

RENATO DEGGAU

**UMA INTERFACE BASEADA EM
CONHECIMENTO PARA
INTERAÇÃO COM DATA
WAREHOUSES ESPACIAIS**

FLORIANÓPOLIS

2010

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA
COMPUTAÇÃO**

**UMA INTERFACE BASEADA EM CONHECIMENTO
PARA INTERAÇÃO COM DATA WAREHOUSES
ESPACIAIS**

Dissertação submetida ao Programa de Pós-Graduação em Ciências da
Computação da Universidade Federal de Santa Catarina para a obtenção do
Grau de Mestre em Ciências da Computação

RENATO DEGGAU

Florianópolis, Agosto de 2010

**UMA INTERFACE BASEADA EM CONHECIMENTO
PARA INTERAÇÃO COM DATA WAREHOUSES
ESPACIAIS**

RENATO DEGGAU

Esta Dissertação foi julgada adequada para obtenção do Título de Mestre em Ciências da Computação, e aprovada em sua forma final pelo Programa de Pós Graduação em Ciências da Computação da Universidade Federal de Santa Catarina.

Prof. Mario Antonio Ribeiro Dantas, Dr.
Coordenador

Banca Examinadora:

Renato Fileto, Dr., Orientador
Presidente

Ana Carolina Brandão Salgado, Dra.,
Universidade Federal de Pernambuco

Eugenio Andrés Díaz Merino, Dr.,
Universidade Federal de Santa Catarina

Ronaldo dos Santos Mello, Dr.,
Universidade Federal de Santa Catarina

AGRADECIMENTOS

Meus agradecimentos:

- a Deus que me deu capacidade e sabedoria;
- à minha esposa Marícia pela compreensão, apoio e companheirismo;
- aos meus filhos Daniel e André pela compreensão nos momentos de ausência;
- ao meu orientador Prof. Renato Fileto, pela sabedoria e paciência com que conduziu todo o meu trabalho, e pelos seus ensinamentos;
- à EPAGRI, que me concedeu esta oportunidade de aprofundar meus conhecimentos;
- à EPAGRI/CEPA e à GIN, pelo apoio, troca de idéias, sugestões e pela participação no experimento;
- a todos os integrantes do LISA (Laboratório para Integração de Sistemas de Informação e Aplicações Avançadas) e ao Grupo de Banco de Dados da UFSC pelo suporte necessário e pelas ótimas discussões, contribuições e sugestões recebidas no decorrer do trabalho;
- ao Prof. Eugênio Merino e ao amigo Danilo Pereira, do Núcleo de Gestão de Design da UFSC (NGD), pelo valoroso suporte na definição e aplicação do experimento com os usuários.
- à Fapesc (contrato 12552-2007-0) e ao CNPq (contrato 48139212007-6) que subsidiaram parte deste trabalho.

Resumo da Dissertação apresentada à UFSC como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências da Computação.

UMA INTERFACE BASEADA EM CONHECIMENTO PARA INTERAÇÃO COM DATA WAREHOUSES ESPACIAIS

Renato Deggau

Agosto / 2010

Orientador: Prof. Renato Fileto, Dr..

Área de Concentração: Engenharia de Software e Bancos de Dados.

Palavras-chave: busca semântica, data warehouses espaciais, descrições semânticas, GUI, OLAP espacial, ontologias, web semântica.

Número de Páginas: 83

Resumo. A análise de informação em um data warehouse espacial (SDW) pode envolver o manuseio de grandes volumes de dados espaciais. Usuários de domínios específicos de aplicação, com habilidades básicas de computação, são geralmente incapazes ou têm sérias dificuldades para resolver suas necessidades de análise de informação interagindo diretamente com SDWs, embora alguns sejam capazes de interagir com data warehouses (DW) convencionais através de uma interface gráfica (GUI). As dificuldades são maiores em um SDW que em um DW convencional, entre outras razões, pela variedade e complexidade dos dados espaciais, operadores espaciais e funções de agregação espacial utilizadas para especificar consultas SOLAP.

Este trabalho propõe um sistema baseado em conhecimento, chamado de S^2DW (Semantic and Spatial Data Warehouses), para auxiliar estes usuários de domínios específicos a efetuar análises de informação em SDWs, acessando descrições semânticas dos data marts espaciais através de uma interface gráfica baseada em conhecimento (GUI). Este trabalho descreve a arquitetura geral do S^2DW e foca em sua GUI.

A interface gráfica baseada em conhecimento do S^2DW permite ao usuário pesquisar data marts relacionados a um determinado assunto, através da especificação de palavras-chave ou pela navegação em uma visão de uma ontologia do domínio. Cada data mart relacionado ao assunto pesquisado é apresentado ao usuário como um grafo representando a estrutura dimensional do cubo de informação. Este grafo é semanticamente enriquecido com descrições do conteúdo dos dados e dos recursos de processamento de dados

do data mart espacial. Consultas espaciais OLAP podem ser especificadas interagindo com a interface gráfica baseada em conhecimento, a qual orienta o usuário a compor adequadamente operadores e funções para tratar os diferentes tipos de dados disponíveis no data mart, visando atender diferentes necessidades de análise. As tabelas, gráficos e mapas fornecidos como resposta as consultas SOLAP também permitem a interação do usuário para gradualmente refinar a análise da informação.

As principais contribuições deste trabalho são a proposta inicial da GUI baseada em conhecimento do S^2DW e o teste de usabilidade desta GUI, em um estudo de caso com usuários reais do domínio agrícola.

Abstract of Dissertation presented to UFSC as a partial fulfillment of the requirements for the degree of Master in Computer Science.

Renato Deggau

Aug / 2010

Advisor: Prof. Renato Fileto, Dr..

Area of Concentration: Software Engineering and Databases.

Keywords: GUI, ontologies, semantic descriptions, semantic search, semantic web, spatial data warehouse, spatial OLAP.

Number of pages: 83

Abstract. Information analysis in a spatial data warehouse (SDW) can handle large volumes of spatial data. Users from specific application domains, with basic computation skills, are usually unable or have serious troubles to solve their information analysis needs by interacting directly with SDW systems, though some of them are able to interact with conventional data warehouses via a graphical user interface (GUI). The difficulties are greater in a SDW than in a conventional data warehouse, among other reasons, because of the variety and complexity of spatial data, spatial operators and spatial aggregation functions used to specify SOLAP queries.

This work proposes a knowledge-based system, called S^2DW (Semantic and Spatial Data Warehouses), to help such users from specific domains to perform information analysis in SDWs, by accessing semantic descriptions of spatial data marts via a knowledge-based GUI. This work describes the general architecture of S^2DW and focuses on its GUI.

The S^2DW 's knowledge-based GUI allows the user to search for data marts related to a given subject, by providing domain specific keywords or browsing a domain ontology view. Each data mart related to the searched subject is presented to the user as a graph representing the dimensional structure of the information cube. This graph is semantically enriched with descriptions of the data contents and the data processing means of the spatial data mart. Spatial OLAP queries can be specified by interacting with the knowledge-based GUI, which helps the user to properly compose operators and functions for handling different kinds of data available in the data mart, for different analysis purposes. The tables, graphs and maps presented in response to SOLAP queries allow further user interaction to gradually refine the information analysis.

The main contributions of this work are the initial proposal of the

S²DW's knowledge-based GUI, and the usability test of this GUI, in a case study with real users of the agriculture domain.

SUMÁRIO

Introdução	1
1.1 Objetivos	3
1.1.1 Objetivo Geral	3
1.1.2 Objetivos Específicos	4
1.2 Justificativa	5
1.3 Delimitações do Trabalho	5
1.4 Metodologia	6
1.5 Organização do Trabalho	7
Fundamentação Teórica	9
2.1 Data Warehouses	9
2.1.1 Modelo Dimensional de Dados	10
2.1.2 Processamento Analítico On-line (OLAP)	11
2.2 Data Warehouses Espaciais	12
2.2.1 Operadores	13
2.2.1.1 Operadores de Data Warehouses Conven- cionais	13
2.2.1.2 Operadores de Dados Espaciais	13
2.2.2 Funções de Agregação	15
2.3 Interfaces Gráficas de Consulta a Bancos de Dados	16
2.3.1 Interfaces para Banco de Dados Convencionais	17
2.3.2 Interfaces para GIS	17
2.4 Web Semântica e Data Warehouses	18
2.5 Considerações	23
Estudo de Caso - Um SDW com Dados Agrícolas	25
3.1 A EPAGRI	25
3.1.1 A Informação na EPAGRI	25
3.1.1.1 A Coleta da Informação	26
3.1.1.2 A Difusão da Informação	26
3.1.1.3 A Infraestrutura de hardware e software	27
3.2 Os Sistemas de Informação da EPAGRI	27
3.2.1 SAFI - Sistema Administrativo-Financeiro	27
3.2.2 SEPLAN - Sistema EPAGRI de Planejamento	27
3.2.3 SEATER - Sistema EPAGRI de ATER - Assistência Técnica e Extensão Rural	28

3.2.4	SEPRO - Sistema EPAGRI de Profissionalização . . .	28
3.2.5	Sistema de gestão de pessoas	28
3.2.6	Preços Agrícolas	28
3.2.7	SIAGRO - Sistema de Informações Agropecuárias . .	29
3.2.8	LAC - Levantamento Agropecuário Catarinense . . .	29
3.2.9	SIGEO - Sistema Integrado de Informações Georef- erenciadas	29
3.2.10	GEURED - Sistema de Gerenciamento de Crédito Agrícola	30
3.2.11	SOLOS - Banco de Dados de Solos	30
3.2.12	As perspectivas de integração	30
3.3	Um SDW sobre Propriedades e Atividades Agrícolas	31
3.3.1	Os sistemas escolhidos para o estudo de caso	31
3.3.2	O Esquema do SDW Proposto	33
3.3.3	Exemplos de Análises a serem Efetuadas	34
3.3.4	Implementação	36
3.4	Considerações	36
Semantic and Spatial Data Warehouses		37
4.1	Arquitetura para Semantic Spatial Data Warehouse	37
4.2	A Ontologia de SDW	39
4.3	A Ontologia de Domínio	43
4.3.1	Exemplo de Ontologia de Domínio - Ontologia Agrícola	43
4.4	Descrição Semântica de Data Marts	43
4.5	A Interface Proposta	44
4.5.1	Busca por data marts espaciais relacionadas a um as- sunto	47
4.6	Executando Análises de Informação Espacial	48
4.6.1	Outros exemplos de análise de informação	52
4.7	Considerações	54
Experimentos		57
5.1	Análise de Usabilidade da GUI Proposta	57
5.2	Detalhamento do Experimento	58
5.3	Resultados do Experimento	60
5.4	Análise dos resultados	65
5.5	Considerações	67
Trabalhos Relacionados		69

6.1	Interfaces gráficas de consulta	69
6.2	Uso de Semântica	70
6.3	Comparações com este trabalho	70
Conclusões e Trabalhos Futuros		73
7.1	Contribuições	73
7.2	Publicações	74
7.3	Trabalhos Futuros	75

LISTA DE FIGURAS

1	Relações topológicas entre objetos espaciais	15
2	O esquema de um SDW de propriedades e atividades agrícolas	34
3	A arquitetura geral do S^2DW e a interface baseada em conhecimento	38
4	Visão geral da ontologia de SDW	39
5	Ontologia de SDW - Componentes Estruturais	40
6	Ontologia de SDW - Tipos de Entidades Espaciais	40
7	Ontologia de SDW - Operadores e Funções de Manipulação de Dados	41
8	Ontologia de SDW - Funções de Agregação Distributivas Espaciais	42
9	Ontologia de SDW - Operadores Espaciais que Geram Nova Geometria	42
10	Visão geral de uma ontologia para domínio agrícola	44
11	Busca baseada em palavra-chave e navegação em uma visão da ontologia de domínio	46
12	Representação gráfica da descrição semântica de um data mart espacial	48
13	Distribuição geográfica das propriedades com casos de intoxicação pelo uso de agrotóxicos e que fazem uso frequente de agrotóxicos no estado de Santa Catarina em 2003 .	49
14	Consulta SOLAP 1 em MDX	50
15	Sedes das propriedades com casos de intoxicação pelo uso de agrotóxicos que fazem uso frequente no município de Alfredo Wagner em 2003.	51
16	Operadores disponíveis para filtrar informação usando o membro espacial <i>Corpos de Água</i>	51
17	Mapa Resultante sem aplicação do operador Buffer	52
18	Mapa Resultante com a aplicação do operador Buffer e Intersects das sedes das propriedades com o resultado do Buffer .	53
19	Consulta SOLAP 2 em MDX	53
20	Propriedades agrícolas com certificação de produção orgânica em Passos Maia localizadas a certa distância de um ponto geográfico	54
21	Propriedades agrícolas do Oeste de SC com morte de animais por doença e que possuem fonte de água	55
22	Percentual de sucesso por tarefa	61

23	Score médio por tarefa	62
24	Tempo médio para a realização de cada tarefa (seg)	63
25	Número médio de cliques do mouse para cada tarefa	64
26	Frequência de resposta por questão	65
27	Resposta média por questão	66

LISTA DE TABELAS

1	Sistemas da EPAGRI, seu potencial para uso de informações georeferenciadas e público alvo	32
2	Visão geral dos mapeamentos para descrever semanticamente os componentes de um SDW sobre propriedades e atividades agrícolas	45
3	Comparação do S^2DW com outros trabalhos	72

LISTA DE ABREVIATURAS E SIGLAS

ASM	Analysis Solution Model.
BCM	Business Concept Model.
CEDAP	Centro de Desenvolvimento em Aqüicultura e Pesca.
CEPA	Centro de Socioeconomia e Planejamento Agrícola.
CEPAF	Centro de Pesquisa para a Agricultura Familiar.
CIRAM	Centro de Excelência em Informações de Recursos Ambientais e de Hidrometeorologia.
DW	Data Warehouses.
EPAGRI	Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina.
ETC	Extração, transformação e carga.
GUI	Graphical User Interface.
HOLAP	Híbrido On-Line Analytical Processing.
MOLAP	Multidimensional On-Line Analytical Processing.
OLAP	On-line Analytical Processing.
OLTP	On-Line Transaction Processing.
PECA	Parque Ecológico Cidade das Abelhas.
ROLAP	Relational On-Line Analytical Processing.
SDW	Spatial Data Warehouses.
SIG	Sistemas de Informação Geográfica.
SOLAP	Spatial On-line Analytical Processing.

1 INTRODUÇÃO

A análise de informação para a tomada de decisão em áreas como planejamento urbano, gerenciamento de recursos naturais, agricultura e prevenção de desastres frequentemente requer a observação de grandes volumes de dados espaciais. Spatial Data Warehouses (SDW) são projetados para atender este tipo de demanda. Eles estendem as possibilidades de Data Warehouses (DW) convencionais e On-line Analytical Processing (OLAP), utilizados para analisar grandes volumes de dados convencionais no modelo dimensional, com outros recursos como os de Sistemas de Informação Geográfica (SIG), para a representação e análise de informação espacial (RAO et al., 2003; RIVEST et al., 2005; MALINOWSKI; ZIMÁNYI, 2007; BIMONTE; TCHOUNIKINE; MIQUEL, 2007).

Objetos geográficos podem aparecer nas dimensões (e.g., polígonos representando estados e cidades) ou como medidas na tabela fato (e.g., pontos geográficos onde ocorrem certos fatos, como casos de intoxicações por uso de agrotóxicos) de um data mart espacial (cubo de informação multidimensional de um SDW) (MALINOWSKI; ZIMNYI, 2008). Assim, além de operadores OLAP tradicionais e funções de agregação de dados escalares, SDWs precisam fornecer suporte a tipos de dados espaciais integrados ao modelo dimensional (FIDALGO et al., 2004), bem como suportar uma grande variedade de operadores e funções para a manipulação de objetos espaciais (SILVA et al., 2008).

Um exemplo de consulta que pode ser especificada em um SDW é a identificação das sedes de propriedades agrícolas do município de Alfredo Wagner, que façam uso frequente de agrotóxicos e onde ocorreram casos de intoxicação pelo uso de agrotóxicos, no ano de 2003; deseja-se a identificação somente das propriedades que estejam dentro do limite de 5 km de distância do principal rio do município, o rio Itajaí do Sul.

Para que um usuário de domínio obtenha o resultado para esta consulta, ele normalmente necessita do auxílio de um especialista de Tecnologia de Informação, que irá especificar esta consulta em uma linguagem apropriada e a submeta ao SDW. Mas estes especialistas também têm dificuldades em compreender a linguagem do domínio e possuem outras atribuições, de forma que a interação adequada com o SDW seja um processo trabalhoso para o usuário de domínio e para o especialista em TI, principalmente em consultas que exigem interações constantes para a obtenção do resultado desejado, comum em procedimentos de análise de informação. Assim, a criação de um sistema que possa fazer o papel de ponte entre o usuário de domínio e

o SDW é uma solução que facilita o processo de interação do usuário com o SDW.

Problemas em aberto na área de SDW incluem: (i) modelagem de SDWs (FIDALGO et al., 2004; MALINOWSKI; ZIMÁNYI, 2007); (ii) extração, transformação e carga (ETC) de dados espaciais em SDWs (SKOUTAS; SIMITSIS, 2006; MARTINO et al., 2009); e (iii) interfaces humano-computador para facilitar a interação do usuário com data marts para suportar busca e entendimento do seu conteúdo e execução de análise de informações (RAO et al., 2003; SELL et al., 2008; XIE et al., 2007).

A variedade de operadores e de funções de agregação utilizadas para manusear objetos espaciais torna a interação do usuário com SDWs muito mais complexa que a interação com DWs convencionais. A análise de informação com objetos espaciais requer conhecimento dos possíveis operadores e funções de agregação disponíveis para cada tipo de dados e para cada propósito. Para a correta utilização dos operadores espaciais e das funções de agregação espacial na análise de informação em SDWs, é necessário o conhecimento de sua semântica, dos tipos de dados de entrada e dos tipos de dados de saída.

Interfaces humano-computador têm sido propostas na literatura para fornecer suporte a interação de usuários com bancos de dados, de forma que estes possam efetuar suas consultas de forma autônoma, isto é, sem sempre precisar de apoio de programadores (TERWILLIGER; DELCAMBRE; LOGAN, 2007). Na área de data warehouses, alguns trabalhos sugerem o uso de ontologias e de descrições semânticas para auxiliar os usuários a resolver algumas de suas necessidades de análise de informação (SELL et al., 2008; XIE et al., 2007). Entretanto, estas abordagens não contemplam aspectos espaciais.

Este trabalho propõe um sistema baseado em ontologias, chamado S^2DW (Semantic and Spatial Data Warehouses), que permite a usuários sem habilidades de programação recuperar descrições semânticas de data marts espaciais relacionados a algum tópico de interesse, e realizar análise de informação utilizando OLAP espacial (SOLAP). Isso é possível mediante a interação com descrições semânticas dos data marts espaciais, em uma interface gráfica baseada em conhecimento (GUI).

As descrições semânticas dos data marts espaciais utilizadas no S^2DW são construídas de acordo com o modelo dimensional espacial, que integra dados convencionais e dados espaciais em dimensões e medidas. Este modelo é descrito em uma ontologia de SDW especialmente desenvolvida para o S^2DW . Esta ontologia define os possíveis componentes estruturais (dimensões, níveis, tabelas fato e medidas), operadores e funções de agregação

usados nos data marts espaciais. A ontologia de SDW descreve as possíveis conexões dos componentes estruturais dos data marts espaciais e os possíveis usos de operadores e funções de agregação, de acordo com seus propósitos e com os tipos de dados manipulados. Ela inclui taxonomias de operadores para filtragem de dados e para funções de agregação de dados, cobrindo a ampla variedade de operadores espaciais e de funções de agregação espacial. Todas essas definições podem ser consultadas e visualizadas graficamente no S^2DW . Esta ontologia de SDW foi desenvolvida como parte deste trabalho.

Na interface baseada em conhecimento do S^2DW o usuário pode fornecer palavras-chave ou navegar sobre um vocabulário obtido a partir de uma ontologia de domínio, de forma a fazer buscas sobre data marts espaciais relacionadas a um determinado assunto. O S^2DW usa técnicas de busca semântica para processar as buscas e recuperar as descrições semânticas dos SDWs relacionados ao tema pesquisado (D'AGOSTINI; FILETO, 2009).

A descrição semântica de um data mart espacial é apresentada ao usuário como um grafo que descreve a estrutura e o conteúdo de um data mart. A GUI baseada neste grafo é enriquecida com descrições semânticas dos componentes estruturais de um data mart espacial, além de operadores e funções de agregação espacial que podem ser aplicados a estes componentes. O usuário pode efetuar consultas pela seleção de medidas da tabela fato, membros de diferentes níveis de cada dimensão e operadores e funções de agregação espacial que podem ser aplicadas sobre estes dados. O sistema auxilia o usuário a efetuar corretamente as combinações de operadores e funções de manipulação de dados espaciais, de forma que sejam compatíveis com os tipos de dados selecionados, para filtrar os dados na tabela fato. Ele também auxilia o usuário a escolher as funções de agregação de dados compatíveis com os tipos de dados das medidas selecionadas, para consolidar estas medidas. O S^2DW permite a visualização dos resultados das consultas em tabelas, gráficos e mapas. O usuário pode também efetuar interações com a descrição semântica do data mart e dos resultados retornados pelo S^2DW em resposta às consultas propostas, visando refinar a análise de informação e a apresentação dos resultados.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo deste trabalho é iniciar o desenvolvimento de um sistema baseado em ontologias, chamado S^2DW (Semantic and Spatial Data

Warehouses). Tal sistema vai possibilitar que usuários comuns recuperem descrições semânticas de data marts espaciais relacionados a algum tópico de interesse e realizem análise de informação utilizando OLAP espacial (SOLAP), interagindo com as descrições semânticas recuperadas dos data marts espaciais em uma interface gráfica baseada em conhecimento (GUI). O foco deste trabalho está na especificação da arquitetura de tal sistema e no desenvolvimento e teste de usabilidade de um protótipo de sua interface gráfica, de modo a testar a viabilidade de tal sistema do ponto de vista do usuário e colher subsídios para aperfeiçoamento da especificação e futuro desenvolvimento de outros módulos do sistema.

