Processo de engenharia reversa em banco de dados: uma comparação escalonada entre o processo manual e o automatizado

Leonardo Teixeira da Silva, Adriana Paula Zamin Scherer

Faculdade Dom Bosco – Porto Alegre – RS – Brasil

leonardotdasilva@yahoo.com.br, adriana.faculdade@dombosco.net

Abstract. This article aims to demonstrate the results obtained in the comparison between the scaled processes, manual and automated reverse engineering database. It began with the application of reverse engineering in ten (10) tables, and after new tables were added till the total of 31 (thirty one) tables. We used metrics defined by Heuser (2001) for verification of conceptual models, beyond the time that each model took to be generated and the correction time for each. The manual process is automated and you need fast. Thus, it is evident that it is best to use the hybrid process, which performs the automated process and corrections are made manually.

Keywords: Database, reverse engineering, tables.

Resumo. Este artigo tem o objetivo de demonstrar os resultados obtidos na comparação escalonada entre os processos, manual e automatizado, de engenharia reversa em banco de dados. Iniciou-se com a aplicação da engenharia reversa em 10 (dez) tabelas, e após foram acrescidas novas tabelas até se atingir o total de 31 (trinta e uma) tabelas. Foram utilizadas métricas definidas por Heuser (2001) para verificação de modelos conceituais, além do tempo que cada modelo levou para ser gerado e o tempo de correção para cada um. O processo manual é preciso e o automatizado é rápido. Desta forma, fica evidenciado que o melhor é utilizar o processo hibrido, onde se realiza o processo automatizado e as correções são feitas manualmente.

Palavras-chave: Banco de dados, engenharia reversa, tabelas.

1. Introdução

Atualmente é imprescindível a utilização de banco de dados nas organizações, para que se atinja o sucesso organizacional no mercado que a cada dia apresenta-se mais competitivo. Assim, a sua construção deve ser realizada através de uma modelagem de dados detalhada, estruturada e clara, para facilitar a implementação e permitir o entendimento do funcionamento do banco de dados mesmo para quem não tenha participado do seu projeto de implementação.

Alterações ao longo do tempo são realizadas nos banco de dados, mas algumas vezes elas não são documentadas pelo Administrador de Dados (AD), e também há casos de sistemas, principalmente os mais antigos, que são implementados sem a existência do modelo conceitual. A ausência desta documentação gera dificuldade de entendimento de como os dados estão organizados e relacionam-se. Para resolver este problema deve-se realizar o processo de engenharia reversa na base de dados. Neste processo, inicia-se no modelo lógico e termina-se no modelo conceitual, conforme a definição de Heuser (2001).

O objetivo deste artigo é relatar o processo utilizado para a realização de engenharia reversa em banco de dados por meio de análise das características dos processos manuais e automatizados. A investigação ocorreu de maneira comparativa e escalonada entre os dois processos, para assim demonstrar em qual momento é indicado utilizar um processo ou outro.

Este artigo está dividido em cinco seções. A primeira é esta introdução, onde estão descritos conceitos gerais sobre banco de dados e engenharia reversa em banco de dados. Na seção de número dois, está destacado o referencial teórico, onde há o aprofundamento das explicações sobre engenharia reversa em banco de dados e ferramentas gratuitas para a realização da engenharia reversa. Na seção de número três, está detalhada a aplicação do grupo de tabelas para a realização do processo manual e automatizado, para que fosse aplicada a engenharia reversa de forma escalonada. Na seção número quatro há o detalhamento dos resultados obtidos através da aplicação das métricas e por fim, na quinta seção conclui-se o artigo.

2. Referêncial Teórico

De acordo com Elmasri (2005), o processo de engenharia reversa em banco de dados é considerado uma técnica que através de um banco de dados já implementado é gerado o seu modelo de forma conceitual.

Heuser (2009) define que a engenharia reversa em banco de dados é composta por um processo definido por quatro etapas, que são:

- a) Identificação da construção ER correspondente a cada tabela;
- b) Mapeamento de relacionamentos 1:n e 1:1;
- c) Mapeamento de atributos;
- d) Definição de identificadores de entidades e relacionamentos.

