em [6] apresenta o sistema *Medical Search*, que se trata de um sistema *web* para extração de informação automática de artigos da área médica. Já o segundo sistema, chamado de DSS-SCD (*Decision Support System for Sickle Cell Disease*), tem como objetivo identificar padrões ou permitir a predição de fatos futuros por meio do uso de *Data Warehouse*.

Arthur Emanuel de Oliveira Carosia é bacharel e mestre em Ciência da Computação pela Universidade de São Paulo. Atualmente é professor do ensino básico, técnico e tecnológico no Instituto Federal de São Paulo, Câmpus São João da Boa Vista (e-mail: arthuremanuel.carosia@gmail.com).

Dessa forma, este trabalho tem como objetivo propor um conjunto de esquemas multidimensionais para o ambiente DSS-SCD, visando fornecer informações e estatísticas relevantes sobre a doença Anemia Falciforme para especialistas médicos. Os esquemas multidimensionais devem, portanto, facilitar a análise de tendências de aspectos relacionadas à doença Anemia Falciforme, tais como tratamentos mais utilizados e seus efeitos mais comuns em pacientes com a doença. O trabalho também visa a construção de um *Data Warehouse*, usando como base os esquemas multidimensionais propostos. Por fim, este trabalho também aborda o desenvolvimento de um aplicativo, denominado *SCD Data Warehouse*, cujo objetivo é permitir ao especialista da área médica acessar as informações geradas pelas análises dos dados contidos no *Data Warehouse*.

Este trabalho está organizado como segue. A Seção II aborda a revisão bibliográfica realizada sobre os seguintes temas: *Data Warehouse*, modelagem multidimensional, e trabalhos relacionados. A Seção III descreve o trabalho realizado durante a proposta dos esquemas multidimensionais e a implementação do *Data Warehouse* a partir desses esquemas. A Seção IV aborda o aplicativo *SCD Data Warehouse* desenvolvido para acesso aos dados armazenados no *Data Warehouse*. Finalmente, a Seção V descreve as conclusões deste trabalho e propostas para trabalhos futuros.

II. REVISÃO BIBLIOGRÁFICA

A. *Data Warehouse*

Um *Data Warehouse* é um banco de dados integrado a partir de diversos provedores de informação, cujo objetivo principal é apoiar os processos de gerência e a tomada de decisão [7]. Ele é a base para o ambiente informacional de uma empresa e, desse modo, a separação entre os ambientes operacional e informacional das empresas é explícita. Enquanto o ambiente operacional é composto por aplicações que dão suporte ao dia a dia do negócio, o ambiente informacional é constituído por aplicações que analisam o negócio. Assim, geralmente, o fluxo de informação em um ambiente de *Data Warehouse* é dado pelos provedores de informação para o *Data Warehouse* [7, 8].

Em se tratando do ambiente operacional, suas características são:

- Voltado ao processamento de transações OLTP (*online analytical processing*), isto é, são as atividades e sistemas que tratam da entrada de dados em um banco de dados.
- Realiza tipicamente operações de atualização, remoção e inserção na base de dados. Estas operações são realizadas em grande volume, contudo, são simples e acessam geralmente poucos registros por vez.
- O projeto do banco de dados é normalizado e oferece suporte às propriedades ACID (atomicidade, consistência, isolamento e durabilidade).
- Como característica principal das consultas elaboradas, o seu tempo de resposta é de poucos segundos.

Por outro lado, o *Data Warehouse* é normalmente

pertencente ao ambiente informacional de uma empresa. Esse ambiente possui as seguintes características:

- Voltado ao processamento de transações OLAP (*online analytical processing*), isto é, uma categoria de processamento analítico sobre um banco de dados histórico voltado para os processos de gerência e tomada de decisão.
- Ambiente que tem como objetivo a realização de consultas longas e complexas, acessando muitos registros e realizando muitas junções e varreduras em tabelas.
- Como característica principal das consultas comumente realizadas, seu tempo de resposta varia de minutos a horas.
- O projeto do banco de dados de um *Data Warehouse* é em esquema multidimensional (descrito na Seção 2.3), refletindo as necessidades de análise dos usuários de sistemas de suporte à decisão.

B. Características dos Dados em um *Data Warehouse*

Em relação à característica dos dados de um *Data Warehouse*, os seguintes fatores são importantes a se considerar [9].