1.1.2 Objetivos Específicos

Os objetivos específicos do trabalho são:

- Efetuar uma revisão bibliográfica sobre a integração de tecnologias de data warehouse (DW) e de bancos de dados espaciais, visando aperfeiçoar o processo de análise de informação em SDW;
- Definir uma proposta de interface gráfica baseada em conhecimento, utilizando recursos de interfaces humano-computador, que auxilie usuários especialistas de um domínio de conhecimento a consultar informação em data marts espaciais e efetuar análise de informação sobre eles;
- Criar um SDW com dados de sistemas de informação da EPAGRI, focando nas necessidades de análise de dados dimensionais e espaciais;
- Aplicar ontologias na descrição semântica de SDWs;
- Efetuar experimentos para avaliar a usabilidade da interface proposta em um estudo de caso com usuários especialistas do domínio agrícola, utilizando o SDW criado com dados da EPAGRI.

Esta estrutura proposta, aplicada às informações da EPAGRI pretende:

- Fornecer uma ferramenta de apoio à tomada de decisão, com base na análise de grandes volumes de dados dimensionais e espaciais;
- Aperfeiçoar o processo de gerenciamento da informação para suporte à tomada de decisão;
- Identificar novos usos para informações já existentes atualmente.

1.2 Justificativa

Usuários comuns, que frequentemente têm dificuldades em interagir com DWs convencionais (que manipulam apenas dados escalares), encontram dificuldades adicionais quando confrontados com objetos espaciais e a grande variedade de recursos que podem ser utilizados na sua manipulação para análise de informação em SDWs.

Estas dificuldades fazem com que seja necessária a participação de um especialista em Engenharia da Informação para a execução dos procedimentos necessários. Porém, estes especialistas também têm dificuldade em extrair os requisitos necessários e entender as demandas dos usuários especialistas do domínio. Além disso, especialistas da área de TI podem ficar sobrecarregados com muitas requisições para análise de dados em adição a normalmente estarem sem tempo devido as outras atividades que realizam no seu trabalho. Já os especialistas do domínio de aplicação conhecem do negócio, mas têm dificuldades na área de TI. Isso exige constantes reuniões e discussões entre especialistas de domínio e de TI, a fim de obter os resultados desejados para os procedimentos de análise de informação.

Assim, a solução mais apropriada é fornecer a usuários finais facilidades adicionais para interação com SDWs. O trabalho aqui apresentado visa propor um sistema inteligente, com interfaces intuitivas e baseadas em conhecimento, para que o usuário final possa especificar suas necessidades e interagir com SDWs, facilitando os processos de análise de informação por parte destes especialistas de domínio.

1.3 Delimitações do Trabalho

Durante a elaboração do trabalho, uma idéia de arquitetura geral foi idealizada. Na busca de um foco de estudo, no tempo disponível para um trabalho de mestrado, este se concentrou na interface gráfica do sistema baseado em conhecimento para permitir que usuários especialistas de um domínio efetuem análises de informação em SDWs. Esta interface foi desenvolvida e um teste de usabilidade foi aplicado junto a especialistas do domínio agrícola em um SDW real.

Os demais itens da arquitetura proposta como geração das descrições semânticas dos data marts, formalização da interface e geração de expressões de consulta a partir das interações do usuário, são deixadas para trabalhos futuros, devido a sua complexidade e dificuldade de resolvê-los no curto tempo de um mestrado.

Para validar a estrutura proposta e a sua utilização no estudo de caso e nos experimentos, o mapeamento entre a ontologia de SDW e a ontologia de domínio foi efetuado manualmente.

1.4 Metodologia

Este trabalho foi desenvolvido através de uma pesquisa aplicada. Foi efetuado um desenvolvimento seguido de avaliações empíricas para verificação das hipóteses estabelecidas.

As etapas para a realização foram as seguintes:

1. Foi efetuada uma revisão bibliográfica nas áreas de DW, GIS, data warehouses espaciais, técnicas de representação de conhecimento para descrição de estrutura, conteúdo, recursos de processamento e interfaces visuais para bancos de dados.
2. Foram analisados e selecionados os sistemas de informação da EPA-GRI e seus parceiros que poderiam servir como fontes de dados para sistemas de suporte a decisão. Foi estudada a documentação destes sistemas, foram realizadas reuniões com usuários e equipe gerencial e foi analisada a funcionalidade e o conteúdo de cada sistema, suas possibilidades de integração e as necessidades de tratamento da informação.
3. Após esta análise dos sistemas existentes, alguns foram selecionados como base e foi desenvolvido um SDW, visando a aplicação como estudo de caso.
4. Foi definida uma proposta de arquitetura para a análise de informação em SDW que usasse técnicas de representação de conhecimento e interfaces visuais para facilitar o seu uso por especialistas de domínio.
5. Foi desenvolvida uma interface visual baseada em conhecimento para facilitar e orientar o uso de recursos de análise de informação em SDWs e permitir a realização de experimentos com usuários do domínio.
6. Foram realizados experimentos de usabilidade e análise de resultados, visando avaliar a viabilidade da interface proposta.

1.5 Organização do Trabalho

O Capítulo 2 apresenta a fundamentação teórica do trabalho, descrevendo Data Warehouses, Data Warehouses Espaciais, Interfaces Gráficas de Consulta e a aplicação de Web Semântica em Data Warehouses.

O Capítulo 3 apresenta um SDW para agricultura que serve como um estudo de caso para a apresentação deste trabalho, criado com dados do domínio agrícola.

O Capítulo 4 descreve a arquitetura e o modelo conceitual do sistema S^2DW , incluindo a ontologia de SDW e uma ontologia de domínio para agricultura, que é utilizada para ilustrar a operação do sistema; descreve também a interface baseada em conhecimento do S^2DW e mostra como o usuário interage com esta interface, através de exemplos de análise de informação sobre o SDW agrícola.

O Capítulo 5 descreve os experimentos que foram efetuados com usuários para avaliar a usabilidade da interface proposta e analisa os resultados.

Finalmente, o Capítulo 6 discute os trabalhos relacionados e o Capítulo 7 apresenta as conclusões e a sugestão de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

A integração dos dados de diversos sistemas dentro de uma organização, pelo uso de tecnologias e metodologias existentes na área da tecnologia da informação, fornece mecanismos que permitem uma visão estratégica integrada do conjunto de informações existentes nestes diversos sistemas.

Sistemas de Apoio à Decisão se baseiam nestes mecanismos, permitindo a recuperação e a disponibilização da informação de maneira rápida, eficiente e segura. Esta visão integrada, combinada com ferramentas de análise de dados e de representação geográfica, fornece à equipe gerencial de uma organização uma importante ferramenta para o processo de tomada de decisão, permitindo um aprimoramento do processo decisório.

De acordo com (FIDALGO, 2005), estas ferramentas podem ser assim resumidas:

- DW - uma base de dados especialmente modelada e carregada para prover dados para suporte a tomada de decisões;
- OLAP - uma categoria de software especializada para processar consultas multidimensionais sobre os dados de um DW;
- SIG - uma ferramenta específica para apoiar a aquisição, manipulação, análise, consulta e representação de dados espacialmente referenciados.

2.1 Data Warehouses

Data Warehouse (DW) (KIMBALL; ROSS, 2002; CHAUDHURI; DAYAL, 1997; INMON et al., 2008) é um dos principais componentes usados em um ambiente de suporte à decisão. As principais características de um DW são (INMON et al., 2008):

1. ser orientado ao assunto (deve armazenar dados correspondentes aos fatos e não sobre as transações que geraram os fatos);
2. ser perfeitamente integrado (deve compatibilizar dados de diferentes origens);
3. ser variante no tempo (deve manter um histórico dos seus dados);
4. não ser volátil (deve evitar atualizações de seus dados, dando suporte basicamente a cargas e consultas).

O processo de criação de um Data Warehouse é composto de diversas fases, sendo que as principais são (FIDALGO, 2005):

1. a extração, transformação e limpeza dos dados do ambiente transacional para a carga no ambiente decisório;
2. o gerenciamento de metadados;
3. a consulta e análise dos dados.

2.1.1 Modelo Dimensional de Dados

Como um DW deve ser implementado é tema de uma grande discussão entre dois autores da área (FIDALGO, 2005): Kimball e Inmon. Estes divergem principalmente sobre como implementar o componente ODS (Operational Data Store). (KIMBALL; ROSS, 2002) defendem que o ODS é um depósito histórico e frequentemente carregado de dados detalhados e integrados que constituem o nível atômico do DW. Por outro lado, (INMON; IMHOFF; BATTAS, 1995) descrevem o ODS como um repositório volátil/temporário de dados correntes e detalhados para o DW. Determinar qual das duas estratégias é a mais apropriada depende diretamente da equipe de implementação e do escopo do DW a ser construído.

Um Data Warehouse é composto por Data Marts. Data Marts são DW setoriais de uma organização, com foco em um segmento ou assunto específico. Como os Data Marts corresponderem a DW setoriais ou departamentais, estes consomem menos recursos e tempo para serem implementados (KIMBALL, 1998; INMON et al., 2008).

O modelo de dados do DW (ou de um Data Mart) é normalmente implantado usando o modelo estrela. O modelo estrela apresenta uma tabela central, chamada de tabela de fatos, e outras tabelas complementares, as tabelas de dimensão. No modelo estrela, as tabelas fatos são compostas pelas chaves primárias das dimensões e por informações de medidas, que correspondem à agregação de valores referentes à combinação das dimensões de uma determinada consulta. A tabela de fatos é normalizada. Já as tabelas de dimensões correspondem às tabelas que descrevem as informações que compõem o conteúdo das tabelas fato. Estas tabelas de dimensões normalmente não são normalizadas.

Além do modelo estrela, existe também o modelo flocos de neve, o qual difere do anterior por ter dimensões totalmente normalizadas. (KIMBALL, 1998) não aconselha o uso do modelo flocos de neve, pois sua economia de armazenamento é insignificante (aproximadamente 1%) e seu desempenho

será sempre inferior ao estrela, pois exige um número maior de junções a serem feitas.

Segundo (KIMBALL; ROSS, 2002; KIMBALL, 1998) as medidas de uma tabela de fatos devem representar, de preferência, entradas continuamente valoradas (diferentes a cada leitura) e aditivas. Porém, podem também existir medidas semi-aditivas e não aditivas. As aditivas são corretamente adicionadas por todas as dimensões e representam medidas de atividades, por exemplo, vendas a dinheiro ou em unidades. As semi-aditivas não são corretamente adicionadas ao longo de certas dimensões (especialmente sobre a dimensão Tempo) e correspondem a medidas de intensidade, por exemplo, saldos bancários, níveis de estoque e temperatura. Por fim, as não-aditivas não podem ser adicionadas e representam valores textuais que só podem ser contados ou no máximo, terem seus registros impressos um-a-um (e.g. Temperatura descrita como frio, agradável ou quente). Porém, o uso de tais medidas não tem sido recomendado (KIMBALL; ROSS, 2002).

2.1.2 *Processamento Analítico On-line (OLAP)*

On-Line Analytical Processing (OLAP) (FIDALGO, 2005; CHAUDHURI; DAYAL, 1997) é uma categoria de software específica para realizar consultas multidimensionais sobre dados previamente agregados e materializados que são extraídos do DW. Estas consultas devem ocorrer com alto desempenho, consistência e interatividade, e seus dados resultantes devem ser interpretados em uma variedade de visões complementares.

OLAP difere do processamento transacional (On-Line Transaction Processing (OLTP)), o qual é destinado a armazenar e recuperar dados de suporte operacional, ou seja, dados frequentemente normalizados e não históricos que essencialmente são manipulados para permitir a análise de registros atômicos, com consistência e eficiência.

As arquiteturas das ferramentas OLAP existentes no mercado são basicamente de três tipos (CHAUDHURI; DAYAL, 1997): 1) Relacional OLAP - (ROLAP), 2) Multidimensional OLAP - (MOLAP) e 3) Híbrido OLAP - (HOLAP). ROLAP realiza a agregação e a materialização dos dados sobre uma estrutura física relacional. MOLAP faz o mesmo que ROLAP usando tecnologia proprietária de banco de dados multidimensionais sobre uma estrutura física baseada em matrizes n-dimensionais. HOLAP integra as duas anteriores, onde os dados atômicos permanecem nas estruturas relacionais, enquanto os dados agregados são replicados e materializados nos bancos de dados multidimensionais.

Independentemente das arquiteturas apresentadas, as ferramentas OLAP manipulam os dados que são extraídos do DW para posteriormente serem agregados e materializados em estruturas multidimensionais lógicas conhecidas como cubos de dados (THOMSEN, 1997).

A estrutura de um cubo é baseada na estrutura dimensional do modelo estrela e, por isso, é definida em termos de dimensões e medidas.

Exemplificando, numa consulta do total de animais abatidos em propriedades agrícolas, por tipo de animal, por município e por ano, identificam-se como dimensões tipo de animal, município (localização geográfica) e ano (tempo). O total de animais abatidos é a medida que comporia o cubo Abate.

Os elementos das dimensões são chamados de membros e estes são organizados, agregados e materializados em níveis hierárquicos (THOMSEN, 1997). Por exemplo, a dimensão tipo de animal pode ser composta por nome do animal, sexo e faixa de idade.

Um cubo de dados pode ser implementado segundo o operador CUBE (GRAY et al., 2007), existente nos SGBD comerciais. Este operador generaliza os construtores relacionais de agregação para oferecer suporte às análises multidimensionais. O operador CUBE produz todas as combinações possíveis de agregações sobre as dimensões especificadas em uma visão multidimensional. Se uma visão engloba n dimensões, então o operador CUBE produzirá um total de 2^n combinações de subconjuntos destas dimensões. Cada um desses subconjuntos representa um minicubo (ou cuboid) que consiste em um nível de agregação das dimensões.

2.2 Data Warehouses Espaciais

Spatial Data Warehouses (SDW) combinam funcionalidades de Data Warehouses (DW) convencionais e de bancos de dados espaciais em sistemas para a representação e análise de informação espacial (RAO et al., 2003; RIVEST et al., 2005; MALINOWSKI; ZIMÁNYI, 2007; BIMONTE; TCHOUNIKINE; MIQUEL, 2007).

Tais sistemas fornecem recursos para coletar, integrar, armazenar e analisar dados espaciais sobre um modelo dimensional estendido (MALINOWSKI; ZIMÁNYI, 2007; FIDALGO et al., 2004). SDWs requerem estruturas de armazenamento e índices adequados para dados espaciais, recursos para seleção e agregação desses dados e visualização dos resultados em mapas, por exemplo.

Extensões geográficas podem ocorrer tanto na tabela fato quanto nas dimensões de um cubo dimensional de um SDW. Membros de níveis

de dimensão referindo-se a entidades geográficas podem ser associados às representações geométricas dessas entidades (e.g., polígonos representando os territórios de estados e municípios). Uma medida espacial eventualmente também pode ser representada por uma geometria (e.g., ponto geográfico onde ocorre um fenômeno).

A análise de dados em SDWs pode envolver predicados espaciais e funções de agregação de dados espaciais. (SILVA et al., 2008) propõe uma classificação das funções de agregação de dados espaciais a partir da combinação de classificações existentes em DWs (KIMBALL; ROSS, 2002) e dados espaciais (EGENHOFER; HERRING, 1992). (RUIZ; TIMES, 2009) propõe uma taxonomia para operadores SOLAP, além de um conjunto de operadores SOLAP e sua sintaxe.

(RAO et al., 2003) propõe métodos para a eficiente agregação de dados espaciais pelo uso de interação de usuários para efetuarem análise de informação em DWs.

(BIMONTE; TCHOUNIKINE; MIQUEL, 2005; BIMONTE; TCHOUNIKINE; MIQUEL, 2006; MALINOWSKI; ZIMÁNYI, 2004; BIMONTE et al., 2006) discutem a definição de modelos para dados multidimensionais e representação de dados espaciais, buscando soluções para problemas de SOLAP que contemplam o suporte para medidas espaciais.

(DAMIANI; DAMIANI; WANG, 2008; MALINOWSKI; ZIMNYI, 2008; MALINOWSKI; ZIMÁNYI, 2006) falam sobre a modelagem de SDWs, abordando aspectos como modelos espaciais para representação de dados multidimensionais com dados convencionais e espaciais e funções de agregação para medidas espaciais.

2.2.1 Operadores

2.2.1.1 Operadores de Data Warehouses Convencionais

A estrutura multidimensional do cubo é manipulada por um conjunto de operadores OLAP, dentre estes, os principais são:

- *Roll-up*: agregação ou generalização dos dados para um nível com maior granularidade;
- *Drill-down*: desagregação ou especialização dos dados para um nível com menor granularidade;
- *Slice, dice*: seleção e projeção de parte de um cubo;

- *Pivoting*: rotação dos eixos do cubo para visualização dos resultados de uma consulta.

2.2.1.2 Operadores de Dados Espaciais

Segundo (EGENHOFER; FRANZOSA, 1991), as relações entre dados espaciais podem ser enquadradas em três categorias: topológicas, métricas e as de ordem parcial ou total.

Mais especificamente dentro da categoria de relações topológicas, as que podem ser efetuadas entre objetos espaciais estão as propostas por (EGENHOFER; HERRING, 1992), descritas a seguir e ilustradas na Figura 1:

- *disjoint* - dois elementos “a” e “b” são disjuntos se a intersecção entre eles é vazia;
- *touches* - dois elementos “a” e “b” se tocam quando existe ao menos um ponto em comum entre “a” e “b”;
- *contains* - um elemento “a” contém um elemento “b” quando “b” está dentro de “a” e suas fronteiras não se tocam;
- *covers* - o elemento “a” cobre o elemento “b” quando suas fronteiras se tocam e “b” está dentro de “a”;
- *equal* - dois elementos “a” e “b” são iguais se “a” está contido em “b” e “b” está contido em “a”;
- *overlaps* - elementos “a” e “b” se sobrepõem se a região que representa a intersecção entre elas tiver a mesma dimensão que as instâncias e as instâncias não forem iguais;
- *inside* - um elemento “a” está dentro de um elemento “b” quando a intersecção entre eles é igual a “a” e a intersecção entre o interior de “a” e o exterior de “b” é vazia;
- *covered by* - um elemento “a” está coberto por um elemento “b” quando suas fronteiras se tocam e “a” está dentro de “b”.

Para SDW, (RIGAUX; SCHOLL; VOISARD, 2002) classificou os operadores pela combinação de duas classificações: de acordo com o número de operandos: unários, binários e n-ários; e de acordo com o tipo de dado retornado: booleano, escalar e espacial. A classificação proposta é a seguinte:

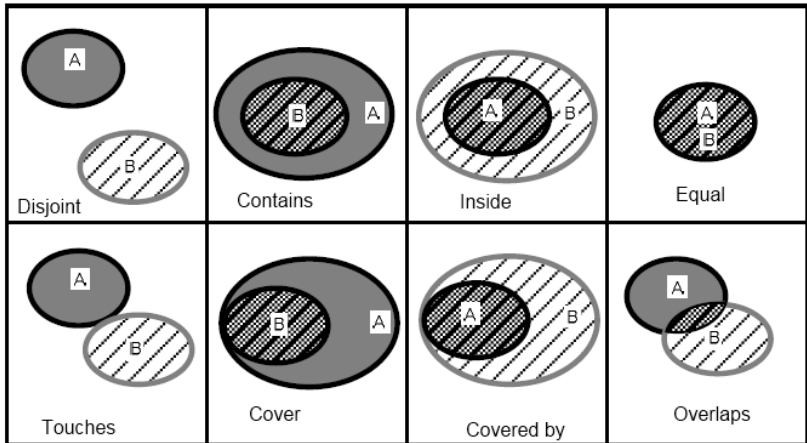


Figura 1: Relações topológicas entre objetos espaciais

- Unário com resultado booleano - Testam um objeto espacial para uma determinada propriedade. Ex.: O objeto é convexo? O objeto é válido? O objeto está conectado ou é um anel?
- Unário com resultado escalar - As operações mais frequentes são o cálculo do comprimento, tamanho da área, perímetro de um objeto espacial. Ex.: Qual a área? Qual o perímetro? Qual o comprimento? Qual a largura de um objeto?
- Unário com resultado espacial - Operações de transformação topológica (rotação, translação, mudança de escala e simetria); transformações de um objeto d-dimensional em um objeto com uma dimensão maior ou menor; extração de objetos e operações de buffer. Ex.: Pontos ao redor de; área recortada de um objeto; boundary; minimal bounding rectangle
- Binário com resultado booleano - Também chamadas de predicados espaciais, são a base para consultas espaciais e para seleção e join espacial. Ex.: Toca? Cruza? Sobrepõe? Intersecta?
- Binário com resultado escalar - Distancia é uma típica operação desta classe. Ex.: Qual a distância? Qual a área?

- Binário com resultado espacial - Operações de conjuntos aplicadas a um conjunto de pontos. Ex.: União, Intersecção, Diferença;
- N-ário com resultado espacial - Operações variadas são parte desta categoria geral. Ex.: Clipping de vários objetos; diagramas de Voronoy;

2.2.2 Funções de Agregação

Em Data Warehouses a agregação de dados se dá através de funções divididas em:

- Distributivas - funções que podem ser calculadas pela divisão dos dados em conjuntos menores, aplicando a função em cada conjunto e ao final aplicando a mesma função aos resultados de cada conjunto. Ex.: contagem, mínimo, máximo, soma, etc.;
- Algébricas- funções que podem ser calculadas pela aplicação de uma função algébrica com n elementos, onde cada elemento é obtido através da aplicação de uma função distributiva. Ex.: média, desvio padrão, etc.;
- Holísticas - funções onde não existe uma função algébrica que defina seu cálculo. Ex.: mediana, maior frequência, rank, etc.

