



Universidade Federal de Itajubá

SIN

BI Aplicações

2018009073 - Ivan Leoni Vilas Boas

Trabalho Individual I

Baseado em: *Coleta de Requisitos e Modelagem de Dados para Data Warehouse:
um Estudo de Caso utilizando Técnicas de Aquisição de Conhecimento*
de LUÍS CLÁUDIO CHAVES ZIULKOSKI

UNIFEI, Itajubá Junho de 2021.

RESUMO

Sobre a ótica de realizar de forma mais assertiva a implantação e, anteriormente a elaboração de um projeto DW, a fim de obter a vantagem de apoiar o processo de tomada de decisão organizacional sobre o negócio é importante considerar o contexto sobre a perspectiva da engenharia de conhecimento com a utilização de técnicas específicas de aquisição de conhecimento para a devida realização da coleta de requisitos e as consequentes definições do projeto. Através de um estudo de caso a UFRGS conseguiu obter satisfatoriamente a coleta e realizar, via análise, as definições de elaboração de um modelo dimensional de dados do DW a ser implantado. Por meio da avaliação de técnicas de entrevistas e de classificação de termos a aquisição do conhecimento, portanto, possibilitou compreender melhor todo o domínio a cerca da produção intelectual, e assim definir o melhor modelo de dados a ser utilizado e, ainda, a estabelecer de forma otimizada a estrutura de dimensões envolvidas no modelo dimensional do novo DW.

Palavras-chave: Tomada de decisão, Negócio, Engenharia de Conhecimento, Aquisição do Conhecimento, Coleta de Requisitos, Projeto DW, Modelo Dimensional, Técnica de Entrevista, Classificação de Termos.

ABSTRACT

Regarding the perspective of carrying out the implementation in a more assertive way and, previously, the elaboration of a DW project, in order to obtain the advantage of supporting the organizational decision-making process about the business, it is important to consider the context from the perspective of the engineering of knowledge with the use of specific knowledge acquisition techniques for the proper accomplishment of the requirements collection and the consequent project definitions. Through a case study, UFRGS was able to obtain the collection satisfactorily and carry out, via analysis, the definitions of the elaboration of a dimensional model of DW data to be implemented. Through the evaluation of interview techniques and the classification of terms, the acquisitions of knowledge, therefore, made it possible to better understand the entire domain around intellectual production, and thus define the best data model to be used and, further, to establish optimized the structure of dimensions involved in the dimensional model of the new DW.

Keywords: Decision making, Business, Knowledge Engineering, Knowledge Acquisition, Requirements Collection, DW Project, Dimensional Model, Interview Technique, Classification of Terms.

SUMÁRIO

	Pág.
SUMÁRIO.....	IV
LISTA DE FIGURAS	V
LISTA DE TABELAS.....	VI
LISTA DE SIGLAS E ABREVIATURAS	VII
CAPÍTULO 1 - AMBIENTE DE NEGÓCIO ENVOLVIDO.....	8
CAPÍTULO 2 - PROBLEMA TRATADO.....	9
CAPÍTULO 3 - SOLUÇÃO UTILIZADA E PORQUE DO USO DA SOLUÇÃO	10
CAPÍTULO 4 - RESUMO DA METODOLOGIA UTILIZADA.....	12
CAPÍTULO 5 - GANHOS OBTIDOS	15
CAPÍTULO 6 - CONSIDERAÇÕES FINAIS	17
REFERÊNCIAS BIBLIOGRÁFICAS.....	18

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 01: Modelo dimensional para Produção Intelectual	14

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 01: Classificação Final de Produção Intelectual	13
Tabela 02: Origens dos dados para o modelo dimensional	14

LISTA DE SIGLAS E ABREVIATURAS

BABi - Sistema de Automação de Bibliotecas

CNP - Conselho Nacional de Pesquisa

DW - *Data Warehouse*

SAI - Secretaria de avaliação institucional

UFRGS - Universidade Federal do Rio Grande do Sul

CAPÍTULO 1

Ambiente de Negócio envolvido

O estudo de caso sobre técnicas de aquisição do conhecimento no processo de modelagem do DW foi realizado na Universidade Federal do Rio Grande do Sul (UFRGS) para melhor tomadas de decisões a cerca da avaliação da produção intelectual. Sendo o processo de negócio definido, então, pela produção intelectual que é um dos principais indicadores do desempenho de uma universidade como a elaboração de artigos científicos, capítulos de livros, trabalhos em anais, partituras musicais, apresentações artísticas, desenvolvimento de aplicativos computacionais e muitas outras formas de expressão do conhecimento e da cultura. Toda a produção criada é de extrema relevância para sociedade e para as tomadas de decisões dentro da universidade, uma vez que, a quantidade de produção intelectual é utilizada diretamente como medição do desempenho dos departamentos e dos docentes da universidade, possibilitando, assim, estabelecer os índices adequados para a distribuição orçamentária, definir de vagas em concurso público para novos docentes, determinar as bolsas de pesquisa, liberar os financiamentos para projetos, entre vários outros importantes processos decisórios na universidade.