- Orientados a assunto: os dados são relativos aos temas de negócio de maior interesse da corporação. Os dados são relevantes ao contexto da tomada de decisão e estruturados de forma desnormalizada.
- Integrados: os dados de um *Data Warehouse* são obtidos a partir do sistema operacional, que em sua maioria é composto de diversos bancos de dados não integrados com informações relativas aos negócios da empresa. Os dados obtidos dos sistemas operacionais são integrados e armazenados sem ambiguidades e com inconsistências já resolvidas no *Data Warehouse*.
- Não-voláteis: o conteúdo de um *Data Warehouse* permanece estável por longos períodos de tempo. Isto porque as operações mais comuns neste ambiente são as de manutenção (carga do *Data Warehouse* visando a consistência dos dados) e leitura (consultas aos dados armazenados no *Data Warehouse*).
- Históricos: os dados armazenados no *Data Warehouse* são relevantes por algum período de tempo, em geral de 5 a 10 anos. Desse modo, consultas que vislumbram análises de tendências podem ser realizadas. No entanto, o volume de dados do *Data Warehouse* é influenciado por esta abordagem, sendo muito superior ao do sistema operacional.

Além disso, em um *Data Warehouse*, os dados são granulares. Granularidade refere-se ao nível de detalhe em que as unidades de dados são mantidas em um *Data Warehouse*. Quanto maior o nível de detalhe, menor a granularidade e vice-versa [10]. A granularidade determina, portanto, a dimensionalidade do banco de dados. Quando o nível de granularidade é pequeno, o *Data Warehouse* é grande, mas é

possível responder uma quantidade grande de consultas. No entanto, se o nível da granularidade for grande, o tamanho do banco de dados e o número de consultas a serem respondidas reduzem.

Por fim, tratando da organização dos dados em um *Data Warehouse*, apresentada na Figura 2, eles possuem diferentes níveis de agregação. O nível inferior é composto dos dados integrados que são obtidos a partir do ambiente operacional. Eles são bastante detalhados e, portanto, servem de base para consultas desconhecidas. À medida que o nível da hierarquia de agregação é incrementado, os dados apresentam-se mais resumidos em relação ao nível anterior. Ou seja, dados do nível “*n*” são alguma forma de agregação dos dados do nível “*n-1*”, sendo que os níveis intermediários utilizam dados do nível inferior para as suas agregações [8].

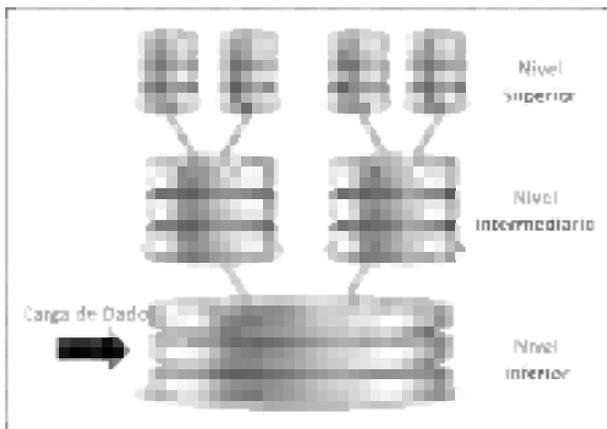


Figura 2. Níveis de agregação em um *Data Warehouse*

C. Modelagem Multidimensional

A modelagem multidimensional é utilizada no projeto do *Data Warehouse* devido principalmente à sua simplicidade em relação à apresentação de informações analíticas ao usuário [10]. Essa modelagem é composta por tabela de fatos e tabelas de dimensões. A tabela de fatos armazena as medidas numéricas de objetos de análise relevantes ao negócio [16, 9]. As tabelas de fatos sempre possuem duas ou mais chaves estrangeiras (FK), representando, dessa forma, relações de muitos para muitos em um esquema multidimensional. Por outro lado, a tabela de dimensão está sempre relacionada a uma tabela de fatos. Sua função é conter atributos que descrevem as medidas numéricas armazenadas na tabela de fatos. Cada dimensão é definida por sua chave primária (PK), que realiza a manutenção da integridade referencial em relação à tabela de fatos à qual está associada.