A etapa de identificação da construção ER correspondente a cada tabela é dividida em três regras. São elas:

- 1. Chave primária composta por mais de uma chave estrangeira: Ocorre quando uma chave primária é composta por mais de uma chave estrangeira e assim caracteriza o relacionamento n:n;
- 2. Chave primária completa forma uma chave estrangeira: Ocorre quando todas as colunas de chave primária formam uma única chave estrangeira. Dessa forma acontece a especialização da entidade;

3. Demais casos: Ocorre quando as regras 1 e 2 não necessitam ser implementadas e dessa forma a tabela é considerada uma entidade.

A etapa de mapeamento de relacionamentos 1:n e 1:1 realiza o tratamento das chaves estrangeiras, que não se enquadram nas regras 1 e 2 da etapa de identificação da construção ER correspondente a cada tabela. O tratamento ocorre da seguinte forma:

- As que possuem colunas que não fazem parte da chave primária;
- ➤ Os que fazem parte de uma chave primária e esta chave primária não possua outras chaves estrangeiras e possua colunas que não são chaves estrangeiras.

A etapa de mapeamento de atributos trata a coluna que não seja chave estrangeira e assim é definido um atributo na entidade e/ou no relacionamento na tabela dona da coluna. É dividida nas seguintes regras:

- > Coluna da chave primária que não é chave estrangeira: Será um atributo identificador da entidade e/ou relacionamento a coluna que fizer parte da chave primária e que não seja chave estrangeira;
- > Coluna da chave primária que é chave estrangeira: Será considerado um atributo identificador externo da entidade a coluna que fizer parte da chave primária e que seja também uma chave estrangeira.

A Figura 1 demonstra o processo de engenharia reversa em banco de dados.

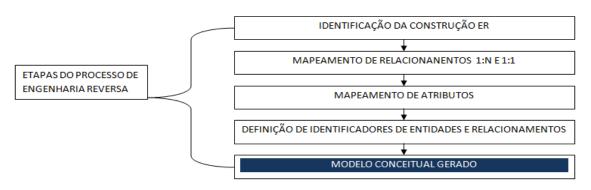


Figura 1. Processo de engenharia reversa em banco de dados

A engenharia reversa em banco de dados proporciona uma série de benefícios. Brognoli (2008) define os seguintes:

- a) Facilita a criação/manutenção da documentação do banco de dados;
- b) Facilita a análise do sistema;
- c) Facilita a manutenção do banco de dados;
- d) Facilita o entendimento do funcionamento do banco de dados.

Já Takaoka (1998) destaca os seguintes benefícios:

a) Recupera informações do banco de dados que se perdeu ao longo do tempo devido à falta de documentação;

- b) Quando ocorrer a necessidade de trocar o Sistema Gerenciador de Banco de Dados (SGBD);
- c) Mudança de arquitetura do banco de dados centralizada para cliente/servidor;
- d) Desenvolvimento de interface mais inteligente aos usuários do banco de dados;
- e) Integração de banco de dados.

E para Vieira (2003), a engenharia reversa traz os seguintes benefícios:

- a) Sistema redocumentado;
- b) Conversão de dados;
- c) Reestruturação do banco de dados;
- d) Manutenção de banco de dados legado;
- e) Extensão de aplicação legada.

Além da possibilidade de realização da engenharia reversa de maneira manual, o mercado tecnológico disponibiliza ferramentas proprietárias e gratuitas que realizam este processo. As proprietárias são pagas e, normalmente, possuem mais recursos técnicos para a realização da engenharia reversa. As ferramentas gratuitas podem oferecer menos recursos técnicos em comparativo com as ferramentas pagas, mas não há custos de aquisição. O presente estudo aborda apenas as ferramentas gratuitas.

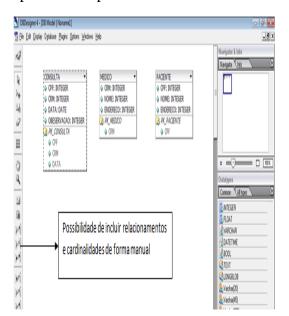
Foram analisadas 02 (duas) ferramentas *freeware* que realizam engenharia reversa em banco de dados e para isto criou-se um cenário para um experimento inicial, composto por três tabelas no SGBD Oracle, a fim de identificar a mais adequada para utilização posterior. Para as avaliações das ferramentas, foram utilizados os seguintes critérios:

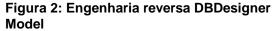
- 1. Tipo de conexão *Open Database Connectivity* (ODBC): Nesse critério, foi avaliado se a ferramenta permite conexão via ODBC, pois para conectar alguns bancos de dados mais antigos como, por exemplo, o Ingres, geralmente é possível somente através de ODBC;
- 2. Escolha da tabela: Nesse critério, foi avaliada se a ferramenta permite selecionar tabela(s) ou se é necessário realizar a engenharia reversa de toda a base de dados. A ferramenta que permite escolher as tabelas oportuniza que se possa realizar a engenharia reversa em apenas algumas tabelas de uma base de dados existente;
- 3. Relacionamentos: Nesse critério, foi avaliado se a ferramenta apresentou relacionamento entre as tabelas na engenharia reversa, pois essa característica da ferramenta facilita o entendimento geral do modelo conceitual gerado;
- 4. Cardinalidades: Nesse critério, foi avaliado se a ferramenta apresentou cardinalidades no(s) relacionamento(s) entre tabelas na engenharia reversa,

pois essa característica da ferramenta facilita o entendimento de forma detalhada do modelo conceitual gerado.

A primeira ferramenta analisada foi a DBDesigner 4. Sua fabricante é a fabForce (2011). Possui conexões nativas para diversos bancos de dados comerciais, tais como, Oracle, MySQL e SQLite. Além disso, ela também permite conexões através de *drivers* ODBC. Com esta ferramenta, é possível escolher a(s) tabela(s) que se deseja realizar a engenharia reversa. Ao obter-se o modelo de forma lógica ele é representado sem os relacionamentos e as cardinalidades. Mas, é possível acrescentar essas informações de forma manual.

A segunda ferramenta analisada foi a Open ModelSphere 3.1 e sua fabricante é a empresa Grandite (2011). A ferramenta possui conexões nativas para diversos bancos de dados comerciais, tais como, PostgreSQL, MySQL, e DB2. Além destas conexões, é possível também conectar-se aos bancos de dados utilizando conexões ODBC. É possível escolher a(s) tabela(s) que se deseja realizar a engenharia reversa. A ferramenta por padrão gera a engenharia reversa do modelo lógico, mas é possível converter para o modelo conceitual, ao clicar com o botão direito do mouse em *Relational Data Model – Convert to a Conceptual Model*. Após a conversão, foi possível verificar que ela foi realizada para o modelo conceitual de forma híbrida, isto é, o modelo apresenta característica do modelo lógico, como por exemplo, o tipo de dado de cada campo das tabelas. Já como característica do modelo conceitual pode-se destacar a representação dos relacionamentos com suas cardinalidades. A Figura 2 demonstra a engenharia reversa gerada na DBDesigner 4 e a Figura 3 demonstra a engenharia reversa gerada na Open ModelSphere 3.1.





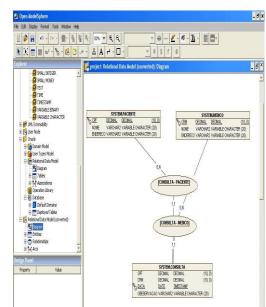


Figura 3: Engenharia reversa Open

A partir da análise dos critérios anteriormente definidos, foi possível avaliar de forma criteriosa cada uma delas e assim escolher a Open ModelSphere 3.1 para a

aplicação da engenharia reversa, pois ela satisfaz todos os critérios de avaliações propostos. A Tabela 1 demonstra o resumo das avaliações das ferramentas.

Tabela 1: Avaliações das ferramentas

Ferramenta	Tipo de conexão ODBC	Escolha da tabela	Relacionamento	Cardinalidades
DBDesigner	Sim	Sim	Não	Não
Open ModelSphere	Sim	Sim	Sim	Sim

3. Aplicação da engenharia reversa

Foi realizado um comparativo entre o processo manual e o automatizado em uma base de dados que contém 31 (trinta e uma) tabelas e estas foram divididas em 05 (cinco) grupos onde, o primeiro grupo possui 10 (dez) tabelas, o segundo, o quarto e o quinto grupo com 05 (cinco) tabelas e o terceiro com 06 (seis) tabelas. A comparação foi realizada de maneira escalonada, a partir do primeiro grupo e desta forma o escalonamento foi aumentando até atingir o total de 31 (trinta e uma) tabelas, para assim demonstrar em qual cenário é melhor utilizar um processo ou outro.