A partir da combinação dos operadores de SIG e das funções de agregação provenientes do DW, (SILVA et al., 2008) propõe uma nova classificação para as funções de agregação em Data Warehouses Espaciais. Esta proposta classifica as funções de acordo com o tipo da função (Distributivas, Algébricas e Holísticas) e com o tipo de retorno (Escalar ou Espacial).

As funções de agregação em Data Warehouses Espaciais se classificam em:

- Distributiva escalar: *CountTouches()*, *CountAtNorth_of()*, *MaxIntersects()*, *MinAtNorth()*, etc;
- Distributiva espacial: *SumTouches()*, *SumAtNorth_Of()*, *SumAtNorthWest_O()*, *SumDisjoint()*, etc.;
- Algébrica escalar: *AvgArea()*, *StdvPerimeter()*, *MaxNArea()*, etc.;
- Algébrica espacial: *MaxNAtNorth_Of()*, *MaxNAtSouth_Of()*, etc.;

- Holística escalar: RankArea(), RankPerimeter(), MedianArea(), ModeArea(), etc.;
- Holística espacial: NNearestNeighbor(), NDistantNeighbor(), Voronoi(), etc.

2.3 Interfaces Gráficas de Consulta a Bancos de Dados

A seguir são descritos trabalhos que abordam o uso de interfaces gráficas de consulta a bancos de dados. Algumas das ideias neles contidas são utilizadas no contexto deste trabalho.

2.3.1 Interfaces para Banco de Dados Convencionais

Na área de interfaces de consulta para bancos de dados, as primeiras ideias surgiram com (ZLOOF, 1976; ZLOOF, 1977). O objetivo desses trabalhos é fornecer ao usuário final um meio de efetuar consultas e manutenções em bancos de dados sem a necessidade de grandes conhecimentos sobre a estrutura de um banco de dados e operadores para acessá-lo. Utilizando exemplos para o conteúdo dos dados, o usuário especifica e resolve as suas necessidades através de interações com o sistema.

(CATARCI et al., 1997; BATINI et al., 1991; BATINI et al., 1992; CATARCI; SANTUCCI, 1994) apresentam sistemas visuais de consulta também com o objetivo de facilitar a interação do usuário com o banco de dados, facilitando o processo de aprendizado do esquema existente. Segundo os autores, o uso de componentes visuais nos procedimentos de visualização de dados e escrita de consultas fornece um meio adequado para que o usuário final especifique suas necessidades e obtenha os resultados desejados.

Propostas mais atuais para permitir aos usuários visualizarem a estrutura e especificarem suas próprias consultas especificamente em bancos de dados foram propostas por (SPAHN et al., 2008; RISHE et al., 2000; TERWILLIGER; DELCAMBRE; LOGAN, 2007).

2.3.2 Interfaces para GIS

Em GIS, (AUFAURE-PORTIER; TRÉPIED, 1996; BONHOMME; TREPIED; LAURINI, 1999; CALCINELLI; MAINGUENAUD, 1994) apresentam propostas de interfaces gráficas.

(AUFAURE-PORTIER; TRÉPIED, 1996) ressalta que pelo fato de GIS armazenar grande quantidade de dados complexos, o uso de interfaces ade-

quadas ao usuário é um aspecto fundamental de pesquisa. O trabalho classifica as linguagens de consulta existentes e suas características: (i) linguagem natural é uma abordagem atrativa, mas encontra problemas em ambiguidade; (ii) o uso de extensões em SQL exige conhecimento técnico do usuário final; (iii) uso de hipermídia facilita a navegação mas não pode ser usada isoladamente e (iv) linguagens visuais são promissoras mas apresentam problemas. Conclui que a solução ideal é o uso de uma interface híbrida com uma combinação harmoniosa destas características.

(CALCINELLI; MAINGUENAUD, 1994) apresenta a linguagem de consulta Cigales, onde a semântica de uma consulta de um usuário final é expressa através de desenhos.

(BONHOMME; TREPIED; LAURINI, 1999) apresenta uma extensão à linguagem Lvis para sistemas de informações geográficas, que utiliza metáforas visuais para a representação visual das consultas, já que dados espaciais são gráficos por natureza.

2.4 Web Semântica e Data Warehouses

Estudos sobre a aplicação das ideias da Web Semântica em DWs são encontrados em diversos trabalhos (XIE et al., 2007; SKOUTAS; SIMITSIS, 2006; SELL et al., 2008).

(XIE et al., 2007) propõe a utilização da Web Semântica para representar metadados do negócio, na definição do DW e data marts, com a utilização de uma OWL (Web Ontology Language) estendida. Apresenta a proposta de uma ferramenta que visa facilitar a rápida adaptação às constantes mudanças existentes no ambiente de negócio. A ideia principal é explicitar semanticamente os metadados do negócio, ou seja, os modelos de dados das diversas fontes e do DW, através de uma linguagem OWL estendida. Com isso, um mapeamento dos metadados é construído, e quando uma análise é requisitada, um data mart customizado pode ser automaticamente gerado com o auxílio deste conhecimento adquirido.

O processo de criação de um DW é uma tarefa custosa e demorada. Além disso, os processos típicos de consulta utilizados através de Business Intelligence (BI) não são muito flexíveis para tratar rapidamente as constantes demandas no negócio de uma organização. Uma das principais dificuldades é que a semântica do negócio está apenas registrada na mente do projetista e codificada nas implementações. Esta semântica do negócio pode ser apenas parcialmente representada nos metadados de negócio que proveem uma descrição orientada ao negócio do conteúdo do DW e de uma representação

formal dos requisitos de análise. Os metadados do negócio são de dois tipos: o modelo de dados conceitual da empresa e o modelo multidimensional.

O modelo de dados conceitual da empresa é um modelo para organizar a terminologia do negócio de uma forma semântica; é uma visão de como o negócio funciona e é constituído de conceitos de negócio, atributos destes conceitos e relacionamentos sobre conceitos. Já o modelo multidimensional é um modelo que define os requisitos analíticos para uma aplicação de BI. É uma visão de como o negócio é medido, sendo composto de medidas e dimensões.

Num processo normal de geração dos dados, os metadados de negócio se tornam artefatos técnicos que não são entendidos por usuários de negócio. Não há distinção entre as transformações necessárias para a semântica do negócio e as necessárias por aspectos técnicos. Com o uso da Web Semântica, a OWL é apropriada para representar a semântica do negócio de uma maneira formal.

A proposta de (XIE et al., 2007) consiste na criação de uma ferramenta (Enterprise Information Asset Workbench (EIAW)), cuja ideia principal é tornar a representação da semântica do negócio de forma explícita no DW através da representação dos metadados de negócio com uma linguagem OWL estendida. Baseada em uma semântica de negócio explícita, a ferramenta suporta a geração de uma aplicação de BI baseada no DW, pela execução de alguns passos, considerando a existência prévia de um modelo de dados conceitual e um DW empresarial. Os passos são os seguintes:

1. Usuários de negócio definem os requisitos de análise usando termos do negócio a partir do modelo de dados conceitual;
2. Usuários de TI criam mapeamentos entre os termos de negócio envolvidos no modelo multidimensional para o esquema do DW;
3. O sistema gera automaticamente um data mart customizado com os dados agregados e um cubo OLAP;

A principal vantagem deste processo de geração é a separação dos conceitos de negócio dos conceitos de TI. Usuários do negócio podem organizar seu conhecimento de negócio e expressar seus requisitos de análise usando termos que lhe são familiares. Usuários de TI têm seu foco na resolução dos mapeamentos dos termos de negócio do DW para a visão técnica e não se preocupar com as medidas e dimensões do negócio.

As abordagens convencionais exigem um conhecimento completo do negócio por parte da equipe de TI e necessitam sempre que o processo seja

refeito cada vez que medidas e dimensões mudam. Nesta abordagem, os usuários de negócio têm um meio eficiente de descrever suas necessidades sem a exigência de constantes reuniões com a equipe de TI. Além disso, no caso de alteração de requisitos, usuários de negócio podem alterar as definições de medidas e dimensões, gerando um novo data mart sem a necessidade de envolvimento da equipe de TI, caso os mapeamentos já estejam criados.

A arquitetura proposta é dividida em três módulos principais: Business Metadata Builder, Mapping Builder e Deployment Engine.

O Business Metadata Builder auxilia o usuário do negócio a criar e editar os metadados de negócio no ambiente do DW. O modelo de dados conceitual é chamado de Business Concept Model (BCM) e o modelo de dados do DW é chamado de Analysis Solution Model (ASM). O ASM é definido usando os termos do BCM.

O Mapping Builder auxilia os usuários de TI a construir os mapeamentos dos termos de negócio para o esquema do DW. São necessários apenas os mapeamentos para os termos de negócio que aparecem nas definições dos ASMs. O construtor de mapeamentos permite o reuso e a construção incremental de mapeamentos. O mapeamento de um termo de negócio pode ser compartilhado se aparecer na definição de outros ASMs. Com os diversos mapeamentos criados ao longo do tempo, cada vez mais os novos ASMs serão compostos de termos já mapeados anteriormente, e no caso em que todos os termos já estiverem definidos previamente para um ASM, este pode ser construído automaticamente, sem envolvimento do pessoal de TI.

O Deployment Engine gera automaticamente o data mart com dados agregados vindos do DW através dos mapeamentos e definições feitas para o ASM. Ele pode também gerar um cubo para este data mart, permitindo as análises OLAP.

A conclusão de (XIE et al., 2007) é que explicitar a semântica de negócio permite aos usuários de negócio organizar o seu conhecimento e expressar os seus requisitos de análise e aos usuários de TI construírem os mapeamentos somente para os aspectos técnicos. Eles entendem que a OWL estendida é uma boa linguagem para expressar a semântica para sistemas de data warehouse, porém identificam que uma formalização completa desta extensão da linguagem é um problema ainda em aberto devido à complexidade de definições de medidas e dimensões.

(SKOUTAS; SIMITSIS, 2006) propõe a utilização de ontologias para definir o correto mapeamento entre os atributos das fontes de dados e das tabelas do DW no processo de ETC. Segundo os autores, no ETC, que é uma

das mais importantes partes da construção de um DW, existe a necessidade de se estabelecer um correto mapeamento entre os atributos das fontes de dados e os atributos das tabelas do DW. Este modelo de ETC proposto deve possuir facilidades na redefinição e revisão, permitindo a comunicação entre as partes envolvidas. Para que isto seja atingido, a proposta é que ontologias sejam utilizadas, possibilitando fornecer um alto grau de automação do processo.

O processo de ETC busca dados de diversas fontes, efetua procedimentos de limpeza, de customização e efetua a inserção no DW. Todos estes procedimentos são executados manualmente pelo projetista ou administrador do DW. Este processo é dirigido pela semântica das fontes de dados e pelas restrições e requisitos do DW. A proposta de usar ontologias neste processo visa facilitar a seleção das informações relevantes das fontes de dados e auxiliar no processo de transformação para a informação necessária ao DW. Com a construção de uma ontologia para criação de um workflow do processo e com a formalização e explicitação da semântica dos esquemas das fontes de dados e do DW, se busca automatizar grande parte do processo.

A ideia do uso de ontologias é que elas fornecem uma forma elegante de prover raciocínio necessário para a automação das transformações exigidas pelo processo; ontologias auxiliam na resolução de um dos grandes desafios da retaguarda de um DW, a heterogeneidade.

A heterogeneidade pode ser de duas formas: (1) estrutural, onde diferentes sistemas de informação armazenam conteúdos em estruturas distintas e (2) semântica, onde o significado dos itens de informação, ou seja, da própria informação tem que ser compreendido entre os sistemas; esta heterogeneidade semântica pode ser por conflitos de interpretação, conflitos de escala ou conflitos de nomenclatura. Para a solução do problema da heterogeneidade e facilitar a construção de um workflow de ETL a proposta é:

1. a criação de um vocabulário comum de aplicação para tratar com os diferentes esquemas das fontes de dados;
2. prover um método para registro das fontes de dados e do DW relacionados ao vocabulário criado;
3. gerar uma ontologia de aplicação que contenha informações sobre os mapeamentos de atributos e as transformações necessárias;
4. gerar um projeto conceitual de ETL, que derive os atributos das fontes de dados para o DW e faça as transformações.

Ainda segundo os autores, a partir da ontologia criada e dos mapeamentos, podem ser descritas as transformações necessárias para integrar

os dados das diversas fontes e gerar os dados no DW. Um procedimento necessário é determinar de quais fontes e de quais atributos/relações as informações precisam ser retiradas para preencher cada atributo/relação do DW. Também devem ser determinadas quais são as transformações necessárias para integrar os dados das fontes para o destino.

A conclusão encontrada por (SKOUTAS; SIMITSIS, 2006) é que a abordagem facilita o processo de seleção de informações relevantes das fontes de dados e o processo de transformação para o preenchimento do DW. O uso de ontologias para formalizar e especificar explicitamente a semântica dos esquemas das fontes de dados possibilita a automação em alto grau do processo de criação do fluxo de ETL.

(SELL et al., 2008) propõem uma arquitetura baseada em Web semântica para ferramentas de análise dimensional, usando ontologias para integrar a semântica do negócio, dos dados e dos serviços oferecidos, com o objetivo de apresentar as informações de acordo com o vocabulário do usuário.

Esta proposta é composta de:

- uma ontologia de domínio, que possui terminologia de acordo com o negócio e que visa dar suporte à ontologia de serviços e contextualizar as fontes de dados na ontologia de BI. Desta forma usuários poderão navegar usando conceitos do negócio em vez de descrições técnicas, já que todas as relações, regras e expressões utilizadas nos processos de análise estão relacionadas aos conceitos do negócio;
- uma ontologia de Business Intelligence, que modela os conceitos usados para descrever a organização dos dados em suas fontes e mapeia estes dados para os conceitos da ontologia de domínio. Ela permite inferências para estender os resultados das consultas e suportar a apresentação dos dados de acordo com o perfil do usuário, além de guiar as alterações das instâncias da ontologia de domínio e permitir a reescrita de consultas;
- o “Instance Manager”, que suporta a replicação dos dados das dimensões dentro da ontologia de domínio, guiado pelos mapeamentos descritos na ontologia de BI. Implementado como uma coleção de classes Java que se conectam à fonte de dados e extraem os dados necessários para gerar as instâncias no servidor de ontologias;
- o “Ontology Manager”, que provê um conjunto de métodos necessários para consultar e fazer inferências sobre o repositório de ontologias,

escondendo a complexidade das linguagens de consulta de ontologias e dos repositórios de ontologias dos desenvolvedores, permitindo uma pequena integração entre as ferramentas analíticas, módulos de consulta e repositórios de ontologia, provendo certa flexibilidade para mudanças;

- o “Service Ontology” que modela os conceitos usados para descrever serviços web semânticos;
- o “Service Manager” que permite o reuso de código para aprimorar a funcionalidade das ferramentas de análise, além de habilitar a composição e orquestração de serviços web semânticos, funcionando como uma interface para uma infraestrutura destes serviços;
- o “Analysis Manager” que provê acesso para todos os componentes: ao “Ontology Manager”, ao “Service Manager” e ao “Instance Manager” e disponibiliza uma série de funcionalidades para OLAP sobre as fontes de dados e implementa o processo de reescrita de consultas.

A proposta foi aplicada em um protótipo em Java, o OntoDSS, que é uma ferramenta analítica para tomada de decisão. Ela possui dois módulos:

- o “Analysis Definition” que dá suporte ao usuário auxiliando-o na definição do conteúdo e no formato das análises. Usuários navegam na Ontologia BI para selecionar os itens de dados a analisar. O vocabulário é baseado nas fontes de dados e no perfil do usuário: um reitor visualiza as pessoas como pesquisadores e estudantes; um presidente de uma companhia visualiza como experientes e novatos;
- o “Exploration Recommender” que dá suporte ao usuário para composição de serviços em explorações mais complexas, recomendando relações, inferências e serviços que possam fornecer diferentes perspectivas ao analista.

O artigo de (SELL et al., 2008) apresenta uma proposta de integração da semântica do negócio, dados e serviços, com informações apresentadas ao usuário no seu vocabulário e permite que as composições efetuadas sejam utilizadas para constituir uma base de conhecimento, permitindo reuso e automação.

Este trabalho baseou-se inicialmente na ideia proposta por (XIE et al., 2007), visando aplicá-la em SDWs.

2.5 Considerações

Este capítulo apresentou a fundamentação teórica que serviu de base para a realização deste trabalho. A representação do conhecimento de um domínio através de ontologias fornece aos usuários de um SDW, uma forma de facilitar a compreensão do seu conteúdo além de auxiliar na identificação e seleção dos operadores e funções de agregação espacial que podem ser aplicados nos processos de análise de informação.

O próximo capítulo apresenta um SDW para agricultura, criado com dados do domínio agrícola e baseado nesta integração das ontologias com SDWs e que foi utilizado como estudo de caso neste trabalho.

3 ESTUDO DE CASO - UM SDW COM DADOS AGRÍCOLAS

O trabalho experimental desta dissertação teve início com a identificação e análise de sistemas da Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (EPAGRI) e a criação de um Data Warehouse Espacial sobre propriedades e atividades agrícolas, gerado com dados selecionados de alguns destes sistemas. Este SDW teve por objetivo consolidar conhecimentos do autor sobre a tecnologia e fornecer um instrumento para aplicação e validação desta proposta. Com a criação deste SDW, foi possível estabelecer um estudo de caso aplicado a dados do domínio agrícola.

A seguir são apresentadas informações sobre a EPAGRI, os sistemas da empresa considerados para servir como fontes de dados para o SDW e os critérios adotados para a sua construção.

3.1 A EPAGRI

A EPAGRI, empresa pública criada em 1991, tem por objetivos promover a preservação, recuperação e conservação dos recursos naturais, buscar a competitividade da agricultura catarinense frente aos mercados globalizados e promover a melhoria da qualidade de vida do meio rural e pesqueiro. Ela possui atualmente 2.136 funcionários e estrutura física nos 293 municípios do estado de Santa Catarina, distribuída em 27 Gerências Regionais, 14 unidades de pesquisa, 2 campos experimentais, 9 Estações Experimentais, 5 Centros de Estudos e Pesquisas - Centro de Pesquisa para a Agricultura Familiar (CEPAF), Centro de Socioeconomia e Planejamento Agrícola (CEPA), Centro de Excelência em Informações de Recursos Ambientais e de Hidrometeorologia (CIRAM), Centro de Desenvolvimento em Aqüicultura e Pesca (CEDAP), Parque Ecológico Cidade das Abelhas (PECA) - 3 unidades de beneficiamento de sementes e 12 centros de treinamento (EPAGRI, 2010).

3.1.1 A Informação na EPAGRI

A EPAGRI possui uma vasta quantidade de informação, referente a pesquisas e projetos necessários ao atendimento de seus clientes. Esta informação precisa ser adequadamente analisada e transmitida ao público interno e externo, para que possa ser utilizada nos procedimentos de atendimento aos clientes.

Para melhor acompanhar e atender às demandas de informação na empresa, ela conta com uma Gerência de Informação, composta de uma equipe

técnica e estrutura física de TI distribuída nos seguintes setores: desenvolvimento e administração de banco de dados, administração de redes, suporte e biblioteca.

A seguir são descritos os processos de coleta e difusão da informação na empresa.

3.1.1.1 A Coleta da Informação

A estrutura de trabalho da EPAGRI na sede e em suas unidades possui a distribuição em programas e projetos, para adequar as ações à sua missão. Os programas agrupam critérios de afinidade das ações e a abrangência de um campo do conhecimento; os projetos podem ser divididos em subprojetos afins e que são formados por experimentos (pesquisa) ou ações (extensão rural). Estes experimentos e ações estão distribuídos em uma grande diversidade de temas e trabalhos, tais como biotecnologia, meteorologia, planejamento agrícola, inclusão digital, capacitação de agricultores, entre outros.

3.1.1.2 A Difusão da Informação

Cada um destes programas gera e trabalha com informação; o volume de informação é significativo e para este conjunto de informações a EPAGRI possui diversos mecanismos para distribuir esta informação ao seu público interno e externo. Informações de outras fontes geradoras também são muito utilizadas na empresa, obtidas através de parcerias ou acesso a outros bancos de dados.

A EPAGRI possui sistemas corporativos que centralizam as informações de execução física e financeira da empresa, informações dos atendimentos aos produtores, informações meteorológicas e informações sócio econômicas. Algumas destas informações são de uso restrito aos técnicos da empresa ou ao nível estratégico, mas existe a preocupação de levar a informação ao público externo. Para isso, a empresa utiliza-se de diversas mídias como programas de rádio e televisão, geração de conteúdo em CD-ROM, bem como a disponibilização no site da empresa de livros, artigos e publicações em formato digital. A busca de novas mídias que possam ser utilizadas como mecanismos de divulgação das informações também merece atenção pela empresa.

3.1.1.3 A Infraestrutura de hardware e software

Este conjunto de informações concentra-se em diversos equipamentos distribuídos na rede interna da empresa; além de servidores de rede que efetuam os serviços de internet, email e serviços de armazenamento dos arquivos da estrutura da empresa, existem dois servidores de banco de dados Oracle 10G, sendo um instalado na sede e um no CIRAM.

3.2 Os Sistemas de Informação da EPAGRI

Em 2008, foi efetuado um estudo para identificar a situação atual da informação na empresa naquele momento, seus mecanismos de geração e difusão das informações e identificados os procedimentos em curso e os em desenvolvimento para o gerenciamento da informação. Este estudo consistiu em reuniões com a gerência de informação da EPAGRI e sua equipe técnica de TI. A documentação dos sistemas foi analisada quando disponível e alguns dados foram acessados e analisados.