A partir de tabela de fatos, relacionada com tabelas de dimensões, tem-se o que é chamado esquema estrela [9, 10]. Este esquema é apresentado na Figura 3, em que pode-se notar a tabela de fatos relacionadas a vendas diárias de um determinado produto com medidas numéricas associadas às tabelas “*Dimensão Data*” e “*Dimensão Produto*”. A vantagem do esquema multidimensional é, principalmente, a sua simplicidade, o que faz com que muitos usuários sejam capazes de entendê-lo facilmente. Além disso, existe ganho de desempenho nas consultas, pois os sistemas gerenciadores de bancos de dados irão processar estes esquemas, que são menos

complexos em relação aos esquemas normalizados, realizando menos junções, que é uma operação muito custosa.

O projeto de um *Data Warehouse*, descrito por [10] e o adotado neste trabalho, abrange os seguintes passos: (i) selecionar o processo a ser modelado; (ii) identificar as tabelas de dimensão e seus atributos; e (iii) identificar os assuntos de interesse. Após a criação das tabelas de dimensões e fatos, deve-se então carregá-las usando o processo denominado ETL (*extraction, transformation and load*). Esse processo consiste de três etapas: (i) extração dos dados dos provedores de dados; (ii) transformação dos dados, de forma a padronizar e corrigir os dados; e (iii) carregamento efetivo dos dados no *Data Warehouse*.



Figura 3. Exemplo de esquema estrela

D. Trabalhos Relacionados

Existe na literatura uma grande quantidade de trabalhos que envolvem *Data Warehouse* e a área médica. O trabalho descrito em [11], tem como foco o uso de um *Data Warehouse* composto por um conjunto de cinco esquemas multidimensionais que auxiliam três instituições hospitalares em vários processos de interesse.

O trabalho descrito em [12] utiliza *Data Warehouse* com o propósito de tornar mais eficazes os processos de informação relacionados à análise dos diagnósticos de saúde realizados pela Secretaria Municipal de Saúde da cidade de Belo Horizonte/MG, Brasil.

Por outro lado, o trabalho de [13] trata da criação de um *Data Warehouse* para a análise de dados sobre diferentes métodos de tratamento de dores. O trabalho de [14] narra a criação de um *Data Warehouse* para prover suporte ao processo de tomada de decisão do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (FMRP).

Além disso, vale destacar que recentemente, os trabalhos [17] e [18] procuram utilizar a ligação entre essas áreas para melhorar a gestão pública de recursos de saúde, enquanto que o trabalho [19] apresenta um estudo de caso em que *Data Warehouse* e modelagem multidimensional são aplicados para gerar relatórios de nutrição.

Apesar das similaridades com os trabalhos correlatos, este trabalho apresenta os seguintes diferenciais:

- Foco especificamente na doença da Anemia Falciforme, uma doença ainda pouco abordada em trabalhos nacionais e que é contextualizada dentro do sistema DSS-SCD.

- Esquemas específicos e otimizados para a doença Anemia Falciforme, que exploram características da doença.
- Exibição dos passos para a criação e o carregamento de um *Data Warehouse* para a doença Anemia Falciforme.
- Apresentação de uma ferramenta que permite a análise dos dados por um especialista da área médica.

II. MODELAGEM MULTIDIMENSIONAL PARA A ANEMIA FALCIFORME

O ambiente para a análise dos dados da Anemia Falciforme é composto por dois sistemas, denominados: DORS-SCD (*Data Organizing and Recovering System for SCD*) e DSS-SCD (*Decision Support System for SCD Data*). O sistema DORS-SCD, ilustrado na Figura 4, tem o objetivo de apoiar a extração de dados de artigos com assunto relacionado à Anemia Falciforme para então, posteriormente, povoar o banco de dados. Para que esse processo ocorra, o módulo de extração possui um dicionário com palavras-chave para auxiliar nesse processo e, além disso, apresenta um módulo de interface com o especialista para que o processo de extração seja avaliado.