É notória a importância da produção intelectual, e considerando que grandes decisões na universidade levam em consideração suas produções é fundamental que as mesmas venham a serem melhores quantificadas, analisadas, visualizadas e controladas. Porém a situação universitária era um tanto caótica neste sentido, os sistemas existentes até então eram incapazes de realizar tal controle e por serem tão diversificados impossibilitavam uma análise aprofundada e detalhada a cerca do processo de negócio em questão. Assim a UFRGS por não apresentar uma integração dos dados e nem conseguir apresentar a visualização direta do conjunto dos dados de forma a permitir relacionar e associar os dados provenientes das variadas fontes percebeu a necessidade de realizar a implantação de um DW para que toda alta administração pudesse fazer uso de um melhor sistema de apoio a cerca das decisões a serem tomadas.

CAPÍTULO 2

Problema Tratado

Na UFRGS todas as informações de produção intelectual eram encontradas em diversos sistemas incompatíveis entre si como, por exemplo, no sistema de bibliotecas (SABi) e também do currículo Lattes que é fornecido pelo CNP um órgão federal que coordena pesquisa no Brasil. Ademais, os dados sobre os departamentos, os docentes e os técnicos, com os quais tem relação direta com a produção intelectual se encontravam ainda em variados sistemas desenvolvidos internamente na instituição, ou seja, os dados eram provenientes de fontes distintas e não se relacionavam devido às incompatibilidades sistêmicas, impossibilitando assim que os sistemas isolados fornecessem a visualização necessária dos dados e assim apoiassem de forma otimizada a alta administração da universidade a tomarem as decisões mais cabíveis e assertivas, e somado a isso os sistemas existentes não consideravam o desempenho que cada departamento e que cada docente envolvido nas criações intelectuais apresentavam, assim era muito difícil, portanto, avaliar criteriosamente a fim de melhor realizar as distribuições de orçamento, de vagas para concurso, de bolsas de pesquisa e de financiamento para os projetos.

Para o problema em questão da UFRGS a solução encontrada foi à aprovação de um projeto para a elaboração e a implantação de um DW no ambiente institucional, uma vez que era inexistente um sistema de apoio à decisão e a universidade não apresentava um repositório único de dados que pudesse viabilizar as informações necessárias para as tomadas de decisões sobre as produções intelectuais disponíveis de forma integrada. Ademais, por parte da alta administração da universidade existia um grande interesse em solidificar um processo de avaliação institucional por meio da utilização da tecnologia com o apoio direto dos benefícios que o DW traria para as tomadas de decisões. E igualmente interessado encontrava-se o setor de sistemas de informação, o centro de processamento de dados, pois visavam fornecer o melhor suporte aos processos de decisão e de avaliação institucional com a implantação do novo DW na instituição.

CAPÍTULO 3

Solução utilizada e porque do uso da solução?

Sabendo da situação atual a qual a universidade se encontrava, onde os dados eram derivados de uma ampla variedade de fontes e de sistemas diversos, e que um DW centralizaria e consolidaria toda essa grande quantidade de dados de forma eficiente e com seus recursos analíticos permitiriam que alta administração visualizasse as informações do negócio para apoiar na tomada de decisões a solução encontrada para o processo de negócio foi à implantação de um DW completo para a universidade a fim de realizar a disponibilização dos dados de toda produção intelectual da instituição para que a alta administração pudesse tomar as melhores decisões cabíveis. O DW ao guardar todas as informações relativas às atividades da organização em bancos de dados de forma consolidada redesenharia a nova base de dados para favorecer diretamente a geração de relatórios, a análise de grandes volumes de dados e a obtenção de informações estratégicas que facilitaria a tomada de decisão pela administração da universidade. Assim o DW poderia trazer entre seus benefícios a realização consultas e análises avançadas que geralmente contêm grandes quantidades de dados históricos.