Naquela oportunidade foram identificados os grandes sistemas da empresa, concluídos, em planejamento ou em fase de implantação. Estes sistemas são descritos a seguir.

3.2.1 SAFI - Sistema Administrativo-Financeiro

Sistematiza os principais processos administrativos, financeiros e orçamentários da EPAGRI; o SAFI permitirá a padronização, integração e melhoria dos processos administrativo-financeiros, assim como, através de seus relatórios e de indicadores financeiros, fornecerá informações gerenciais ao nível estratégico da EPAGRI. O sistema está em desenvolvimento, possuindo alguns módulos em operação.

3.2.2 SEPLAN - Sistema EPAGRI de Planejamento

Sistematiza a elaboração, aprovação, acompanhamento, reprogramação, avaliação de programas, projetos, experimentos e ações de extensão. Para cada figura programática podem ser armazenadas informações textuais, orçamento e metas. A partir dessas informações é possível a consolidação do orçamento técnico e do Programa Anual de Trabalho-PAT. O sistema está em operação e o planejamento técnico do próximo ano deverá ser elaborado usando este novo sistema.

3.2.3 SEATER - Sistema EPAGRI de ATER - Assistência Técnica e Extensão Rural

Sistematiza as atividades da assistência técnica e extensão rural. Cria um banco de dados integrado de produtores rurais, com informações de identificação do produtor, atendimentos recebidos, capacitações realizadas etc. A partir dos registros de ATER é possível a geração de diversos indicadores de desempenho da extensão rural, como produtos, família, atividades e resultados alcançados.

O sistema está com a implantação piloto na regional de Tubarão e em alguns Escritórios Municipais.

3.2.4 SEPRO - Sistema EPAGRI de Profissionalização

Sistematiza as atividades da profissionalização, permitindo a elaboração de um calendário de cursos, controle de inscrições através dos Centros de Treinamento ou dos Escritórios Municipais, cadastro de participantes, emissão de certificados, estatísticas, indicadores e relatórios para convênios da profissionalização. O sistema está em uso nos centros de treinamento, com implantação piloto nos escritórios municipais (inscrição do produtor)

3.2.5 Sistema de gestão de pessoas

Organiza todas as informações relativas ao quadro funcional da empresa; além das informações cadastrais e o histórico funcional dos funcionários, tem também o foco estratégico voltado ao público externo, na disponibilização da base curricular dos funcionários através da integração com a Plataforma Lattes Institucional-CV. O sistema está em fase de contratação da empresa para o desenvolvimento.

3.2.6 Preços Agrícolas

O Sistema de Preços Agrícolas é um sistema que permite o acompanhamento de preços pagos e recebidos pelos agricultores de Santa Catarina, numa larga variedade de produtos e insumos agropecuários. Os preços são coletados em diversos locais do estado, em pontos considerados pólos regionais. Têm periodicidades distintas de registro de preços, como diária, semanal, mensal ou anual. Possui uma grande série histórica destes preços permitindo o acompanhamento da evolução dos mesmos.

Está operacional, mas uma nova versão do mesmo está sendo elaborada dentro do SIAGRO, com a incorporação de novos recursos e reutilização total dos dados históricos.

3.2.7 SIAGRO - Sistema de Informações Agropecuárias

O SIAGRO visa integrar as diversas informações socioeconômicas relacionadas ao meio agrícola, fornecendo uma ferramenta de acesso padronizada a informações de fontes distintas, externas ou internas, permitindo ao público interno e externo um acesso simplificado, seguro e rápido a um grande volume de informações.

Este sistema permitirá a incorporação de dados de diversos trabalhos de pesquisa elaborados no EPAGRI/CEPA, relativos a pesquisas sócio-econômicas, cujos dados encontram-se em outros sistemas específicos ou em planilhas eletrônicas. Um módulo deste sistema incorporará o sistema de Preços Agrícolas, sendo que está sendo prevista a incorporação para este sistema também dos dados do LAC - Levantamento Agropecuário Catarinense. O sistema está em desenvolvimento, com conclusão prevista para breve. Grande parte do conjunto de dados que formará a série histórica de dados já existe e está cadastrado em outro sistema, do qual os dados serão convertidos para o SIAGRO.

3.2.8 LAC - Levantamento Agropecuário Catarinense

O LAC contém os dados de um censo efetuado no ano de 2003 em todos os estabelecimentos agropecuários de Santa Catarina, contendo uma série de informações destes estabelecimentos. O LAC é composto dos dados deste levantamento, não possuindo uma interface própria de consulta, com os dados sendo consultados diretamente no banco de dados, ficando isso ao encargo dos usuários que necessitam deste acesso.

3.2.9 SIGEO - Sistema Integrado de Informações Georeferenciadas

O SIGEO visa fornecer a sistematização de informações com representação através de mapas e gráficos, possibilitando uma melhor visualização e agrupamento de informações. O sistema está em desenvolvimento, sendo utilizado internamente para a geração de alguns mapas, e tem a conclusão prevista para breve. O SIGEO acessa dados de diversos bancos de dados para compor o seu conteúdo.

3.2.10 GECRED - Sistema de Gerenciamento de Crédito Agrícola

O GECRED é um sistema desenvolvido e mantido pela EPAGRI para o acompanhamento financeiro dos contratos de crédito celebrados entre a Secretaria da Agricultura e os produtores rurais de Santa Catarina. Controla todos os programas de financiamento gerenciados pela Secretaria da Agricultura, permitindo o acompanhamento financeiro de cada um dos contratos.

O sistema está em operação, sendo realizadas algumas eventuais etapas de manutenção, estando pendente ainda a conversão de alguns programas de crédito já encerrados ou em fase de devolução dos recursos recebidos, mas que precisam ficar registrados para efeito histórico e financeiro.

3.2.11 SOLOS - Banco de Dados de Solos

Este sistema é integrante de um projeto nacional desenvolvido pela EMBRAPA através da EMBRAPA Informática Agropecuária e EMBRAPA Solos, que visa possibilitar o armazenamento, manipulação e disponibilização das informações sobre os solos brasileiros. Conforme descreve (EMBRAPA, 2008), “O sistema de solos vai armazenar dados detalhados sobre esse recurso natural, para que possam ser acessados pela internet, combinados e analisados, sob vários pontos de vista. O banco reunirá informações de solos coletados e analisados de todas as regiões do Brasil. A partir desta base de dados serão desenvolvidas aplicações para a tomada de decisões do agronegócio, em zoneamento agrícola, na estimativa da produtividade de culturas, no ensino e na pesquisa. A base será, ainda, continuamente alimentada por pesquisadores da Embrapa e de outras instituições.” A EPAGRI é parceira do projeto e responsável pela alimentação dos dados relativos a Santa Catarina.

3.2.12 As perspectivas de integração

A busca de uma maior integração deste vasto conjunto de informações da empresa é uma preocupação constante, conforme afirmação da própria gerência de informação da empresa. As possibilidades fornecidas pelas novas tecnologias têm sido objeto de constante análise pela empresa, no sentido de racionalizar e aperfeiçoar o uso da tecnologia de informação reduzindo custos e aumentando a qualidade das informações geradas.

A preocupação atual é com a conclusão dos sistemas de apoio operacional e incrementar a elaboração de sistemas com o foco estratégico. Existe o objetivo de incluir a empresa no portal da Inovação do Ministério da Ciência e Tecnologia, para melhorar difundir o conhecimento disponível na empresa.

Com a identificação destes sistemas, percebe-se a possibilidade de avançar nos procedimentos de integração das informações através da construção de um SDW para a EPAGRI.

Muitos dos sistemas existentes têm um bom potencial para uso de informações georeferenciadas e esta utilização permitiria aos usuários destes sistemas um novo recurso na análise de dados auxiliando no processo de tomada de decisão.

A seguir é apresentada a Tabela 1 que descreve os sistemas e seu potencial para uso de informações georeferenciadas. Esta identificação do potencial foi fruto do processo de análise de cada sistema de informação e das discussões com a gerência de informação da EPAGRI.

3.3 Um SDW sobre Propriedades e Atividades Agrícolas

3.3.1 *Os sistemas escolhidos para o estudo de caso*

Os sistemas foram escolhidos após uma etapa de análise de diversos sistemas de informação em utilização na instituição e a partir de determinadas características como:

- abrangência de assuntos - busca por sistemas com leque abrangente de conteúdo, de forma a atender demandas de diversas pessoas do público alvo nos experimentos;
- abrangência geográfica - busca por sistemas com conteúdo de informação com maior abrangência geográfica para possibilitar vasta utilização de operadores espaciais e de consultas distintas;
- granularidade da informação geográfica - busca por sistemas com alto nível de detalhe para as informações geográficas, para possibilitar o uso de maior número de operadores espaciais e sua combinação.

Após esta análise, os sistemas escolhidos foram:

- SIAGRO - escolhido por ser um sistema de informação com dados históricos de produção agrícola, comercialização, população urbana e rural, com dados de séries históricas e com dados municipais, estaduais e internacionais.
- LAC - escolhido por ser um sistema de informação originado de um censo relacionado ao setor agropecuário, onde todos os estabelecimentos agropecuários de Santa Catarina tiveram suas informações coletadas, inclusive com suas coordenadas geográficas, fator importante

Tabela 1 : Sistemas da EPAGRI, seu potencial para uso de informações georeferenciadas e público alvo

Sistema	Potencial para uso de informações georeferenciadas	Público Alvo
SAFI	Grande potencial, na comparação de informações entre as regionais, os municípios	Gerencial
SEPLAN	Médio potencial, para acompanhamento de execução de metas	Gerencial
SEATER	Bom potencial, para análises da equipe técnica e dos atendimentos efetuados	Gerencial/técnico
SEPRO	Pequeno potencial, na análise de números de capacitação	Gerencial/técnico
Gestão de pessoas	Médio potencial, para acompanhamento de gestão das pessoas e suas funções	Gerencial/técnico
Preços Agrícolas	Médio potencial, na análise de preços	Gerencial, técnico e público
SIAGRO	Grande potencial, para todo o seu conteúdo.	Gerencial, técnico e público
LAC	Grande potencial, com restrição para informações que caracterizem a propriedade	Gerencial, técnico e público
SIGEO	Grande potencial; sistema desenvolvido para uso de informações georeferenciadas, com arquitetura própria e conteúdo com informações de outros bancos de dados	Gerencial, técnico e público
GEURED	Grande potencial, para análise do uso dos recursos de crédito no estado	Gerencial/técnico
SOLOS	Grande potencial	Gerencial/técnico

para sua escolha. O sistema tem informações completas relativas ao ano de coleta do censo. Estabelecimento agropecuário neste levantamento foi definido (EPAGRI/CEPA, 2010) como: “É toda unidade de produção dedicada, total ou parcialmente, a atividades agropecuárias, subordinadas a uma única administração (do produtor ou de um administrador), independentemente de tamanho, forma jurídica, situação (urbana ou rural) ou finalidade da produção (subsistência ou mercado)”.

- SOLOS - escolhido por ser um sistema de informação que contempla informações coletadas em diversos pontos geográficos de Santa Catarina e por apresentar variedade de conteúdo e nomenclatura específica de domínio.

Estes sistemas possuem grande potencial para uso de informações geográficas, pois abrangem dados de todo o estado de Santa Catarina e no caso específico do SIAGRO, dados de outros estados do Brasil e de outros países. Estes sistemas possuem séries históricas de informação e possuem um conjunto de dados com interesse desde o nível gerencial da empresa até usuários finais.

O SDW construído após a etapa de análise dos sistemas consistiu de dados provenientes do SIAGRO e do LAC. O sistema de SOLOS não teve a sua conclusão em tempo hábil para disponibilizar seus dados para a incorporação neste trabalho.

3.3.2 O Esquema do SDW Proposto

A Figura 2 mostra um data mart do SDW sobre atividades em propriedades agrícolas e sua produção agrícola em Santa Catarina. Este SDW foi produzido como parte deste trabalho, com dados da EPAGRI. A tabela fato inclui medidas escalares *AreaPlantada*, *AreaColhida*, *QuantidadeProduzida*, e *NumeroPropriedades*, além da medida espacial *SedesPropriedades* (coordenada geográfica da sede da propriedade). Estas medidas são agrupadas de acordo com níveis e membros das dimensões *Local* (com objetos geográficos representando o *Estado*, suas *Regiões*, *Municípios*, e *CorposDagua*), *Tempo*, *ProdutoAgricola*, and *CaracteristicasPropriedade*.

Este SDW permite análises de informação pela exploração dos componentes espaciais, ou seja, análises que manipulam objetos espaciais da dimensão *Local* e/ou a medida espacial *SedesPropriedades*. Os objetos na forma de multi-polígonos apresentados na dimensão espacial *Local* podem ser usados como parâmetros das operações espaciais para selecionar

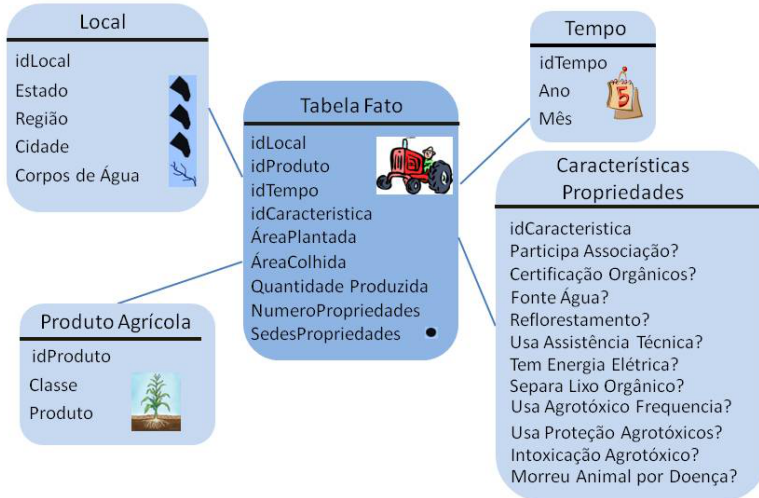


Figura 2: O esquema de um SDW de propriedades e atividades agrícolas

informação como, por exemplo, sedes das propriedades de um determinado município, e plotar esta informação em mapas temáticos, isto é, a distribuição das propriedades sobre as cidades. A agregação da medida espacial *SedesPropriedades* pode ser efetuada pela construção de coleções de propriedades ou pela aplicação de funções de agregação espacial compatíveis com o tipo de dado espacial e a semântica deste objeto geográfico. O resultado da agregação desta medida espacial também pode ser exibido em mapas. Diferentes membros de dimensões espaciais e valores de medidas espaciais pré-agregados podem referenciar a mesma representação de objeto espacial, para evitar cópias de representações potencialmente grandes de objetos espaciais como, por exemplo, perímetros de cidades, que podem ser representados por grandes sequências de coordenadas.

3.3.3 Exemplos de Análises a serem Efetuadas

A seguir são descritas algumas análises de informação que podem ser efetuadas sobre o SDW descrito acima:

Exemplo 1: Distribuição da Produção Agrícola

- Determinar a distribuição da produção de cebola nos municípios do estado de Santa Catarina no ano de 2005;
- Determinar a distribuição da produtividade (quantidade produzida/área colhida) de cebola nos municípios do estado de Santa Catarina no ano de 2005;
- Identificar o município com maior produtividade de cebola no ano de 2005, na tabela ou mapa resultante da execução da consulta anterior;

Exemplo 2: Problemas envolvendo agrotóxicos

- Determinar a distribuição das sedes de propriedades agrícolas nos municípios do estado de Santa Catarina que fazem uso de agrotóxicos com frequência e tiveram casos de intoxicação por agrotóxicos no ano de 2003;
- Identificar o município com o maior número de propriedades onde houve uso de agrotóxicos com frequência e casos de intoxicação por agrotóxicos no ano de 2003;
- Identificar, em tal município, as sedes de propriedades onde houve uso de agrotóxicos com frequência e casos de intoxicação por agrotóxicos no ano de 2003 a até 5 quilômetros do rio principal.

Exemplo 3: Certificação de Produtores

- Determinar a distribuição das sedes de propriedades agrícolas nos municípios do estado de Santa Catarina que possuíam certificação de produtor de produtos orgânicos no ano de 2003;
- Identificar o município com maior número de propriedades que possuíam certificação de produtor de produtos orgânicos no ano de 2003;
- Identificar, em tal município, as sedes das propriedades que estão a uma distância inferior a 5 km da sede do município.

3.3.4 Implementação

O banco de dados escolhido para a criação do SDW foi o ORACLE, pelo fato deste ser o banco de dados da EPAGRI, onde os dados originais estão armazenados, por oferecer suporte a objetos espaciais e por ser do conhecimento do autor.

Os procedimentos de ETC (extração, transformação e carga) no SDW foram desenvolvidos após a análise da estrutura e do conteúdo dos dados destes sistemas. Rotinas foram desenvolvidas para a execução dos procedimentos de transferência dos dados em scripts de transferência de dados entre os bancos de dados.

Por se tratar de um protótipo, alguns mapas foram gerados previamente para exibir os resultados que seriam esperados durante o experimento realizado com usuários. Não foi utilizado servidor OLAP nem o servidor de mapas para sua geração.

Os mapas foram gerados com auxílio da Mapoteca Topográfica Digital de Santa Catarina (EPAGRI/IBGE, 2004).

3.4 Considerações

Este capítulo apresentou o SDW criado para este trabalho e que foi utilizado como estudo de caso. Os dados foram obtidos após a análise de sistemas da EPAGRI e o SDW foi construído utilizando objetos espaciais nas dimensões e na tabela fato, para permitir diferentes análises espaciais. O próximo capítulo descreve a arquitetura e o modelo conceitual do S^2DW e as ontologias de domínio e de SDW desenvolvidas. Descreve também a interface baseada em conhecimento e interação do usuário com esta interface para a execução de análises de informação.

4 SEMANTIC AND SPATIAL DATA WAREHOUSES

S^2DW (Semantic and Spatial Data Warehouses) é um sistema para suportar análises de informação em SDWs (DEGGAU; FILETO, 2009). Ele permite ao usuário especialista de um domínio identificar SDWs relacionados a um tema de interesse e efetuar consultas nesses SDWs através de uma interface gráfica baseada em conhecimento. Interações adicionais podem ser realizadas sobre os resultados das consultas na forma de tabelas, gráficos ou mapas.

Este capítulo apresenta a arquitetura geral do S^2DW , as ontologias usadas na confecção de descrições semânticas de SDWs e a análise de informação suportada por tais descrições, em uma interface gráfica baseada em conhecimento.

4.1 Arquitetura para Semantic Spatial Data Warehouse

A Figura 3 ilustra a arquitetura do S^2DW e destaca a função da sua interface gráfica baseada em conhecimento. O S^2DW utiliza uma ontologia de SDW e uma ontologia de domínio para descrever semanticamente a estrutura e o conteúdo de SDWs.

A ontologia de SDW é fixa e inclui definições conceituais de componentes estruturais de data marts espaciais (dimensões, níveis, medidas) e seus relacionamentos, além da descrição de operadores e funções de agregação de dados em SDWs.

A ontologia de domínio, que varia com a aplicação, permite descrever o conteúdo de SDWs utilizando vocabulário específico do domínio. As duas ontologias foram desenvolvidas no escopo deste trabalho.

A interface baseada em conhecimento do S^2DW fornece uma visão gráfica do conhecimento contido nas ontologias e nas descrições semânticas dos data marts. Ela permite que os próprios usuários explorem e analisem a informação destes data marts, para produzir tabelas com dados, mapas e gráficos consolidados.

A implementação da arquitetura completa do S^2DW envolve a solução de diversos problemas. Primeiramente, são necessárias pesquisas sobre a representação semântica da estrutura e conteúdo de SDWs. (XIE et al., 2007) propõem que tal representação seja feita na forma de mapeamentos entre componentes estruturais de SDWs e termos definidos em uma ontologia de domínio. Este trabalho segue tal linha, definindo uma ontologia de componentes estruturais de SDWs baseada em um modelo dimensional espacial

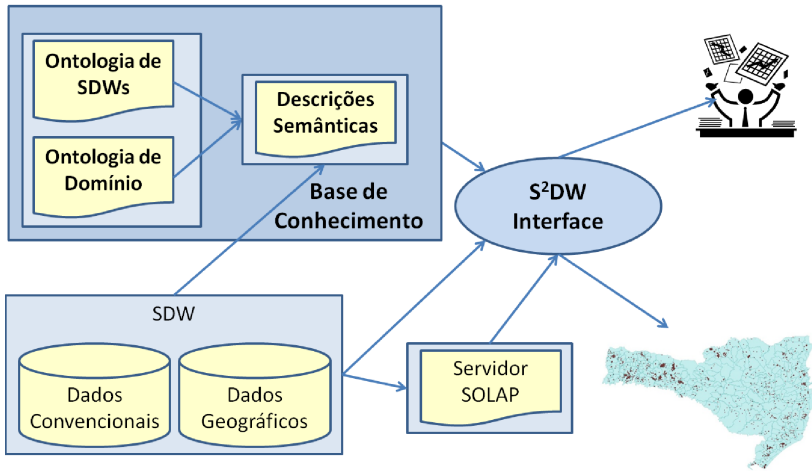


Figura 3: A arquitetura geral do S^2DW e a interface baseada em conhecimento

e enfatizando a classificação de operadores e funções de agregação de dados espaciais. O objetivo é utilizar tal ontologia para orientar o usuário na composição correta destes operadores e funções, para executar suas análises.

A geração das descrições semânticas de data marts não foi considerada no escopo deste trabalho. Obviamente esta é uma questão relevante e que precisa ser considerada, para viabilizar o sistema aqui proposto. A geração de descrições semânticas é um processo complexo que requer a cooperação do pessoal de TI com especialistas de domínio. A quantidade de trabalho envolvido na descrição semântica completa de um data mart poderia inviabilizar o cumprimento dos demais objetivos desta dissertação. Porém, espera-se que tais descrições de data marts sejam reusadas como modelos que possam ser adaptados de modo a construir outros data marts para suportar outras análises. Com o reuso de especificações espera-se reduzir o trabalho de construção de descrições semânticas de data marts e a própria geração de data marts para suportar análise de informação em SDWs (XIE et al., 2007).