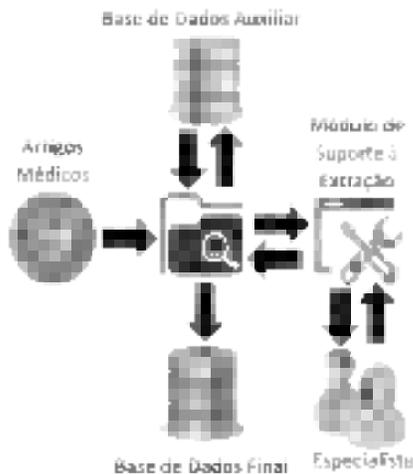


Figura 4. Sistema DORS-SCD

Por outro lado, o sistema DSS-SCD, ilustrado na Figura 5, tem a função de apoiar a análise de dados da Anemia Falciforme visando o processo de tomada de decisão. Este sistema é composto de dois módulos importantes. O primeiro baseia-se no desenvolvimento de um *Data Warehouse* para organizar e armazenar dados relativos à Anemia Falciforme (indicado por *a*). O segundo módulo é relativo à implementação de análise quantitativa nos dados por meio de ferramentas OLAP (indicado por *b*) e de mineração de dados (indicado por *c*) para identificar relacionamentos de causa e efeito. Também há no sistema a necessidade de um módulo que realize a interface com o especialista, por meio da qual o especialista pode fazer as consultas quantitativas e identificar padrões e relacionamentos sobre a doença Anemia Falciforme.

Neste trabalho, relacionado à construção do *Data Warehouse* do sistema DSS-SCD, foram propostos quatro esquemas multidimensionais, abordando três diferentes

assuntos pertinentes à Anemia Falciforme: (i) tratamento de um grupo de pacientes com Anemia Falciforme; (ii) efeitos no tratamento da doença; e (iii) publicações sobre a doença. Essa Seção é organizada da seguinte forma: a Seção *A* descreve o desenvolvimento dos esquemas multidimensionais, enquanto que a Seção *B* exemplifica o processo de povoamento do *Data Warehouse*.

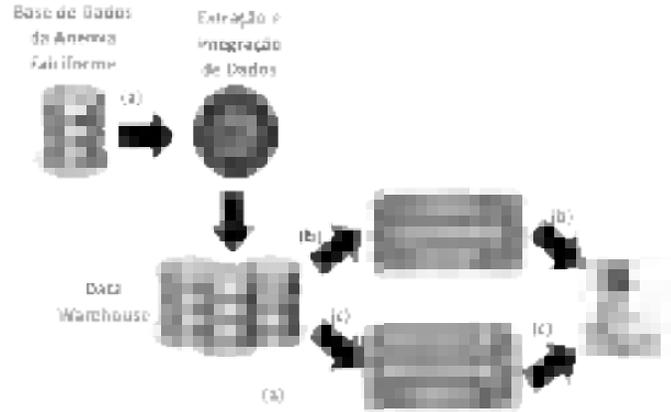


Figura 5. Sistema DSS-SCD

A. Esquemas Multidimensionais

Com relação ao assunto *tratamento de um grupo de pacientes com Anemia Falciforme*, cujo esquema multidimensional é ilustrado na Figura 6, é importante notar que o tratamento de um grupo de pacientes abordado em um artigo pode ser realizado por uma ou várias drogas ou terapias, podendo apresentar, após um tempo, efeito positivo ou negativo. Por meio da análise desse assunto, pode-se conhecer melhor os resultados obtidos nos tratamentos realizados em pacientes com Anemia Falciforme. O fato escolhido é, portanto, quantidade de pacientes. As dimensões são: grupo de pacientes, artigo, período, tratamento e efeito. Por meio do esquema multidimensional proposto, é possível responder perguntas como:

- Quantos pacientes obtiveram efeito negativo com o uso de uma determinada droga?
- Quantos pacientes faleceram durante o tratamento?
- Qual a média de pacientes tratados pelo periódico *American College of Medicine*?