A engenharia reversa iniciou no primeiro grupo de forma manual. A aplicação da engenharia reversa manual foi realizada de acordo com o processo explicado na seção 2, Figura 1. Após, nestas dez tabelas do grupo inicial, foi efetuado o processo de engenharia reversa de maneira automatizada. Uma vez que, o processo manual e escalonado para o primeiro grupo tenha sido finalizado, foram acrescidas ao escopo inicial as tabelas pertencentes ao segundo grupo. Após a finalização do processo manual e escalonado para o segundo grupo foi acrescido, ao escopo existente, as tabelas pertencentes ao terceiro grupo. Ao término do processo manual e escalonado para o terceiro grupo, foram acrescidas as tabelas pertencentes ao quarto grupo. Ao se concluir o processo manual e escalonado para o quarto grupo, acrescentou-se as tabelas pertencentes ao quinto grupo.

Dentro do processo de validação, a variável conhecimento da estrutura do banco poderia gerar resultados diferentes, uma vez que com este conhecimento a geração manual seria facilitada. Desta forma, tornou-se necessário que as implementações fossem realizadas por um indivíduo que conhecesse apenas o negócio da instituição, mas não a estrutura das tabelas. Para a validação do trabalho, foram utilizadas cinco métricas, sendo que dessas, as três primeiras são definidas por Heuser (2001). São elas:

- a) Completude da modelagem conceitual gerada: Esta métrica teve o objetivo de avaliar o quão completa a modelagem foi gerada. Foi subdivida em: relacionamentos entre tabelas, cardinalidades, atributos e identificadores de entidades e relacionamentos;
- b) Tratamento adequado para conversão de tabelas: Esta métrica teve o objetivo de realizar a identificação de qual construção ER correspondeu a uma tabela através da composição de sua chave primária;

- c) Correto: Utilizou-se esta métrica com o objetivo de evitar erros na modelagem de dados, e foi subdividida em: sintático, para analisar o respeito às regras do modelo ER e semântico, para refletir a realidade de forma consistente;
- d) Tempo de geração: Esta métrica teve o objetivo de avaliar o tempo em que cada um dos processos levou para ser executado. Foi validada em horas e minutos;
- e) Tempo de correção: Esta métrica teve o objetivo de avaliar o tempo de correção das informações das modelagens geradas. Foi validada em horas e minutos.

Ao final da engenharia reversa realizada em cada grupo de tabelas, ocorreram as validações, primeiramente foram aplicadas as validações sobre o processo manual e após no processo automatizado, para que pudesse ser realizada a comparação entre os dois processos. O processo de validação se repetiu nos cinco grupos de tabelas. A Figura 4 demonstra o processo de validação da engenharia reversa.



Figura 4: Processo de validação da engenharia reversa

4. Resultados

A primeira métrica avaliada, que é a completude da modelagem conceitual gerada, cujo objetivo é a avaliação do quão completa a modelagem conceitual gerada está, teve os mesmos resultados obtidos no processo manual e no automatizado, ou seja, foram representados os relacionamentos entre as tabelas, cardinalidades, atributos e identificadores de entidades. Desta forma, é possível afirmar que a ferramenta se comportou de maneira adequada e para esta métrica os processos, manual e automatizado, geram resultados corretos e que, para o escopo analisado, a quantidade de tabelas não influenciou o resultado.

A métrica de tratamento adequado para conversão de tabelas, que tem como objetivo realizar a identificação de qual construção ER corresponderá uma tabela, obteve resultados diferentes. No processo de engenharia reversa manual, foi identificado que uma tabela do grupo quatro era originária de um relacionamento n:n, assim como outro do grupo cinco. Além disso, as duas deveriam ser transformadas em entidades associativas, pois o relacionamento deveria relacionar-se com outras entidades. Na engenharia reversa gerada pela ferramenta, observou-se que as tabelas não se

transformaram em entidades associativas e este fato demonstra que a ferramenta reconhece apenas os componentes básicos do modelo ER.

A métrica de correto tem o objetivo de avaliar a presença de erros dos tipos sintáticos e semânticos. Os erros sintáticos representam desrespeito às regras do modelo ER, já os erros semânticos são aqueles que não refletem a realidade de forma adequada. No processo de engenharia reversa manual, não foram apresentados erros sintáticos e nem semânticos. Na engenharia reversa automatizada foram apresentados erros, mas somente do tipo semântico. Os principais erros apresentados foram de relacionamentos com cardinalidades invertidas e duplicidade de relacionamentos. O grupo 01 (um) de tabelas apresentou 13 (treze) erros de cardinalidades invertidas e 02 (dois) erros de relacionamentos duplicados.