No entanto, antes da implantação do DW foi necessário à obtenção de requisitos e a transferência do conhecimento através da aplicação de técnicas de aquisição de conhecimento. Realizou-se o uso da técnica de entrevista livre, ou não estruturada com gravação de áudio, o que permitiu e facilitou obter mais informações relevantes do processo de tomada de decisão ou da análise do negócio dos especialistas que ao apresentarem maior dificuldade de descrever verbalmente tudo que é de fato importante para resoluções de cada uma de suas decisões frente às informações que já possui ou que melhor facilitaria tomar caso tivesse tal informação faltante. Também foi utilizada a técnica de classificação de termos por possuir uma característica particularmente desejável para situações como esta, onde consegue levar o especialista a visualizar uma organização mental das informações que ele mesmo utiliza, mas que nunca havia articulado de forma consciente anteriormente.

Com o resultado do trabalho de coleta de requisitos foi possível compreender o domínio da atividade de pesquisa e especificamente da produção intelectual na universidade. Assim o tipo de modelo de dados utilizado foi o modelo dimensional, técnica de modelagem de dados para DW voltada especialmente para a implementação de um modelo que permita a visualização de dados de forma intuitiva e com altos índices de desempenho na extração de dados. A visualização de dados é intuitiva porque os usuários (analistas de negócio) facilmente conseguem entender e navegar pelo modelo, vendo as questões às quais desejam que o DW responda. Portanto as razões para a escolha de modelo dimensional na UFRGS orientou-se pelos seguintes benefícios: simplicidade, facilidade de navegação pelo usuário final e modelagem

voltada aos indicadores usados na tomada de decisão e não na relação de dependência entre os dados.

CAPÍTULO 4

Resumo da Metodologia utilizada

O desenvolvimento do estudo de caso constituiu-se das seguintes etapas:

1. **Definição do método a aplicar:** Utilizou-se a técnica de entrevista não estruturada para conhecer o domínio estudado e obter os termos mais importantes. Em seguida através da técnica de classificação de termos para definição dos objetos de domínio e as suas relações permitindo assim definir o modelo dimensional de dados.
2. **Escolha dos participantes na modelagem:** Para a escolha dos entrevistados e participação no projeto foram eleitos a diretora e vice do SAI e também a vice Pró-Reitora de pesquisa da UFRGS, representando assim os especialistas e analistas de negócios que estavam mais diretamente ligados e interessados ao processo de negócio da instituição.
3. **Realização de entrevistas livres:** Realizaram-se duas entrevistas livres, ou não estruturadas, com os participantes eleitos e com duração de uma hora cada, onde toda conversa foi gravada e posteriormente transcrita para ser analisada e assim auxiliar na próxima fase.
4. **Aplicação da técnica de Classificação de Termos:** Analisou-se a frequência e relevância dos termos mais utilizados na entrevista resultando em 63 termos que foram transformados em fichas e um novo encontro com os entrevistados foi realizado e proposto realizar classificações da seguinte ordem e maneira:
 - I. Dividir o conjunto completo em 2 ou 3 grupos;
 - II. Para cada grupo criado anteriormente dividir em 2 ou 3 subgrupos;
 - III. Repetir I e II enquanto o especialista conseguir classificar

Depois de seguir o processo descrito acima, a classificação final dada pelos especialistas entrevistados foi definida em três subgrupos: Iniciativas internas de divulgação contendo 4 itens, Iniciativas externas de divulgação também contendo 4 itens e como subgrupo final temos a produção intelectual contendo um total de 16 itens categóricos, assim como pode ser observado na *Tabela 01*.

Iniciativas Externas de Divulgação	Iniciativas Internas de Divulgação	Produção Intelectual
CAPES CNPq Grupo de Pesquisa Lattes	Comissão Pesquisa Pós-Graduação Pró-Reitoria de Pesquisa Sistema Pesquisa	Artigo Nacional Artigo Internacional Livro Patente Periódico Produção Artística Produção Bibliográfica Produção Científica Produção Intelectual Produção Técnica Produto Produto Tecnológico Projeto Publicação Resumo SABI

Tabela 01: Classificação Final de Produção Intelectual

Elaboração do modelo de dados: Após a escolha do processo de negócio e a técnica de classificação de termos determinou-se o domínio da produção intelectual da universidade e a sua granularidade. Cada registro na tabela de fatos foi visto como um produto intelectual apresentado à sociedade. Com a devida análise verificou-se a ausência de fatos neste caso específico da UFRGS, onde o único fato importante é a existência do produto intelectual, que será utilizada então para as contagens e geração de índices para relatórios. Nesta fase definiram-se também as dimensões do cubo em Autor, Categoria de Produção, Área de Conhecimento e Ano (como dimensão tempo). Após definir as dimensões foi elaborado o modelo dimensional ideal do projeto para o novo DW deste estudo de caso, assim como representado na *Figura 01*.