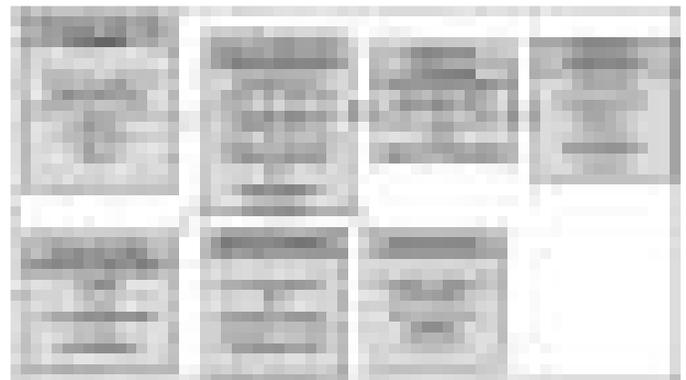


Figura 6. Esquema Tratamento de Pacientes

O segundo esquema multidimensional proposto aborda os efeitos no tratamento de um grupo de pacientes com Anemia Falciforme. Diferentemente do esquema anterior, o foco neste caso é a quantidade de efeitos positivos e negativos abordada nos artigos. Assim, a importância deste assunto está na possibilidade de entender quais tratamentos obtiveram maior sucesso (maior quantidade de efeitos positivos) e devem, portanto, ser recomendados e quais tratamentos levaram a um grande número de falhas (maior quantidade de efeitos negativos) e devem, portanto, ser evitados.

No esquema ilustrado na Figura 7, o foco escolhido são efeitos por artigo. Os fatos escolhidos são a quantidade de efeitos positivos e quantidade de efeitos negativos. As dimensões são: grupo de pacientes, artigo, período e tratamento. Por meio do esquema multidimensional proposto, é possível responder perguntas como:

- Quantos efeitos positivos foram verificados para um determinado medicamento?
- Qual medicamento possui maior quantidade de efeitos positivos?
- Quais efeitos estão associados ao tratamento de pacientes do sexo masculino?



Figura 7. Esquema Efeitos

O terceiro esquema multidimensional proposto aborda novamente o assunto *tratamento de um grupo de pacientes com Anemia Falciforme*. Porém, nesse caso o foco é a quantidade de tratamentos que são descritas em um artigo e que podem levar a um resultado positivo ou negativo. Assim, pode-se obter informações sobre o tratamento mais utilizado ou com maior índice de efeitos positivos. Esse esquema inclui as dimensões que podem ser preenchidas com dados que o projeto da Anemia Falciforme extrai atualmente dos artigos da área médica. O esquema proposto é ilustrado na Figura 8. O foco escolhido para o esquema são tratamentos relatados em artigos. O fato escolhido é a quantidade de tratamentos e as dimensões são: artigo, tratamento e efeito. Por meio desse esquema, é possível responder consultas como:

- Qual o número de tratamentos por tipo de efeito (positivo ou negativo)?
- Qual o número de tratamentos abordados por periódico?
- Qual o número de tratamentos por artigo em cada ano, por tipo de efeito (positivo ou negativo)?

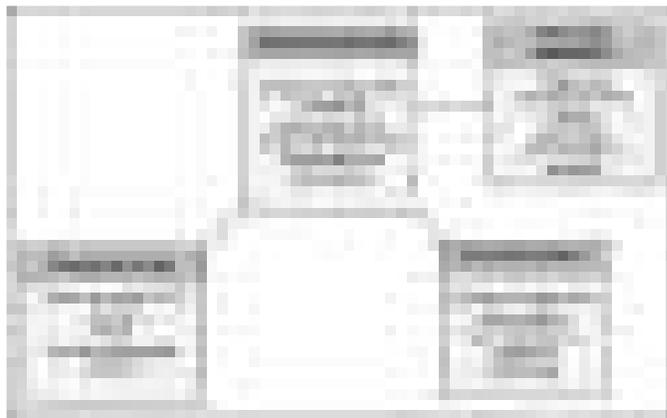


Figura 8. Esquema Tratamentos em Artigos

Por fim, o quarto esquema multidimensional, ilustrado na Figura 9, aborda um assunto diferente dos anteriores: *publicações sobre a doença Anemia Falciforme*. O foco é a quantidade de publicações relatando um tratamento que resulta em um efeito positivo ou negativo. Assim, pode-se obter informações sobre o tratamento mais utilizado em publicações da área médica ou mesmo a verificação de tendências em tratamentos utilizados em artigos recentes, como exemplo, tratamentos publicados na última década. A granularidade escolhida são conjunto de artigos que abordam um tratamento. O fato escolhido é, portanto, a quantidade de artigos e as dimensões são: artigos, tratamento e efeito. Por meio desse esquema, é possível responder perguntas como:

- Qual o número de artigos por tipo de efeito (positivo ou negativo) para cada tratamento?
- Qual o número de artigos por tipo de efeito (positivo ou negativo) para cada periódico?
- Qual o número de artigos por tipo de efeito, (positivo ou negativo) para cada periódico, em cada ano, para cada tratamento?