Este trabalho foca no desenvolvimento, teste de usabilidade e validação da interface gráfica baseada em conhecimento para suportar especialistas de domínio na execução de suas análises de informação em SDWs. O objetivo é testar a viabilidade de tal interface, e assim de uma boa parte da abordagem proposta, do ponto de vista de especialistas de domínio, antes de investir mais esforço na pesquisa e desenvolvimento na

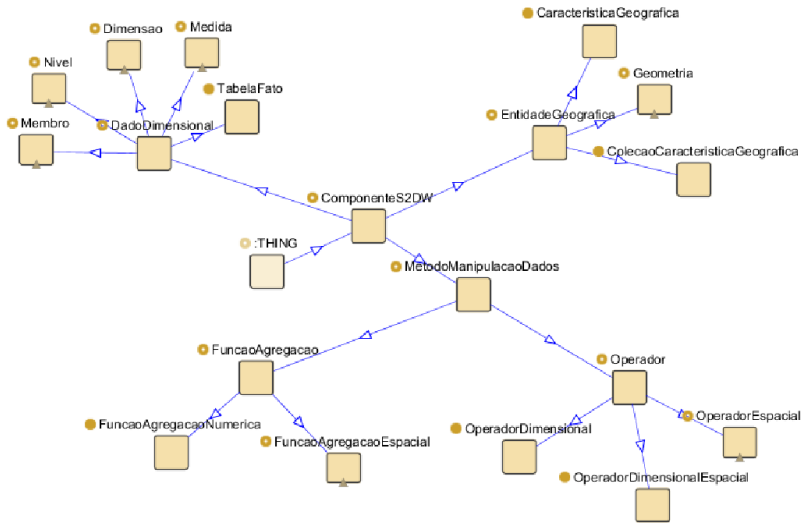


Figura 4: Visão geral da ontologia de SDW

implementação completa da arquitetura e da própria interface aqui proposta.

4.2 A Ontologia de SDW

A ontologia de SDW descreve as estruturas de composição permitidas e as facilidades de manipulação em um data mart espacial acessado via S^2DW , de acordo com o modelo dimensional com extensões espaciais de (FIDALGO et al., 2004).

A Figura 4 ilustra a hierarquia dos conceitos de nível mais alto desta ontologia e omite diversos conceitos dos níveis mais detalhados das hierarquias apresentadas e relacionamentos entre esses conceitos e apresenta apenas relacionamentos do tipo ISA (subsumption).

A Figura 5 descreve os componentes estruturais de data marts espaciais. Os conceitos relacionados são *TabelaFato*, *Medida*, *Dimensão*, *Nível*, e *Membro*. Uma *TabelaFato* pode incluir diversas *Medidas*. Uma *Medida* pode estar associada a várias *Dimensões*. Todos estes componentes no modelo podem ser escalares ou espaciais.

A Figura 6 descreve os tipos de entidades espaciais; as entidades ge-

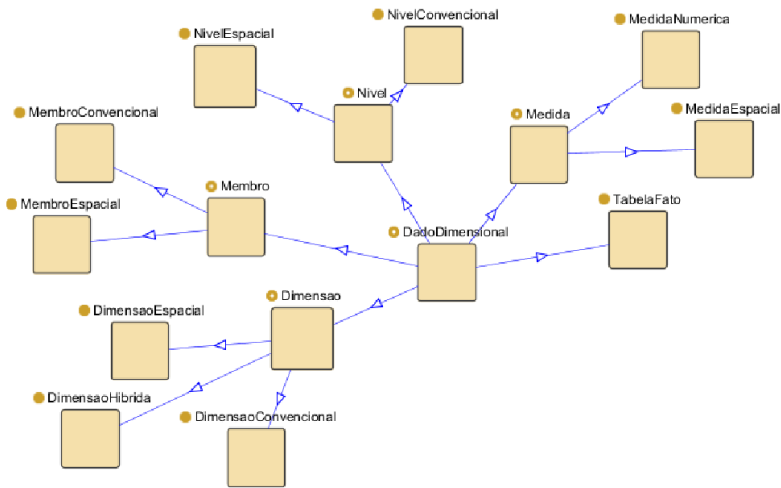


Figura 5: Ontologia de SDW - Componentes Estruturais

ográficas foram implementadas de acordo com a especificação da OGC Geographic Objects Implementation Specification¹.

A Figura 7 descreve os métodos de manipulação de dados divididos em duas classes:

1. operadores de manipulação de dados, usados para selecionar e agrupar informações de acordo com os dados das dimensões e
2. funções para agregação de dados contidos nas medidas das tabelas fato. Estas classes são definidas pelo operador e natureza das funções e os tipos de seus parâmetros de entrada e as saídas podem ser escalares ou espaciais.

A Figura 8 apresenta o detalhamento da ontologia até o menor nível das instâncias que descrevem as funções classificadas como funções distributivas espaciais. A Figura 9 apresenta os operadores que estão classificados dentro do grupo de operadores que geram novas geometrias.

Esta versão da ontologia de SDW utiliza a classificação de operadores e funções de agregação espacial de (SILVA et al., 2008), mas a ontologia de

¹<http://www.opengeospatial.org/standards/go>

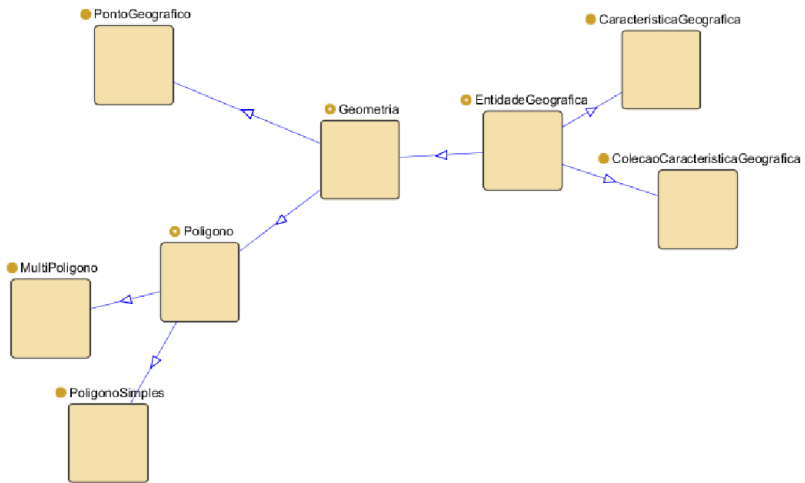


Figura 6: Ontologia de SDW - Tipos de Entidades Espaciais

SDW pode ser adaptada para descrever outras classificações de operadores e funções de agregação.

4.3 A Ontologia de Domínio

A ontologia de domínio descreve os conceitos existentes no domínio de aplicação em que o S^2DW é usado. O S^2DW é independente de domínio, permitindo adaptação mediante a substituição da ontologia de domínio, sendo necessária neste caso a redefinição dos mapeamentos entre a nova ontologia e a ontologia de SDW. Como a ontologia de domínio usa um vocabulário específico ela permite que o sistema apresente os data marts, efetue consultas e exiba os resultados de acordo com o conhecimento do domínio e no seu vocabulário específico. Isso facilita o processo de especificação das necessidades de informação pelo usuário.

4.3.1 Exemplo de Ontologia de Domínio - Ontologia Agrícola

Para o estudo de caso do S^2DW , foi utilizada uma ontologia do domínio agrícola, cujo detalhamento é descrito a seguir.

A Figura 10 descreve uma porção da ontologia de agricultura utilizado

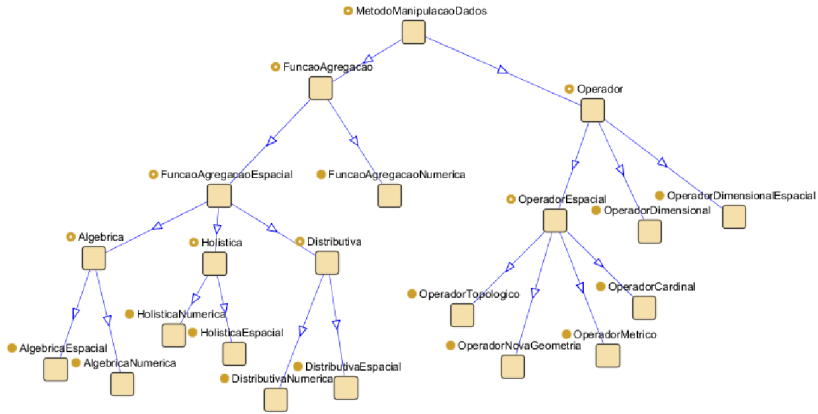


Figura 7: Ontologia de SDW - Operadores e Funções de Manipulação de Dados

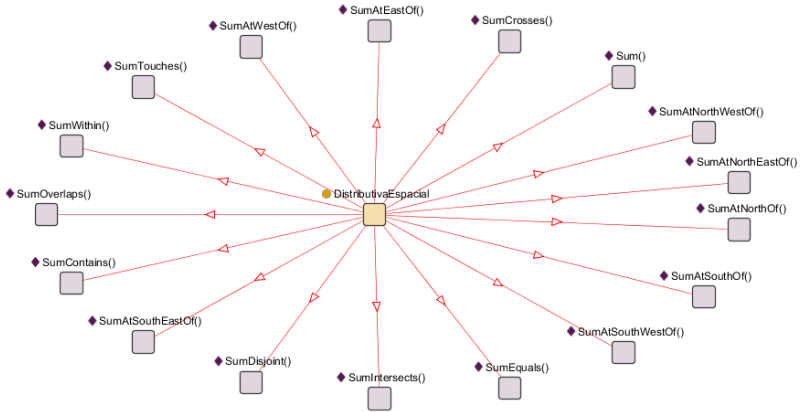


Figura 8: Ontologia de SDW - Funções de Agregação Distributivas Espaciais

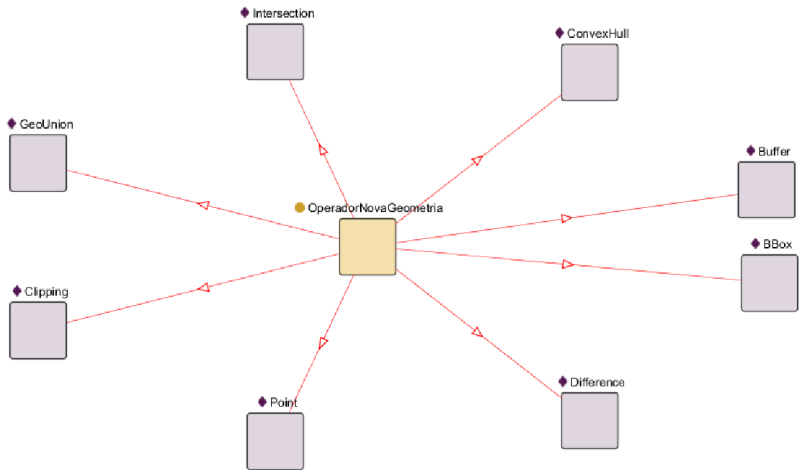


Figura 9: Ontologia de SDW - Operadores Espaciais que Geram Nova Geometria

no estudo de caso e nos exemplos de interação do usuário com data marts espaciais neste trabalho. Ela inclui classificação de medidas de produção agrícola, de produtos agrícolas, produtos animais, propriedades agrícolas e população rural. No lado esquerdo da figura 10 a relação da medida *produtividade*, com as medidas *quantidade produzida* e *área colhida* é destacada ($produtividade = quantidade\ produzida / \acute{a}rea\ colhida$). Uma porção do FAO Agrovoc Thesaurus ² foi utilizada para servir como ontologia de domínio neste trabalho, com o objetivo de validar neste domínio a proposta aqui apresentada.

4.4 Descrição Semântica de Data Marts

A descrição semântica de um data mart espacial é baseada nas ontologias de domínio e de SDW. Um SDW pode ser descrito semanticamente através da definição de mapeamentos entre conceitos das duas ontologias (XIE et al., 2007). Esses mapeamentos habilitam o sistema a informar e auxiliar o usuário no processo de busca, exploração e entendimento dos conteúdos (medidas, dimensões, etc.) do SDW e nos procedimentos de análise do conteúdo

²<http://www.fao.org/agrovoc/>

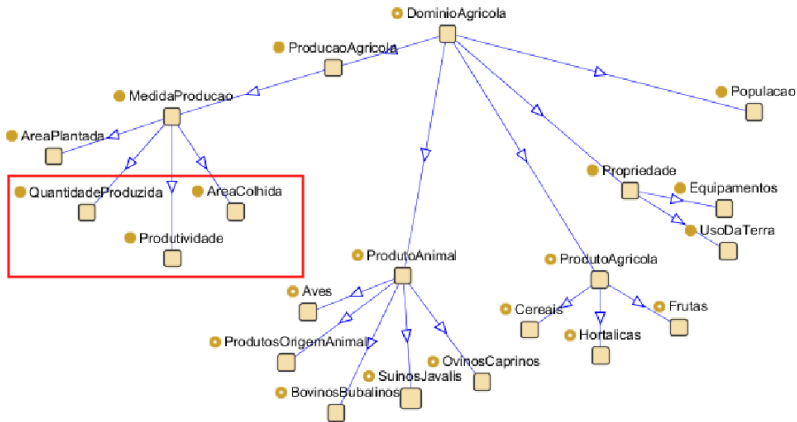


Figura 10: Visão geral de uma ontologia para domínio agrícola

(operadores e funções de agregação). Isto contribui para uma interação mais efetiva com o sistema, tornando a análise de informação de acordo com os significados SOLAP.

A geração de mapeamentos formais completos está fora do escopo deste trabalho, sendo deixada para trabalhos futuros. Este trabalho projeta, desenvolve e analisa o uso de uma interface para SDWs baseada somente nas ontologias descritas acima (incluindo as classificações de operadores e funções de agregação de dados espaciais) e mapeamentos entre definições de componentes estruturais de SDWs e termos do domínio utilizados para descrevê-los, como ilustrado de maneira simplificada na Tabela 2.

4.5 A Interface Proposta

A Figura 11 apresenta a tela inicial do S2DW customizada com uma ontologia agrícola ilustrada na Figura 10. Esta tela inclui um campo para entrada de palavras-chave, uma visão das hierarquias presentes na ontologia de domínio e um quadro para apresentar descrições semânticas de conceitos ou instâncias presentes na base de conhecimento do S^2DW .

O usuário inicia o uso do sistema informando uma palavra-chave ou navegando na hierarquia de conceitos do domínio para encontrar um ou mais conceitos relativos ao assunto que ele está interessado. Um campo texto do

Tabela 2: Visão geral dos mapeamentos para descrever semanticamente os componentes de um SDW sobre propriedades e atividades agrícolas

Classe Ontologia Domínio	Classe Ontologia SDW
Quantidade Produzida	MedidaNumerica
AreaPlantada	MedidaNumerica
Produtividade	MedidaNumerica
....
Sede de Propriedade Agrícola	MedidaEspacial em SDWAgrícola PontoGeografico
....
Produto Agrícola	DimensaoConvencional
Tempo	DimensaoConvencional
Características Propriedades	DimensaoConvencional
....
Local	DimensaoEspacial
....
Cidade	NivelEspacial de Local MultiPoligono
Estado	NivelEspacial de Local MultiPoligono

lado esquerdo da janela na Figura 11 apresenta informações provenientes da base de conhecimento, acerca do conceito selecionado, caso ele seja encontrado na base de conhecimento. Quando o usuário indica a conclusão dos assuntos de seu interesse, ele solicita ao sistema que efetue uma busca semântica por data marts que contenham informações relacionadas ao assunto de interesse, através do pressionamento da tecla < Enter > ou através da seleção do botão < Pesquisa >, que são ícones da interface.

A interação do usuário com a interface do S^2DW para buscar data marts espaciais e efetuar análise de informação sobre eles segue a seguinte sequência de passos:

1. busca por data marts relacionados ao assunto através do preenchimento de palavras-chave ou pela navegação na ontologia de domínio para selecionar palavras relativas ao assunto de interesse;
2. especificar uma ou mais consultas sobre os data marts encontrados;
 - (a) selecionar as medidas a recuperar, os membros das dimensões para filtragem e agregação destas medidas, os operadores para filtrar dados e funções de agregação para consolidar os dados;

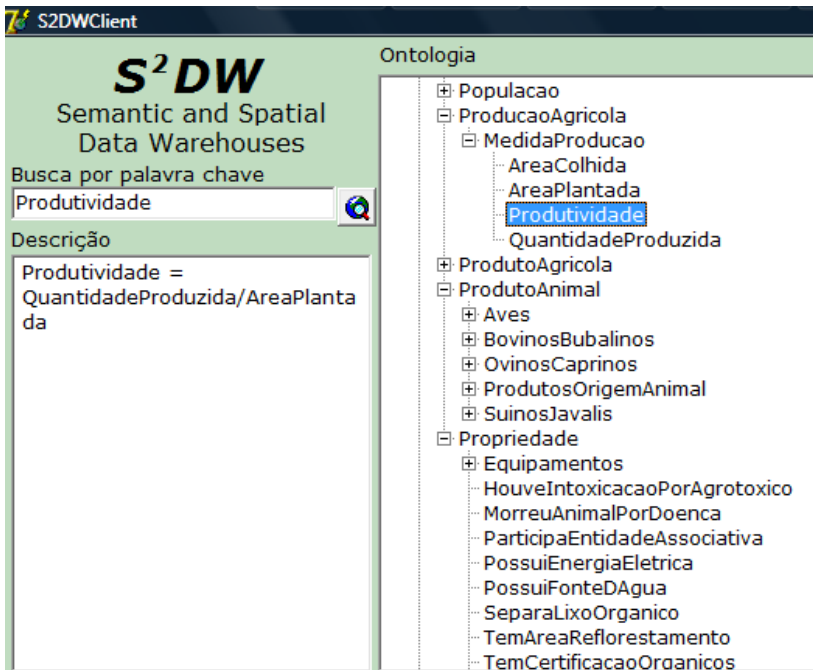


Figura 11: Busca baseada em palavra-chave e navegação em uma visão da ontologia de domínio

- (b) solicitar a execução das consultas, pressionando o botão < *ExibirResultados* >;
- (c) visualizar os resultados em tabelas, gráficos e/ou mapas;
- (d) se necessário, refinar a análise interagindo com os resultados apresentados ou com a representação gráfica do data mart espacial (retorno ao início do passo 2).

A seguir o processo é demonstrado através de exemplos, onde o usuário interage com a interface do S^2DW customizada para o domínio agrícola.

4.5.1 Busca por data marts espaciais relacionadas a um assunto

Supondo que um usuário esteja interessado em analisar o uso de agrotóxicos em propriedades agrícolas e seu possível impacto em fontes de água. Com algumas palavras simples ou compostas em sua mente (e.g. “agrotóxico”), ele pode informá-la como palavra-chave. Ele pode também navegar na ontologia de domínio e encontrar e selecionar um ou mais conceitos do seu interesse. Quando ele solicita a execução da busca com as palavras fornecidas, o sistema executa uma busca semântica (MANGOLD, 2007; D’AGOSTINI; FILETO, 2009) nas ontologias e nas descrições semânticas dos data marts espaciais. Após processar a busca, o S^2DW apresenta as descrições semânticas de cada data mart espacial com algumas informações relacionadas à palavra-chave em uma aba.

A Figura 12 mostra como a interface baseada em conhecimento do S^2DW apresenta a descrição semântica de um data mart para analisar informações sobre propriedades e produção agrícola conforme a Figura 2. A tabela fato do data mart descrito contém as medidas numéricas *AreaPlantada*, *AreaColhida*, *Quantidade Produzida*, *Numero Propriedades* e a medida espacial *Sedes Propriedades*. Este data mart possui quatro dimensões: *Local* com os níveis espaciais *Estado*, *Região*, *Cidade* e *Corpos de Água*; *Produto Agrícola* com os níveis convencionais *Classe* e *Produto*; *Tempo* com os níveis *Ano* e *Mês*; e a dimensão *Característica Propriedade*, com diversas características das propriedades.

Agora, o usuário pode escolher medidas deste data mart espacial, junto com os membros dos diversos níveis de suas dimensões, para filtrar e agregar as informações. A ordem em que as seleções são feitas pode variar. O usuário especifica uma consulta SOLAP nesta interface de forma gradual. Após o término das seleções desejadas, o usuário clica no botão

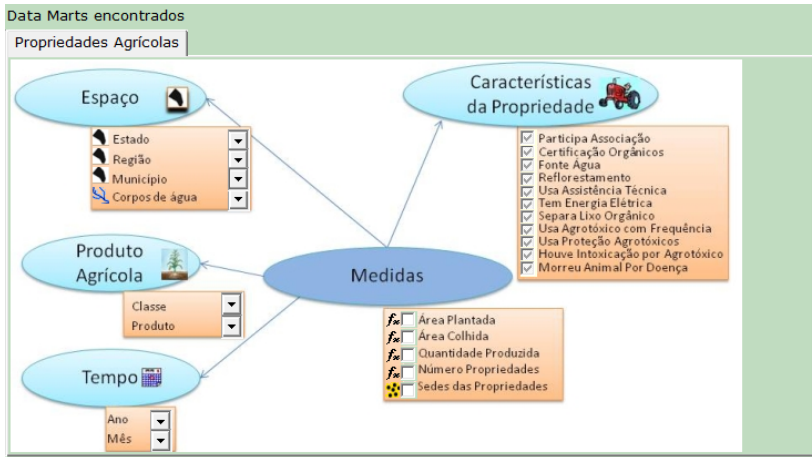


Figura 12: Representação gráfica da descrição semântica de um data mart espacial

< *ExibirResultados* > para o sistema gerar e executar a consulta SOLAP como parte do processo de análise de informação desejado pelo usuário.