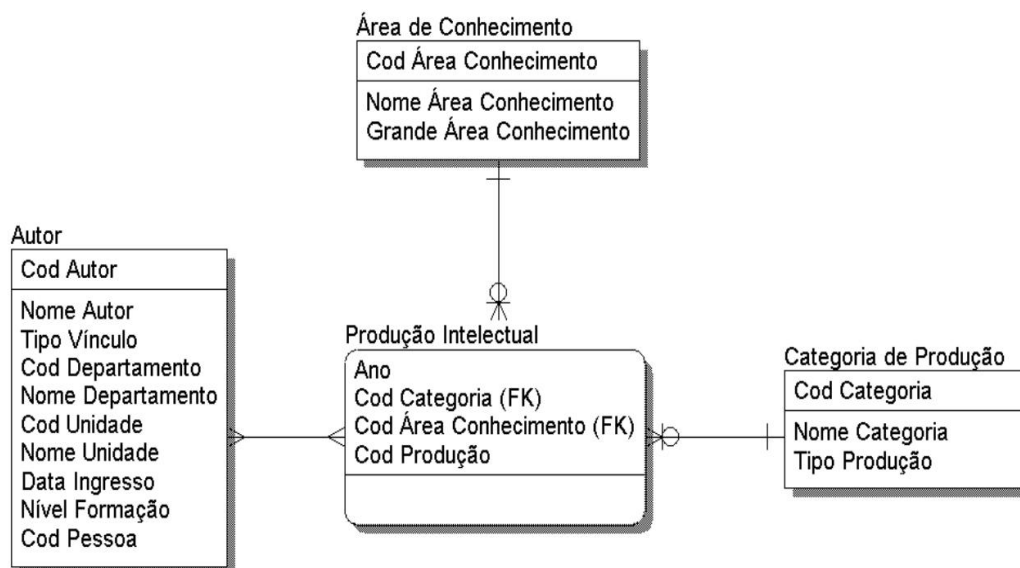


Figura 01: Modelo dimensional para Produção Intelectual

5. **Levantamento de origens dos dados presentes no modelo de dados:** A identificação dos repositórios, local onde são encontrados os dados necessários à implantação do modelo foi realizada pela CPD da universidade e apresenta a seguinte relação de origem conforme mostra a *Tabela 02* abaixo:

Sistema de Origem dos Dados	Modelo Proposto
Banco ADMINI	Autor (dimensão)
Criação manual	Categoria de Produção (dimensão)
Criação manual	Área de Conhecimento (dimensão)
SABi	Produção Intelectual (fato)

Tabela 02: Origens dos dados para o modelo dimensional

CAPÍTULO 5

Ganhos Obtidos

O estudo de caso da UFRGS envolvendo modelagem de dados foi desenvolvido e avaliado focando principalmente na análise das vantagens obtidas com a aplicação de técnicas de aquisição de conhecimento na coleta de requisitos em DW. O uso de técnicas de aquisição de conhecimento na coleta de requisitos trouxe benefícios, especialmente neste caso onde é difícil para o usuário do DW descrever verbalmente quais informações são empregadas no processo de tomada de decisão ou da análise do negócio. O uso de técnicas de aquisição de conhecimento para a coleta de requisitos em DW é recomendável justamente quando o usuário não tem ainda uma organização mental das informações suficientemente clara para assim conseguir expô-las verbalmente. Normalmente os executivos e gerentes de uma organização tomam suas decisões baseando-se em questões pessoais e subjetivas, como sua experiência e seu *feeling*, ao invés de se basear em fatos objetivos. Em uma situação como esta, o executivo dificilmente será capaz de relatar como toma decisões e que informações são usadas no processo, pois ele não tem uma organização mental das informações. Neste caso, por exemplo, o uso de técnicas de aquisição de conhecimento para a coleta de requisitos de um DW se faz necessária.