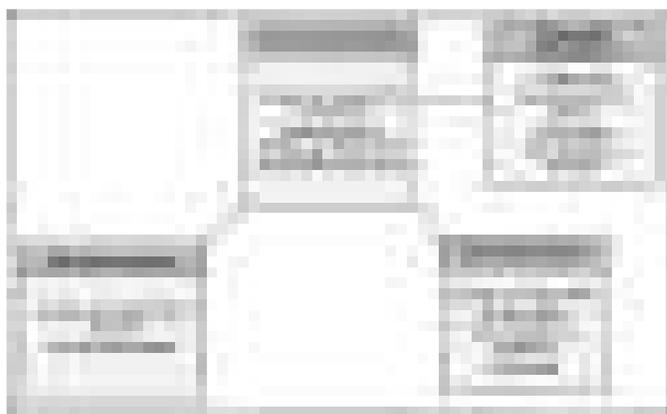


Figura 9. Esquema Publicações

Além das tabelas de dimensão particulares, os esquemas implementados possuem tabelas de dimensão em comum (por exemplo, dimensões tratamento e efeito). Essas tabelas de dimensão em comum são descritas usando-se os mesmos atributos, ou seja, em conformidade. O objetivo dessa abordagem é evitar que tabelas de dimensão distintas sejam armazenadas contendo dados iguais ou semelhantes.

Os esquemas propostos para o *Data Warehouse* da Anemia Falciforme pertencem ao ambiente DSS-SCD. Desse modo, o *Data Warehouse* foi implementado separadamente do ambiente DORS-SCD, o qual realiza o povoamento do banco de dados a partir da extração de dados de artigos da área médica. Esse banco de dados é tido como o provedor de informações para o *Data Warehouse* implementado. No entanto, como o projeto sobre a Anemia Falciforme ainda está em andamento, apenas dados reais sobre efeitos positivos, negativos e artigos puderam ser utilizados.

B. Carregamento do Data Warehouse

Para o carregamento do *Data Warehouse*, foi desenvolvido um conjunto de procedimentos armazenados (i.e., *stored procedures*) implementados na linguagem *PL/PGSQL* do SGBD *PostgreSQL* [15], os quais foram responsáveis pela extração dos dados das tabelas do banco de dados da Anemia Falciforme, tradução para o formato dos dados do *Data Warehouse* por meio de procedimento armazenado, e, finalmente, carga das tabelas do *Data Warehouse*. A Figura 10 ilustra a implementação realizada.



Figura 10. Processo de carga das tabelas do *Data Warehouse*

III. O APLICATIVO *SCD DATA WAREHOUSE*

Nesta seção é descrito o aplicativo *SCD Data Warehouse*, desenvolvido na linguagem de programação Java. O aplicativo tem como objetivo de realizar a comunicação com o *Data Warehouse* desenvolvido, além de permitir ao especialista da área médica acessar análises estatísticas dos dados por meio de uma interface gráfica. A Figura 11 ilustra a interface inicial da ferramenta, que em sua parte superior apresenta três abas:

- Início: aba que trata da aba inicial da ferramenta;
- Cadastro: aba que permite ao usuário inserir dados no *Data Warehouse*;
- Consultas DW: aba que permite ao usuário consultar dados contidos no *Data Warehouse*.

A Figura 12 exibe a aba da ferramenta que é responsável pelo cadastro de informações no *Data Warehouse*. Assim, o cadastro de dados no *Data Warehouse* pode ser feito de duas formas: automaticamente por meio do módulo de extração e integração de dados do sistema DSS-SCD (conforme mostrado na Seção 3) e ou por meio da ferramenta *SCD Data Warehouse*. Nessa aba, existe a possibilidade de cadastrar informações em quatro tabelas diferentes do *Data Warehouse*: (i) artigo; (ii) paciente; (iii) tratamento; e (iv) efeito. Dessa forma, as tabelas que compõem os esquemas estrela, apresentados na Seção 3, podem ser preenchidas com dados digitados manualmente pelo especialista.