4.6 Executando Análises de Informação Espacial

A seguir são descritos alguns exemplos de análises de informação espacial e a sua execução na interface gráfica proposta.

Consulta SOLAP 1:

Mostre a distribuição geográfica dos casos de intoxicação por agrotóxicos entre as propriedades agrícolas de Santa Catarina no ano de 2003

O usuário pode especificar esta consulta inicialmente selecionando a medida espacial *Sedes Propriedades* e a medida numérica *Número de Propriedades*. Então, ele pode selecionar o valor “TRUE” para a característica *Houve Intoxicação por uso de Agrotóxico* e *Uso Frequente de Agrotóxico*, na dimensão *Características Propriedades*, a fim de indicar a condição de filtragem nesta dimensão. Ele pode também selecionar o ano “2003” na dimensão *Tempo* e o estado de “Santa Catarina” na dimensão *Local*, para in-

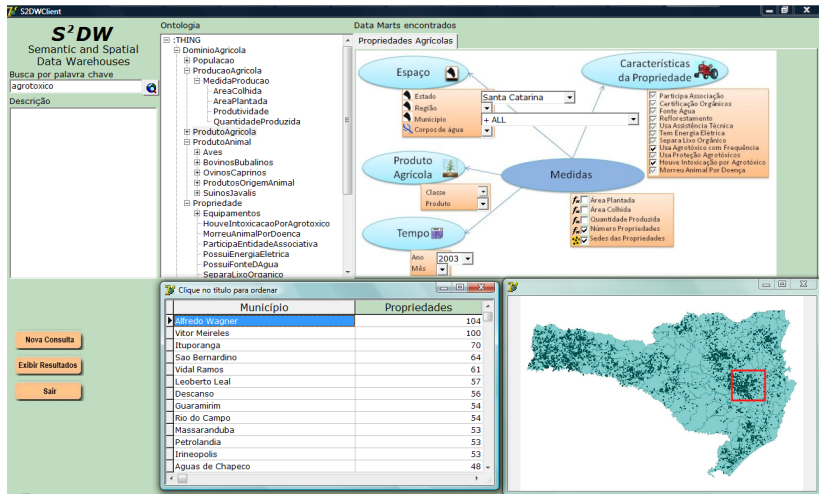


Figura 13: Distribuição geográfica das propriedades com casos de intoxicação pelo uso de agrotóxicos e que fazem uso frequente de agrotóxicos no estado de Santa Catarina em 2003

dicar as condições de filtragem para esta dimensão. A especificação do valor “ALL” para o nível *Cidade* indica que a medida precisa ser agregada por cidade. Neste exemplo, não há necessidade de selecionar uma sequência de operadores para filtrar os dados, porque o *S²DW* usa o operador *GeoIntersects* e a função de agregação *Collection* como default. Após finalizar as seleções desejadas, o usuário pressiona o botão < *Exibir Resultados* >.

Os componentes de dados selecionados e os resultados para esta consulta estão demonstrados na Figura 13. A informação solicitada é apresentada em uma tabela, com os valores numéricos para o número de propriedades de cada cidade, no lado esquerdo da Figura 13. Esta informação também é apresentada em um mapa temático, que permite ao usuário visualizar facilmente a distribuição das sedes das propriedades que fazem uso frequente de agrotóxico e que possuem casos de intoxicação pelo seu uso, em 2003 no estado de Santa Catarina. Esta consulta em MDX é apresentada na Figura 14

Consulta SOLAP 2:

No município com maior número de propriedades com casos de

```

SELECT Count ( [Measures].[SedesPropriedades] ) AS Propriedades,
           Collection ( [Measures].[SedesPropriedades] ) AS Sedes
ON ROWS { [Local].[Cidade] AS Município },
FROM SDW_Agrícola
WHERE ( [Características Propriedades].[Usa Agrotóxico Freqüência].[True]
AND [Características Propriedades].[Intoxicação Agrotóxico].[True]
AND [Tempo].[Ano].[2003])

```

Figura 14: Consulta SOLAP 1 em MDX

intoxicação pelo uso de agrotóxico e que o usam frequentemente, exibir as propriedades mais próximas do rio Itajaí do Sul.

O município com maior número de ocorrências de intoxicação pelo uso de agrotóxicos em Santa Catarina em 2003, “Alfredo Wagner”, pode ser facilmente identificado na Figura 13. O usuário pode interagir com a descrição semântica do data mart espacial, ou com a tabela ou com o mapa apresentado pelo S^2DW , para refinar suas análises de informação. Por exemplo, ele pode selecionar o município de Alfredo Wagner, pela escolha deste município na descrição semântica do data mart, clicando no município que está no topo da tabela apresentada ou clicando na representação do município destacado no retângulo da Figura 13. Então, o usuário pode refinar mais ainda a sua análise, a fim de avaliar o risco de contaminação de *Corpos de Água* próximos a propriedades neste município. Para fazer isto, ele pode posicionar o mapa para focar ao redor do rio “Itajaí do Sul”. Este rio é o maior corpo de água que cruza o município de Alfredo Wagner, como mostrado na Figura 15.

O usuário pode selecionar o rio desejado clicando em sua representação no mapa ou selecionando-o no nível *Corpo de Água* do dimensão *Local*, na descrição semântica do data mart espacial. Então, ele pode clicar no ícone do nível da dimensão correspondente, para verificar os operadores existentes que podem ser usados para filtrar informações da tabela fato, usando os objetos espaciais do tipo *Corpos de Água*.

Como resultado a caixa de diálogo exibida na Figura 16 aparece na tela. Ela mostra os operadores que podem ser usados com este tipo de objeto espacial, classificados como operadores “Topologic”, “Geometry”, e “Cardinal”. As explicações que podem ser acessadas através do clique no respectivo operador ou classe de operador podem auxiliar o usuário a identificar que ele primeiro necessita aplicar o operador “Buffer”, com um certo valor para seu parâmetro distância, a fim de criar uma geometria expandida que pode ser us-

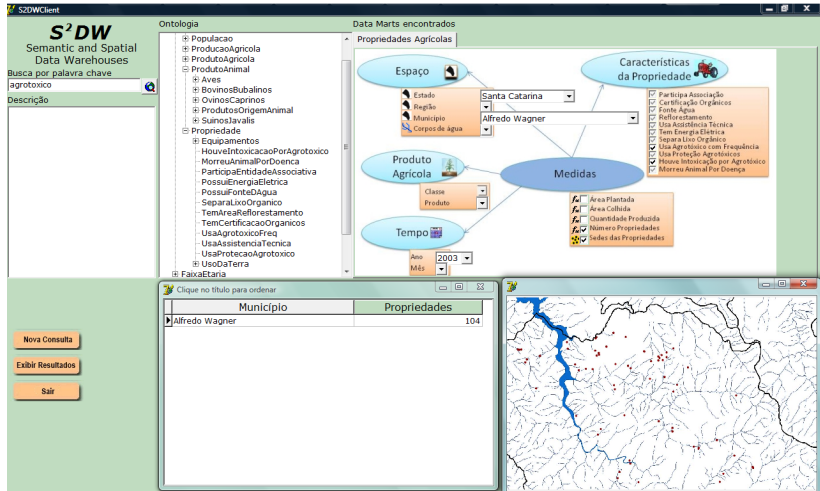


Figura 15: Sedes das propriedades com casos de intoxicação pelo uso de agrotóxicos que fazem uso frequente no município de Alfredo Wagner em 2003.

ada com o operador GeoIntersects para encontrar as sedes das propriedades próximas ao rio Itajaí do Sul.

O resultado sem a aplicação do “Buffer” é melhor detalhado na Figura 17. Após o operador “Buffer” ser aplicado, a execução do operador GeoIntersects gera o resultado final desejado que é ilustrado na Figura 18.

Esta consulta expressa em MDX está descrita na Figura 19.

4.6.1 Outros exemplos de análise de informação

A seguir são apresentados outros resultados de consultas elaboradas com a interface gráfica, as quais também foram objeto do experimento realizado com os usuários.

Uma consulta realizada em 3 etapas:

1. Determinar a distribuição das sedes de propriedades agrícolas nos municípios do estado de Santa Catarina que possuíam certificação de produtor de produtos orgânicos no ano de 2003.
2. Identificar o município com maior número de propriedades que

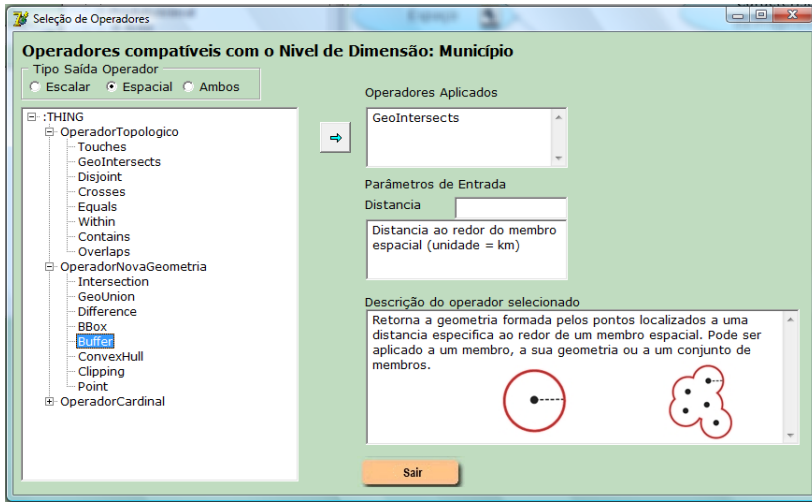


Figura 16: Operadores disponíveis para filtrar informação usando o membro espacial *Corpos de Água*

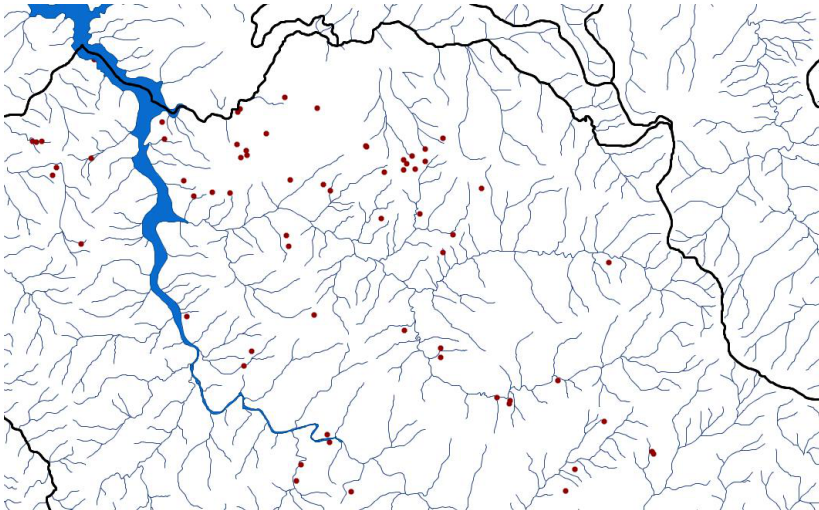


Figura 17: Mapa Resultante sem aplicação do operador Buffer

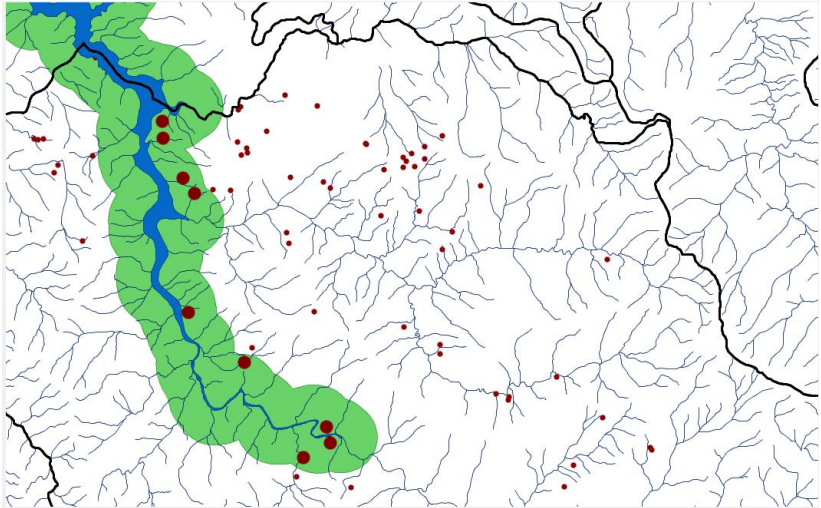


Figura 18: Mapa Resultante com a aplicação do operador Buffer e Intersects das sedes das propriedades com o resultado do Buffer

```

SELECT Count ( [Measures].[SedesPropriedades] ),
      Collection ( [Measures].[SedesPropriedades] AS Sedes )
ON ROWS { [Local],[Cidade] },
ON COLUMNS { [ProdutoAgrícola].[Classe].[“Hortaliças”],
              [ProdutoAgrícola].[Classe].[“Frutas”],
              [ProdutoAgrícola].[Produto] },
FROM SDW_Agrícola
WHERE ( [Caracteristica Propriedade].[Usa Agrotóxico Freqüência].[True] AND
        [Tempo].[Ano].[2003]
        AND { buffer( [Local].[Corpos de Água].[“Rio Itajaí do Sul”,5 )
              intersects
              Sedes )

```

Figura 19: Consulta SOLAP 2 em MDX

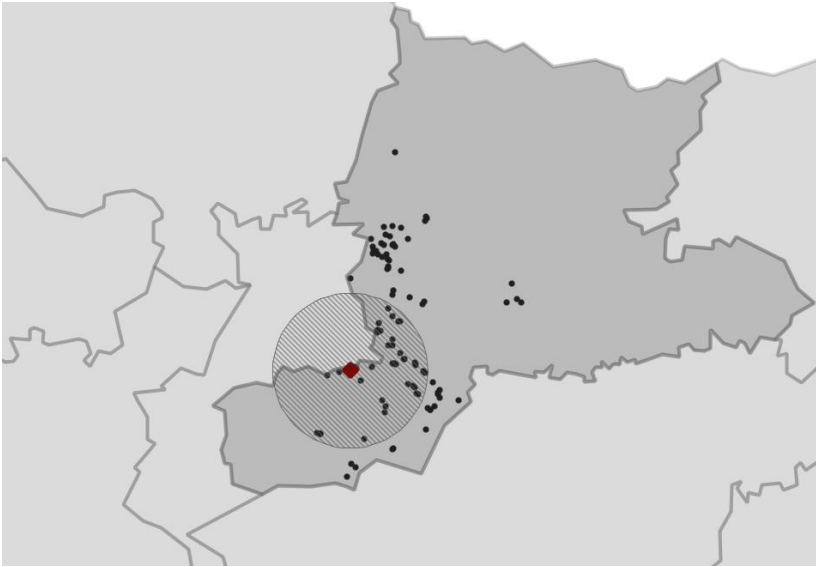


Figura 20: Propriedades agrícolas com certificação de produção orgânica em Passos Maia localizadas a certa distância de um ponto geográfico

possuíam certificação de produtor de produtos orgânicos no ano de 2003.

3. Identificar, em tal município, as sedes das propriedades que estão a uma distância inferior a 5 km de um certo ponto geográfico.

O resultado desta consulta está demonstrado na Figura 20 onde o estado de Santa Catarina é representado e o município de Passos Maia é destacado; o ponto geográfico desejado está identificado em destaque e a área representada pelo círculo corresponde ao resultado desejado.

A Figura 21 exhibe o resultado da consulta: “Analisar a distribuição geográfica das propriedades agrícolas que possuem fonte de água e tiveram registros de casos de morte de animais por doença na região Oeste do estado de Santa Catarina.”

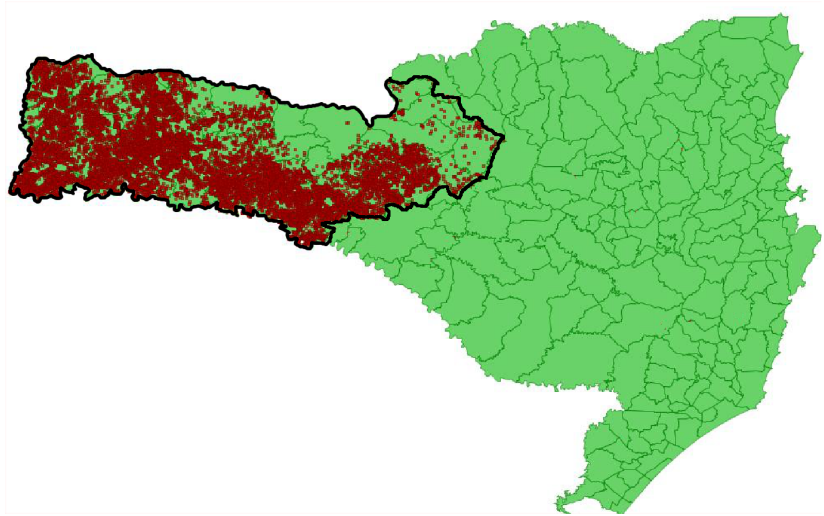


Figura 21: Propriedades agrícolas do Oeste de SC com morte de animais por doença e que possuem fonte de água

4.7 Considerações

Este capítulo apresentou a arquitetura e o modelo conceitual do S^2DW , além de sua interface, com exemplos de análise de informação e a forma como elas devem ser especificadas. O próximo capítulo descreve os experimentos efetuados com usuários para avaliar a usabilidade da interface e analisa os resultados deste experimento.

5 EXPERIMENTOS

5.1 Análise de Usabilidade da GUI Proposta

Um protótipo do sistema S^2DW foi desenvolvido com o objetivo de analisar a usabilidade da interface proposta. Ela foi implementada em uma aplicação stand-alone na linguagem Delphi, em ambiente Windows. Este protótipo foi utilizado na execução de experimentos com usuários do domínio agrícola.

Nestes experimentos os usuários foram desafiados a efetuar algumas análises de informação sobre o SDW utilizando o data mart apresentado na Figura 2, com o objetivo de avaliar a sua performance na execução destas análises. A ontologia de domínio agrícola apresentada na Figura 10 foi usada para customizar o S^2DW para este domínio. Ao final do experimento foi aplicado um questionário semi-estruturado aos participantes do experimento, perguntando algumas questões sobre a execução das tarefas realizadas, visando avaliar sua opinião sobre o sistema proposto.

A seguir são descritos alguns conceitos básicos utilizados no experimento, segundo a norma ISO 9241-11 (ISO/IEC, 1991). :

- usuário: “pessoa que interage com o produto”;
- usabilidade: “a capacidade de um produto ser usado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso”;
- Contexto de uso: “usuários, tarefas, equipamentos (hardware, software e materiais), ambiente físico e social em que o produto é usado”;
- Eficácia: “precisão e completeza com que os usuários atingem objetivos específicos, acessando a informação correta ou gerando os resultados esperados. A precisão é uma característica associada à correspondência entre a qualidade do resultado e o critério especificado, enquanto a completeza é a proporção da quantidade-alvo que foi atingida”;
- Eficiência: “precisão e completeza com que os usuários atingem seus objetivos, em relação à quantidade de recursos gastos”;
- Satisfação: “conforto e aceitabilidade do produto, medidos por meio de métodos subjetivos e/ou objetivos”.

5.2 Detalhamento do Experimento

O experimento buscou identificar se a interface baseada em conhecimento do S^2DW fornecia meios apropriados para que os usuários executassem análise de informação no SDW. O experimento usou métodos de inspeção de acordo com a norma ISO 9241-11 (ISO/IEC, 1991). O teste empírico de usabilidade, também conhecido como teste da técnica de observação, foi escolhido como método de inspeção.

De acordo com (NIELSEN, 1993), testes com 6 a 12 usuários são suficientes para identificar 85% dos problemas de usabilidade. O experimento foi aplicado a 17 usuários:

- 8 são profissionais do domínio agrícola que usam sistemas de informação regularmente (Analistas de Mercado, Bibliotecários, Economistas);
- 7 são da área de tecnologia da informação da EPAGRI (Analistas de sistemas e Administradores de Rede)
- 2 são da área de design gráfico do Núcleo de Gestão de Design da UFSC (Graduando e Mestrando).

Para aplicação do experimento aos usuários, visando observar suas ações, captura de seus dados e obtenção de suas respostas ao questionário, foi utilizado o software Morae (TECHSMITH, 2010). Esse software é composto de um conjunto de ferramentas para a avaliação de usabilidade, dando suporte a vários tipos de técnicas, tais como testes empíricos com usuários em ambiente web e aplicativos stand-alone, focusgroups, testes com dispositivos móveis (celulares, Palm, ipod, etc) e prototipação em papel. O software é composto por três aplicativos, o Morae Recorder, o Morae Observer e o Morae Manager.

Com o auxílio do produto, é registrada previamente a sequência de tarefas a ser realizada, com indicações para marcação do início e fim de cada tarefa. As perguntas de avaliação são também registradas no Morae e ao final deste processo um arquivo no formato Morae é gerado com o registro de todas estas informações. Este arquivo será então utilizado no experimento pelos usuários, permitindo que o Morae gerencie o andamento do experimento e registre as informações desejadas.

A execução do experimento foi dividida em três etapas, descritas a seguir:

Passo 1 - Apresentação do Sistema S^2DW : O objetivo desta apresentação foi fornecer uma visão geral do processo de análise de

informação com o S^2DW , apresentando a execução de procedimentos similares àqueles que seriam executados pelos usuários no experimento. Ela descrevia os objetivos do S^2DW , sua arquitetura geral, sua GUI baseada em conhecimento e dava exemplos de como alguém poderia especificar as suas consultas para análise de informação espacial no SDW existente.