As duas técnicas de aquisição do conhecimento utilizadas, tanto a de entrevistas não estruturadas quanto a de classificação de termos foram empregadas em sequência e de forma complementar a idealização do projeto. As técnicas em conjunto atingiram satisfatoriamente os objetivos buscados. Objetivos estes de definirem corretamente os requisitos e abstrair o modelo dimensional de dados. Abaixo são destacados os ganhos observados no projeto de implantação do DW da UFRGS no que tange as duas técnicas de aquisições do conhecido utilizadas:

- **A técnica de entrevistas não estruturadas:** trouxeram como vantagens uma maior liberdade ao entrevistado. Conseguiu-se estabelecer um bom relacionamento entre engenheiro e o especialista. O engenheiro pode facilmente ter uma visão ampla do tópico da entrevista e o especialista pode descrever o domínio da forma como lhe é familiar, enfatizava sempre os tópicos que considerava mais importante e, ainda, ignorava aquele que julgava desinteressante.
- **A técnica de classificação de termos:** trouxe rapidez na aplicação e facilidade na análise dos resultados para implantação do projeto. Os especialistas não estavam conseguindo de forma verbal expor seus conhecimentos da maneira mais ideal, então esta técnica os auxiliou a colocar no formato explícito as construções mentais. Na maioria das vezes a técnica foi instrutiva até mesmo para o especialista, pois o levava a visualizar uma organização mental que ele mesmo utilizava, mas que nunca havia articulado conscientemente antes. Como resultados obtidos com a classificação de termos destacaram-se o reconhecimento do

relacionamento entre os termos (como hierarquias), os atributos comuns e sinonímia, a obtenção de termos não evidenciados anteriormente, as diferentes formas de visualizar o domínio e por fim forneceram uma melhor compreensão global do domínio.

CAPÍTULO 6

Considerações Finais

Segundo POE para atender as necessidades provenientes da evolução no processo de tomada de decisão e do aumento da competitividade no mercado, as organizações necessitam de um sistema automatizado de apoio à decisão. Um sistema que provê informações para seus usuários de forma que eles possam analisar uma situação e tomar as decisões mais assertivas. E que seja flexível para se adaptar às constantes mudanças impulsionadas pela evolução do processo de tomada de decisão. Sobre esta perspectiva a UFRGS decidiu realizar um projeto de implantação de DW, pois seria o ideal para apoiar nas decisões que contemplariam a produção intelectual de todos os professores e departamentos da universidade.

O processo de tomada de decisão organizacional baseado em DW na perspectiva do contexto da engenharia de conhecimento através do uso das técnicas de aquisição de conhecimento com aplicação da entrevista e da classificação de termos para a obtenção de requisitos de DW e para a sua melhor definição de modelo dimensional a fim de realizar sua implantação conseguiu atender satisfatoriamente para este estudo de caso.

No entanto vale ressaltar duas importantes considerações a serem destacadas sobre a coleta de requisitos para que seja realizada de maneira correta e sem causar futuros transtornos. Primeiro a coleta de requisitos deve ser realizada antes da construção do DW, pois evita um esforço que possivelmente não traria benefícios por não estar em conformidade com as expectativas dos usuários, uma vez que os dados no ambiente operacional podem ser encontrados em formato inadequado para o suporte à decisão. E segundo o escopo da coleta de requisitos para DW deve ser definido com base no setor da organização que se deseja contemplar, selecionando-se os usuários adequados. Ao final do processo deve-se obter um amplo entendimento dos negócios do usuário. Assim, sendo, deve-se limitar o escopo da coleta de requisitos apenas pelos usuários que serão atendidos, e não pelos processos cujos resultados originam os dados usados nas medidas de desempenho dos negócios. Somente após conhecer os negócios do usuário, e identificar quais são as atividades que geram as informações desejadas, é que devem ser escolhidos os processos para os quais serão construídos modelos dimensionais.

REFERÊNCIAS BIBLIOGRÁFICAS

Ziulkoski, L. C. C. **Coleta de requisitos e Modelagem de dados para Data Warehouse: um estudo de caso utilizando Técnicas de Aquisição de conhecimento.** Projeto de Diplomação (Bacharelado em Ciências da computação) – Universidade Federal do rio Grande Do Sul Instituto de Informática. Porto Alegre, 2003. 63 p.

Poe, V., Klauer, P., Brobst, S. **Building a Data Warehouse for Decision Support.** New Jersey, Prentice Hall. 1998.