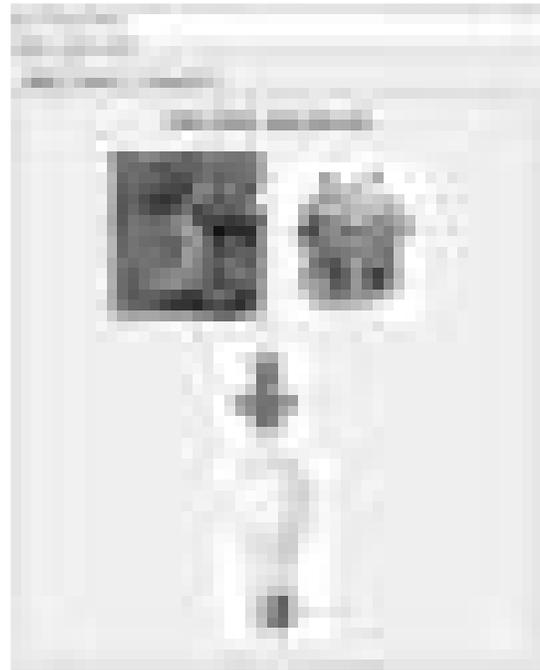


Figura 11. Aplicativo *SCD Data Warehouse*



Figura 12. *SCD Data Warehouse*: Aba de Cadastro

Além disso, para o tema artigos, a seguinte consulta foi implementada:

- Qual o número de óbitos relatados por artigo?

Por fim, a Figura 14 ilustra, por meio de uma tabela, um exemplo de resultado para a consulta “Qual o número de óbitos relatados por artigo?”.

Dessa forma, é possível mostrar diferentes análises sobre artigos da doença Anemia Falciforme. Se houver a necessidade de acrescentar à ferramenta novas análises sobre o *Data Warehouse* desenvolvido, basta acrescentar um novo módulo de consulta à ferramenta de acordo com a necessidade do especialista. Vale destacar que os dados utilizados na ferramenta *SCD Data Warehouse* são provenientes da ferramenta descrita em [6], *Medical Search*, que faz parte deste projeto e pertence ao ambiente *DORS-SCD*. Essa ferramenta se trata de um sistema *web* para extração de informação automática de artigos da área médica.



Figura 13. SCD Data Warehouse: Aba de Consultas



Figura 14. SCD Data Warehouse: Análise do Data Warehouse

IV. CONCLUSÃO

Este trabalho teve como objetivo descrever a proposta de um conjunto de esquemas multidimensionais para a doença Anemia Falciforme. As principais atividades detalhadas foram:

- Criação de quatro esquemas multidimensionais para a doença Anemia Falciforme.
- Implementação desses esquemas multidimensionais, gerando o *Data Warehouse* da Anemia Falciforme.
- Desenvolvimento de uma ferramenta, denominada *SCD Data Warehouse*, que permite ao especialista

médico realizar consultas no *Data Warehouse*.

A principal contribuição deste trabalho é, dado o domínio dos dados armazenados, o auxílio à área médica no entendimento de questões relacionadas a tratamentos de pacientes que possuem a Anemia Falciforme. Dessa forma, são fornecidos meios para analisar estatisticamente os melhores tratamentos obtidos e que são descritos na literatura científica. Além disso, os esquemas multidimensionais elaborados procuram responder questões que envolvem muitas publicações sobre a doença Anemia Falciforme e que seriam difíceis de serem respondidas por um especialista sem o auxílio de um sistema computadorizado. Por fim, este trabalho também descreveu um aplicativo, chamado *SCD Data Warehouse*, cujo objetivo é permitir a especialistas da área médica um meio de acessar os dados armazenados no *Data Warehouse* por meio de consultas e realizar análises estatísticas.

Como trabalhos futuros, serão realizadas as seguintes tarefas:

- O preenchimento do *Data Warehouse* com uma quantidade maior de dados por meio da extração de termos de artigos relacionados à doença Anemia Falciforme.
- Generalização dos esquemas propostos, adaptando-os a outras doenças.
- Desenvolvimento de mais consultas disponíveis para o especialista por meio da ferramenta *SCD Data Warehouse*.