Passo 2 - Execução das Tarefas de Análise de Informação: Após a apresentação do sistema S^2DW aos usuários, eles foram solicitados a utilizar a GUI baseada em conhecimento para especificar algumas consultas SOLAP para análise de informação no SDW. A sequência de tarefas foi previamente armazenada em um roteiro no MORAE. Este roteiro define o fluxo de tarefas a ser executado, terminando com o questionário de avaliação. Ao iniciar o experimento, o sistema solicitava ao usuário que se identificasse. A seguir, cada tarefa era apresentada ao usuário. Ele lia a tarefa e indicava ao sistema quando ele estava pronto para iniciar a tarefa. Ao final de cada tarefa, ele indicava a sua conclusão e o sistema lhe apresentava a próxima tarefa.

Passo 3 - Aplicação do Questionário: Após a conclusão de todas as tarefas de análise de informação pelo usuário, o software Morae solicitava ao usuário que respondesse ao questionário e expressasse sua opinião sobre a execução das tarefas propostas com o S^2DW .

Todos estes procedimentos foram executados por cada usuário durante o experimento. Utilizando o arquivo previamente preparado, o Morae coletou as informações sobre as interações de cada usuário com o S^2DW , incluindo a manipulação dos componentes da GUI, o tempo gasto e o número de cliques no mouse em cada tarefa. Também foi registrado em vídeo o comportamento de cada usuário durante a realização do experimento.

As tarefas de análise de informação solicitadas aos usuários nos experimentos foram as seguintes:

1. Encontrar um data mart com informação sobre algum assunto.
 - (a) Fornecer (no formulário) ou selecionar (na visão da ontologia agrícola) uma ou mais palavras-chave para pesquisar por data marts contendo informação sobre produção agrícola.
2. Analisar a produção e a produtividade (Quantidade Produzida/Área Colhida) de um certo produto agrícola em um espaço e tempo usando a descrição do data mart encontrado.

- (b) Determinar a distribuição da produção de cebola nos municípios de Santa Catarina em 2005.
 - (c) Determinar a distribuição da produtividade de cebola nos municípios de Santa Catarina em 2005.
 - (d) Identificar o município com maior produtividade de cebola em 2005, na tabela ou no mapa resultante da execução da consulta anterior.
3. Executar análise de informação sobre a certificação de produtores orgânicos usando a descrição do data mart encontrado.
- (e) Determinar a distribuição das sedes das propriedades agrícolas nos municípios de Santa Catarina que possuíam certificação de produtor de produtos orgânicos em 2003.
 - (f) Identificar o município com maior número de propriedades com certificação de produtor de produtos orgânicos em 2003.
 - (g) Identificar, neste município, as sedes das propriedades que estão a uma distância inferior a 5 km de um certo ponto geográfico.

5.3 Resultados do Experimento

Após a execução do experimento com todos os usuários, os vídeos e os dados registrados foram manualmente analisados. Um score foi atribuído para cada tarefa e para cada usuário, de acordo com o nível de execução da tarefa, conforme o seguinte critério:

- **Sem sucesso (score = 0):** quando o usuário não foi capaz de executar as interações necessárias para a realização da tarefa.
- **Realizou parcialmente (score = 1):** quando o usuário foi capaz de realizar algumas das interações necessárias para a realização da tarefa.
- **Realizado com dificuldade (score = 2):** quando o usuário conseguiu executar os passos necessários para a realização da tarefa, mas em um tempo superior a 4 minutos.
- **Realizado com facilidade (score = 3):** quando o usuário conseguiu executar os passos necessários para a realização da tarefa, em tempo inferior a 4 minutos.

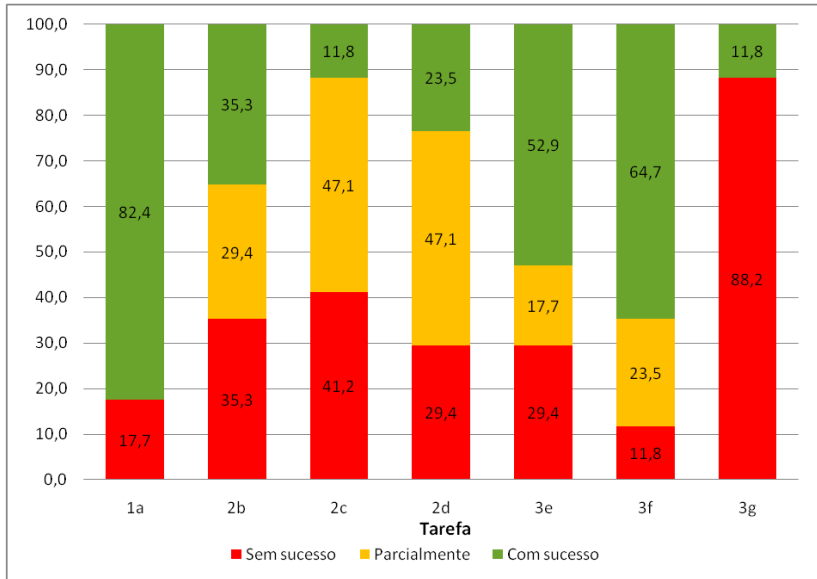


Figura 22: Percentual de sucesso por tarefa

As Figuras 22 e 23 mostram as medidas do sucesso dos usuários na realização de cada tarefa. A Figura 22 apresenta o percentual dos usuários com diferentes níveis de sucesso na realização de cada tarefa. A Figura 23 apresenta o score médio dos usuários na realização de cada tarefa.

As Figuras 24 e 25 mostram o grau de dificuldade encontrado pelos usuários na realização das tarefas. A Figura 24 apresenta o tempo gasto médio em segundos para a execução de cada tarefa. A Figura 25 apresenta o número médio de cliques no mouse que os usuários efetuaram na realização de cada tarefa.

O questionário preenchido pelos usuários após o teste era composto das seguintes questões:

1. A interface proposta é clara e fácil de entender?
2. O processo de construção das consultas é simples e de fácil aprendizado?
3. Os termos utilizados na interface são claros?

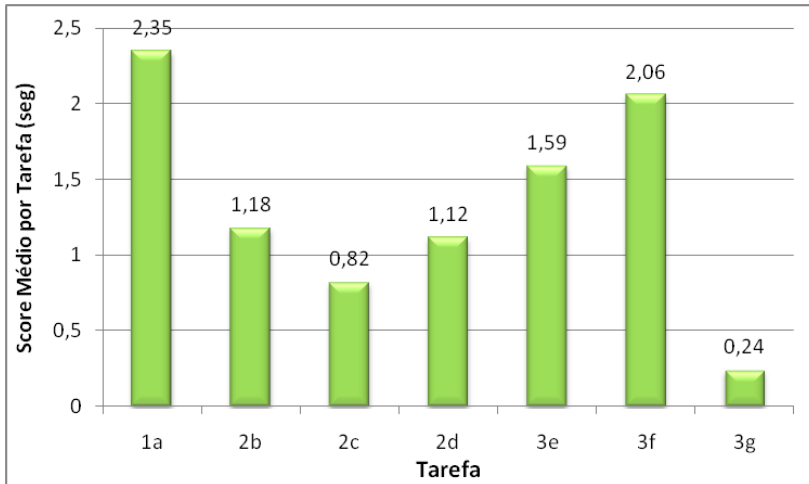


Figura 23: Score médio por tarefa

4. Os resultados finais das consultas foram alcançado?
5. A interface proposta atende necessidades de consulta do seu trabalho?

Para cada questão acima, o usuário selecionava um dos seguintes scores: (1) Discordo totalmente; (2) Discordo; (3) Nem concordo nem discordo; (4) Concordo; (5) Concordo totalmente.

As figuras 26 e 27 mostram os resultados obtidos com a aplicação do questionário. A Figura 26 apresenta para cada uma das cinco questões representadas no eixo X, o percentual de cada resposta fornecido pelos usuários. A Figura 27 apresenta a resposta média para cada questão de acordo com o score estabelecido.

Ao final, um campo adicional foi fornecido para que os usuários expressassem comentários e observações gerais sobre a interface e o experimento.

A seguir são transcritos alguns dos comentários fornecidos pelos usuários no próprio sistema, para cada uma das questões propostas:

1. A interface proposta é clara e fácil de entender.
 - “Num primeiro contato fica difícil seguir os passos. Parece que com a prática fica mais fácil”;

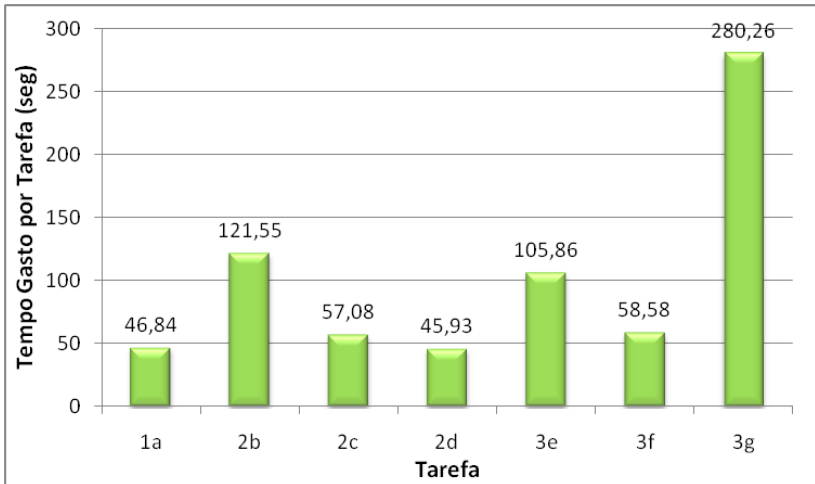


Figura 24: Tempo médio para a realização de cada tarefa (seg)

- “à primeira vista, sem treinamento de como utilizar, parece um pouco difícil para quem não tem muita facilidade com o uso de sistemas”;
 - “As consultas mais simples são fáceis, já as mais elaboradas são difíceis de executar.”;
 - “mas confundi muito os botões de nova consulta com exibir resultados”;
 - “A interface é clara e intuitiva, mas na última tarefa, não foi possível localizar a área para entrada dos dados referentes à distância e localização”;
2. O processo de construção das consultas é simples e de fácil aprendizado.
- “Se torna simples a medida que se entende a lógica de funcionamento, mas não à primeira vista”;
 - “As consultas que envolvem operações espaciais não estão intuitivas”;
 - “O aprendizado é rápido. São necessários poucos instantes para o usuário se acostumar com o programa”.

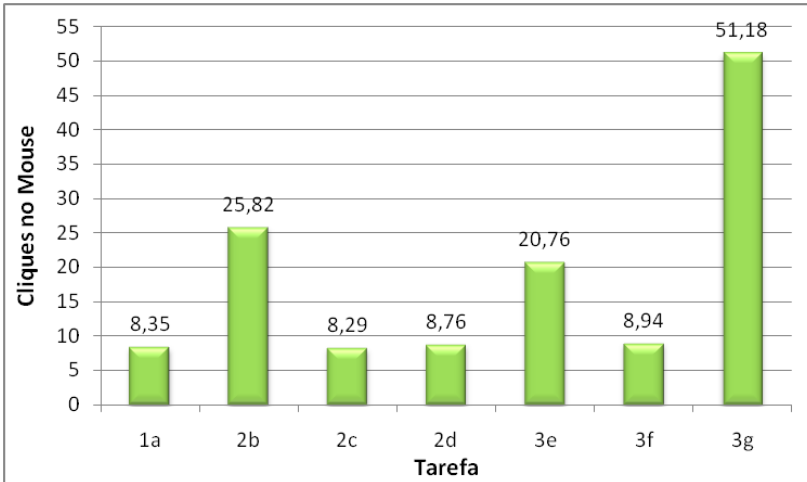


Figura 25: Número médio de cliques do mouse para cada tarefa

3. Os termos utilizados na interface são claros.
 - “Os termos utilizados nas operações espaciais não ficaram totalmente claros, apesar do campo de descrição que existe para explicar cada termo”;
 - “Sim, foram claros, poderia ter uma dica dizendo que ao selecionar uma opção para filtro espacial a descrição seria mostrada logo abaixo.”;
4. Os resultados finais das consultas foram alcançados.
 - “Não encontrei modo de usar o critério de distancia na ultima consulta”;
 - “sim, com exceção da última tarefa que foi parcialmente completada pois não encontrei como filtrar propriedades a uma certa distância da sede do município ”;
 - “As tarefas iniciais foram realizadas com rapidez e precisão. Mas houve dúvidas na execução da última delas. Não foi possível localizar a área para entrada de medidas e distâncias”.
5. A interface proposta atende necessidades de consulta do seu trabalho.

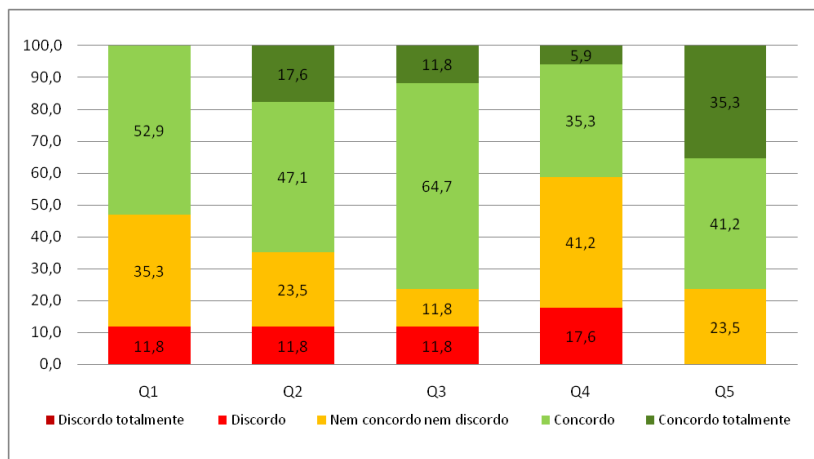


Figura 26: Frequência de resposta por questão

- “Aplica-se a várias situações problema.”;
- “Sim atende”;
- “É uma interface clara, intuitiva”.

5.4 Análise dos resultados

O percentual de usuários que obtiveram score “Sem sucesso” na execução das tarefas cresce nas 3 primeiras tarefas, como demonstra a Figura 22. Isto ocorre porque as tarefas 1a e 2b são consideradas relativamente fáceis. A partir daí os percentuais começam a decrescer. A sequência de tarefas entre as tarefas 2c e 3f apresentou acréscimo no score médio (Figura 23), indicando que um processo de aprendizado ocorreu durante a execução destas tarefas. Na execução das atividades do grupo 3, que são similares àquelas do grupo 2, os usuários já possuem maior familiaridade com o sistema e conseguem identificar padrões para a realização de algumas tarefas.

O percentual de sucesso na execução da tarefa 2c foi o menor, porque a tarefa envolveu a medida “produtividade”, que não está explícita na tabela fato do data mart, mas teria que ser calculada a partir de medidas já existentes (quantidade produzida / área colhida). Muitos usuários fizeram buscas baseadas na quantidade produzida e na área colhida, mas não através da

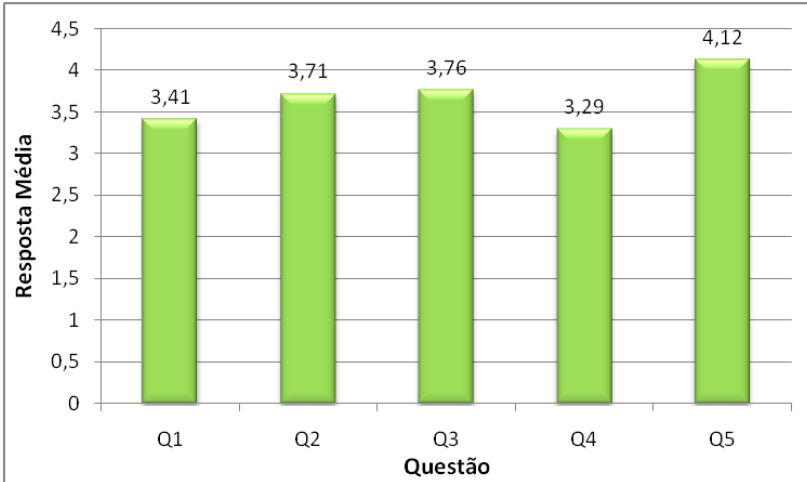


Figura 27: Resposta média por questão

palavra-chave “produtividade”. Isto indica que o S^2DW deve ser aperfeiçoado para apresentar medidas derivadas aos usuários, baseadas nas descrições das palavras-chave da ontologia de domínio.

O score médio da tarefa 3g foi o menor (Figura 23). Houve usuários que não conseguiram encontrar o botão para selecionar os operadores espaciais que poderiam ser aplicados ao membro selecionado, tornando impossível a realização da tarefa. Este botão está localizado do lado esquerdo de cada nível da dimensão. O tempo gasto e o número de cliques na realização da tarefa 3g demonstra que os usuários sabiam o que queriam e efetuavam algumas ações para tentar encontrar a maneira de selecionar as operações espaciais a serem usadas.

A Figura 24 mostra a diminuição do tempo necessário para a execução das tarefas subsequentes no mesmo grupo, com exceção da tarefa 3g. Com o refinamento de uma análise de informação, a interface mantém as seleções prévias, acelerando os procedimentos para a especificação das consultas SOLAP subsequentes. Na tarefa 1a, que supostamente seria a mais rápida, observou-se que muitos usuários inicialmente indicaram suas próprias palavras-chave ao sistema, ignorando as instruções fornecidas no experimento, apenas para explorar o conteúdo do sistema.

Analisando as respostas fornecidas pelos usuários no questionário,

observa-se que há um total ou parcial concordância da maioria dos usuários (nove ou mais usuários) com todas as questões, com exceção da questão 4. A questão 4 indica a correlação com os resultados desejados, expressando o sentimento de alguns usuários e de sua correspondência com os resultados na tarefa de análise. Os usuários expressam a sensação de que alguns dos resultados desejados não foram obtidos. A análise das observações e sugestões fornecidas pelos usuários também comprova a concordância com o sistema e a indicação da existência de alguns problemas a resolver.

5.5 Considerações

Este capítulo apresentou os experimentos efetuados com os usuários e analisou os seus resultados. Observa-se que há uma concordância geral de que o sistema proposto atende às necessidades de análise de informação dos usuários e que alguns ajustes são necessários para aperfeiçoar o sistema e aumentar a eficiência dos usuários no seu uso. O próximo capítulo discute os trabalhos relacionados com esta proposta.

6 TRABALHOS RELACIONADOS

6.1 Interfaces gráficas de consulta

Diversos trabalhos citam as dificuldades enfrentadas por usuários com pouco conhecimento de tecnologia da informação para efetuarem consultas sobre bancos de dados em geral. Eles propõem diferentes abordagens para os usuários visualizarem a estrutura e o conteúdo de bases de dados e especificarem consultas sobre as mesmas.

(RISHE et al., 2000) propõem um modelo de dados (Sem-ODM) que visa proporcionar flexibilidade, interfaces de usuário mais amigáveis, programas menores e facilitar o processo de consulta nos bancos de dados.

(TERWILLIGER; DELCAMBRE; LOGAN, 2007) citam as dificuldades que usuários especialistas de um domínio têm para efetuar consultas em bancos de dados: ou precisam de uma interface de consultas especial ou de SQL para expressar as suas consultas. Com uma interface especial, exige-se tempo e esforço para sua construção; e com SQL o especialista precisa aprender SQL e compreender a semântica dos dados e a sua estrutura. Usuários especialistas do domínio normalmente não têm conhecimentos para uso de SQL; usuários de tecnologia não necessariamente conhecem os conceitos do domínio. Sua proposta é criar uma interface de consulta que seja útil para usuários experientes que permita que expressem suas consultas sobre elementos de contexto da interface do usuário, em vez de sobre o esquema do banco.

(SPAHN et al., 2008) apresentam a ferramenta Semantic Query Designer (SQD), baseada em uma arquitetura que faz uso de ontologias para criação de consultas por usuários de negócio. Utilizam um middleware semântico que integra dados de diversos sistemas e fornece um modelo de dados na forma de uma ontologia de negócio. O SQD fornece uma ferramenta flexível e visa habilitar os usuários de negócio a melhorar a sua tomada de decisão. O vocabulário da ontologia de negócio é familiar a usuários finais e a interface gráfica facilita a navegação na estrutura da ontologia e na criação de consultas.

Algumas ferramentas de análise de informação se propõem a auxiliar o usuário nos procedimentos de análise de informação em SDWs: Pentaho - (PENTAHO, 2010b), Mondrian - (PENTAHO, 2010a), SpagoBI - (SPAGOWORLD, 2010b), GeoBI - (SPAGOWORLD, 2010a) e JMap - (KHEOPSTECHNOLOGIES, 2010).

(SHEKHAR et al., 2001) propõem a ferramenta MapCube para a visualização em SDWs. Eles ampliam o conceito de cubo de dados para o

domínio espacial.

(SCOTCH; PARMANTO, 2005) apresenta uma proposta de integração OLAP e GIS, que juntou técnicas de visualização espacial com a análise de dados multidimensionais.

(ESCRIBANO et al., 2007) apresenta o Piet, uma ferramenta para análise multidimensional de dados onde uma dimensão geográfica é composta de um conjunto de grafos, cada um descrevendo um conjunto de geometrias de uma camada temática.

Em interfaces para GIS, foram considerados os trabalhos de (AUFAURE-PORTIER; TRÉPIED, 1996; BONHOMME; TREPIED; LAURINI, 1999; CALCINELLI; MAINGUENAUD, 1994).

6.2 Uso de Semântica

Alguns outros trabalhos consideram especificamente o uso de semântica para suportar análise de informação em data warehouses (SELL et al., 2008; XIE et al., 2007; DIAMANTINI; POTENA, 2008).

(SELL et al., 2008) usam ontologias e inferência para integrar a semântica do negócio com os dados dimensionais, de modo a suportar serviços de análise de informação.

(XIE et al., 2007) usam uma extensão de OWL para especialistas em TI construírem mapeamentos entre o esquema de um data warehouse e termos utilizados no modelo de análise de informação, com o objetivo de permitir aos usuários especialistas de domínio especificar suas necessidades de análise e gerar automaticamente data marts para atendê-las.