O grupo 02 (dois) herdou os erros do grupo (01) um e apresentou novos erros, foram 12 (doze) erros de cardinalidades invertidas e 02 (dois) de relacionamentos duplicados. O grupo 03 (três) recebeu todos os erros dos grupos anteriores e apresentou novos erros, foram 09 (nove) erros de cardinalidades invertidas e 03 (três) relacionamentos duplicados. O grupo 04 (quatro) herdou todos os erros dos grupos anteriores e apresentou novos erros, foram 17 (dezessete) erros de cardinalidades invertidas e 04 (quatro) de relacionamentos duplicados.

O último grupo recebeu os erros dos grupos anteriores e apresentou erros novos, foram 11 (onze) erros de cardinalidades invertidas e 03 (três) de relacionamentos duplicados. Como não surgiram erros sintáticos, nos cinco grupos, tanto na engenharia reversa manual, como na automatizada, é possível afirmar que a ferramenta se comportou de maneira adequada e para esta métrica os processos, manual e automatizado, geram resultados corretos e que, para o escopo analisado, a quantidade de tabelas não influencia o resultado. Já acerca dos erros semânticos, pode-se observar que na engenharia reversa manual, nos cinco grupos de tabelas, não foi identificado nenhum erro.

No entanto, na engenharia reversa automatizada eles ocorreram nos cinco grupos de tabelas e foi possível observar que em todas as vezes que a base de dados foi aumentada, mais erros apareceram. Essa situação deve-se pelo fato de que em cada grupo de tabela, quando concluída sua engenharia reversa além de surgirem erros do próprio grupo foram herdados os erros dos grupos anteriores o que ao final do processo aumentou consideravelmente a quantidade de erros.

A métrica tempo de geração, que tem o objetivo de avaliar o tempo em que cada um dos processos levou para ser executado, obteve resultados diferentes, entre os grupos e os processos. No primeiro grupo, o processo manual teve a duração de 1h e 50min e o automático 05 minutos. No segundo grupo, o processo manual teve a duração de 2h e 40min e o automático 07 minutos. No terceiro grupo, o processo manual teve a duração de 3h e 45min e o automático 09 minutos. No quarto grupo, o processo manual teve a duração de 4h e 22min e o automático 12 minutos.

No grupo cinco, o processo manual teve a duração de 5h e 04min e o automático 18 minutos. Ao comparar-se a engenharia reversa manual com a automatizada dentro dos grupos, observou-se que manualmente o tempo de geração é sempre muito maior e aumenta consideravelmente a medida que aumenta a quantidade de tabelas. Observou-se

também que a diferença de tempo de geração de um grupo manual para o outro grupo também manual é bem maior do que a diferença ocorrida de um grupo automatizado para o outro grupo automatizado. Esta situação é plenamente explicável, pois ferramentas de maneira geral tendem a ser mais velozes na geração de resultados em comparação com a velocidade apresentada pelos seres humanos.

A métrica tempo de correção, que tem o objetivo de avaliar o tempo de correção das informações das modelagens geradas com erro, somente foi coletada nos modelos gerados pelo processo automatizado, pois no modelo manual não foram apresentados erros. No grupo 01 (um) o processo de correção teve a duração de 25 minutos. No grupo 02 (dois) a correção ocorreu em 30 minutos. No grupo 03 (três) a correção foi realizada em 38 minutos. No grupo 04 (quatro) a correção foi realizada em 51 minutos. Já no grupo 05 (cinco) a correção ocorreu em 58 minutos. Como apenas as modelagens automatizadas tiveram que ser corrigidas, os tempos de gerações não foram muito elevados e a diferença de tempo de um grupo para o outro também não apresentou grande distanciamento.

Ficou demonstrado através dos tempos coletados que é melhor realizar a engenharia reversa de forma automatizada quando se deseja maior velocidade na geração, no entanto o resultado pode conter erros. Já quando o objetivo é a precisão o mais indicado é a engenharia reversa de forma manual, embora o tempo de geração seja elevado. No entanto, essa consideração é válida desde que o processo explicado na Figura 1 seja seguido corretamente. Pode-se exemplificar essa situação na engenharia reversa gerada no primeiro grupo, onde foram utilizadas apenas dez tabelas. O tempo de geração manual foi de 1h e 50min, mas gerou um modelo preciso, pois seguiu o processo de engenharia reversa explicada na seção 2. Já o tempo total de geração da engenharia reversa de maneira automatizada foi de 05 minutos. No entanto, se somando a este tempo de geração automática, o de correção, o tempo total foi de apenas 35 minutos, portanto, menor do que o tempo necessário para o processo manual.