REFERÊNCIAS

- [1] Charache, S., et al. Effect of Hydroxyurea on the Frequency of Painful Crises in Sickle Cell Anemia. *The New England Periódico of Medicine*, v. 332, n. 20, p. 1317-1322, 1995.
- [2] Covas, D.T., et al. Effects of Hydroxyurea on the Membrane of Erythrocytes and Platelets in Sickle Cell Anemia. *Haematologica*, v. 89, n.3, 2004.
- [3] Gulbis, B., et al. Hydroxyurea for Sickle Cell Disease in Children and for Prevention of Cerebrovascular Events: the Belgian Experience. *Clinical Observations, Interventions, and Therapeutic Trials*, v. 105, n. 7, p. 2685-2690, 2005.
- [4] Hankins, J.S., et al. Long-term Hydroxyurea Therapy for Infants with Sickle Cell Anemia: the HUSOFT Extension Study. *Clinical Observations, Interventions, and Therapeutic Trials*, v.106, n.7, p. 2269-2275, 2005.
- [5] Rodrigues, C. M. Hemopatias na gestação. 2008. Disponível em: <<http://www.paulomargotto.com.br/documentos/Hemopatias%20na%20Gesta%C3%A7%C3%A3o.ppt>>. Acesso em: 03 out. 2015.
- [6] CAROSIA, A. E. O.. *MedicalSearch: Um Sistema Web para Extração de Informação de Artigos da Área Médica*. *Revista Junior de Iniciação Científica em Ciências Exatas e Engenharia*, v. 1, p. 16, 2016.
- [7] Chaudhuri, S.; DAYAL, U. An Overview of Data Warehousing and OLAP Technology. *SIGMOD Record*, v.26, n.1, p.65-74, 1997.
- [8] Ciferri, Cristina Dutra de Aguiar. "Distribuição dos Dados em Ambientes de Data Warehousing: O Sistema WebD²W e Algoritmos voltados à Fragmentação Horizontal dos Dados." Tese de Doutorado. Recife, PE, Brasil, 2002.

- [9] Inmon, W H. Building the *Data Warehouse*. 2nd edition. John Wiley & Sons, 1996.
- [10] Kimball, Ralph, e Margy Ross. The *Data Warehouse Toolkit*. Vol. I. Editora Campus, 2002.
- [11] Ciferri, C. D. A. ; Ciferri ; Ogata, Reinaldo Jiunji ; Lima, André Moraes Paula ; TRAINA, Agma Juci Machado . Data Warehousing na Saúde: Melhorando a Tomada de Decisão Médico-Analítica. In: XXXII Conferencia Latinoamericana de Informática, 2006, Santiago.
- [12] Miranda, R. M. Utilização do Modelo Dimensional para Diagnóstico de Saúde no Município de Belo Horizonte. Monografia de Especialização: Prodebel/IRT-PUC/MG, Brasil, (2000).
- [13] Ferreira, Carlos, et al. "O data mining na compreensão do fenômeno da dor: uma proposta de aplicação." (2006).
- [14] Góes, Wilson Moraes. Análise dos sistemas de informação do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto: rumo ao sistema de informação gerencial. Diss. Universidade de São Paulo, 2007.
- [15] Douglas, Korry, and Susan Douglas. PostgreSQL: a comprehensive guide to building, programming, and administering PostgreSQL databases. SAMS publishing, 2003.
- [16] Elmasri, Ramez, e Shamkant B. Navathe. Fundamentals os Database Systems. 2004.
- [17] de Campos Souza, Paulo Vitor, and José Araújo Damasceno. "UTILIZAÇÃO DE TÉCNICAS DE BUSINESS INTELLIGENCE NA BASE DE CONSULTAS ESPECIALIZADAS DO SISTEMA ÚNICO DE SAÚDE." ÚNICA Cadernos Acadêmicos 2 (2016).
- [18] da Silva, Joaquim Cardoso. UTILIZAÇÃO DAS FERRAMENTAS EXTRATOR DE DADOS E DATA WAREHOUSE NAS UNIDADES DE RECURSOS HUMANOS DA FUNDAÇÃO NACIONAL DE SAÚDE. Diss. Universidade Católica de Brasília, 2014.
- [19] Ribeiro, Cleidiane Silva, Silas César da Silva Ribeiro, and Moisés Henrique Ramos Pereira. "MODELAGEM DIMENSIONAL E IMPLEMENTAÇÃO DE DATA WAREHOUSE PARA GERAÇÃO DE RELATÓRIOS SOBRE DADOS DE TERAPIA NUTRICIONAL: ESTUDO DE CASO." e-xacta 6.1 (2013): 131-144.