(DIAMANTINI; POTENA, 2008) propõem um modelo para anotação semântica de data warehouses, que leva em consideração uma ontologia de domínio e uma ontologia matemática para descrever consultas.

Nenhum desses trabalhos apresenta interfaces visuais para o usuário efetuar consultas dimensionais, nem considera objetos espaciais e sua manipulação.

6.3 Comparações com este trabalho

Considerando as interfaces gráficas propostas para análise de informação em SDWs, a interface proposta neste trabalho considerou aspectos de interfaces já existentes e procurou inserir alguns aspectos que auxiliassem o usuário especialista de domínio a especificar suas necessidades de análise no SDW.

A inserção de representações visuais da ontologia de domínio e do data mart espacial, com descrições de conteúdo e dos mecanismos de análise de informação aplicáveis, de acordo com a natureza das medidas e dimensões consideradas, facilita o processo de análise de informação. Ferramentas como (PENTAHO, 2010b; SPAGOWORLD, 2010b; SCOTCH; PARMANTO, 2005) exibem os dados de acordo com a estrutura das tabelas do SDW, não contemplando aspectos da linguagem do especialista do domínio. Espera-se que as descrições semânticas possibilitem e agilizem os procedimentos de interação de um especialista de domínio com o SDW, pois elas evitam a necessidade dele consultar documentações e dicionários de dados para compreender a estrutura do SDW.

Da mesma forma, a exibição da ontologia de SDW com alguns aspectos gráficos que auxiliem o usuário na compreensão da função e objetivos dos operadores facilita a análise por parte do usuário no processo de seleção de operadores espaciais que podem ser aplicados durante os procedimentos de análise.

A representação abstrata dos data marts através de grafos semanticamente enriquecidos também fornece ao usuário uma melhor visualização da estrutura e do conteúdo disponível para as consultas, exibindo as tabelas fato, medidas, dimensões, níveis das dimensões, membros de níveis, operadores e funções de agregação de dados. Esta visualização auxilia o usuário no processo de especificação de suas consultas nestes data marts.

A estrutura proposta contempla e busca atender diversos tipos de usuários, desde aqueles que querem efetuar análises de informação mais simples, com resultados em tabelas ou mapas básicos, bem como aqueles usuários que pretendem efetuar análises espaciais mais aprofundadas com aplicação de conjuntos sobrepostos de operadores.

A Tabela 3 apresenta uma comparação entre os diversos trabalhos analisados e o S^2DW .

Este trabalho teve por base a arquitetura proposta por (XIE et al., 2007) e a estendeu, aplicando ideias dos demais trabalhos.

Tabela 3: Comparação do S^2DW com outros trabalhos

Autor	Interface gráfica para análise de informação	Uso de Ontologia de domínio	Representação visual do conteúdo	Uso de guagem domínio Usuário	Lin- de do	Objetos Espaciais
Rishe	x					
Terwilliger	x		x	x		
Spahn	x	x	x	x		
Shekar	x		x			x
Scotch	x		x			x
Escribano	x		x			x
Sell		x	x	x		
Xie		x	x	x		
Diamantini		x	x	x		
S^2DW	x	x	x	x		x

7 CONCLUSÕES E TRABALHOS FUTUROS

O S^2DW utiliza ontologias para descrever a estrutura, conteúdo de informação e as capacidades de análise de informação de SDWs. Data marts espaciais são semanticamente descritos no S^2DW através de definições de correlação da ontologia de SDW com a ontologia de domínio. Esta última pode ser trocada na base de conhecimento S^2DW , de forma a permitir a adaptação do sistema a diferentes domínios de modo a suportar a interação de usuários com SDWs de acordo com vocabulários específicos.

7.1 Contribuições

As principais contribuições deste trabalho são:

1. Proposta de uma arquitetura de sistema baseado em conhecimento para permitir a usuários especialistas de domínio realizarem análises de informação em SDWs.
2. Desenvolvimento de um SDW sobre propriedades e produção agrícola de Santa Catarina para servir como estudo de caso na especificação e validação parciais do sistema proposto.
3. Desenvolvimento e prototipação de uma interface gráfica do sistema proposto.
4. Descrição das interações a serem realizadas sobre tal interface para a realização de análises de informação (envolvendo o uso de dados, operadores e funções de agregação espaciais) no SDW usado como estudo de caso.
5. Realização de estudos empíricos sobre a viabilidade e usabilidade da interface proposta.

A interface gráfica baseada em conhecimento do S^2DW provê abstrações para permitir que usuários comuns efetuem consultas na busca de data marts espaciais com informações sobre assuntos de interesse e efetuar análise de informação nos data marts espaciais retornados. Possui as seguintes capacidades:

- busca semântica em data marts espaciais relacionados a um tópico de interesse, através do uso de palavras-chave fornecidas pelo usuário ou

escolhidas em uma visão da ontologia de domínio fornecida pelo sistema;

- apresenta descrição semântica da estrutura e das possibilidades de análise de informação dos data marts espaciais de interesse, em uma interface baseada em conhecimento;
- permite interações com o data mart através desta interface, auxiliando o usuário a utilizar corretamente as capacidades de análise de informação, especificando consultas SOLAP sobre o data mart espacial;
- permite interações do usuário com a descrição semântica do data mart e com as tabelas, gráficos e mapas apresentados em resposta a consultas fornecidas pelo usuário, de maneira a refinar gradualmente sua análise de informação.

O principal benefício esperado com este trabalho é prover facilidade para a exploração e consulta em SDWs por usuários especialistas do domínio, em vez de apenas engenheiros de informação e programadores. Os benefícios do sistema proposto se aplicam a data warehouses convencionais, mas são destacados em SDWs, devido a variedade e a complexidade dos objetos espaciais e das possibilidades de manipulação de dados espaciais. O resultado do experimento realizado com a interface baseada em conhecimento do S^2DW , para a execução de análise de informação em um SDW do domínio agrícola, indica que o sistema proposto necessita de alguns aperfeiçoamentos e ajustes, mas atende às necessidades do usuário. Os usuários consideram que a interface é clara e fácil de entender. O processo de consulta é fácil de aprender e permitiu que a maioria dos usuários especificasse suas consultas SOLAP. Assim, os experimentos dão uma indicação de que a proposta é promissora.

7.2 Publicações

Durante a realização do trabalho, foram publicados os seguintes artigos relacionados às pesquisas realizadas:

- (DEGGAU; FILETO, 2009) - DEGGAU, R.; FILETO, R. Enriquecendo Data Warehouses Espaciais com Descrições Semânticas. In: WTDBD (Workshop de Teses e Dissertações em Banco de Dados). [S.l.: s.n], 2009. p. 61-66).

- (DEGGAU; FILETO, 2010)- DEGGAU, R.; FILETO, R. Interagindo com data warehouses espaciais através de descrições semânticas. In: VI Escola Regional de Banco de Dados. [S.l.: s.n], 2010. ISSN: 2177-4226.
- (PEREIRA et al., 2010) - PEREIRA, D.; MERINO, E; DEGGAU, R.; FILETO, R. Avaliação de usabilidade de interface humano computador para análise de informação em data warehouses espaciais. In: ABERGO 2010 - Congresso Brasileiro de Ergonomia. [S.l.: s.n.], 2010.
- DEGGAU, R.; FILETO, R.; PEREIRA, D; MERINO, E. Interacting with Spatial Data Warehouses via Semantic Descriptions In: GEOINFO. 2010. (in submission)

7.3 Trabalhos Futuros

Como trabalhos futuros pode-se destacar:

- aperfeiçoar a interface visando melhorar a interação do usuário com o sistema, corrigindo aspectos identificados no experimento e acrescentando exemplos de análises complexas com diversos operadores em exemplos disponíveis na interface;
- descrever formalmente os data marts espaciais com os mapeamentos entre a ontologia de SDW e ontologia de domínio;
- descrever um processo que auxilie os procedimentos de mapeamentos entre a ontologia de SDWs e outras ontologias de domínio, visando facilitar o uso do S^2DW em outros domínios de conhecimento;
- transformar as interações do usuário com a interface baseada em conhecimento do S^2DW em consultas SOLAP;
- incluir capacidades de interpretação com linguagem natural, regras e inferência para melhor auxiliar os usuários na escolha adequada e na composição dos membros das dimensões, operadores, funções de agregação e as medidas para sua análise de informação;
- aprofundar-se na questão de geração de descrições semânticas de data marts espaciais, definindo a estrutura formal de tais descrições e um processo para a sua geração, permitindo o reuso de descrições para a

criação de novos data marts visando suportar análises diferentes mas análogas.

REFERÊNCIAS

- AUFAURE-PORTIER, M.-A.; TRÉPIED, C. A survey of query languages for geographic information systems. In: *IDS*. [S.l.: s.n.], 1996. p. 3.
- BATINI, C. et al. Visual strategies for querying databases. In: *In IEEE Workshop on Visual Languages*. [S.l.: s.n.], 1991. p. 183–189.
- BATINI, C. et al. Visual query systems: A taxonomy. In: *Proceedings of the IFIP TC2/WG 2.6 Second Working Conference on Visual Database Systems II*. Amsterdam, The Netherlands, The Netherlands: North-Holland Publishing Co., 1992. p. 153–168. ISBN 0-444-89609-0.
- BIMONTE, R.; TCHOUNIKINE, A.; MIQUEL, M. Spatial olap: Open issues and a web based prototype. In: *Proceedings of the 10th AGILE International Conference on Geographic Information Science*. [S.l.: s.n.], 2007. p. 1–11.
- BIMONTE, S.; TCHOUNIKINE, A.; MIQUEL, M. Towards a spatial multidimensional model. In: SONG, I.-Y.; TRUJILLO, J. (Ed.). *DOLAP*. [S.l.]: ACM, 2005. p. 39–46. ISBN 1-59593-162-7.
- BIMONTE, S.; TCHOUNIKINE, A.; MIQUEL, M. Geocube, a multidimensional model and navigation operators handling complex measures: Application in spatial olap. In: YAKHNO, T. M.; NEUHOLD, E. J. (Ed.). *ADVIS*. [S.l.]: Springer, 2006. (Lecture Notes in Computer Science, v. 4243), p. 100–109. ISBN 3-540-46291-0.
- BIMONTE, S. et al. Gewolap: A web based spatial olap proposal. In: MEERSMAN, R.; TARI, Z.; HERRERO, P. (Ed.). *OTM Workshops (2)*. [S.l.]: Springer, 2006. (Lecture Notes in Computer Science, v. 4278), p. 1596–1605. ISBN 3-540-48273-3.
- BONHOMME, C.; TREPIED, C.; LAURINI, R. A visual language for querying spatio-temporal databases. In: *ACM GIS '99*. [S.l.]: ACM Press, 1999. p. 34–39.
- CALCINELLI, D.; MAINGUENAUD, M. Cigales: A visual query language for geographical information system: The user interface. *Journal of Visual Languages and Computing*, v. 5, p. 113–132, 1994.
- CATARCI, T. et al. Visual query systems for databases: A survey. *Journal of Visual Languages and Computing*, v. 8, p. 215–260, 1997.

- CATARCI, T.; SANTUCCI, G. Query by diagram: a graphical environment for querying databases. In: *SIGMOD '94: Proceedings of the 1994 ACM SIGMOD international conference on Management of data*. New York, NY, USA: ACM, 1994. p. 515. ISBN 0-89791-639-5.
- CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and olap technology. 1997. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=?doi=10.1.1.133.6667>>.
- D'AGOSTINI, C. S.; FILETO, R. Capturing users' preferences and intentions in a semantic search system. In: *Proceedings of the 21st International Conference on Software Engineering & Knowledge Engineering (SEKE), Boston, Massachusetts, USA, July 1-3*. [S.l.]: Knowledge Systems Institute Graduate School, 2009. p. 587–591.
- DAMIANI, M. L.; DAMIANI, S. S. M.; WANG, J. *Spatial Data Warehouse Modelling*. Information Science Reference, 2008. Disponível em: <<http://hdl.handle.net/2434/36498>>.
- DEGGAU, R.; FILETO, R. Enriquecendo Data Warehouses Espaciais com Descrições Semânticas. In: *WTDBD (Workshop de Teses e Dissertações em Banco de Dados)*. [S.l.: s.n.], 2009. p. 61–66.
- DEGGAU, R.; FILETO, R. Interagindo com data warehouses espaciais através de descrições semânticas. In: *VI Escola Regional de Banco de Dados*. [S.l.: s.n.], 2010. ISSN: 2177-4226.
- DIAMANTINI, C.; POTENA, D. Semantic enrichment of strategic datacubes. In: *DOLAP '08: Proceeding of the ACM 11th international workshop on Data warehousing and OLAP*. New York, NY, USA: ACM, 2008. p. 81–88. ISBN 978-1-60558-250-4.
- EGENHOFER, M. J.; FRANZOSA, R. D. Point set topological relations. *International Journal of Geographical Information Systems*, v. 5, p. 161–174, 1991.
- EGENHOFER, M. J.; HERRING, J. Categorizing binary topological relationships between regions, lines and points in geographical databases. Department of Surveying Engineering, University of Maine, Orono, 1992.
- EMBRAPA. *Embrapa desenvolve banco de dados de solos na internet*. Maio 2008. <http://migre.me/UQ0P>.

- EPAGRI. *EPAGRI*. Junho 2010. <http://www.epagri.sc.gov.br/>.
- EPAGRI/CEPA. *Dados do LAC*. Maio 2010. http://www.cepa.epagri.sc.gov.br/Dados_do_LAC/lac_conceitos.htm.
- EPAGRI/IBGE. *Mapoteca Topográfica Digital de Santa Catarina*. 2004. <http://ciram.epagri.sc.gov.br/mapoteca/>.
- ESCRIBANO, A. et al. Piet: a gis-olap implementation. In: *DOLAP '07: Proceedings of the ACM tenth international workshop on Data warehousing and OLAP*. New York, NY, USA: ACM, 2007. p. 73–80. ISBN 978-1-59593-827-5.
- FIDALGO, R. do N. Uma infra-estrutura para integração de modelos, esquemas e serviços multidimensionais e geográficos. Biblioteca Digital de Teses e Dissertações da UFPE, 2005. Disponível em: <<http://www.btd.ufpe.br/tedeSimplificado/tdebusca/arquivo.php?codArquivo=560>>.
- FIDALGO, R. do N. et al. Geodwframe: A framework for guiding the design of geographical dimensional schemas. In: *DaWaK*. [S.l.: s.n.], 2004. p. 26–37.
- GRAY, J. et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. 2007. Disponível em: <<http://arxiv.org/abs/cs/0701155>>.
- INMON, W. H.; IMHOFF, C.; BATTAS, G. *Building the operational data store*. New York, NY, USA: John Wiley & Sons, Inc., 1995. ISBN 0-471-12822-8.
- INMON, W. H. et al. Data warehouse third edition building the data warehouse third edition. 2008. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.91.6134>>.
- ISO/IEC. *Quality characteristics and guidelines for use. - ISO/IEC 9126*. [S.l.]: ISO, Geneva, Switzerland, 1991.
- KHEOPSTECHNOLOGIES. *JMap Spatial OLAP*. Julho 2010. <http://www.kheops-tech.com/en/jmap/solap.jsp>.
- KIMBALL, R. The data warehouse lifecycle toolkit : expert methods for designing, developing, and deploying data warehouses / r. kimball...

- [et al.]. Nueva York, EUA : Wiley, 1998. Disponível em: <<http://148-201.96.14/dc/ver.aspx?ns=000117337>>.
- KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. New York, NY, USA: John Wiley & Sons, Inc., 2002. ISBN 0471200247.
- MALINOWSKI, E.; ZIMÁNYI, E. Representing spatiality in a conceptual multidimensional model. In: PFOSER, D.; CRUZ, I. F.; RONTHALER, M. (Ed.). *GIS*. [S.l.]: ACM, 2004. p. 12–22. ISBN 1-58113-979-9.
- MALINOWSKI, E.; ZIMÁNYI, E. Requirements specification and conceptual modeling for spatial data warehouses. In: MEERSMAN, R.; TARI, Z.; HERRERO, P. (Ed.). *OTM Workshops (2)*. [S.l.]: Springer, 2006. (Lecture Notes in Computer Science, v. 4278), p. 1616–1625. ISBN 3-540-48273-3.
- MALINOWSKI, E.; ZIMÁNYI, E. Logical representation of a conceptual model for spatial data warehouses. *GeoInformatica*, v. 11, n. 4, p. 431–457, 2007.
- MALINOWSKI, E.; ZIMNYI, E. *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications (Data-Centric Systems and Applications)*. [S.l.]: Springer Publishing Company, Incorporated, 2008. ISBN 3540744045, 9783540744047.
- MANGOLD, C. A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontologies*, Inter-science, v. 2, n. 1, p. 23–34, 2007.
- MARTINO, S. D. et al. Integrating google earth within olap tools for multidimensional exploration and analysis of spatial data. In: *ICEIS*. [S.l.: s.n.], 2009. p. 940–951.
- MEERSMAN, R.; TARI, Z.; HERRERO, P. (Ed.). *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, OTM Confederated International Workshops and Posters, AWeSOMe, CAMS, COMINF, IS, KSinBIT, MIOS-CIAO, MONET, OnToContent, ORM, PerSys, OTM Academy Doctoral Consortium, RDDS, SWWS, and SeBGIS 2006, Montpellier, France, October 29 - November 3, 2006. Proceedings, Part II*, v. 4278 de *Lecture Notes in Computer Science*, (Lecture

- Notes in Computer Science, v. 4278). [S.l.]: Springer, 2006. ISBN 3-540-48273-3.
- NIELSEN, J. *Usability Engineering*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 0125184050.
- PENTAHO. *Pentaho Analysis Services: Mondrian Project*. Julho 2010. <http://mondrian.pentaho.org/>.
- PENTAHO. *Pentaho Open Source Business Intelligence*. Julho 2010. <http://www.pentaho.com/>.
- PEREIRA, D. et al. Avaliação de usabilidade de interface humano-computador para análise de informação em data warehouses espaciais. In: *ABERGO 2010 - Congresso Brasileiro de Ergonomia*. [S.l.: s.n.], 2010.
- RAO, F. et al. Spatial hierarchy and olap-favored search in spatial data warehouse. In: *DOLAP*. [S.l.: s.n.], 2003. p. 48–55.
- RIGAUX, P.; SCHOLL, M.; VOISARD, A. *Spatial databases with application to GIS*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2002. ISBN 1-55860-588-6.
- RISHE, N. et al. Semantic access: Semantic interface for querying databases. In: *VLDB*. [S.l.: s.n.], 2000. p. 591–594.
- RIVEST, S. et al. Solap technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 60, n. 1, p. 17 – 33, 2005. ISSN 0924-2716. Including Theme Section: - Advances in Spatio-temporal Analysis and Representation. Disponível em: <<http://www.sciencedirect.com/science/article/B6VF4-4HKCYTV-1/2/faa4d3336c65b9f80f7a4c6458fb23e1>>.
- RUIZ, C. V.; TIMES, V. C. A taxonomy of solap operators. In: BRAYNER, A. (Ed.). *SBBD*. [S.l.]: SBC, 2009. p. 151–165.
- SCOTCH, M.; PARMANTO, B. Sovat: Spatial olap visualization and analysis tool. In: *HICSS '05: Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 6*. Washington, DC, USA: IEEE Computer Society, 2005. p. 142.2. ISBN 0-7695-2268-8-6.

- SELL, D. et al. Sbi: a semantic framework to support business intelligence. In: *OBI '08: Proceedings of the first international workshop on Ontology-supported business intelligence*. New York, NY, USA: ACM, 2008. p. 1–11. ISBN 978-1-60558-219-1.
- SHEKHAR, S. et al. *Map Cube: A Visualization Tool for Spatial Data Warehouses*. 2001.
- SILVA, J. da et al. A set of aggregation functions for spatial measures. In: *DOLAP '08: Proceeding of the ACM 11th international workshop on Data warehousing and OLAP*. New York, NY, USA: ACM, 2008. p. 25–32. ISBN 978-1-60558-250-4.
- SKOUTAS, D.; SIMITSIS, A. Designing etl processes using semantic web technologies. In: *DOLAP '06: Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*. New York, NY, USA: ACM, 2006. p. 67–74. ISBN 1-59593-530-4.
- SPAGOWORLD. *Geo BI initiative*. Julho 2010. <http://www.spagoworld.org/xwiki/bin/view/GeoBI/>.
- SPAGOWORLD. *SpagoBI Open Source Business Inteligence*. Julho 2010. <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/>.
- SPAHN, M. et al. Supporting business intelligence by providing ontology-based end-user information self-service. In: *OBI '08: Proceedings of the first international workshop on Ontology-supported business intelligence*. New York, NY, USA: ACM, 2008. p. 1–12. ISBN 978-1-60558-219-1.
- TECHSMITH. *MORAE*. Abril 2010. <http://www.techsmith.com/morae.asp>.
- TERWILLIGER, J. F.; DELCAMBRE, L. M. L.; LOGAN, J. Querying through a user interface. *Data Knowl. Eng.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 63, n. 3, p. 774–794, 2007. ISSN 0169-023X.
- THOMSEN, E. *OLAP Solutions : Building Multidimensional Information Systems*. [S.l.]: Nueva York, EUA : Wiley, 1997.
- XIE, G. T. et al. Eiaw: Towards a business-friendly data warehouse using semantic web technologies. In: *ISWC/ASWC*. [S.l.: s.n.], 2007. p. 857–870.

- ZLOOF, M. M. Query-by-example: operations on hierarchical data bases. In: *AFIPS '76: Proceedings of the June 7-10, 1976, national computer conference and exposition*. New York, NY, USA: ACM, 1976. p. 845–853.
- ZLOOF, M. M. Query-by-example: a data base language. *IBM Syst. J.*, IBM Corp., Riverton, NJ, USA, v. 16, n. 4, p. 324–343, 1977. ISSN 0018-8670.