Com esta análise, surge a oportunidade de gerar a engenharia reversa através de um processo hibrido. Neste caso, inicialmente a engenharia reversa é realizada de maneira automatizada e após, de forma manual, os erros gerados no processo automatizado são corrigidos. O processo híbrido oportuniza a construção do modelo de forma correta e rápida e assim demonstra ser uma excelente solução a ser aplicada.

5. Conclusão

Os bancos de dados fazem parte da realidade das organizações, mas muitas vezes não é possível entender como os dados estão relacionados dentro deles. Um dos motivos é que alguns sistemas são implementados sem a criação do modelo conceitual. Outro motivo é que com o passar do tempo, muitas alterações são realizadas nos bancos de dados, mas algumas vezes elas não são documentadas. Com a aplicação da engenharia reversa em banco de dados torna-se possível resolver essas dificuldades. Ela pode ser realizada de forma manual ou automatizada, neste caso, com a utilização de uma ferramenta proprietária ou livre para tal.

O objetivo do artigo foi analisar através da comparação escalonada qual processo de engenharia reversa, manual ou automatizado, é mais vantajoso. Através dos resultados obtidos na aplicação da engenharia reversa, é possível afirmar que, para o

escopo definido, quando a organização deseja apenas precisão na engenharia reversa gerada, o indicado é realizá-la de maneira manual, pois não apresenta erros por seguir um processo definido, porém haverá um acréscimo considerável de tempo de geração conforme a quantidade de tabelas é aumentada.

Já para uma organização que se preocupa apenas com a velocidade de geração da engenharia reversa, de acordo com o escopo analisado, indica-se a utilização de uma ferramenta, pois a geração ocorre de forma mais rápida do que manualmente, no entanto apresentará erros e os mesmos necessariamente terão que ser corrigidos. Com a correção, o tempo de geração total do processo automatizado aumentará, mas ainda assim será menor que o tempo manual, mostrando-se como uma boa alterativa. Assim, pode-se afirmar que é possível utilizar o processo de forma hibrida, onde se realiza a engenharia reversa de maneira automatizada e as correções necessárias são realizadas de forma manual. Esse processo demonstra ser o mais interessante, pois une a velocidade da engenharia reversa automatizada com a precisão da manual.

Como ocorreram muitos erros semânticos na ferramenta analisada, abre-se a possibilidade de em trabalhos futuros, analisar a aplicabilidade da engenharia reversa automatizada com a utilização de outras ferramentas livres disponíveis no mercado e assim analisar se será possível obter resultados diferentes.

Referencias

- BROGNOLI, Samuel Nicolau; CUNHA, Daniel Pezzi da; GIORDANI, Fabrício *Sistema de apoio à engenharia reversa de banco dados relacionais por meio da extração de metadados.*Disponível

 em: http://infociencia.info/obra_detalhes_ok.php?codigo=486&tipo=c%20Brognoli>.

 Acesso em: 22 de setembro de 2011.
- *DBDESIGNER*. Disponível em < http://fabforce.net/dbdesigner4/>. Acesso em: 05 de outubro de 2011.
- ELMASRI, Ramez; NAVATHE, Shamkant B. *Sistemas de banco de dados*. 4.ed.São Paulo. Pearson Addison Wesley, 2005.
- GRANDITE. Disponível em:<<u>www.modelsphere.com/interfaces_matrix.html</u>>. Acesso em: 05 de outubro de 2011.
- HEUSER, Carlos Alberto. *Projeto de banco de dados*. 4.ed.Porto Alegre, Sagra Luzzato, 2001.
- HEUSER, Carlos Alberto. *Projeto de banco de dados*. 6.ed.Porto Alegre, Bookman, 2009.
- TAKAOKA, Hiroo. *Engenharia reversa em banco de dados relacionais*. Disponível em http://www.ead.fea.usp.br/semead/3semead/pdf/MQI/Art040.PDF>. Acesso em: 22 de setembro de 2011.
- VIEIRA, André Accioly. *Ontoextract: Uma ferramenta para extração de ontologias a partir de banco de dados relacionais*. Disponível em: . Acesso em: 03 de outubro de 2011.